

# The Roles of Cancer-Associated Fibroblasts in Colorectal Carcinogenesis

Hiroki Kobayashi

Joint Degree Program

Adelaide Medical School

The University of Adelaide

Nagoya University

Graduate School of Medicine

April 2021

## Table of contents

|  |     |
|--|-----|
| Included publications .....  | 3   |
| Abstract.....  | 4   |
| Declaration.....   | 6   |
| Acknowledgements.....  | 7   |
| Chapter 1: Introduction: Cancer-associated fibroblasts in gastrointestinal cancer .....                          | 8   |
| Chapter 2: The balance of stromal BMP signaling mediated by GREM1 and ISLR drives colorectal carcinogenesis..... | 45  |
| Chapter 3: The origin and contribution of cancer-associated fibroblasts in colorectal carcinogenesis .....       | 152 |
| Chapter 4: Discussion.....   | 262 |
| Bibliography .....   | 264 |

## Included publications

Kobayashi H, Enomoto A, Woods SL, Burt AD, Takahashi M, Worthley DL. Cancer-associated fibroblasts in gastrointestinal cancer. *Nat Rev Gastroenterol Hepatol*. 2019;16(5):282-95.

Springer Nature. doi: 10.1038/s41575-019-0115-0.

<https://www.nature.com/articles/s41575-019-0115-0>

### (Included in Chapter 1)

Kobayashi H, Gieniec KA, Wright JA, Wang T, Asai N, Mizutani Y, Iida T, Ando R, Suzuki N, Lannagan TRM, Ng JQ, Hara A, Shiraki Y, Mii S, Ichinose M, Vrbanac L, Lawrence MJ, Sammour T, Uehara K, Davies G, Lisowski L, Alexander IE, Hayakawa Y, Butler LM, Zannettino ACW, Din MO, Hasty J, Burt AD, Leedham SJ, Rustgi AK, Mukherjee S, Wang TC, Enomoto A\*, Takahashi M\*, Worthley DL\*, Woods SL\* The balance of stromal BMP signaling mediated by *GREM1* and *ISLR* drives colorectal carcinogenesis. *Gastroenterology*. 2021;160(4):1224-39 e30. \*co-correspondence

Elsevier. doi:10.1053/j.gastro.2020.11.011.

[https://www.gastrojournal.org/article/S0016-5085\(20\)35400-7/fulltext](https://www.gastrojournal.org/article/S0016-5085(20)35400-7/fulltext)

### (Included in Chapter 2)

Kobayashi H\*, Gieniec KA\*, Lannagan TRM\*, Wang T, Asai N, Mizutani Y, Iida T, Ando R, Sakai A, Suzuki N, Ichinose M, Wright JA, Vrbanac L, Ng JQ, Goyne J, Lawrence MJ, Sammour T, Hayakawa Y, Klebe S, Shin AE, Asfaha S, Bettington M, Arpaia N, Rieder F, Arpaia N, Danino T, Butler LM, Burt AD, Leedham SJ, Rustgi AK, Mukherjee S, Takahashi M, Wang TC, Enomoto A\*\*, Woods SL\*\*, Worthley DL\*\*. The origin and contribution of cancer-associated fibroblasts in colorectal carcinogenesis. **Submitted**. \*co-first; \*\*co-correspondence

### (Included in Chapter 3)

## Abstract

Cancer-associated fibroblasts (CAF), key constituents of the tumour microenvironment, either promote or restrain tumour growth. Attempts to therapeutically target CAFs have been hampered by our rudimentary understanding of the functions and origins of these heterogeneous cells.

In this thesis, I first addressed the functional heterogeneity of CAFs involving bone morphogenetic proteins (BMPs), a key growth factor in CRC progression. Using human CRC RNA expression data, I identified Gremlin 1 (*Grem1*) and immunoglobulin superfamily containing leucine-rich repeat (*Islr*) as CAF-specific genes involved in BMP signalling. Functionally, *GREM1* and *ISLR* acted to inhibit and promote BMP signalling, respectively. *GREM1* and *ISLR* marked *ACTA2*<sup>high</sup> and *ACTA2*<sup>low</sup> colorectal CAFs, respectively. *Grem1* and *Islr* expression were differentially regulated by transforming growth factor- $\beta$  (TGF- $\beta$ ) and FOXL1, providing an underlying mechanism to explain fibroblast biological dichotomy. In CRC patients, high *GREM1* and *ISLR* expression were associated with poor and favourable survival, respectively. A *GREM1*-neutralizing antibody or fibroblast *Islr* overexpression reduced CRC tumoroid growth and promoted *Lgr5*<sup>+</sup> intestinal stem cell differentiation. Finally, adeno-associated virus 8 (AAV8)-mediated delivery of *Islr* to hepatocytes increased BMP signalling and improved mouse survival in our preclinical model of hepatic metastasis. Stromal BMP signalling predicts and modifies CRC progression and survival, and can be therapeutically targeted by novel AAV-directed gene delivery to the liver.

Next, I examined the origins and contributions of colorectal CAFs. Using five different fate-mapping models with BrdU dosing, this study revealed that half of *ACTA2*<sup>+</sup> CAFs emerge through proliferation in a mouse model of CRC. Intestinal pericryptal Leptin receptor (*Lepr*)<sup>+</sup> cells were the major origins of the proliferating CAFs. These *Lepr*-lineage CAFs, in turn, express melanoma cell adhesion molecule (MCAM), a CRC stroma-specific marker we identified using RNA-sequencing. High MCAM expression induced by TGF- $\beta$  was inversely associated with patient survival in human CRC. In mice, stromal *Mcam* knockout attenuated orthotopically injected colorectal tumoroid growth and improved mouse survival through decreased tumour-associated macrophage recruitment. Mechanistically, fibroblast MCAM interacted with interleukin-1 receptor 1 to augment nuclear factor-

$\kappa$ B-IL34/CCL8 signaling that promotes macrophage chemotaxis. Preventing the expansion/differentiation of Lepr-lineage CAFs or inhibiting MCAM activity could be effective therapeutic approaches for CRC.

These data indicate that targeting these CAF subpopulations could be novel potential therapeutic strategies to inhibit CRC progression.

## Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines. Embargo has been approved by the University to restrict access for two years after submission of my thesis.

Hiroki Kobayashi

01/04/2021

## Acknowledgements

I would like to thank all members at Gut Cancer, the University of Adelaide, South Australian Health and Medical Research Institute and Department of Pathology, Nagoya University Graduate School of Medicine: Krystyna A. Gieniec, Tamsin RM. Lannagan, Josephine A. Wright, Tongtong Wang, Nobumi Suzuki, Jia Q Ng, Mari Ichinose, Laura Vrbanac, Daniel L. Worthley, and Susan L. Woods (Gut Cancer, the University of Adelaide); Naoya Asai, Yasuyuki Mizutani, Tadashi Iida, Ryota Ando, Akihiro Sakai, Akitoshi Hara, Yukihiro Shiraki, Shinji Mii, Kaori Ushida, Kozo Uchiyama, Atsushi Enomoto and Masahide Takahashi (Department of Pathology, Nagoya University Graduate School of Medicine).

Especially, I thank my supervisors, Daniel Worthley, Susan Woods, Alastair Burt, Masahide Takahashi and Atsushi Enomoto, for their lots of support throughout my PhD candidature.

My work was supported by grants from Japan Society for the Promotion of Science Overseas Challenge Program for Young Researchers, Takeda Science Foundation Fellowship, Greaton International PhD Scholarship.

## **Chapter 1: Introduction: Cancer-associated fibroblasts in gastrointestinal cancer**

Colorectal cancer (CRC) is a major cause of cancer mortality worldwide<sup>1</sup>. One hallmark of colorectal cancer is a prominent desmoplastic reaction, which is mainly composed of cancer-associated fibroblasts (CAFs)<sup>2</sup>. Recent studies have elucidated the significance of CAFs in the regulation of gastrointestinal cancer progression. CAFs are now known to be heterogeneous cells in terms of markers and functions. Here, in the review article<sup>3</sup>, I have outlined our accumulating understanding of CAF heterogeneity, markers and functions as well as their potential cellular origins **(Chapter 1)**.

Previous papers have indicated that some CAFs promote, while others restrain, cancer progression<sup>4,5</sup>. However, it remains unknown what marker and signalling defines this CAF biological dichotomy. Moreover, the origin and contribution of CAFs remains elusive, making it challenging to therapeutically target CAFs and their progenitors.

In this thesis, I investigated the role of bone morphogenetic protein (BMP) signalling in shaping the functional heterogeneity of colorectal CAFs **(Chapter 2)**<sup>6</sup>. Next, through comprehensive lineage tracing experiments in a mouse model of CRC, this study uncovered a major source of colorectal CAFs that are crucial for CRC progression **(Chapter 3)**.

## Statement of Authorship

|                     |   |
|---------------------|---|
| Title of Paper      | Cancer-associated fibroblasts in gastrointestinal cancer  |
| Publication Status  | Published   |
| Publication Details | Kobayashi H, Enomoto A, Woods SL, Burt AD, Takahashi M, Worthley DL. Cancer-associated fibroblasts in gastrointestinal cancer. Nat Rev Gastroenterol Hepatol. 2019;16(5):282-95. Epub 2019/02/20. doi: 10.1038/s41575-019-0115-0. |

## Candidate

|                           |   |      |          |
|---------------------------|---|------|----------|
| Name of Candidate         | Hiroki Kobayashi  |      |          |
| Contribution to the Paper | Literature review. Wrote the manuscript.  |      |          |
| Overall Percentage (%)    | 60%   |      |          |
| Certification:            | This paper is a review article I wrote during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. |      |          |
| Signature                 |   | Date | 30/10/20 |

## Co-author contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all the co-author contributions is equal to 100% less the candidate's stated contribution.

Note that multiple authors have given written permission for my primary supervisor (Dr. Daniel Worthley) to sign on their behalf.

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Atsushi Enomoto                                |      |         |
| Contribution to the paper | Wrote the manuscript. Co-corresponding author. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Susan Woods   |      |         |
| Contribution to the paper | Discussed manuscript content and edited the manuscript. |      |         |
| Signature                 |   | Date | 6/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Alastair Burt   |      |         |
| Contribution to the paper | Discussed manuscript content and edited the manuscript. |      |         |
| Signature                 |   | Date | 9/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Masahide Takahashi                                      |      |         |
| Contribution to the paper | Discussed manuscript content and edited the manuscript. |      |         |
| Signature                 |   | Date | 9/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Daniel Worthley   |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation and manuscript evaluation, wrote the manuscript. Co-corresponding author. |      |         |
| Signature                 |   | Date | 6/11/20 |

## **Cancer-associated fibroblasts in gastrointestinal cancer**

Hiroki Kobayashi<sup>1,2</sup>, Atsushi Enomoto<sup>2\*</sup>, Susan L. Woods<sup>1</sup>, Alastair D. Burt<sup>3</sup>, Masahide Takahashi<sup>2</sup>  
and Daniel L. Worthley<sup>1\*</sup>

<sup>1</sup>Adelaide Medical School, University of Adelaide and South Australian Health and Medical Research Institute, Adelaide, Australia

<sup>2</sup>Department of Pathology, Nagoya University Graduate School of Medicine, Nagoya, Japan

<sup>3</sup>Faculty of Health and Medical Sciences, University of Adelaide, Adelaide, Australia

\*e-mail: [Dan.Worthley@sahmri.com](mailto:Dan.Worthley@sahmri.com); [enomoto@iar.nagoya-u.ac.jp](mailto:enomoto@iar.nagoya-u.ac.jp)

## Key points

- Cancer-associated fibroblasts (CAFs) include all fibroblasts in the tumour, and are involved in functionally controlling cancer progression.
- CAFs are composed of heterogeneous subpopulations arising from distinct cellular origins such as local fibroblasts and mesenchymal stem cells.
- Distinct CAFs influence cancer cell proliferation, tumour immunity, angiogenesis, extracellular matrix-remodelling and metastasis.
- Functionally, CAFs can be classified into subpopulations such as tumour-promoting CAFs and tumour-retarding CAFs.
- An improved understanding of CAF biology could lead to the development of novel stroma-based diagnostics, prognostics and therapeutics.

## Abstract

The tumour microenvironment, also termed the tumour stroma or tumour mesenchyme, includes fibroblasts, immune cells, blood vessels and the extracellular matrix, and substantially influences the initiation, growth and dissemination of gastrointestinal cancer. One of the critical components of the tumour mesenchyme are cancer-associated fibroblasts (CAFs), which not only provide physical support for epithelial cells, but are also key functional regulators in cancer, promoting and retarding tumourigenesis in a context-dependent manner. In this Review, we outline the emerging understanding of gastrointestinal CAFs with a particular emphasis on their origin and heterogeneity, as well as their function in cancer cell proliferation, tumour immunity, angiogenesis, extracellular matrix-remodelling and drug resistance. Moreover, we discuss the clinical implications of CAFs as biomarkers and potential targets for prevention and treatment of patients with gastrointestinal cancer.

## Introduction

The classic concept of carcinomas as ‘wounds that never heal’ or ‘organs that never develop’ has its limitations, but as an analogy it provides a helpful framework for understanding the inflammatory and developmental signalling between cancer cells and the activated tumour microenvironment (TME)<sup>7</sup>. Also termed the tumour stroma or tumour mesenchyme, the TME is composed of fibroblasts, inflammatory cells, blood vessels, extracellular matrix (ECM) and basement membrane (FIG. 1). Although most previous research has focused on the biology of cancer cells themselves, it is clear that the TME is a major contributor to cancer development<sup>8-11</sup>. For example, pancreatic ductal adenocarcinoma (PDAC) is characterized by a prominent desmoplastic reaction, a fibrotic stromal reaction accompanied by activated CAFs and extensive deposition of ECM, accounting for up to 90% of the tumour<sup>12,13</sup>. Among the heterogeneous components of the cancer mesenchyme, cancer-associated fibroblasts (CAFs) are probably one of the most relevant cell types, but unfortunately also one of the least understood in terms of their origins, subtypes, biology and even definition. However, we are now developing the necessary understanding to help apply CAF biology to the treatment of patients with gastrointestinal cancer.

Studies investigating the function of CAFs are largely based on preclinical gastrointestinal cancer models. For instance, the role of CAFs in cancer restraint was first established in sophisticated studies using transgenic mouse models such as *Kras*<sup>LSL-G12D/+</sup>; *Trp53*<sup>LSL-R172H/+</sup>; *Pdx1-Cre* (KPC) mice, which recapitulate the desmoplastic features of human PDAC<sup>4,5</sup>, and then also later in studies using a colitis-associated colon cancer model induced by azoxymethane and dextran sulfate<sup>14,15</sup>. Contrary to previous studies which showed that CAFs promote tumour growth<sup>16,17</sup>, depletion of  $\alpha$ -smooth muscle actin ( $\alpha$ SMA)<sup>+</sup> CAFs in the mouse model of PDAC or blockade of hedgehog signalling, a key signalling pathway necessary for activation of CAFs, in the mouse models of PDAC and colon cancer accelerated cancer progression<sup>4,5,14,15</sup>. The concept that mesenchymal stem cells (MSCs) recruited from the bone marrow develop into CAFs which promote cancer progression was proposed on the basis of experiments using the mouse model of gastric carcinogenesis induced by *Helicobacter felis* infection<sup>18</sup>. Previous studies using subcutaneous injections of cancer cells lacked

the effects of the TME in which gastrointestinal CAFs co-evolve with tumour cells. However, the development of physiologically accurate autochthonous cancer models and orthotopic injection of genetically edited organoids is providing the necessary understanding to translate basic gastrointestinal CAF research into the clinic<sup>19,20</sup>.

Structurally and functionally, CAFs make a substantial contribution to the development of cancer through a variety of mechanisms. For instance, CAFs release various tumour-promoting factors such as cytokines and chemokines, which support cancer cell growth and angiogenesis<sup>16,21</sup>. Previous studies using genetically engineered mouse models (GEMMs) of pancreatic cancer have shown that CAFs, and ECM produced by CAFs, confer resistance to chemotherapy by impairing efficient drug delivery<sup>22-24</sup>. Furthermore, it has been demonstrated that CAFs could contribute to poor responses to immunotherapy in PDAC and colorectal cancer (CRC) mouse models<sup>19,25</sup>. Notably, CAFs are recruited to metastatic lesions at the nano-metastases stage<sup>26</sup> and also appear to create a favourable microenvironment for cancer growth at the secondary site<sup>10,27,28</sup>, suggesting that CAFs could be a potential target for the development of new therapeutics against human malignancies.

The molecular subtyping of gastrointestinal cancers has highlighted the clinical significance of stroma-related genes as prognostic and predictive markers. In many types of gastrointestinal cancer including CRC, PDAC and hepatocellular carcinoma (HCC), stromal activation gene signatures are associated with poor prognosis<sup>29-35</sup>. Strikingly, the stromal gene signature, rather than epithelial gene signature, was found to more closely inform outcome in patients with CRC<sup>32</sup>.

Notwithstanding the accumulating evidence showing the critical roles of CAFs in tumour progression, it has been challenging to therapeutically target CAFs, or at least the right CAFs or the right CAF-related factors. One example of this situation was the failure of a much-anticipated clinical trial of a hedgehog inhibitor<sup>36</sup>, which in combination with gemcitabine was initially shown to improve survival in preclinical mouse models of PDAC by improving drug. Hedgehog ligands, especially Sonic hedgehog (Shh), secreted by cancer cells were shown to play a central role in activation of CAFs, leading to increased desmoplasia and PDAC progression<sup>37</sup>. However, consistent with the failure of the hedgehog inhibitor in the PDAC clinical trial, a subsequent study using a mouse model of PDAC revealed that inhibition of the hedgehog pathway unexpectedly resulted in increased

PDAC progression with predominantly undifferentiated cancer cell histology, suggesting that some populations of CAFs activated by hedgehog inhibit cancer progression<sup>5,38</sup>. One of the critical challenges in targeting CAFs for cancer treatment is the functional heterogeneity of CAFs<sup>5,10,12</sup>. For many years, CAFs were considered to be a uniform entity that exerted a tumour-promoting effect as an ‘accomplice’ of cancer cells by secreting pro-tumourigenic factors<sup>16,17</sup>. Interestingly, however, a growing number of studies have demonstrated that certain populations of CAFs actually inhibit tumour growth<sup>4,5,39-42</sup>. Here, we offer a novel nomenclature for CAFs based on function: tumour-promoting CAFs (pCAF), tumour-retarding CAFs (rCAF) and neutral CAFs (nCAF) that neither promote or retard tumour progression.

In this Review, we summarize the current advances in our understanding of CAF origin and heterogeneity with a particular emphasis on local and recruited mesenchymal progenitor cells as one probable origin. We describe how CAFs can affect tumour progression from the viewpoint of stromal to epithelial interactions, tumour immunity, angiogenesis and ECM-remodelling, particularly focusing on gastrointestinal cancers such as gastric cancer, CRC, PDAC and HCC. In addition, the mechanism by which CAFs confer resistance to chemotherapy, immunotherapy and radiotherapy will be discussed. Lastly, we provide an overview of the clinical significance of gastrointestinal CAFs as biomarkers and therapeutic targets.

## **Definition of fibroblasts and CAFs**

Fibroblasts are spindle-shaped, non-epithelial and non-immune cells embedded in the ECM, which are easily propagated in adherent cell culture<sup>10,43</sup>. They are a major constituent of the stroma in gastrointestinal organs and, as in other tissues, they are highly organized. Throughout the gastrointestinal tract, a reticular network of stromal cells lies coincident with the epithelial basement membrane<sup>44</sup>. The subepithelial plexus, composed of reticular stromal cells, entirely surrounds the glandular axis from the stomach to the rectum<sup>45</sup>. This compartment is dynamic with a radial axis of proliferation and differentiation, analogous to the epithelium, developing from gremlin 1-expressing intestinal reticular stem cells<sup>46</sup>. These cells give rise to intestinal reticular cells<sup>46</sup>, probably

overlapping with FOXL1<sup>+</sup> subepithelial telocytes and GLI1<sup>+</sup> mesenchymal cells which constitute an essential mesenchymal niche to support the intestinal stem cells<sup>45,47</sup>. Beneath this highly compartmentalized population exist loose arrangements of fibroblasts within the lamina propria that interact with each other and deeper stromal elements including smooth muscle, vessels, nerves and inflammatory cells<sup>44,48,49</sup>. Functionally, fibroblasts are fundamental regulators of ECM synthesis, paracrine and juxtacrine signalling to nearby epithelium to regulate growth and differentiation, and they are also ready to respond to tissue injury, either in wounding or tumourigenesis<sup>10,50</sup>.

CAFs are generally accepted to be all of the fibroblasts found within and surrounding a cancer<sup>51</sup>. This group includes native, normal fibroblasts and activated, proliferating (Ki67<sup>+</sup>) or recruited fibroblasts in response to stimuli from cancer. These new CAFs could, in turn, have originated via a number of possible mechanisms that we discuss below. Despite the rapid evolution of immunophenotyping and subtyping of immune cells, there is no single, precise positive discriminator of CAFs<sup>10,51-53</sup>. This lack of understanding has led to different studies reporting on overlapping, incomplete or discrete populations of CAFs and the use of markers that label both CAFs and other cell populations. These difficulties have complicated interpretation of several studies, which will be discussed below.

## **Heterogeneity of CAFs**

### **Marker heterogeneity**

Representative CAF markers include, but are not limited to,  $\alpha$ SMA, the serine protease fibroblast activation protein (FAP; also known as prolyl endopeptidase FAP), fibroblast-specific protein 1 (FSP1; also known as S100A4), platelet-derived growth factor receptor  $\alpha$  (PDGFRA) and PDGFRB. Some functions of well-established CAF markers and several cell types in which they are expressed are briefly summarized in Table 1. One of the most well-established CAF markers,  $\alpha$ SMA, fails to distinguish all CAFs in the TME<sup>52,54</sup> and none of these CAF markers is specific to CAFs, as they are also expressed in other cell types and healthy tissues. For instance,  $\alpha$ SMA expression is observed in smooth muscle cells in the muscular layer of the gastrointestinal tract and in vascular smooth

muscle cells<sup>11,49</sup>. FAP<sup>+</sup> CD45<sup>+</sup> cells also correspond to a subset of tumour-associated macrophages<sup>55,56</sup> and FSP1 has been demonstrated to mark epithelial cells undergoing epithelial-to-mesenchymal transition (EMT)<sup>57-59</sup> and inflammatory macrophages, but not  $\alpha$ SMA<sup>+</sup> myofibroblasts, in liver fibrosis models<sup>60</sup>. Experimentally, using fluorescence-activated cell sorting, CAFs are isolated by their lack of expression of an epithelial marker (epithelial cell adhesion molecule, EpCAM), a haematopoietic cell marker (CD45) or an endothelial marker (CD31), and/or their expression of CAF markers such as FAP and PDGFRA<sup>25,32,54,61-64</sup>. Moreover, CAFs are a heterogeneous population on the basis of both markers and functions with a broad spectrum of different CAFs existing simultaneously in the cancer mesenchyme<sup>52,54,65</sup>, adding further complexity to CAF definitions. Notably, contradictory results as to whether CAFs promote or retard cancer progression can be obtained depending on the specific CAF markers used. Thus, future work is required to identify the right CAF marker(s) for the right therapy to make a major breakthrough in this area of study. Fundamentally, CAFs will be best understood and subtyped by biology and by function, with subgroups previously suggested including tumour-restraining CAFs, tumour-promoting CAFs, secretory CAFs, inflammatory CAFs and myofibroblastic CAFs<sup>10,54</sup>.

It is possible that CAFs identified by a single marker are composed of a range of distinct CAF subtypes that have functionally opposing roles in cancer progression. Accordingly, it will be necessary to subdivide CAFs by the combination of several marker proteins to help better and prospectively characterize their biology and thus their therapeutic relevance. Interestingly, single-cell RNA-sequencing analyses from human CRC samples revealed the presence of two major subtypes of CAFs<sup>30</sup>. On the basis of TGF $\beta$  pathway gene expression, CAFs in CRC could be divided into 'CAF-A', characterized by high expression of matrix metalloproteinase 2 (MMP2), decorin, Col1a2 and FAP, and 'CAF-B', which were characterized by high expression of myofibroblastic markers such as  $\alpha$ SMA, transgelin and PDGFA<sup>30</sup>.

## Functional heterogeneity

Studies suggest that CAFs are composed of various functionally heterogeneous subsets that either promote or restrain cancer growth<sup>10</sup>. Most previous studies have focused only on the pro-tumourigenic functions of CAFs, on the basis that co-culture or in vivo co-implantation of cancer cells with CAFs facilitated tumour growth<sup>16,17,62,66</sup>. For instance, CAFs co-injected into mice with human breast cancer cells promote tumour growth and angiogenesis more than normal fibroblasts, through secretion of CXC chemokine ligand 12 (CXCL12, also known as stromal cell-derived factor 1, SDF1)<sup>16</sup>. However, much of the previous work failed to address the functions of CAFs from the viewpoint of complicated TME interactions. In the past 5 years, the development of sophisticated GEMMs that spontaneously develop cancer has enabled CAFs to be fully incorporated into the complex TME interactions (FIG. 1) and has shed light on novel tumour-inhibiting roles of CAFs. For example, specific depletion of  $\alpha$ SMA<sup>+</sup> cells, including CAFs, led to the progression of PDAC in mice by inducing immunosuppression, implying that  $\alpha$ SMA<sup>+</sup> cells include a subset of rCAF<sup>s</sup> at least in this experimental model<sup>4</sup>. Although Hedgehog signalling was shown to promote PDAC progression in an initial short-term assessment<sup>22</sup>, subsequent analysis revealed that long-term genetic and pharmacological inhibition of the Hedgehog pathway and stromal desmoplasia unexpectedly accelerated PDAC growth<sup>5</sup>. The anti-tumourigenic role of the Hedgehog pathway was corroborated by the failure of clinical studies of Hedgehog inhibitors<sup>36,67</sup> and a further preclinical study that used three distinct mouse models of PDAC<sup>68</sup>. In agreement with these results, more recent work has shown that blockade of Hedgehog signalling accelerated cancer progression in colitis-associated colon cancer models<sup>14,15</sup>, further supporting the notion that a subset of CAFs marked by GLI1, a transcriptional factor involved in the Hedgehog signalling pathway, are a population of rCAF<sup>s</sup><sup>69</sup>.

CAF<sup>s</sup> are more than inert cells; they actively modulate their environment. Several CAF-derived proteins have been suggested to have tumour-inhibiting functions, but conflicting results have also been reported<sup>39,40,64,70,71</sup>. For instance, I $\kappa$ B kinase  $\beta$  (IKK $\beta$ )-mediated nuclear factor- $\kappa$ B (NF- $\kappa$ B) activation in CAF<sup>s</sup> is responsible for inducing tumour-promoting inflammation in a mouse model of

skin carcinogenesis<sup>70</sup>. Consistent with this finding, the genetic deletion of IKK $\beta$  in collagen type VI (ColVI)<sup>+</sup> fibroblasts resulted in reduced tumour growth and immune cell infiltration in a mouse model of colitis-associated cancer (CAC) via decreased IL-6 production by IKK $\beta$ -deficient CAFs<sup>64</sup>. However, genetic deletion of IKK $\beta$  in a larger population of collagen type I alpha 2 (Col1a2)<sup>+</sup> fibroblasts in a similar CAC model unexpectedly accelerated tumour growth through enhanced hepatocyte growth factor (HGF) secretion<sup>39</sup>.

Asporin, a CAF marker, has been suggested to promote the coordinated invasion of gastric cancer cells and CAFs through activation of RAC1<sup>71</sup>. Asporin in breast cancer CAFs, however, exerts a tumour-restraining effect by inhibiting the TGF $\beta$  pathway and EMT of cancer cells, and high expression of asporin in human breast cancer stroma is associated with better clinical outcome in patients with breast cancer<sup>40</sup>. Taken together, these conflicting results underline a broad spectrum of CAF functions with one molecule exerting pleiotropic effects in distinct CAF subpopulations. Thus, caution is warranted in generalizing CAF therapies, as context in terms of both native organ and tumour stage is probably critical.

Conversely, consistent evidence exists for the tumour-promoting function of FAP. Specific depletion of FAP<sup>+</sup> cells using transgenic ablation or targeting FAP<sup>+</sup> cells via chimeric antigen receptor (CAR) T-cells inhibited tumour growth in a mouse model of PDAC by enhancing antitumour immunity and reducing desmoplasia and vascular density<sup>25,72,73</sup>. Indeed, FAP-knockout impaired development of PDAC in KPC mice and subcutaneously injected colon cancer in mice<sup>74,75</sup>, suggesting that FAP has a tumour-promoting function. In humans, elevated expression of FAP in the stroma of CRC and PDAC has been shown to correlate with poor patient prognosis<sup>32,74</sup>. In this regard, FAP could be a candidate marker for pCAFs.

## Intratumoural heterogeneity

Analogous to phenotypic heterogeneity among cancer cells<sup>76</sup>, CAF phenotypes are not only different between tumours (intertumoural heterogeneity) but also within tumours (intratumoural heterogeneity)<sup>77</sup>. Notably, Ohlund and colleagues identified two spatially and functionally distinct

subtypes of CAFs in human and mouse PDAC;  $\alpha$ SMA<sup>high</sup> IL-6<sup>low</sup> “myofibroblastic CAFs” which are marked by expression of myofibroblast genes and TGF $\beta$ -responsive genes such as Acta2, Ctgf and Col1a1, and are located adjacent to cancer cells, and  $\alpha$ SMA<sup>low</sup> IL-6<sup>high</sup> “inflammatory CAFs” which secrete inflammatory cytokines and chemokines such as IL-6, IL-11, CXCL1 and LIF, and are located distantly from cancer cells<sup>54</sup>. Subsequent study has revealed that TGF $\beta$  signalling and IL-1/JAK/STAT signalling are responsible for inducing differentiation of pancreatic stellate cells (PSCs), a precursor of CAFs into myofibroblastic CAFs and inflammatory CAFs, respectively<sup>78</sup>.

Currently, the heterogeneity of CAFs and their multifaceted roles remain to be fully elucidated. Further studies using single-cell RNA sequencing, translatable *in vivo* cancer models, discrete transgenic targeting and new stromal reagents, such as specific CAR T-cell approaches, will provide novel insights into these different types of CAF heterogeneity.

## Origin of CAFs

Although studies have begun to illustrate the heterogeneous nature of CAFs, little is known about the origins of CAFs. Different pathways probably exist for development of different CAF subpopulations. Fundamentally, cancer develops within an initially normal organ<sup>8</sup>. Depending on the stage of tumorigenesis, there will, at least in very early stages, exist some remnant native fibroblasts<sup>10,11,65</sup>. Thereafter, these cells are increasingly replaced by new CAFs that are different from native fibroblasts within normal tissue<sup>10,11,65</sup>. These new CAFs arise through one of several of the following processes (FIG. 2): transdifferentiation, in which CAFs can develop from a non-fibroblastic lineage such as epithelial cells<sup>58,59</sup>, blood vessels<sup>79</sup> or serosa<sup>80-82</sup>, with gene-expression and biology changed to adopt a fibroblastic phenotype (FIG. 2a); activation, in which altered gene expression and phenotype is induced within the existing resident fibroblasts in response to the TME (FIG. 2b, BOX1); recruitment, in which cells can arise from remote circulating populations, most often suggested to be bone marrow mesenchymal stem cells (MSCs) (FIG. 2c)<sup>18</sup>; and finally differentiation, in which CAFs might arise in a typical stem cell-progenitor cell hierarchy, as has been shown to occur within the periepithelial mesenchymal sheath of the mouse intestine (FIG. 2d)<sup>46</sup>.

Elucidating the contribution of these four routes is vital for understanding the therapeutic challenges and opportunities to influence mesenchymal remodelling in cancer. The reality is that these pathways are not mutually exclusive and all might be operating in the development of CAFs. The key consideration is whether CAF ontogeny informs CAF biology, a question that requires further study.

Numerous publications show that some CAFs originate from local fibroblasts, others from bone marrow MSCs and, depending on how they were experimentally defined, pericytes<sup>10,49,83</sup>. Interestingly, Arina *et al.* have shown using bone marrow transplantation, parabiosis and skin graft models, that Col1a1<sup>+</sup> and  $\alpha$ SMA<sup>+</sup> CAFs predominantly derive from local precursors and not bone marrow precursors<sup>84</sup>. Using traditional, transgenic lineage tracing experiments, others have shown that epithelial cells and endothelial cells differentiate to CAFs through EMT and endothelial-to-mesenchymal transition, respectively<sup>58,59,79</sup>.

Studies using bone marrow transplantation have shown that MSCs (BOX 2) have a remarkable feature called 'tumour-specific tropism', in which they actively migrate to tumour sites<sup>18,85,86</sup>. In fact, bone marrow transplantation of  $\alpha$ SMA-reporter MSCs revealed that at least 20% of CAFs arise from bone marrow-derived MSCs in a mouse model of inflammation-driven gastric carcinoma<sup>18</sup>. Notably, the presence of bone marrow-derived cells in tumour mesenchyme is confirmed in human gastric adenocarcinoma and rectal adenoma in patients who developed tumours following bone marrow transplantation<sup>86</sup>. The recruitment of MSCs to the TME is dependent on CXCL12, CXCL16, TGF $\beta$ , CC chemokine ligand 2 (CCL2) and CCL5 secreted by cancer cells, inflammatory cells and CAFs<sup>18,87,88</sup>. In response to the soluble factors secreted from the TME, these recruited MSCs are converted to pCAF<sup>s</sup> expressing high levels of IL-6, WNT5A, bone morphogenetic protein 4 (BMP4) and CCL5<sup>18,87,89</sup>. It is also conceivable that specific immature CAF subpopulations arising from MSCs could potentially exert a tumour-inhibitory effect. Indeed, it was indicated that high stromal expression of CD271 (also known as nerve growth factor receptor, NGFR), a human bone marrow MSC marker, predicts a favourable prognosis in human PDAC<sup>90</sup>. In the foreseeable future, lineage tracing of MSC markers in the development of cancer will help elucidate the cellular origin and evolution of CAFs and could identify MSC or CAF markers of therapeutic value.

## Function of CAFs

In this section, we discuss how CAFs functionally modulate cancer progression through interaction with other compartments in the TME.

### Stromal and epithelial interactions

CAFs directly confer growth advantages to cancer cells via paracrine signalling, exosome transfer and physical interaction<sup>10,49,91</sup> (FIG. 3).

**Paracrine signalling.** Chemokines, cytokines and growth factors secreted by CAFs, such as CXCL12, HGF, epithelial growth factor (EGF), insulin-like growth factor (IGF), IL-6, IL-8 and IL-11 have an essential role in stimulating epithelial cell growth and maintaining cancer stem cells (CSCs) (FIG. 3a)<sup>16,92-104</sup>. For instance, CXCL12 produced by activated CAFs enhances breast cancer growth by binding to its cognate receptor, CXCR4, which is expressed in cancer cells<sup>16</sup>. Activation of the HGF-receptor MET induces cancer stemness and chemoresistance in models of HCC and colon cancer<sup>92-94</sup> and upregulates keratin 19 expression in HCC, which is a predictor of poor patient survival<sup>95</sup>. It has been demonstrated that hepatic myofibroblasts and colon cancer CAFs secrete EGF family proteins and promote cancer progression through activation of ERBB receptors, including EGFR<sup>96,97</sup>. IGF2 secreted by CAFs maintains the stemness of cancer cells<sup>98,99</sup>. Additionally, IL-6 and IL-11 increase cancer cell proliferation and liver metastases, respectively, in models of colon cancer by augmenting STAT3 signaling<sup>49,64,103</sup>. Notably, CD10<sup>+</sup> GPR77<sup>+</sup> CAFs were identified as a novel subset of stemness-sustaining CAFs that provide IL-6 and IL-8 to maintain CSCs and promote chemoresistance<sup>104</sup>.

The WNT pathway and BMP signalling have a crucial role in controlling intestinal stem cell fate in health and cancer<sup>50</sup>. HGF expressed by CAFs plays a vital role in maintaining colon CSCs by augmenting WNT signalling<sup>93</sup>. The expression of a ligand-sequestering BMP antagonist, gremlin 1, distinguishes intestinal reticular stem cells that give rise to the periepithelial mesenchymal sheath<sup>46</sup>

and is upregulated in CAFs of human gastrointestinal cancers including oesophageal cancer, PDAC and CRC<sup>105</sup>. Interestingly, disruption of BMP morphogen gradients by aberrant epithelial gremlin 1 expression induces ectopic crypt formation and progressive intestinal polyps in a transgenic mouse model<sup>106</sup>. Furthermore, gremlin 1-knockout ameliorated tumourigenesis in the mutant APC mouse model of intestinal cancer<sup>106</sup>.

**Exosome transfer.** Several studies offer evidence that bidirectional communication between CAFs and cancer cells is mediated in part by exosomes<sup>10,107</sup>. For instance, TGF $\beta$ <sup>+</sup> exosomes released by gastric cancer cells can convert MSCs to  $\alpha$ SMA<sup>+</sup> activated CAFs<sup>108</sup>. Intriguingly, it has been delineated that similar secretion of exosomal microRNA-1247-3p by metastatic HCC cells results in activation of fibroblasts, which in turn secrete inflammatory cytokines such as IL-6 and IL-8 and foster lung metastasis of HCC<sup>109</sup>. Conversely, exosome transfer from CAFs to cancer cells confers a survival advantage to the cancer cells<sup>110-112</sup>. In this regard, *in vitro* experiments have shown that gemcitabine treatment increases the release of exosomes containing microRNA-146a and Snail from human PDAC CAFs, leading to increased cancer cell proliferation and chemoresistance<sup>112</sup>.

Given that the classic concept that soluble factors secreted by CAFs promote cancer cell proliferation has already been well established, further research should concentrate on identification of candidate molecules that can be taken advantage of for cancer treatment. In view of the heterogeneity of CAFs, it is also vital to identify the CAF subpopulations that produce each soluble factor.

**Physical interaction.** In addition to the aforementioned biochemical crosstalk, direct physical interactions between CAFs and cancer cells play a critical part in cancer cell migration. Co-culture of CAFs and skin cancer cells has shown that leading CAFs can generate a 'track' by ECM-remodelling to facilitate the collective migration of cancer cells behind the CAFs<sup>66,113</sup>. CAFs can also directly lead the collective invasion of cancer cells by generating 'pulling forces' on cancer cells

through a mechanism mediated by N-cadherin and E-cadherin in the CAFs and cancer cells, respectively (FIG. 3a)<sup>91</sup>.

## Tumour immunology

CAFs are major contributors to the immunosuppressive TME<sup>10</sup>, which might act as a restitution programme to help support epithelium in acute injury, but promotes cancer growth in a tumour setting<sup>10,114</sup>. Mechanistically, cytokines and chemokines such as IL-6 and CCL2 produced by CAFs directly recruit immune cells and modulate both innate and adaptive immune systems<sup>10</sup>. Furthermore, CAFs impede trafficking of T cells indirectly by remodelling ECM, thereby suppressing antitumour immunity<sup>115</sup>. Several studies suggest that CAFs are one of the mediators of response to immune checkpoint inhibitors<sup>12,19,25,61</sup>.

CAFs and tumour-associated MSCs produce chemokines such as CXCL1, CCL2, CCL5 and CXCL12, and induce recruitment of polymorphonuclear myeloid-derived suppressor cells (MDSCs), monocytic MDSCs and regulatory T cells (T<sub>reg</sub> cells), all of which restrain tumour immunity and promote tumour progression (FIG. 3b)<sup>30,61,85,116,117</sup>. The expression of the granulocytic chemokine, CXCL1, by CAFs is negatively regulated by crosstalk between cancer cells and CAFs through colony-stimulating factor-1 (CSF1)–CSF1 receptor signaling<sup>30</sup>. CSF1 receptor inhibitors have been used as anti-tumour-associated macrophage agents in clinical trials for solid tumours including gastrointestinal cancers, however, they have shown limited response<sup>30,118</sup>. This outcome was attributed to the ability of CAFs to neutralize the therapeutic effect of the agents by recruiting polymorphonuclear MDSCs to tumour sites and shaping an immunosuppressive TME<sup>30</sup>.

Some CAFs appear to preferentially recruit CD4<sup>+</sup>CD25<sup>+</sup> T cells by secreting CCL5 and CXCL12, and increase their differentiation to tumour-promoting CD25<sup>high</sup> FOXP3<sup>high</sup> T<sub>reg</sub> cells by CXCL12<sup>61,117</sup>, a finding consistent with the histological observation that  $\alpha$ SMA<sup>+</sup> CAFs are in close proximity to FOXP3<sup>+</sup> T<sub>reg</sub> cells<sup>117</sup>. Another example of the immunosuppressive roles of CAFs in immune cell recruitment is that activated pancreatic stellate cells (PSCs which are equivalent to activated CAFs in PDAC sequester antitumour CD8<sup>+</sup> cytotoxic T cells and prevent their migration to pancreatic

cancer cells<sup>119</sup>. However, this finding was challenged by a computational analysis of multiplex immunohistochemistry using human PDAC samples, which demonstrated that high or low cytotoxic T cell infiltration around the tumour cells was not associated with  $\alpha$ SMA and Col1 expression levels<sup>120</sup>. These seemingly contradictory observations might be explained in part by the heterogeneity of CAFs.

In addition to recruiting immune cells, CAFs also modulate the immunosuppressive properties of these cells. For example, CAFs and colonic myofibroblasts both express programmed cell death ligand 1 (PDL1), an immune checkpoint molecule that plays an essential role in inhibiting activation and proliferation of T cells through binding to programmed cell death protein 1 (PD1) on T cells<sup>114,121,122</sup>. Moreover, in vitro experiments indicate that CAFs induce an immunosuppressive TME by secreting CXCL12, IL-6 and IL-8 and promoting M2 polarization of macrophages (Fig. 3b)<sup>123,124</sup>. In the liver, granulins secreted by metastasis-associated macrophages convert quiescent hepatic stellate cells to periostin<sup>+</sup> pCAF, resulting in increased metastatic tumour burden in a PDAC model<sup>125</sup>. Notably, CAFs are also one of the major producers of the immunosuppressive cytokine TGF $\beta$  in the cancer mesenchyme<sup>19,114</sup>, and multiple studies have underscored the significance of TGF $\beta$  in shaping the immunosuppressive TME (FIG. 3b)<sup>19,126,127</sup>. The development of a novel mouse model that recapitulates the scarce T-cell infiltration of human microsatellite-stable (MSS) CRC has revealed that TGF $\beta$  inhibits the type 1 T helper cell (T<sub>H</sub>1)-effector phenotype, thereby limiting benefit from anti-PD1–PDL1 therapy<sup>19</sup>. Notably, combination therapy with the TGF $\beta$  receptor inhibitor (galunisertib) and an anti-PDL1 antibody unleashed the cytotoxic immune response and eradicated most liver metastases in this model<sup>19</sup>. The combination therapy with galunisertib and an anti-PDL1 antibody is being assessed in a phase 1b clinical trial for metastatic pancreatic cancer<sup>128</sup>.

Among a variety of CAF subtypes, the immunosuppressive role of FAP<sup>+</sup> CAFs has been investigated by multiple groups<sup>25,61,73,116</sup>. FAP<sup>+</sup> CAFs induce immunosuppression by secreting CXCL12<sup>25,61</sup>. Specific depletion of FAP<sup>+</sup> cells in a mouse model of PDAC resulted in enhanced antitumour immunity, and the combination of the FAP<sup>+</sup> CAF depletion and immune checkpoint inhibitors (anti-cytotoxic T lymphocyte-associated protein 4 (anti-CTLA-4) or anti-PDL1) exerted a synergistic effect in reducing tumour volume<sup>25</sup>. Consistent with this finding, Costa *et al.*<sup>61</sup> have

identified four different subsets of human breast cancer CAFs by fluorescent-activated cell sorting and found that the CAF-S1 subset, characterized by high FAP expression, is responsible for generating an immunosuppressive TME by recruiting CD4<sup>+</sup>CD25<sup>+</sup> T cells and promoting their differentiation to T<sub>reg</sub> cells.

Although there has been growing interest in cancer immunology, especially in regards to immune checkpoint inhibitors, we are only beginning to understand how CAFs participate in tumour immunosurveillance. Further CAF research is required to identify promising target molecules or CAF subpopulations and develop novel therapeutics that can improve clinical responses to current immunotherapies.

## Angiogenesis

Neovascularization in cancer is regulated not only by tumour cells but also by stromal cells<sup>129</sup>. Indeed, CAFs promote tumour angiogenesis directly by secreting pro-angiogenic factors (FIG. 3c) and indirectly by producing ECM<sup>10,129</sup>. Besides cancer cells, CAFs are a major source of vascular endothelial growth factor A (VEGFA), the most potent pro-angiogenic factor that promotes angiogenesis by acting on its cognate receptor, VEGF receptor 2 (VEGFR2), expressed on endothelial cells<sup>129,130</sup>. Moreover, CAFs induce angiogenesis by secreting several pro-angiogenic factors such as CXCL12, fibroblast growth factor 2 (FGF2) and PDGFC<sup>16,131-133</sup>. In turn, the leaky vasculature in tumours results in platelet extravasation and the subsequent degranulation of pro-angiogenic factors such as PDGF and TGFβ, which in turn activate fibroblasts<sup>129,134</sup>. In addition to paracrine signalling, CAFs also contribute to angiogenesis indirectly via remodelling ECM proteins such as periostin, tenascins, fibronectin, osteopontin and collagens<sup>10,129,135</sup>.

Importantly, the crosstalk between CAFs and endothelial cells confers resistance to anti-VEGF therapy and chemotherapy<sup>133,136,137</sup>. PDGFC and FGF2 secreted by CAFs and bone marrow-derived Col1a1<sup>+</sup> CXCR4<sup>+</sup> fibrocyte-like cells, respectively, are crucial mediators in the acquisition of resistance to anti-VEGF therapy in several mouse models of solid tumours<sup>133,136</sup>. Microfibrillar-associated protein 5 (MFAP5) has also been described as a novel CAF-derived pro-angiogenic

factor that upregulates lipoma-preferred partner (LPP) in endothelial cells and confers resistance to chemotherapy by increasing microvessel leakiness<sup>137</sup>. Collectively, these studies provide a rationale for targeting CAFs as a potential therapeutic strategy to alter tumour vasculature and improve drug delivery to tumour cells.

## ECM-remodelling

Activated CAFs are the main producers of ECM constituents such as collagen, fibronectin, proteoglycans, periostin and tenascin C, and the ECM is degraded primarily by CAF-derived MMPs (FIG. 3d)<sup>10,138</sup>. CAFs can enhance tumour progression indirectly by generating a mechanically stiff ECM<sup>139</sup>. A considerable amount of the literature suggests that ECM stiffness plays a central part in cancer progression<sup>13,138,140,141</sup>. Collagen crosslinking and increased ECM stiffness promote cancer cell proliferation, EMT, metastasis and resistance to chemotherapy<sup>142-144</sup>. Additionally, increased ECM stiffness leads to the generation of a dysregulated, leaky vasculature<sup>145</sup>, and a dense ECM impedes migration of T cells to cancer cells<sup>115</sup>. One study has revealed that SPOCK1, a member of the SPARC family, predominantly expressed in the PDAC stroma, facilitates invasive pancreatic cancer cell growth by modifying collagen fibre patterns<sup>146</sup>. The ECM-remodelling mediated by YAP activation in CAFs, and collagen crosslinking by the lysyl oxidase (LOX) family, substantially contributes to tissue stiffness<sup>139,142,147,148</sup>. In turn, matrix stiffness elevates YAP activity in CAFs, resulting in activation of CAFs and further matrix stiffening<sup>139,147</sup>. Notably, ECM stiffness is also crucial for inducing differentiation of MSCs to CAFs, which then support cancer cell proliferation<sup>149</sup>. Importantly, targeting the ECM has shown to improve the effectiveness of standard chemotherapy in KPC mice<sup>23,24,150</sup>. For instance, enzymatically depleting hyaluronic acid, an ECM component, in combination with gemcitabine, significantly prolongs the median overall survival of KPC mice compared to gemcitabine monotherapy (91.5 versus 55.5 days)<sup>23</sup> through improved drug delivery<sup>23,24</sup>. Combining a LOX-neutralizing antibody with gemcitabine was shown to reduce fibrillar collagen and increase tumour-free survival in KPC mice with early-stage tumours<sup>150</sup>. Further

development of methods to manipulate the ECM and tissue stiffness will probably improve the therapeutic response to conventional chemotherapy.

## **CAFs and therapeutic resistance**

Despite advances in chemotherapy, molecularly targeted drugs and immunotherapy, these treatments offer survival benefits only to a small group of patients<sup>114,151</sup>. Emerging evidence has demonstrated that CAFs confer substantial resistance to cancer therapeutics via impaired drug delivery and biochemical signalling<sup>10,87</sup>. The ECM produced by CAFs acts as a physical barrier to prevent the penetration of drugs by increasing interstitial fluid pressures and inducing vascular collapse<sup>23</sup>. One study using mass spectrometry revealed that pancreatic cancer CAFs entrap an active gemcitabine metabolite<sup>152</sup>, demonstrating another mechanism to impair drug delivery to cancer cells. Important CAF-derived soluble factors that mediate resistance to chemotherapy include IL-6, IL-17A, IGF1, IGF2, nitric oxide and platinum-induced polyunsaturated fatty acids<sup>153-157</sup>. In regards to molecularly targeted drugs, HGF and IGF2 released by CAFs contribute to the resistance to tyrosine kinase inhibitors<sup>94,99</sup>. Indeed, dual inhibition of EGFR and HGF receptor, or insulin and IGF1 receptors (IR/IGF1R) that mediate the IGF2/IR/IGF1R signalling axis, enhanced the therapeutic response to an EGFR inhibitor in xenograft models of colon cancer and cholangiocarcinoma, respectively<sup>94,99</sup>. Interestingly, Wang and colleagues found that CAF-released cysteine and glutathione lead to reduced intracellular cisplatin content in ovarian cancer cells, conferring resistance to platinum-based chemotherapy<sup>158</sup>. Furthermore, CAFs influence the responsiveness to immune checkpoint inhibitors by shaping the immunosuppressive TME, as discussed above. Remarkably, a high “compound stromal score” defined by three stromal components (CAFs, leukocytes and endothelial cells) can predict resistance to radiotherapy in patients with rectal cancer<sup>35</sup>. In oesophageal squamous cell carcinoma, the expression of long non-coding RNA DNMT3OS in cancer cells is increased by CAFs through PDGF $\beta$ /PDGFR $\beta$ /FOXO1 signalling, leading to radioresistance by regulating DNA damage response<sup>159</sup>. The radioresistance is

attributed, in part, to an altered interaction between cancer cells and CAFs following radiation<sup>160</sup>. Indeed, radiation induces CRC CAFs to secrete IGF1, thereby supporting cancer cell growth<sup>161</sup>.

## Clinical implementation

### Biomarkers

In addition to the functional contribution of CAFs to cancer progression described above, CAFs and their gene expression patterns have diagnostic and prognostic value in clinical oncology<sup>162</sup>. Surprisingly, the presence of circulating FAP<sup>+</sup> CAFs was confirmed in peripheral blood of patients with cancer, including those with metastatic CRC<sup>163</sup>. Furthermore, increased levels of circulating stroma-related molecules such as MMP7 and connective tissue growth factor (CTGF) can help discriminate patients with PDAC from healthy individuals or patients with chronic pancreatitis when combined with CA19.9, a commonly used PDAC biomarker<sup>164</sup>. Gene expression analyses and proteome profiling of cancer tissues have revealed that stromal gene signatures predict poor patient outcome in multiple types of gastrointestinal cancer<sup>29-35,165,166</sup>. In particular, it has been shown that elevated stromal expression of TGF $\beta$ -related genes is associated with poor prognosis in CRC<sup>32,103</sup>. In support of this finding, histological observations revealed that high expression of  $\alpha$ SMA or high stromal proportion are predictive of poor clinical outcome in patients with CRC, PDAC and HCC<sup>167-170</sup>. As CAFs accumulate at the tumour site at an early stage of tumourigenesis<sup>10</sup>, future investigations will probably identify valuable CAF markers that might facilitate early detection of cancer.

### CAF-targeting therapy

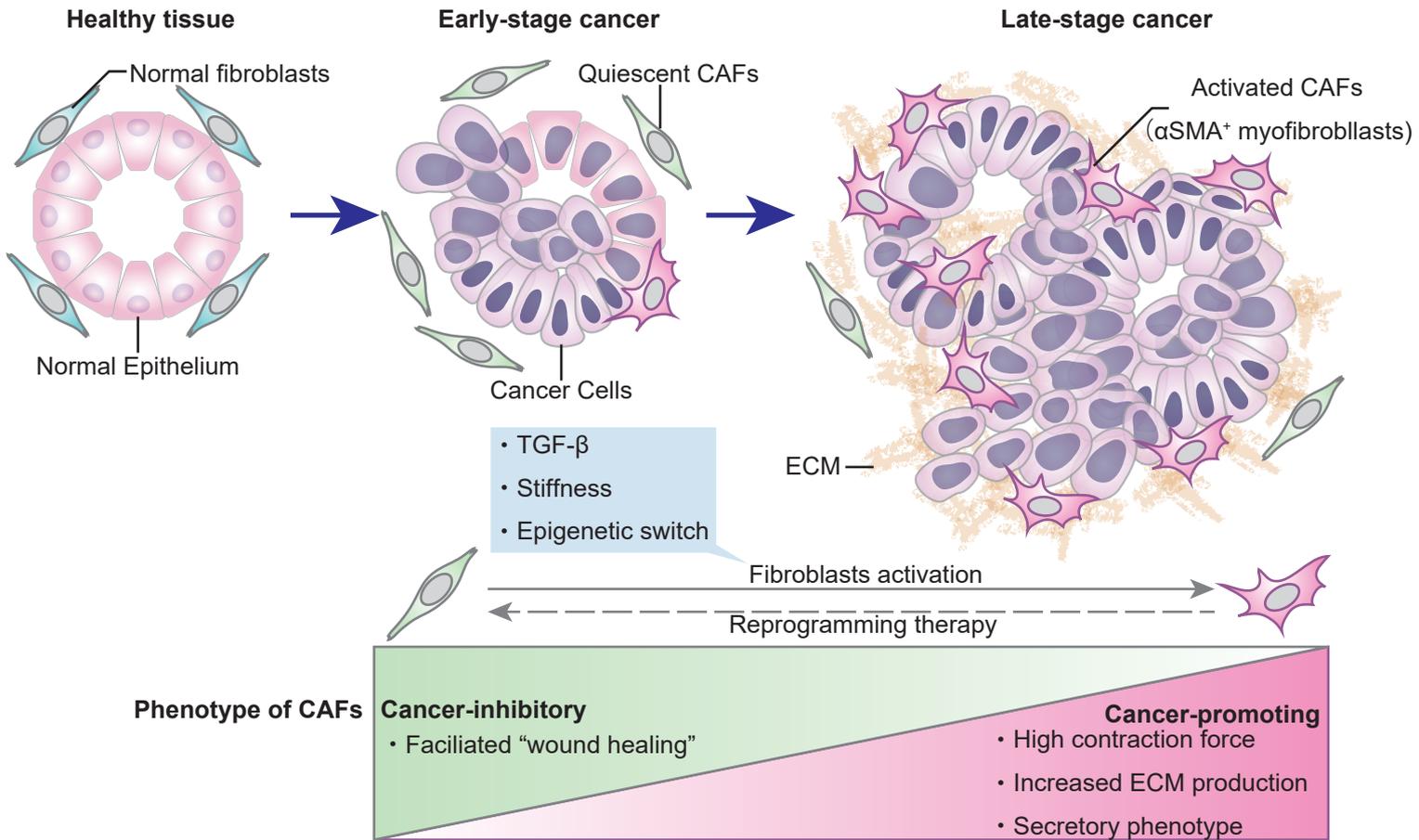
In the past 5 years, there has been considerable interest in therapeutic strategies to target CAFs, and numerous clinical trials for gastrointestinal cancers are ongoing to assess their benefits<sup>13,49</sup>. Elimination of FAP<sup>+</sup> pCAF by CAR T-cells or vaccination has been shown to inhibit tumour progression in several different mouse models of cancers including PDAC and CRC<sup>72,171</sup> (FIG. 4).

CAF reprogramming by vitamin D and vitamin A, which revert pCAFs to rCAFs, has attracted much attention in the field of PDAC and colon cancer<sup>172-175</sup>. Administration of a vitamin D analogue inhibits tumour-promoting signalling in activated PSCs, resulting in substantially improved therapeutic efficacy of gemcitabine in KPC mice<sup>172</sup>. A phase 2 clinical trial of concomitant treatment with a PD1 inhibitor and a vitamin D analogue in PDAC is now underway<sup>176</sup>. FAP has been used in a xenograft model of PDAC and clinical trials for solid tumours to target drug delivery specifically to tumour sites using antibody-drug conjugates or immunocytokines, in which, for example, a cytotoxic drug (maytansinoid DM1, a tubulin inhibitor) or an IL-2 variant is conjugated with FAP antibodies<sup>177-179</sup>. A highly anticipated approach, which utilizes the tumour-specific tropism of MSCs, is the administration of MSCs that are engineered to express enzymes that metabolize pro-drugs to active drugs or to secrete tumour-inhibitory molecules such as TRAIL and IFN $\alpha$ <sup>87,180-183</sup>. Acceptable safety and tolerability have been reported in the first phase 1 clinical trial for gastrointestinal cancers employing autologous MSCs genetically engineered to express herpes simplex virus-thymidine kinase, which converts the prodrug ganciclovir into its active cytotoxic metabolite<sup>184</sup>. Investigators are also using TGF $\beta$  inhibitors or Hedgehog inhibitors in combination with standard chemotherapies or immunotherapies in an attempt to block pro-tumourigenic signalling relevant to CAFs in gastrointestinal cancers<sup>13,19,49</sup>.

## Conclusions

CAFs are important in the development of gastrointestinal cancers, both in their promotion and, as we increasingly appreciate, in their antagonism. CAFs are not one entity but rather contain heterogeneous functional subpopulations including pCAFs, rCAFs, and probably also a neutral subset that neither promotes or retards (nCAFs). CAF biology is mediated through their direct and paracrine interactions with both cellular (tumour cells, immune cells and vascular cells) and acellular (ECM) compartments. Despite the importance of CAFs, we are still in the infancy of CAF-directed approaches to cancer care. To help accelerate the integration of CAF science into CAF clinical care, we encourage future work in this field to precisely define the CAF population being studied through

careful characterization including, where appropriate, immunophenotyping, multiplex immunofluorescence, discrete transgenic markers, single-cell transcriptional analysis, and to combine these characteristics with CAF biology. We offer the biological nomenclature of pCAF, rCAF and nCAF, but expect these terms, as in the case of lymphocyte and myeloid cellular classifications, to be replaced by a precise, biologically rooted and clinically translatable immunophenotypic classification. Furthermore, future studies should carefully consider the tumour context of their experimental models. We speculate that CAFs have considerable plasticity in terms of their function and marker expression, analogous to other important cell populations including CSCs<sup>76</sup>. The interconversion of CSCs and the inherent difficulties in reproducing specific CSC markers has hindered progress in this area of cancer research<sup>76</sup>. We are encountering similar obstacles in CAF research. Future genetic fate-mapping of CAFs, more widely accepted CAF subclassifications and a better understanding of the context-specific behaviour of CAFs will offer novel insights into the heterogeneity, hierarchy and plasticity of CAFs. These scientific discoveries will help drive the more rapid translation of CAF basic science into CAF clinical practice and the development of new diagnostics, prognostics, preventatives and therapeutics for patients with gastrointestinal cancer.

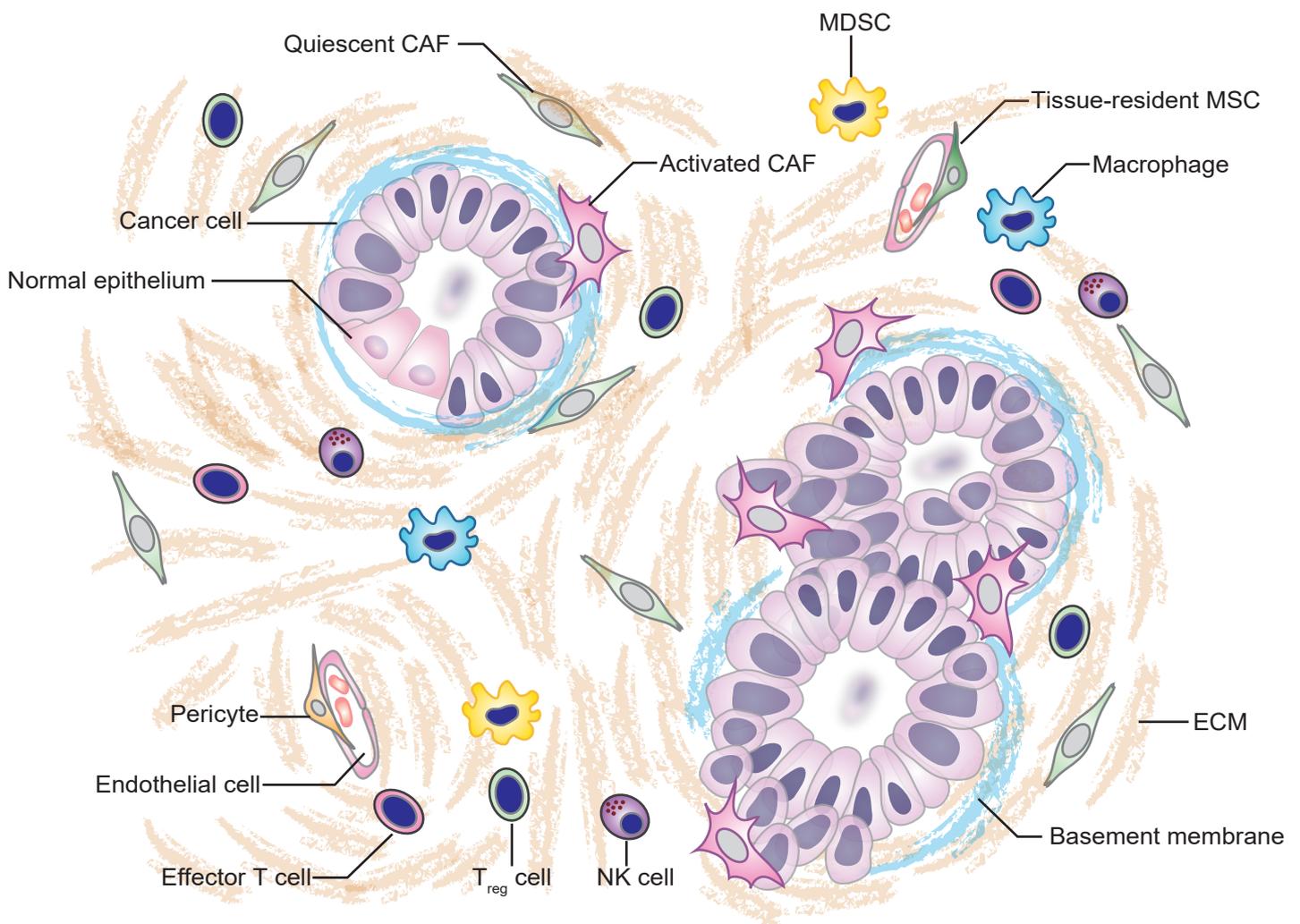


### **BOX1: Fibroblast activation**

Tissue-resident fibroblasts in health are activated in response to a plethora of stimuli from tumour microenvironment (TME). Transforming growth factor- $\beta$  (TGF $\beta$ ), platelet-derived growth factor (PDGF), Sonic Hedgehog (Shh), bone morphogenetic protein (BMP), IL-1, IL-6, TNF and reactive oxygen species (ROS) are important biochemical activators of fibroblasts<sup>10,70,185,186</sup>. Notably, Kras<sup>G12D</sup> mutation, one of the most common driver mutations in PDAC induces Shh secretion by pancreatic cancer cells, thereby activating pancreatic stellate cells (PSCs), fibroblast-like cells in the pancreas, and supporting cancer cell growth<sup>186</sup>. The coevolution of cancer-associated fibroblasts (CAFs) with cancer cells seems to be analogous to the developmental biological crosstalk between mesenchymal cells and epithelial cells. Indeed, soluble factors necessary for CAF activation, such as Hedgehog and BMP, have a critical role in defining the epithelial and stromal niche in developmental gastrointestinal organs<sup>48,69,187</sup>. Mechanical stiffness generates a positive feedback loop for CAF activation<sup>139</sup>. Activated CAFs express  $\alpha$ -smooth muscle actin ( $\alpha$ SMA), a cytoskeleton molecule required for cell contraction, and are therefore regarded as myofibroblasts<sup>10,51</sup>. Activated CAFs acquire a highly contractile, ECM-synthesizing, proliferative and secretory phenotype and are epigenetically and metabolically distinct from quiescent fibroblasts<sup>10,188,189</sup>. Functionally, fibroblasts initially facilitate wound healing, a biological response beneficial for tissue regeneration<sup>10</sup>. However, chronic activation of fibroblasts leads to organ fibrosis<sup>10</sup>, which is deleterious for maintaining organ function, as exemplified by liver cirrhosis. In this regard, perpetually activated CAFs might play a key part in fostering cancer progression. In contrast to the mechanism by which activated CAFs promote tumour progression, how quiescent CAFs restrain the development of cancer is largely unknown. Importantly, CAFs with different degrees of activation (for example, quiescent CAFs and activated CAFs) exist in the TME, contributing to the heterogeneity of CAFs<sup>49,52</sup>.

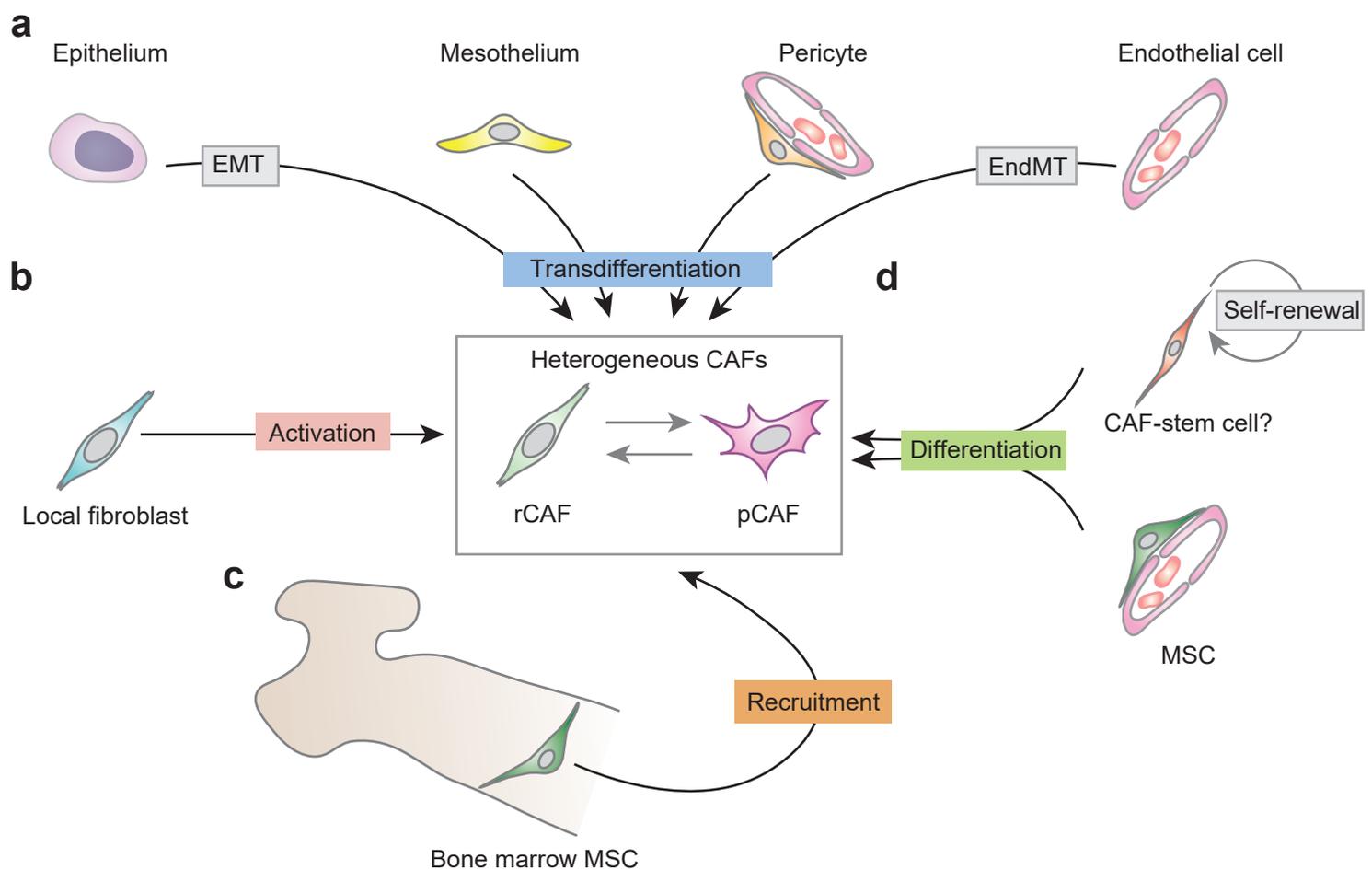
## **BOX2: Mesenchymal stem cells and their markers**

Mesenchymal stem cells (MSCs) have stellate morphology and represent a rare subset of stromal cells, which are localized mainly in the bone marrow, adipose tissue and umbilical cord<sup>10,190</sup>. Importantly, MSCs or MSC-like cells are also observed in the perivascular regions of many organs, implicating their potential overlap with pericytes classically labelled by neuron-glial antigen 2 (NG2) and platelet-derived growth factor receptor  $\beta$  (PDGFR $\beta$ )<sup>191,192</sup>. MSCs play a vital part in the maintenance of haematopoietic stem cells (HSCs) in the bone marrow, and interest has been growing in how MSCs maintain epithelial stem cell niches in other tissues<sup>193</sup>. MSCs are developmentally derived mainly from embryonic mesoderm<sup>190</sup> and demonstrate a self-renewal capability and a capacity to differentiate into osteocytes, adipocytes and chondrocytes<sup>194</sup>. Numerous studies have shown that MSCs give rise to  $\alpha$ -smooth muscle actin ( $\alpha$ SMA)<sup>+</sup> myofibroblasts in fibrosis in various organs such as the liver, heart and lung<sup>190,195,196</sup>. Notably, it has been shown that administration of MSCs is a promising therapeutic strategy for acute graft-versus-host disease, intestinal ulcers and inflammatory bowel diseases, owing to the immunomodulatory and pro-angiogenic property of MSCs<sup>197,198</sup>.



**Figure 1: Cellular components of the tumour microenvironment.**

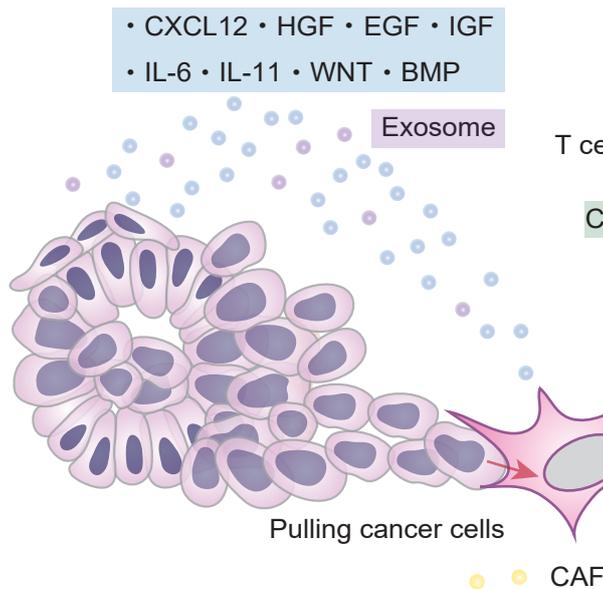
Fibroblasts are a vital component of the tumour microenvironment (TME). The TME is comprised of cancer-associated fibroblasts (CAFs), cancer cells, normal epithelial cells, endothelial cells, pericytes, mesenchymal stem cells (MSCs), the extracellular matrix (ECM), basement membrane and inflammatory cells such as T cells, natural killer (NK) cells, macrophages and myeloid-derived suppressor cells (MDSCs). CAFs are a highly heterogeneous population and include quiescent CAFs and activated CAFs.



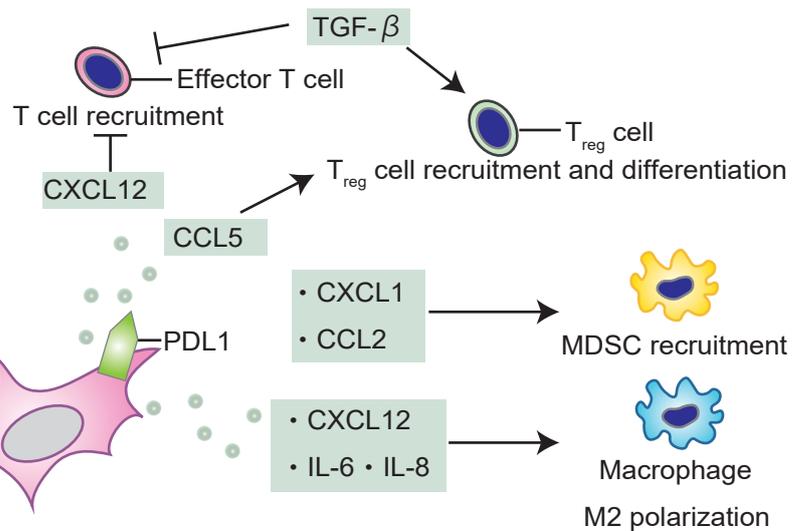
**Figure 2: The origins of CAFs.**

Cancer-associated fibroblasts (CAFs) are considered to arise through four non-mutually exclusive mechanisms. Non-fibroblast lineage cells such as epithelial cells and endothelial cells become a part of the CAF population through transdifferentiation (a). Local fibroblasts acquire CAF phenotypes via activation (BOX1) that are distinct from a normal fibroblast phenotype (b). Mesenchymal precursor cells, typically bone marrow mesenchymal stem cells (MSCs), are recruited to the tumour by cytokines and chemokines secreted from the tumour microenvironment, such as transforming growth factor- $\beta$  (TGF $\beta$ ) and CXC chemokine ligand 12 (CXCL12) (c)<sup>10,49</sup>. It is also conceivable that a CAF-stem cell exists (d). In this scenario, a minor subpopulation of CAF progenitors in a hierarchical organization could have self-renewal capacity and could also give rise to progeny CAFs, such as cancer-promoting CAFs (pCAFs) and cancer-restraining CAFs (rCAFs). Some of the CAF-stem cells probably overlap with subpopulations of MSCs whose lineage is committed to CAFs during cancer progression. We speculate that CAFs from different cellular origins have functionally distinct phenotypes; however, the relationship between the CAF subtypes and their cellular origins is not fully elucidated. EMT, epithelial-mesenchymal transition; EndMT, endothelial-mesenchymal transition.

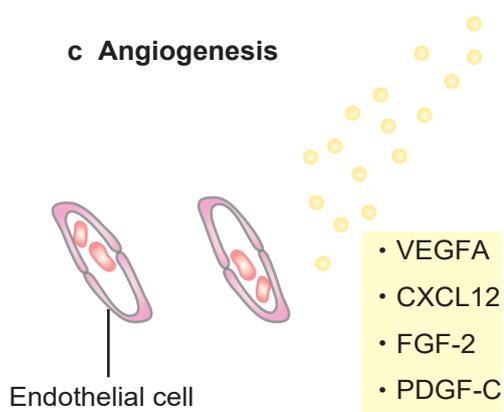
**a Interaction with cancer cells**



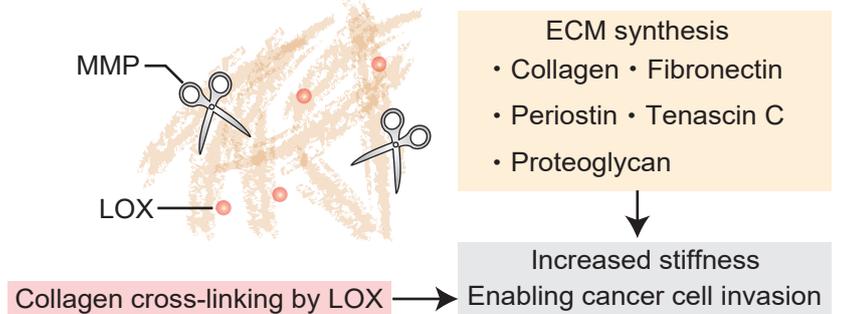
**b Tumour immunology**



**c Angiogenesis**

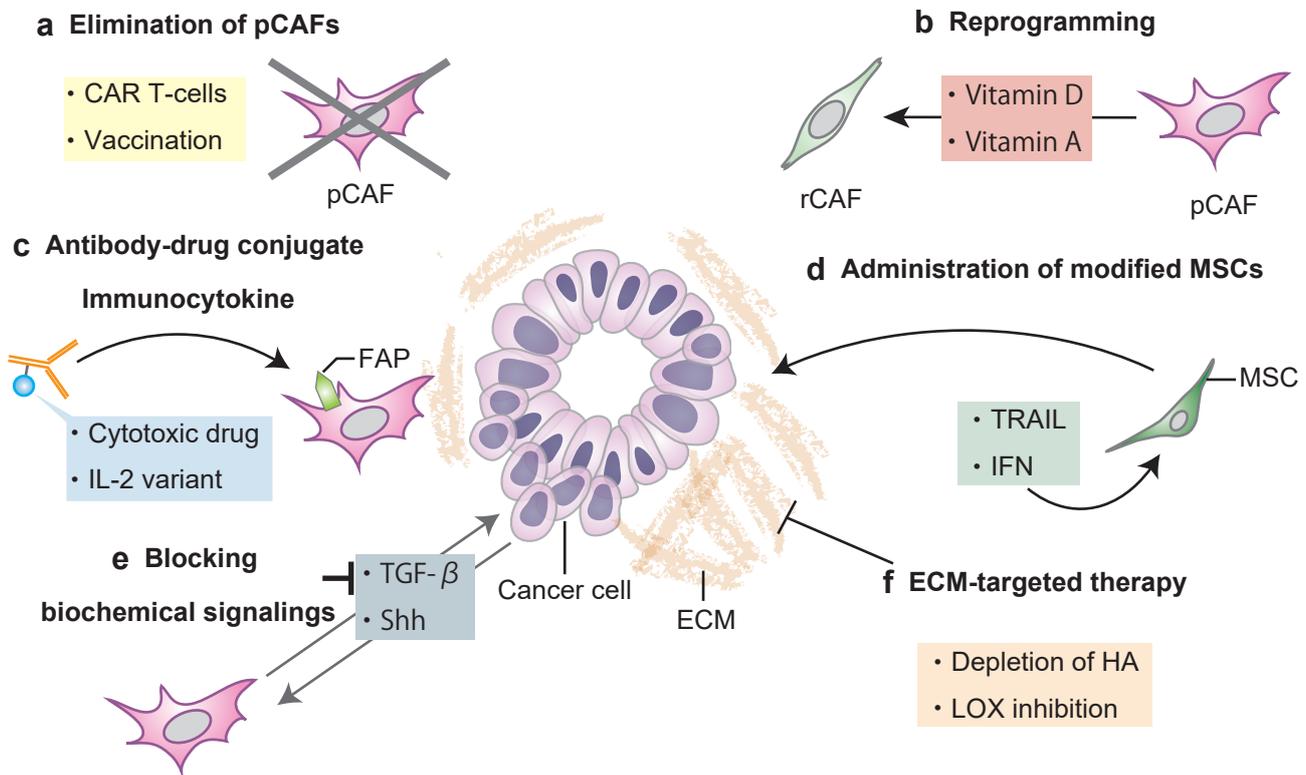


**d ECM remodelling**



### Figure 3: The functions of CAFs

Cancer-associated fibroblasts (CAFs) orchestrate the development of cancer. CAFs promote cancer cell proliferation by secreting a plethora of pro-tumourigenic factors. Transfer of proteins and RNAs is mediated in part by exosomes. CAFs also physically pull cancer cells to guide collective cancer cell migration (a). CAFs secrete numerous chemokines and cytokines such as CXC chemokine ligand 12 (CXCL12) and transforming growth factor- $\beta$  (TGF- $\beta$ ), thereby inducing immunosuppression in the tumour microenvironment (TME). Of note, CAFs express programmed death ligand 1 (PDL1), a target protein for immune checkpoint inhibitors (b). Vascular endothelial growth factor A (VEGFA), CXCL12, fibroblast growth factor-2 (FGF2) and platelet-derived growth factor-C (PDGFC) produced by CAFs facilitate the formation of new blood vessels in the TME (c). CAFs synthesize ECM components such as collagen and fibronectin, and the ECM is degraded by matrix metalloproteinases (MMPs) secreted by CAFs. Collagen cross-linking is mediated by the lysyl oxidase (LOX) family produced, in part, by CAFs<sup>142,199</sup>. CAFs contribute to increased ECM stiffness, which in turn promotes cancer progression, for example, by increasing cancer cell proliferation and invasion<sup>138,139,142</sup> (d). BMP, bone morphogenetic protein; CCL5, C-C motif chemokine ligand 5; EGF, epithelial growth factor; HGF, hepatocyte growth factor; IGF, insulin-like growth factor; IL-6, interleukin-6;; T<sub>reg</sub> cell, regulatory T cell.



**Figure 4: Therapies that target CAFs**

Cancer-associated fibroblasts (CAFs) and mesenchymal stem cells (MSCs) present multiple therapeutic approaches for cancer treatment. Fibroblast activation protein (FAP)<sup>+</sup> cancer-promoting CAFs (pCAFs) can be eliminated by chimeric antigen receptor (CAR)-T cell or DNA vaccination (a). A reprogramming therapy such as vitamin D and vitamin A can be used to dedifferentiate pCAFs to cancer-restraining CAFs (rCAFs) (b). Further studies are necessary to identify molecules that can effectively reprogram CAFs. An antibody–drug conjugate (ADC) or immunocytokine against FAP, a membrane marker expressed in CAFs, enables effective delivery of drugs to tumour sites (c). Administration of MSCs engineered to express anti-tumourigenic molecules such as TRAIL, IFN $\alpha$  and herpes simplex virus-thymidine kinase (HSV-TK) leads to the accumulation of the MSCs in the tumour site, thereby inducing cancer cell death (d). Blocking the biochemical interaction between cancer cells and CAFs by transforming growth factor- $\beta$  (TGF $\beta$ ) inhibitors or hedgehog inhibitors might prevent cancer progression (e). Manipulation of the extracellular matrix (ECM), which in the TME is produced predominantly by CAFs, leads to improved drug delivery (f). HA, hyaluronic acid; LOX, lysyl oxidase;

**Table1: Some representative CAF markers in cancer and homeostasis**

| CAF marker                          | Fibroblast type       | Description                                       | Expression pattern (excluding fibroblasts)         | Refs               |
|-------------------------------------|-----------------------|---|--|--------------------|
| <b>Membrane proteins</b>            |                       |   |  |                    |
| FAP                                 | Activated fibroblasts | Serine protease                                   | Macrophages  | 55,56              |
| PDGFR $\beta$<br>(CD140 $\beta$ )   | Activated fibroblasts | Growth factor receptor                            | Pericytes, cancer cells                            | 10,83,200          |
| Podoplanin<br>(GP38 in mice)        | Activated fibroblasts | Transmembrane glycoprotein                        | LECs, cancer cells                                 | 201,202            |
| PDGFR $\alpha$<br>(CD140 $\alpha$ ) | Quiescent fibroblasts | Growth factor receptor                            | BM-MSCs, cancer cells                              | 70,200,203-205     |
| <b>Intracellular proteins</b>       |                       |   |  |                    |
| $\alpha$ SMA                        | Activated fibroblasts | Cytoskeletal protein crucial for cell contraction | SMCs, pericytes                                    | 10,49              |
| Desmin                              | Activated fibroblasts | Intermediate filament                             | Skeletal muscle cells, SMCs, pericytes             | 10,49              |
| Vimentin                            | Activated fibroblasts | Intermediate filament                             | SMCs, endothelial cells, neural cells, cancer cell | 10,11,49,206       |
| FSP1<br>(S100A4)                    | Quiescent fibroblasts | Calcium-binding protein                           | Cancer cells, macrophages                          | 52,58-60           |
| Gli1                                | Miscellaneous         | Transcription factor in                           | Perivascular fibroblasts, BM-MSCs, cancer cells    | 5,22,36,67,195,207 |

|                          |                       |                                |  |                |
|--------------------------|-----------------------|--------------------------------|--|----------------|
|                          |                       | Hedgehog signalling            |  |                |
| <b>Secreted proteins</b> |                       |                                |  |                |
| CXCL12                   | Activated fibroblasts | Chemokine                      | BM-MSCs, BECs, osteoblasts, haematopoietic cells | 16,208         |
| Gremlin 1                | Activated fibroblasts | BMP antagonist, VEGFR2 agonist | BM-MSCs, iRSCs                                   | 46,105,106,209 |
| <b>ECM proteins</b>      |                       |                                |  |                |
| Col1a1                   | Activated fibroblasts | Component of type I collagen   | Osteoblasts, tendon                              | 193,210        |
| Periostin                | Activated fibroblasts | Matricellular protein          | Periosteum, osteoblasts, tendon                  | 211,212        |

$\alpha$ SMA,  $\alpha$ -smooth muscle actin; BECs, blood endothelial cells; BM-MSCs, bone marrow-mesenchymal stem cells; BMP, bone morphogenetic protein, CAFs, cancer-associated fibroblasts; Col1a1, collagen type I alpha 1; CXCL12, C-X-C motif chemokine ligand 12; FAP, fibroblast activation protein; FSP1, fibroblast-specific protein 1; Gli1, glioma-associated oncogene homolog 1; iRSCs, intestinal reticular stem cells; LECs, lymphatic endothelial cells; PDGFR, platelet-derived growth factor receptor; SMCs, smooth muscle cells; VEGFR2, vascular endothelial receptor-2

## **-Glossary**

### **Extracellular matrix**

(ECM). An intricate network of fibrous proteins in the extracellular space, such as collagen, laminin and fibronectin.

### **Basement membrane**

Highly specialized ECM that separates epithelial cells or endothelial cells from underlying connective tissue.

### **Desmoplastic reaction**

Increase in a stromal component especially with prominent fibrous tissue in cancer.

### **Angiogenesis**

Formation of new blood vessels to satisfy increased demand for nutrients and oxygen.

### **Gene signature**

A gene expression pattern characteristic of a certain biological process

### **Telocytes**

Mesenchymal cells that have extending cytoplasmic processes termed “telopodes”.

### **Tumour-associated macrophages**

(TAMs). A heterogeneous population of macrophages in the tumour, which contributes to tumour progression.

### **Epithelial-to-mesenchymal transition**

(EMT). A process by which epithelial cells gain a mesenchymal phenotype, leading to their migration and invasion.

### **Myofibroblasts**

A specific type of fibroblasts that are characterized by high expression of  $\alpha$ -smooth muscle actin.

### **Chimeric antigen receptor T-cells**

(CAR T-cells). T cells engineered to recognize a tumour-associated antigen and induce target-specific killing.

### **Stellate cells**

Fibroblast-like cells characterized by their vitamin A storage. They are found in the pancreas and liver.

### **Single-cell RNA-sequencing**

Gene expression analysis of an individual cell, instead of diverse cell populations.

### **Pericytes**

Fibroblast-like cells that wrap around the wall of capillaries.

### **Parabiosis**

Two organisms joined together surgically to share blood circulation.

### **Lineage tracing**

A method to genetically label cells of interest and all of their progenies. Also known as genetic fate-mapping.

### **Endothelial-to-mesenchymal transition**

(EndMT). A process by which endothelial cells lose an endothelial phenotype and acquire a mesenchymal phenotype.

### **Exosome**

An Extracellular vesicle (30-150 nm in size) that is released from many types of cells and contains proteins and RNAs.

### **Cancer stem cells**

(CSCs). A minor subpopulation of cancer cells that has self-renewal capability and drives cancer progression, metastasis and resistance to treatment.

### **Immunosuppressive TME**

TME in which antitumour immunity is inhibited and cancer immunotherapy is ineffective.

### **Immune checkpoint inhibitors**

Agents that unleash antitumour immunity through blocking an immune checkpoint which is a ligand-receptor-mediated pathway to suppress an immune response.

### **Myeloid-derived suppressor cells**

(MDSCs). A heterogeneous population of bone marrow-derived immune cells that suppresses T cell activity.

### **Regulatory T cells**

(T<sub>reg</sub> cells). A subset of immunosuppressive T cells that express CD4, CD25 and FOXP3 and maintains immune tolerance to self-antigens and prevent activation of effector T cells.

### **M2 polarization of macrophages**

The M2 macrophage is a subtype of tumour-associated macrophages (TAMs) that suppresses antitumour immunity and promotes cancer progression.

### **Extravasation**

Leakage of blood cells from capillaries to the surrounding tissue.

### **Fibrocytes**

Bone marrow-derived circulating cells that have features of both fibroblasts and monocytes.

### **Antibody-drug conjugate**

(ADC). A small molecule drug linked to a monoclonal antibody that recognizes a tumour-associated antigen.

### **Immunocytokines**

A cytokine fused to a monoclonal antibody that recognizes a tumour-associated antigen.

## **-Acknowledgments**

This work is supported by Cure Cancer Australia/Cancer Australia (APP1102534) and the Australian National Health and Medical Research Council (NHMRC) (APP1140236 and APP1143414). D.L.W. is supported by a NHMRC Career Development Fellowship, H.K. by Japan Society for the Promotion of Science (JSPS) Overseas Challenge Program for Young Researchers and Takeda Science Foundation Fellowship.

## **-Contributions**

All authors contributed substantially to discussion of content for the article, and reviewed and/or edited the manuscript before submission. H.K., A.E., and D.L.W. researched data and wrote the manuscript.

## **-Competing interests statement**

The authors declare no competing interests.

## **Chapter 2: The balance of stromal BMP signalling mediated by *GREM1* and *ISLR* drives colorectal carcinogenesis.**

In this chapter, I addressed the functional dichotomy of CAFs from a perspective of BMP signalling, a key signalling pathway that defines the intestinal stem cell niche. Our Gut Cancer laboratory at Adelaide University focuses on the biological role of a BMP antagonist, Gremlin 1 (*Grem1*)<sup>46</sup>. At our Nagoya pathology laboratory, we recently identified immunoglobulin superfamily containing leucine-rich repeat (*Islr*) as a functionally opposing regulator of BMP signalling (i.e., potentiator of BMP signaling)<sup>213</sup>. Taking advantage of the Joint Degree Program collaboration between the two universities, I have shown that *GREM1* and *ISLR* are colorectal CAF-specific genes with roles in BMP signalling. Next, I characterized their distinct expression patterns in the normal colon and tumour microenvironment. Finally, I have examined whether adeno-associated virus (AAV)-mediated gene delivery of *Islr* to hepatocytes could augment BMP signalling and inhibit CRC hepatic metastasis growth.

## Statement of Authorship

|                     |  |
|---------------------|--|
| Title of Paper      | The balance of stromal BMP signalling mediated by GREM1 and ISLR drives colorectal carcinogenesis.   |
| Publication Status  | Published  |
| Publication Details | Kobayashi, H., Gieniec, K. A., Wright, J. A., Wang, T., Asai, N., Mizutani, Y., Iida, T., Ando, R., Suzuki, N., Lannagan, T. R. M., Ng, J. Q., Hara, A., Shiraki, Y., Mii, S., Ichinose, M., Vrbanc, L., Lawrence, M. J., Sammour, T., Uehara, K., Davies, G., Lisowski, L., Alexander, I. E., Hayakawa, Y., Butler, L. M., Zannettino, A. C. W., Din, M. O., Hasty, J., Burt, A. D., Leedham, S. J., Rustgi, A. K., Mukherjee, S., Wang, T. C., Enomoto, A., Takahashi, M., Worthley, D. L., Woods, S. L. (2020). The balance of stromal BMP signalling mediated by GREM1 and ISLR drives colorectal carcinogenesis. <i>Gastroenterology</i> , doi: <a href="https://doi.org/10.1053/j.gastro.2020.11.01">https://doi.org/10.1053/j.gastro.2020.11.01</a> |

## Candidate

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of Candidate         | Hiroki Kobayashi   |      |          |
| Contribution to the Paper | Performed most of the experiments, data interpretation and statistical analyses, performed animal experiments and wrote the manuscript. Principal author.  |      |          |
| Overall Percentage (%)    | 70%  |      |          |
| Certification:            | This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. |      |          |
| Signature                 |  | Date | 30/10/20 |

## Co-author contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all the co-author contributions is equal to 100% less the candidate's stated contribution.

Note that multiple authors have given written permission for my primary supervisor (Dr. Daniel Worthley) to sign on their behalf.

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of principal author  | Krystyna Gieniec  |      |         |
| Contribution to the paper | Discussed manuscript content and helped in data interpretation. Performed the experiment examining the effect of the GREM1 neutralising antibody on the survival of mice with hepatic colorectal metastases and analysed the associated data. |      |         |
| Signature                 |   | Date | 6/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Josephine Wright   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, generated genetically engineered colorectal tumouroids, performed an <i>in vitro</i> tumouroid experiment using <i>Islr</i> -YH2 cells. |      |         |
| Signature                 |  | Date | 6/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Tongtong Wang  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, analysed RNA-sequencing datasets. |      |         |
| Signature                 |  | Date | 6/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Naoya Asai  |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation, performed animal experiments with reporter mice. |      |         |
| Signature                 |   | Date | 9/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Yasuyuki Mizutani  |      |          |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, performed smFISH. |      |          |
| Signature                 |  | Date | 17/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Tadashi Iida   |      |          |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, performed smFISH. |      |          |
| Signature                 |  | Date | 18/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Ryota Ando   |      |          |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, performed smFISH. |      |          |
| Signature                 |  | Date | 17/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Nobumi Suzuki   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, generated genetically engineered colorectal tumours. |      |         |
| Signature                 |   | Date | 9/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Tamsin Lannagan  |      |          |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, generated genetically engineered colorectal tumouroids. |      |          |
| Signature                 |  | Date | 17/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Jia Ng  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, performed animal experiments with reporter mice. |      |         |
| Signature                 |   | Date | 6/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Akitoshi Hara  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 8/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Yukihiro Shiraki   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Shinji Mii   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Mari Ichinose  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Laura Vrbanac  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 6/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Matthew Lawrence   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Tarik Sammour  |      |          |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |          |
| Signature                 |  | Date | 10/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Kay Uehara   |      |          |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, collected human colorectal cancer clinical data and provided samples. |      |          |
| Signature                 |  | Date | 10/11/20 |

|                           |   |      |                           |
|---------------------------|---|------|---------------------------|
| Name of co-author         | Gareth Davies   |      |                           |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, provided the GREM1 neutralising antibody and intellectual input. |      |                           |
| Signature                 |   | Date | 10 <sup>th</sup> Nov 2020 |

|                           |   |      |            |
|---------------------------|---|------|------------|
| Name of co-author         | Leszek Lisowski   |      |            |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, produced and tittered AAV vectors. |      |            |
| Signature                 |   | Date | 9 Nov 2020 |

|                           |  |      |            |
|---------------------------|--|------|------------|
| Name of co-author         | Ian Alexander  |      |            |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |            |
| Signature                 |  | Date | 9 Nov 2020 |

|                           |  |      |            |
|---------------------------|--|------|------------|
| Name of co-author         | Yoku Hayakawa  |      |            |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |            |
| Signature                 |  | Date | 11/18/2020 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Lisa Butler  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 6/11/20 |

|                   |                   |  |  |
|-------------------|-------------------|--|--|
| Name of co-author | Andrew Zannettino |  |  |
|-------------------|-------------------|--|--|

|                           |  |      |          |
|---------------------------|--|------|----------|
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |          |
| Signature                 |  | Date | 11/11/20 |

|                           |   |      |            |
|---------------------------|---|------|------------|
| Name of co-author         | M. Omar Din   |      |            |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation.                      |      |            |
| Signature                 |  | Date | 2020-11-09 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Jeff Hasty   |      |          |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |          |
| Signature                 |  | Date | 19/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Alastair Burt  |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation, wrote the manuscript. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Simon Leedham  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                   |             |  |  |
|-------------------|-------------|--|--|
| Name of co-author | Anil Rustgi |  |  |
|-------------------|-------------|--|--|

|                           |  |      |         |
|---------------------------|--|------|---------|
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 1/12/10 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Siddhartha Mukherjee   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Timothy Wang   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

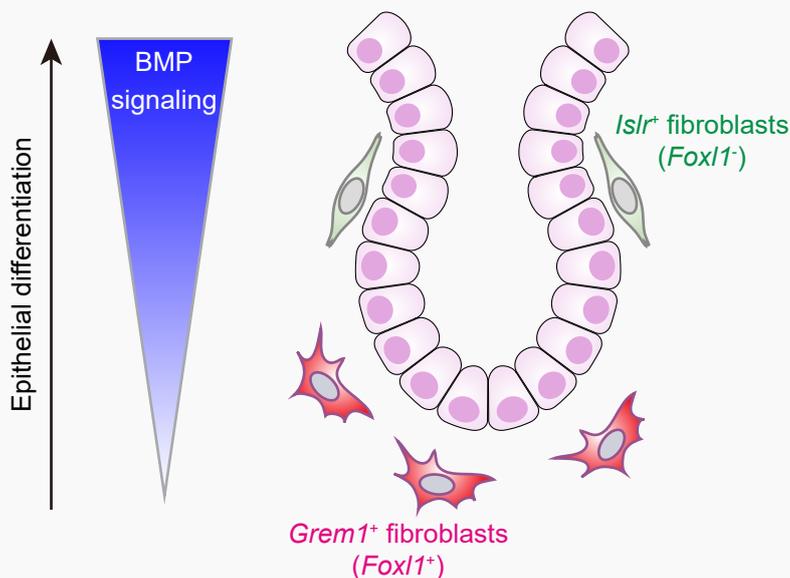
|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Atsushi Enomoto  |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation and manuscript evaluation, wrote the manuscript. Will act as corresponding author. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Masahide Takahashi   |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation and manuscript evaluation. Will act as corresponding author. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Daniel Worthley  |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation and manuscript evaluation, wrote the manuscript. Will act as corresponding author. |      |         |
| Signature                 |  | Date | 6/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Susan Woods  |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation and manuscript evaluation, generated genetically engineered colorectal tumouroids, wrote the manuscript. Will act as corresponding author. |      |         |
| Signature                 |  | Date | 6/11/20 |

## Normal colon



## Colorectal cancer

Stromal BMP-Low

Stromal BMP-High

*Grem1*<sup>+</sup> cancer-promoting CAFs  
(TGFβ-responsive myfibroblastic CAFs)

*Islr*<sup>+</sup> cancer-inhibitory CAFs

GREM1-neutralizing antibody

GREM1

BMP signaling

AAV-*Islr*

Inhibiting tumor growth

Promoting tumor cell differentiation

Poorly differentiated

Tumor cell differentiation

Well differentiated

*Grem1*

*Islr*

BMP signaling

# The balance of stromal BMP signaling mediated by *GREM1* and *ISLR* drives colorectal carcinogenesis.

Short title: The role of BMP signaling in the CRC stroma

Hiroki Kobayashi<sup>1, 2, 3, 4</sup>, Krystyna A. Gieniec<sup>1,2</sup>, Josephine A. Wright<sup>2</sup>, Tongtong Wang<sup>1,2</sup>, Naoya Asai<sup>5</sup>, Yasuyuki Mizutani<sup>3, 6</sup>, Tadashi Iida<sup>3, 6</sup>, Ryota Ando<sup>3</sup>, Nobumi Suzuki<sup>1,2,7</sup>, Tamsin RM. Lannagan<sup>1,2</sup>, Jia Q Ng<sup>1,2</sup>, Akitoshi Hara<sup>8</sup>, Yukihiro Shiraki<sup>3</sup>, Shinji Mii<sup>3,4</sup>, Mari Ichinose<sup>1,2</sup>, Laura Vrbnac<sup>1,2</sup>, Matthew J. Lawrence<sup>9</sup>, Tarik Sammour<sup>1,2,9</sup>, Kay Uehara<sup>10</sup>, Gareth Davies<sup>11</sup>, Leszek Lisowski<sup>12,13,14</sup>, Ian E. Alexander<sup>15,16</sup>, Yoku Hayakawa<sup>7</sup>, Lisa M. Butler<sup>1,2</sup>, Andrew C. W. Zannettino<sup>1,2</sup>, M. Omar Din<sup>17</sup>, Jeff Hasty<sup>18</sup>, Alastair D. Burt<sup>1,19</sup>, Simon J. Leedham<sup>20</sup>, Anil K. Rustgi<sup>21</sup>, Siddhartha Mukherjee<sup>22</sup>, Timothy C. Wang<sup>22</sup>, Atsushi Enomoto<sup>3\*</sup>, Masahide Takahashi<sup>3,4,23\*</sup>, Daniel L. Worthley<sup>2\*</sup>, and Susan L. Woods<sup>1,2, 24\*</sup>

<sup>1</sup>Adelaide Medical School, University of Adelaide, Adelaide, SA, 5000, Australia.

<sup>2</sup>South Australian Health and Medical Research Institute, Adelaide, SA, 5000, Australia.

<sup>3</sup>Department of Pathology, <sup>4</sup>Division of Molecular Pathology, Center for Neurological Disease and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Aichi, 466-8550, Japan.

<sup>5</sup>Department of Molecular Pathology, Graduate School of Medicine, Fujita Health University, Toyoake, Aichi, 470-1192, Japan.

<sup>6</sup>Department of Gastroenterology and Hepatology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, 466-8550, Japan.

<sup>7</sup>Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, 113-0033, Japan.

<sup>8</sup>Department of Cardiology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, 466-8550, Japan.

<sup>9</sup>Colorectal Unit, Department of Surgery, Royal Adelaide Hospital, Adelaide, SA, 5000, Australia.

<sup>10</sup>Division of Surgical Oncology, Department of Surgery, Nagoya University Graduate School of Medicine, Nagoya, Aichi, 466-8550, Japan.

<sup>11</sup>UCB Pharma, Slough, Berkshire, UK.

<sup>12</sup>Translational Vectorology Research Unit, Children's Medical Research Institute, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia.

<sup>13</sup>Vector and Genome Engineering Facility, Children's Medical Research Institute, Faculty of Medicine and Health, The University of Sydney, Westmead, NSW 2145, Australia.

<sup>14</sup>Military Institute of Hygiene and Epidemiology, The Biological Threats Identification and Countermeasure Centre, 24-100 Puławy, Poland.

<sup>15</sup>Gene Therapy Research Unit, Sydney Children's Hospitals Network and Children's Medical Research Institute, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia.

<sup>16</sup>Discipline of Child and Adolescent Health, Faculty of Medicine and Health, The University of Sydney, NSW, Australia.

<sup>17</sup>GenCirq, Inc., San Diego, CA, USA.

<sup>18</sup>Department of Bioengineering, University of California, San Diego, La Jolla, CA USA.

<sup>19</sup>Precision and Molecular Pathology, Newcastle University, Newcastle upon Tyne NE2 4HH, UK.

<sup>20</sup>Intestinal Stem Cell Biology Lab, Wellcome Trust Centre Human Genetics, University of Oxford, Oxford, UK.

<sup>21</sup>Herbert Irving Comprehensive Cancer Center, Division of Digestive and Liver Diseases, Department of Medicine, Columbia University, New York, NY, USA.

<sup>22</sup>Department of Medicine and Irving Cancer Research Center, Columbia University, New York, NY, USA.

<sup>23</sup>International Center for Cell and Gene Therapy, Fujita Health University, Toyoake, Aichi, 470-1192, Japan.

<sup>24</sup>Lead contact.

\*Co-corresponding authors.

## Grant Support

This study was supported by grants from the National Health and Medical Research Council (APP1156391 to D.L.W., S.L.W.; APP1108311 and APP1161583 to L.L.); Cancer Council SA Beat Cancer Project on behalf of its donors and the State Government of South Australia through the Department of Health (MCF0418 to S.L.W., D.L.W., and PRF1117 to L.M.B.); a Grant-in-Aid for Scientific Research (S) (26221304 to M.T.) commissioned by the Ministry of Education, Culture, Sports, Science and Technology of Japan; AMED-CREST (Japan Agency for Medical Research and Development, Core Research for Evolutional Science and Technology (19gm0810007h0104 and 19gm1210008s0101 to A.E.); the Project for Cancer Research and Therapeutic Evolution (P-CREATE) from AMED (19cm0106332h0002 to A.E.); Japan Society for the Promotion of Science Overseas Challenge Program for Young Researchers (to H.K.), Takeda Science Foundation Fellowship (to H.K.), Greaton International Ph.D. Scholarship (to H.K.).

## Abbreviations

AAV, adeno-associated virus; ANOVA, analysis of variance; AOM, azoxymethane; BMP, bone morphogenetic protein; CAFs, cancer-associated fibroblasts; CRISPR, clustered regularly interspaced short palindromic repeats; CM, conditioned medium; CMS, consensus molecular subtype; CRC, colorectal cancer; DSS, dextran sodium sulfate; FACS, fluorescence-activated cell sorting; GFP, green fluorescent protein; IF, immunofluorescence; IHC, immunohistochemistry; ISH, in-situ hybridization; IVIS, *in vivo* imaging system; PC, principal component; MSCs, mesenchymal stem/stromal cells; qRT-PCR, quantitative reverse-transcription polymerase chain reaction; scRNA-seq, single-cell RNA-sequencing; smFISH, single-molecule fluorescent in-situ hybridization; TGF $\beta$ , transforming growth factor  $\beta$ ; TSS, transcriptional start site

## Correspondence

Lead contact; S.L.W.

S.L.W.; [susan.woods@adelaide.edu.au](mailto:susan.woods@adelaide.edu.au), Group Leader, Gut Cancer, the University of Adelaide,

D.L.W.; [Dan.Worthley@sahmri.com](mailto:Dan.Worthley@sahmri.com), Group Leader, Gut Health, SAHMRI

Address: SAHMRI (5 South), North Terrace, Adelaide SA 5000, AUSTRALIA, 08-8128-4386

M.T.; [mtakaha@med.nagoya-u.ac.jp](mailto:mtakaha@med.nagoya-u.ac.jp), Professor, International Center for Cell and Gene Therapy, Fujita Health University

Address: 1-98 Dengakugakubo, Kutsukake-Cho, Toyoake, Aichi, 470-1192, JAPAN. 05-6293-2884

A.E.; [enomoto@iar.nagoya-u.ac.jp](mailto:enomoto@iar.nagoya-u.ac.jp), Professor, Department of Pathology, Nagoya University School of Medicine

Address: 65 Tsurumai-Cho, Showa-Ku, Nagoya, Aichi, 466-8550, JAPAN, 05-2744-2093

## Disclosures

These authors disclose the following: Gareth Davies is a researcher at UCB Pharma (Slough, Berkshire, UK). A GREM1-neutralizing antibody was generated by the UCB Pharma and provided to researchers at the Gut Cancer Group, the University of Adelaide (Australia), at no cost. L.L. has consulted on technologies addressed in this paper and has stock and/or equity in companies with technology broadly related to this paper. M.O.D. and J.H. have financial interest in GenCirq Inc. (San Diego, CA, USA). All other authors declare no conflicts of interest.

**Transcript Profiling:** N/A

**Writing Assistance:** N/A

## Author Contributions

H.K., K.A.G., A.E., M.T., D.L.W., and S.L.W. conceived and designed the study. H.K. performed most of the experiments and statistical analyses. K.A.G. conducted an *in vivo* GREM1-neutralizing antibody experiment. J.A.W. performed an *in vitro* tumoroid experiment using *Islr*-YH2 cells. T.W. (Tongtong Wang) analyzed RNA-seq datasets. J.A.W., T.R.M, N.S., and S.L.W. generated genetically engineered CRC organoids. Y.M., T.I., and R.A. performed smFISH. H.K., N.A., and J.Q.N. performed animal experiments with reporter mice. K.U. collected human CRC clinical data and provided the samples. L.L. produced and titered AAV vectors. G.D. provided a GREM1-

neutralizing antibody and intellectual input. N.A., A.D.B., A.E., M.T., D.L.W., and S.L.W. supervised the project. H.K., A.D.B., A.E., M.T., D.L.W., and S.L.W. wrote the manuscript. All authors contributed substantially to the discussion of content for the article, reviewed and/or edited the manuscript before submission.

**Word count:**

Abstract; 259 words

Main manuscript characters; 6997 words

(inclusive of main text; references; figure legends)

**Total number of figures and tables:** 6 figures

## Abstract

### Background and aims

Cancer-associated fibroblasts (CAF), key constituents of the tumor microenvironment, either promote or restrain tumor growth. Attempts to therapeutically target CAFs have been hampered by our incomplete understanding of these functionally heterogeneous cells. Key growth factors in the intestinal epithelial niche, bone morphogenetic proteins (BMPs), also play a critical role in colorectal cancer (CRC) progression. However, the crucial proteins regulating stromal BMP balance and the potential application of BMP signaling to manage CRC remain largely unexplored.

### Methods

Using human CRC RNA expression data, we identified CAF-specific factors involved in BMP signaling, then verified and characterized their expression in the CRC stroma by *in situ* hybridization. CRC tumoroids and a mouse model of CRC hepatic metastasis were used to test approaches to modify BMP signaling and treat CRC.

### Results

We identified *Grem1* and *Islr* as CAF-specific genes involved in BMP signaling. Functionally, *GREM1* and *ISLR* acted to inhibit and promote BMP signaling, respectively. *Grem1* and *Islr* marked distinct fibroblast subpopulations, and were differentially regulated by TGF $\beta$  and FOXL1, providing an underlying mechanism to explain fibroblast biological dichotomy. In CRC patients, high *GREM1* and *ISLR* expression were associated with poor and favorable survival, respectively. A *GREM1*-neutralizing antibody or fibroblast *Islr* overexpression reduced CRC tumoroid growth and promoted *Lgr5*<sup>+</sup> intestinal stem cell differentiation. Finally, adeno-associated virus 8 (AAV8)-mediated delivery of *Islr* to hepatocytes increased BMP signaling and improved survival in our mouse model of hepatic metastasis.

### Conclusions

Stromal BMP signaling predicts and modifies CRC progression and survival, and can be therapeutically targeted by novel AAV-directed gene delivery to the liver.

### Keywords:

colorectal cancer; cancer-associated fibroblasts; tumor microenvironment; Bone morphogenetic protein; Meflin.

## Introduction

Colorectal cancer (CRC) is a major cause of cancer mortality<sup>1</sup>. Despite advances in surgical techniques and medical therapies targeting tumor cells, endothelial cells, and immune cells, the majority of metastatic CRC patients still die from their disease<sup>1</sup>.

Cancer-associated fibroblasts (CAFs), a key constituent of the tumor microenvironment, influence CRC initiation, progression, and dissemination, and can promote drug resistance via secretion of growth factors, chemokines, extracellular matrix, and pro-angiogenic factors<sup>3</sup>. CAFs are not a uniformly pro-tumorigenic entity; rather, CAFs are composed of functionally heterogeneous subpopulations including tumor-promoting CAFs and tumor-retarding CAFs<sup>3</sup>. However, the markers and mechanisms underlying this CAF biological dichotomy are largely unknown, which has hampered therapeutic attempts to exploit these differences<sup>3</sup>.

One key family of growth factors secreted by CAFs, as well as cancer cells, are the BMPs (bone morphogenetic proteins)<sup>214</sup>. BMPs belong to the TGF $\beta$  (transforming growth factor- $\beta$ ) superfamily. Binding of BMP ligands such as BMP 2, 4, and 7, to type I and II BMP receptors, induces phosphorylation of SMAD1/5/8, which in turn binds SMAD4 to increase target gene expression such as *ID1*, *2*, *3* and *4*<sup>214</sup>. BMP gradients partly define the intestinal epithelial stem cell niche in the normal colon and serve to promote or retard cancer progression in a context-dependent manner<sup>214</sup>. In the normal colon, the epithelial stem cell niche is maintained by low BMP and high Wnt at the crypt base, whereas epithelial cell differentiation is driven by increasing BMP and low Wnt towards the luminal surface<sup>215</sup>. The BMP gradient is finely tuned by BMP inhibitors, such as GREM1 and NOGGIN, which are secreted by fibroblasts near the crypt base<sup>46,215,216</sup>. In CRC, inactivation of BMP signaling through germline or sporadic mutations in BMP receptors and *SMAD4* contributes to CRC predisposition and progression<sup>215</sup>. Numerous studies have demonstrated the tumor-retarding role of BMP signaling in CRC cells themselves<sup>214,217,218</sup>. BMP signaling has been shown to reduce stemness of intestinal stem cells (ISCs) such as *Lgr5*<sup>+</sup> ISCs, leading to epithelial differentiation<sup>217,218</sup>. Most CRC studies, however, have failed to address the function of stromal BMP signaling in CRC progression.

Here, we identified gremlin 1 (*GREM1*) and immunoglobulin superfamily containing leucine-rich repeat (*ISLR*), specifically and distinctly expressed by different types of CRC CAFs, as important regulators of BMP signaling within the tumor microenvironment.

*GREM1*, a ligand-sequestering antagonist for BMP2, 4, and 7, is expressed by mesenchymal stem/stromal cells (MSCs) in the bone marrow and intestinal fibroblasts<sup>46</sup>. *GREM1* expressed by CAFs can accelerate tumor cell proliferation via inhibition of BMP signaling<sup>105</sup>. In patients with CRC or breast cancer, high *GREM1* expression is associated with poor prognosis<sup>219-221</sup>. *GREM1* secreted by glioma cancer stem cells blocks BMP2-induced differentiation of glioma cells, thereby maintaining their proliferation and stemness<sup>222</sup>. Furthermore, overexpression of *GREM1* in intestinal epithelial cells initiates colonic tumorigenesis, supporting the pro-tumorigenic role of *GREM1*<sup>106,223</sup>.

*ISLR* (also known as Meflin), a GPI (glycosylphosphatidylinositol)-anchored membrane protein, that is also secreted, was recently identified as a specific marker for MSCs and fibroblasts in various organs including the bone marrow, heart, and pancreas<sup>213,224</sup>. In contrast to the tumor-promoting role of *GREM1*, our recent study found that *ISLR* defines a subset of tumor-retarding CAFs, which are distinct from  $\alpha$ -smooth muscle actin ( $\alpha$ SMA)<sup>+</sup> CAFs in pancreatic cancer<sup>224</sup>. *ISLR* interacts with BMP7 to augment BMP7-Smad1/5 signaling<sup>213</sup>. Although a recent report has indicated that *ISLR* is highly expressed by fibroblasts in the inflamed colon and CRC<sup>225</sup>, the biological role of *ISLR* in CRC related to BMP signaling remains unknown.

In this study, following the identification and validation of *GREM1* and *ISLR* as two functionally opposing BMP-related genes specifically expressed by CRC CAFs, we examined the prognostic significance of *GREM1* and *ISLR* expression levels in human CRC. Then, we characterized distinct expression patterns of *GREM1* and *ISLR* in both normal colon and CRC and the potential mechanism by which this CAF polarization occurs. Next, we explored whether a *GREM1*-neutralizing antibody or conditioned medium transfer from *Islr*-overexpressing colonic fibroblasts could restrain CRC organoid growth. Finally, we investigated whether AAV (adeno-associated virus) 8-mediated ectopic overexpression of *Islr* in hepatocytes could retard CRC liver metastasis progression in mice, as a novel therapeutic approach to restrain CRC metastasis progression.

## Materials and Methods

### Statistical analysis

A comparison of 2 groups was performed using two-tailed unpaired t-tests or Mann–Whitney U tests. For multiple comparisons, we used analysis of variance (ANOVA) with subsequent Tukey or Sidak's post-hoc analysis (for parametric tests), or Kruskal-Wallis test followed by Dunn's post-hoc multiple comparisons (for non-parametric tests). For survival analyses, Kaplan-Meier survival estimation with a Log-rank (Mantel-cox) test was performed. Statistical analyses were conducted using GraphPad Prism 8.00 (GraphPad) or SPSS Statistics ver. 25 (IBM). P-values of less than 0.05 were considered statistically significant.

For all other Materials and Methods, see **Supplementary Materials**.

## Results

### ***Identification of CAF-specific expression of the BMP antagonist *GREM1* and the BMP potentiator *ISLR* in CRC***

To identify which BMP-related genes are specifically expressed by CAFs, we first analyzed expression microarray data from a study of fluorescence-activated cell sorting (FACS)-purified cells from human primary CRC tissues<sup>103</sup>. The top 150 differentially expressed gene probes upregulated in CAFs in each group (FAP<sup>+</sup> CAFs vs. EpCAM<sup>+</sup> cancer cells, FAP<sup>+</sup> CAFs vs. CD31<sup>+</sup> endothelial cells, and FAP<sup>+</sup> CAFs vs. CD45<sup>+</sup> immune cells) were selected for our analysis, resulting in the identification of 34 genes specifically expressed in human CRC CAFs (**Figure 1A; Supplementary Figure 1A and B; Supplementary Table 1**). Next, to examine for genes involved in BMP signaling, we compared the 34 CAF-specific genes, with BMP-relevant genes listed in Gene Ontology and identified by literature review (BMP signaling pathway; GO0030509)<sup>213</sup> (**Figure 1B**). This analysis identified two genes, *GREM1* and *ISLR*, as human CRC CAF-specific genes relevant to BMP signaling.

To validate *GREM1* and *ISLR* expression in human CRC CAFs, we performed RNA *in situ* hybridization (ISH) on CRC patient samples and confirmed that both *GREM1* and *ISLR* are highly expressed by fibroblastic cells in the human CRC stroma compared to the normal colorectal stroma (**Figure 1C; Supplementary Figure 2**). Consistent with the microarray data, single-molecule fluorescent ISH (smFISH) for *GREM1* and *ISLR* followed by FAP immunofluorescence showed that *GREM1* and *ISLR* were expressed by FAP<sup>+</sup> CAFs in human CRC sections (**Supplementary Figure 3**). Moreover, the fibroblast-specific expression pattern of *GREM1* and *ISLR* was corroborated by analyses of publicly available single-cell RNA-seq (scRNA-seq) data from human normal colon mucosa and primary CRC<sup>226</sup> (**Supplementary Figure 4A and B**).

We next sought to verify the functional roles of *GREM1* and *ISLR* in the regulation of BMP signaling, using lentivirus-mediated overexpression of *Grem1* or *Islr* in a mouse colonic fibroblast cell line, YH2 cells. *GREM1* and *ISLR* overexpression were detected in the conditioned medium from *Grem1*-overexpressing and *Islr*-overexpressing YH2 cells, respectively, suggesting that *GREM1* and *ISLR* were secreted into the medium (**Figure 1D**). Luciferase assays of BMP-responsive elements revealed that *GREM1* overexpression suppressed, whereas *ISLR* overexpression augmented, BMP signaling (**Figure 1E**). Furthermore, *GREM1* overexpression inhibited BMP7-mediated phosphorylation of Smad1/5, a downstream effector of BMP signaling, thereby preventing the BMP7-induced increase in expression of BMP target genes, *Id2* and *Id4* (**Figure 1F and G**). Conversely, these surrogates for the BMP signaling pathway were increased by *ISLR* overexpression. When *Grem1*-overexpressing YH2 cells were admixed with *Islr*-overexpressing YH2 cells, *GREM1* and *ISLR* counteracted each other's effect on BMP7 signaling (**Supplementary Figure 5**). Similar to the *GREM1* antagonism of BMP7 signaling, *GREM1* overexpression also prevented the BMP2-induced increase in *Id2* and *Id4*. In contrast, *ISLR* overexpression promoted the BMP2-mediated increase in *Id4*, but not *Id2* (**Supplementary Figure 6**). Collectively, these data indicate that two functionally opposing stromal regulators of BMP signaling, *GREM1* and *ISLR*, are upregulated in CRC CAFs and may contribute to fine-tuning of BMP signaling within the CRC stroma.

### ***GREM1 and ISLR expression are upregulated during colorectal carcinogenesis in humans***

Next, we investigated whether expression of *GREM1* and *ISLR* is upregulated during CRC progression. ISH for *GREM1* and *ISLR*, as well as scRNA-seq and expression microarray analyses, revealed that *GREM1* and *ISLR* expression were increased during human colorectal carcinogenesis (**Figure 2A and B; Supplementary Figure 4A and 7A-C**). *GREM1* and *ISLR* upregulation was also observed in the stroma of liver metastases of human CRC compared to the normal liver tissues (**Supplementary Figure 8A-C**). Furthermore, in line with the fibroblast-specific expression of *GREM1* and *ISLR*, analyses of The Cancer Genome Atlas (TCGA) and expression microarray data showed that the highest expression of *GREM1* and *ISLR* was observed in a stroma-rich molecular subtype of CRC (Consensus Molecular Subtype 4; CMS4) (**Supplementary Figure 9A and B**). This is consistent with a recent paper showing that CMS4 tumors displayed the highest *GREM1* transcript levels<sup>219</sup>. Overall, these data suggest that *GREM1* and *ISLR* expression are upregulated in the CRC stroma during colorectal carcinogenesis.

### ***GREM1 and ISLR expression levels are associated with poor and favorable clinical outcomes in patients with CRC, respectively***

To investigate the clinical significance of *GREM1* and *ISLR* expression in CRC CAFs, we evaluated *GREM1* and *ISLR* expression by ISH in 53 rectal cancer surgical samples (**Figure 2C; Supplementary Table 2**). Survival analyses demonstrated that high *GREM1* expression (Score  $\geq 3$ ) and high *ISLR* expression (Score  $\geq 2$ ) were independent prognostic factors for poor and favorable disease-free survival, respectively, in patients with rectal cancer (**Figure 2D: Supplementary Table 2 and 3**).

Furthermore, analysis of expression microarray data from 556 primary colon cancer patient confirmed that the *GREM1*-high and *ISLR*-low groups each independently exhibited poor overall survival (**Figure 2E and Supplementary Table 4**). No patients in this cohort had both *GREM1*-high and *ISLR*-low, suggesting that *GREM1*-high patients and *ISLR*-low patients were two separate patient subgroups (**Supplementary Table 5**). Together, these data indicate that *GREM1* and *ISLR*

expression levels may serve as prognostic biomarkers in human CRC with *GREM1* expression associated with poorer and *ISLR* expression associated with improved survival.

***Grem1*<sup>+</sup> fibroblasts are distinct from *Islr*<sup>+</sup> fibroblasts in the normal mouse colon, with the majority of *Grem1*<sup>+</sup> fibroblasts marked by *Foxl1***

We next sought to characterize the specific stromal cell types expressing *Grem1* and *Islr* in the normal colon. To this end, we performed *Grem1* and *Islr* ISH using normal mouse colons (**Figure 3A and B**). As shown elsewhere<sup>106</sup>, our ISH data confirmed that *Grem1* expression was observed in fibroblastic cells near the base of the colonic crypts in the lamina propria, as well as in muscularis mucosae cells. Interestingly, however, *Islr*<sup>+</sup> fibroblasts were located near the middle of the colonic crypts, suggesting that *Grem1*<sup>+</sup> intestinal fibroblasts were topographically distinct from *Islr*<sup>+</sup> intestinal fibroblasts in the normal colonic crypts.

To further define the fibroblast subpopulations expressing *Grem1* and *Islr*, we studied *Grem1*-CreERT2<sup>46</sup>; Rosa26-LSL-tdtomato mice and *Islr*-CreERT2<sup>213</sup>; Rosa26-LSL-tdtomato mice and co-stained with intestinal fibroblast markers (**Figure 3C and D**). A recent report has illustrated that subepithelial telocytes identified by expression of *Foxl1* and platelet-derived growth factor receptor alpha (*Pdgfra*), provide key intestinal stem cell niche signaling molecules such as Wnts and *Grem1*<sup>227</sup>. Indeed, analysis of RNA-seq data from FACS-purified *Foxl1*-lineage intestinal telocytes and non-*Foxl1*-lineage intestinal mesenchymal cells<sup>227</sup> suggested that *Grem1* expression was observed in *Foxl1*-lineage<sup>+</sup> telocytes, but not in *Foxl1*-lineage<sup>-</sup> cells (**Supplementary Figure 10**). Consistent with this, *Foxl1* smFISH or PDGFR $\alpha$  immunofluorescence (IF) using *Grem1*-CreERT2 mice showed that the majority of the *Grem1*<sup>+</sup> fibroblasts, also known as *Grem1*<sup>+</sup> intestinal reticular stem cells (iRSCs)<sup>46</sup>, expressed the telocyte markers, *Foxl1* and PDGFR $\alpha$ , in the lamina propria of the normal colon (**Figure 3C and D**). In contrast, *Islr*<sup>+</sup> fibroblasts exhibited lower positivity for *Foxl1* and PDGFR $\alpha$  than *Grem1*<sup>+</sup> iRSCs. These data implied a high degree of overlap between *Grem1*<sup>+</sup> iRSCs and *Foxl1*<sup>+</sup> telocytes. This prompted us to investigate whether *Foxl1* might drive *Grem1* expression at the expense of *Islr* expression.

### ***FOXL1 directly upregulates Grem1 transcription while repressing Islr expression in mouse colonic fibroblasts***

To assess the effect of FOXL1 on the regulation of *Grem1* and *Islr* expression, human FOXL1-overexpressing YH2 cells were generated by lentiviral transduction (**Figure 3E**). Consistent with our earlier co-localization analyses, FOXL1 overexpression in YH2 cells induced *Grem1* upregulation at the expense of *Islr* expression, accompanied by decreased *Id2* expression (**Figure 3F**). Similarly, luciferase reporter assays showed that FOXL1 overexpression increased the activity of the *GREM1*-promoter reporter and reduced the activity of the *ISLR*-promoter reporter in comparison to control empty YH2 lines (**Supplementary Figure 11**). Conversely, CRISPR/Cas9-mediated *Foxl1*-knockdown in mouse primary colonic fibroblasts attenuated *Grem1* expression while inducing upregulation of *Islr* and BMP target genes (**Figure 3G and Supplementary Figure 12**).

To explore whether FOXL1 is directly involved in regulating *Grem1* expression, we generated FOXL1-Hemagglutinin (HA) tag-overexpressing YH2 cells (**Figure 3H**) and performed chromatin immunoprecipitation (CHIP) using an anti-HA antibody. CHIP-qPCR showed enrichment of FOXL1-HA binding to the *Grem1* promoter (transcriptional start site +2008 to +2014 bps; an intronic region) compared to immunoglobulin controls (**Figure 3I and J**). Consistent herewith, a luciferase assay using the human *GREM1* promoter region in YH2 cells confirmed that FOXL1-mediated augmentation of *GREM1* expression was abrogated by truncation of the *GREM1* promoter to remove the FOXL1-binding region (**Figure 3K**). Collectively, these data indicate that FOXL1 is recruited to the *Grem1/GREM1* promoter to drive *Grem1* expression, providing mechanistic insight into the overlap between *Grem1*<sup>+</sup> iRSCs and *Foxl1*<sup>+</sup> telocytes.

### ***GREM1<sup>+</sup> CAFs are myofibroblastic CAFs, which are distinct from ISLR<sup>+</sup> CAFs, in mouse and human CRC***

Next, we sought to characterize *GREM1*<sup>+</sup> CAFs and *ISLR*<sup>+</sup> CAFs in the CRC mesenchyme. Using azoxymethane (AOM)/ dextran sulfate sodium (DSS) mouse CRC and human CRC samples, we carried out *Grem1/GREM1* and *Islr/ISLR* smFISH as well as IF for  $\alpha$ SMA, a well-established marker of myofibroblastic CAFs. We found that, both in mouse and human CRC, *Grem1/GREM1*<sup>+</sup>

CAFs were distinct from *Islr/ISLR*<sup>+</sup> CAFs and that  $\alpha$ SMA positivity was higher in *Grem1/GREM1*<sup>+</sup> CAFs than *Islr/ISLR*<sup>+</sup> CAFs (**Figure 4A-G**). A collagen gel contraction assay also revealed that *Grem1*-overexpressing YH2 cells exhibited increased contraction, a hallmark of activated myofibroblasts<sup>3</sup> (**Supplementary Figure 13A and B**). Consistent with our smFISH data, scRNA-seq data from human CRC tissues<sup>226</sup> confirmed that *GREM1* expression levels were inversely correlated with *ISLR* expression (**Supplementary Figure 14A**) and that *GREM1*, but not *ISLR*, transcripts were positively correlated with *ACTA2* expression in CRC CAFs (**Supplementary Figure 14B**). Furthermore, the scRNA-seq dataset revealed that *GREM1* was predominantly expressed by myofibroblasts. High *ISLR* transcripts were observed not only in myofibroblasts but also in Stromal 2 fibroblasts that are characterized by spatial proximity to epithelial cells and high expression of BMP ligands including BMP7<sup>226</sup> (**Supplementary Figure 14C and D**). Interestingly, *GREM1*<sup>+</sup> CAFs were spatially distinct from *ISLR*<sup>+</sup> CAFs in desmoplastic human CRC, with *ISLR*<sup>+</sup> CAFs located in closer proximity to cancer cells than *GREM1*<sup>+</sup> CAFs (**Figure 4H and I**).

As TGF- $\beta$ 1 has well-characterized functions in inducing myofibroblastic differentiation of CAFs<sup>3</sup>, we examined whether TGF- $\beta$ 1 is involved in controlling the differential expression of *Grem1* and *Islr*. Stimulation of YH2 cells with recombinant TGF- $\beta$ 1 increased transcript levels of TGF- $\beta$  target genes, *Serpine1* and *Acta2*, but also *Foxl1* and *Grem1*, while decreasing *Islr* expression. This was rescued by co-treatment with Galunisertib, a specific inhibitor for TGF- $\beta$  receptor 1 (**Figure 4J**). In keeping with the TGF $\beta$ -induced upregulation of *Foxl1* and *Grem1* *in vitro*, *GREM1*<sup>+</sup> CAFs showed a higher degree of co-localization with *FOXL1* in human CRC sections than *ISLR*<sup>+</sup> CAFs did (**Supplementary Figure 15**). Together, our data suggest that TGF- $\beta$ 1 upregulates *Foxl1* in fibroblasts that, in turn, binds to the *Grem1* promoter to upregulate *Grem1* expression. TGF- $\beta$ 1 and *Foxl1* also reduce the expression of *Islr*. This signaling pathway may be involved in the selective development or differentiation of  $\alpha$ SMA<sup>+</sup>*Grem1*<sup>+</sup> myofibroblastic CAFs, which are distinct from *Islr*<sup>+</sup>  $\alpha$ SMA<sup>-</sup> CAFs, in mouse and human CRC.

***Blocking BMP antagonism using a GREM1-neutralizing antibody promotes CRC organoid differentiation and restrains growth***

What relevance do these stromal changes have on the cancer? We investigated whether augmenting BMP signaling either by *Grem1* inhibition or *Islr* overexpression could retard CRC progression. To this end, we took advantage of CRISPR/Cas9 genome engineering and generated luciferase-expressing *Apc*<sup>ΔΔ</sup>, *Trp53*<sup>ΔΔ</sup> mouse CRC organoids (henceforth referred to as AP tumoroids) (**Supplementary Figure 16A and B**). To disrupt BMP signaling, we sequentially mutated *Smad4*, a downstream effector of BMP signaling, to generate Smad4-mutant AP tumoroids (hereafter termed APS tumoroids) (**Supplementary Figure 16C and D**).

Firstly, to test the role of GREM1 in CRC organoid growth, conditioned medium (CM) from *Grem1*-YH2 cells or control GFP-YH2 cells was transferred to either AP or APS tumoroids (**Supplementary Figure 17A**). As expected, the expression of BMP target genes was repressed by CM transfer from *Grem1*-YH2 cells only in AP tumoroids, but not in APS tumoroids (**Supplementary Figure 17B**). Luciferase activity was used to assess an effect of the treatment on viable cell number in luciferase-expressing AP or APS tumoroid cultures. CM from *Grem1*-YH2 cells increased the tumoroid-derived luciferase signals in AP tumoroids (**Supplementary Figure 17C**). This was ameliorated by the loss of *Smad4* in APS tumoroids, indicating that the pro-proliferation effect of GREM1 occurred via antagonism of BMP signaling. Consistent with the role of BMP signaling in promoting epithelial cell differentiation<sup>215,228</sup>, CM from *Grem1*-YH2 cells decreased the expression of *Krt20*, a marker for differentiated CRC cells, in AP tumoroids, but not in APS tumoroids (**Supplementary Figure 17D**).

We next examined whether restoring BMP signaling with a GREM1-neutralizing antibody could repress CRC organoid growth. For this purpose, either a GREM1-neutralizing antibody or an IgG isotype was added to the AP and APS tumoroids incubated in CM from *Grem1*-YH2 cells (**Figure 5A**). The effect of the GREM1-neutralizing antibody to abolish GREM1-mediated BMP antagonism was validated by quantitative reverse-transcription PCR (qRT-PCR) of *Id2* in YH2 cells (**Supplementary Figure 18A and B**). Blocking the antagonism of BMP signaling by GREM1 using the GREM1-neutralizing antibody restored BMP target gene expression (**Figure 5B**) and reduced tumoroid-derived luciferase signals and tumoroid size in AP tumoroids, but not in APS tumoroids (**Figure 5C and D; Supplementary Figure 19A**). Moreover, treatment with the GREM1-neutralizing antibody decreased *Lgr5* expression with a concomitant increase in *Krt20* expression in AP

tumoroids. This effect was abrogated in APS tumoroids (**Figure 5E**). Collectively, our *in vitro* data suggest that restoring BMP signaling with the GREM1-neutralizing antibody promoted differentiation of *Lgr5*<sup>+</sup> intestinal stem cells and attenuated tumoroid growth of *Smad4*-wild-type, but not *Smad4*-mutant CRC.

***Conditioned medium from *Islr*-overexpressing intestinal fibroblasts increases BMP signaling, facilitates CRC organoid differentiation, and attenuates growth***

Next, we tested whether CM from *Islr*-overexpressing YH2 cells could augment BMP signaling in CRC tumoroids and thus inhibit CRC tumoroid growth. Given that ISLR overexpression promoted BMP signaling in the presence of recombinant BMP7 (**Figure 1F and G**), we collected CM from *Islr*- or GFP-overexpressing YH2 cells incubated with recombinant BMP7 and transferred the CM to AP tumoroids (**Figure 5F**). In keeping with our hypothesis of the opposing roles of GREM1 and ISLR, CM from *Islr*-YH2 cells increased *Id1* expression in AP tumoroids, suggesting that ISLR overexpression in fibroblasts enhanced BMP signaling in CRC tumoroids in a paracrine manner (**Figure 5G**). Moreover, medium conditioned by *Islr*-YH2 cells decreased AP tumoroid-derived luciferase signals and tumoroid size (**Figure 5H and I; Supplementary Figure 19B**). Similar to the GREM1-neutralizing antibody, this tumoroid growth inhibition by *Islr*-YH2 cell media was accompanied by a reduction in *Lgr5* transcripts and increased *Krt20* expression in AP tumoroids (**Figure 5J**). Taken together, our *in vitro* studies indicate that modulating stromal, secreted BMP regulators of BMP signaling through either GREM1-neutralizing antibody or *Islr*-overexpression facilitates *Lgr5*<sup>+</sup> intestinal stem cell differentiation and diminishes CRC tumoroid growth.

***Modifying the metastatic niche by AAV8-mediated in vivo overexpression of *Islr* in hepatocytes retards CRC hepatic metastasis***

Hepatic metastasis is the major cause of CRC death<sup>1</sup>, and, through portal vein dissemination, complicates most advanced gastrointestinal adenocarcinomas. Therefore, we explored whether enhancing BMP signaling either by the GREM1-neutralizing antibody or *Islr* overexpression could

impair hepatic metastagenesis of CRC *in vivo*. For this purpose, we generated a mouse model of CRC hepatic metastasis using intraportal injection of AP tumoroids (**Figure 6A**).

Initially, we examined whether treatment with the GREM1-neutralizing antibody could retard CRC hepatic metastasis and improve survival in comparison to an IgG isotype-treated control group (**Supplementary Figure 20A**). Consistent with our earlier observation in human CRC liver metastases (**Supplementary Figure 8A and B**), ISH for *Grem1* confirmed that *Grem1* was expressed by fibroblastic cells in the stroma of mouse CRC hepatic metastases (**Supplementary Figure 20B**). Immunohistochemistry for pSmad1/5/8 demonstrated that the treatment with GREM1-neutralizing antibody restored BMP signaling in metastatic CRC (**Supplementary Figure 20C and D**). The GREM1-neutralizing antibody-treated group showed a trend towards prolonged overall survival and a trend towards decreased tumor growth, albeit that the groups, as dosed in this study, did not show significant differences in survival (**Supplementary Figure 20E and F**).

Thus, we then focused on the other key stromal BMP signaling regulator, *Isr*. Inspired by recent advances in AAV-mediated gene therapy in human diseases<sup>229,230</sup>, we reasoned that augmenting BMP signaling, via ectopic overexpression of *Isr* in hepatocytes, a liver cell type shown to contribute to a metastatic niche<sup>231</sup>, could potentially ameliorate the progression of CRC hepatic metastasis. We injected AAV8 (an AAV serotype with tropism for murine hepatocytes<sup>232</sup>) encoding either *Isr* or, as a control, a red fluorescent protein (mRuby2), via mouse tail vein (**Figure 6A**) to generate ectopic *Isr* overexpression in hepatocytes *in vivo* (**Figure 6B and C**). Two weeks after the tail vein injection, AP tumoroid cells were injected directly into the portal vein to generate CRC hepatic metastases. The liver-directed delivery of *Isr* enhanced BMP signaling in CRC hepatic metastasis as well as in normal hepatocytes (**Figure 6D and E; Supplementary Figure 21A-C**), and significantly prolonged mouse survival compared to the AAV8-mRuby2 control group (**Figure 6F**). Notably, there was no histological evidence of liver injury induced by AAV8-*Isr* (**Supplementary Figure 22A and B**).

To evaluate alterations in growth kinetics and histopathology by AAV8-*Isr*, we next monitored the growth of CRC hepatic metastases with an *in vivo* imaging system (IVIS) and harvested all mice 3 to 4 weeks after tumor injection (**Figure 6A**). In line with improved survival, the AAV8-*Isr* group showed reduced tumor-derived luminescence signal, histological tumor area, and Ki-67 cell

proliferation index (**Figure 6G-L**). Consistent with our earlier *in vitro* data showing that stromal *Islr* overexpression promoted CRC tumoroid differentiation, *Islr*-overexpression in hepatocytes yielded more differentiated CRC histology (**Figure 6M and N**). Furthermore, in agreement with anti-fibrotic roles of BMP7 signaling<sup>213</sup>,  $\alpha$ SMA immunostaining and Picro-Sirius red staining revealed that fibrosis was reduced in the AAV8-*Islr* group (**Supplementary Figure 23A-D**). Taken together, these data suggest that BMP modulation could be an attractive target in CRC metastasis and that leveraging hepatocytes to augment BMP signaling by AAV8-*Islr* could represent an exciting, novel therapeutic opportunity in metastatic CRC.

## Discussion

Initially identified more than 50 years ago<sup>233</sup>, BMP is now known to be important in regulating intestinal epithelial homeostasis and cancer cell proliferation<sup>214,215</sup>. The regulation of BMP in the tumor microenvironment and the role of BMP in tumor management, however, are still largely unknown. In the present study, we have shown that *GREM1* and *ISLR* are CAF-specific factors that exert opposing effects on BMP signaling in colonic fibroblasts and define distinct subpopulations of fibroblasts in the normal colon and CRC. FOXL1 and TGF $\beta$  may explain, at least in part, the polarization of CAFs into tumor-promoting *GREM1*<sup>+</sup> CAFs and tumor-retarding *ISLR*<sup>+</sup> CAFs. Moreover, *GREM1* and *ISLR* expression levels were associated with poor and favorable outcomes in patients with CRC. Using organoid culture and a preclinical mouse model, our data support that BMP signaling imbalance, regulated by *Grem1* and *Islr*, drives CRC progression and is a key target for cancer treatment. We provided the therapeutic proof-of-principle that augmenting BMP signaling, either by using a *GREM1*-neutralizing antibody or AAV8-*Islr*, represents an attractive future approach to treat CRC.

Previous studies have shown that stromal deletion of *BMP2* could facilitate carcinogenesis in CRC and breast cancer, suggesting tumor-suppressive functions of mesenchymal BMP signaling<sup>234,235</sup>. Moreover, augmentation of BMP signaling mediated either by increased BMP ligand or decreased BMP antagonist secretion from the tumor mesenchyme, which occurs as a consequence of stromal Hedgehog signaling activation, restrains bladder or colorectal cancer progression<sup>14,228</sup>. Our work supports these findings and provides novel insights into regulatory mechanisms of BMP signaling in the CRC tumor microenvironment. Given that loss of stromal *BMP2* expression increases cytokine production in a mouse model of breast cancer<sup>234</sup>, further studies are warranted to investigate whether mesenchymal BMP signaling modulated by *GREM1* and *ISLR*, could also be involved in shaping the immunosuppressive tumor microenvironment.

High stromal TGF $\beta$  signaling is associated with worse outcomes for CRC patients<sup>103</sup>. Here, we provide a potential mechanism underlying CAF heterogeneity, initiated by TGF $\beta$  and FOXL1, that polarizes fibroblasts towards *Grem1*<sup>high</sup>*Islr*<sup>low</sup> CAFs or *Islr*<sup>high</sup>*Grem1*<sup>low</sup> CAFs. TGF $\beta$  drives *Foxl1*

expression in CAFs (**Figure 4J**). FOXL1, in turn, directly upregulates *Grem1* expression to antagonize BMP signaling (**Figure 3E-K**) and promote cancer progression by suppressing differentiation of *Lgr5*<sup>+</sup> stem cells while inducing epithelial proliferation. In contrast, relatively low levels of TGF $\beta$  in the microenvironment of some tumors results in lower levels of FOXL1 that permits higher *Islr* expression and a relative tumor-suppressive and higher BMP signaling milieu (**Figure 3E-G and 4J**). This potential TGF $\beta$ -FOXL1-*Grem1*/*Islr*-axis that modulates BMP signaling in the colon provides a novel mechanism to help understand the polarization of CAFs within the tumor microenvironment and presents a promising target for future cancer treatment.

Using scRNA-seq from normal mouse small intestines, one recent paper suggested that *Grem1* is expressed mainly by *Pdgfra*<sup>low+</sup>*Cd81*<sup>+</sup> trophocytes that are distinct from *Foxl1*<sup>+</sup> telocytes<sup>216</sup>. Consistent with this, our *Foxl1* smFISH in *Grem1*-CreERT2 mice showed that there were less *Foxl1* and *Grem1* double-positive fibroblasts in the small intestine than the colon (**Supplementary Figure 24**). This observation raises the possibility of organ-dependent *Foxl1* expression patterns within the context of gastrointestinal fibroblast heterogeneity, which warrants further research.

One limitation of the present study is that we have not unraveled the origins and lineage hierarchy of *Grem1*<sup>+</sup> iRSCs and *Islr*<sup>+</sup> fibroblasts despite presenting a potential mechanism for fibroblast polarization. Whether *Grem1*<sup>+</sup> CAFs and *Islr*<sup>+</sup> CAFs arise from their local progenitors, are recruited from the bone marrow, or are simply new expression profiles within existing cells, requires further investigation. Considering the possible plasticity of CAFs<sup>3</sup>, it is plausible that *Grem1*<sup>+</sup> CAFs and *Islr*<sup>+</sup> CAFs could undergo phenotypic interconversion during tumor development, a state of dynamic flux between a relatively polarized cancer-retarding or cancer-promoting microenvironment.

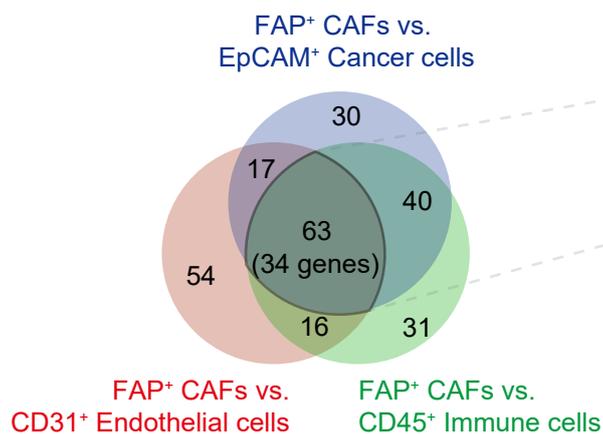
We also provided the first experimental evidence that the GREM1-neutralizing antibody promoted differentiation of *Lgr5*<sup>+</sup> intestinal stem cells and retarded CRC tumoroid growth only in a *Smad4*-wild-type setting, but not in *Smad4*-mutant tumoroids. Our data reinforce the importance of stratifying patients that may benefit from the GREM1-neutralizing antibody according to the mutation status of BMP-related genes such as *SMAD4* and BMP receptors in future preclinical and clinical trials. Our preliminary data with an *in vivo* experimental model of CRC hepatic metastases implied that the GREM1-neutralizing antibody still requires further optimization of therapeutic setting

(metastatic prevention vs. treatment), dosage regimens, route of administration, and in combination with other agents. However, the combined *in vitro* and *in vivo* findings are encouraging.

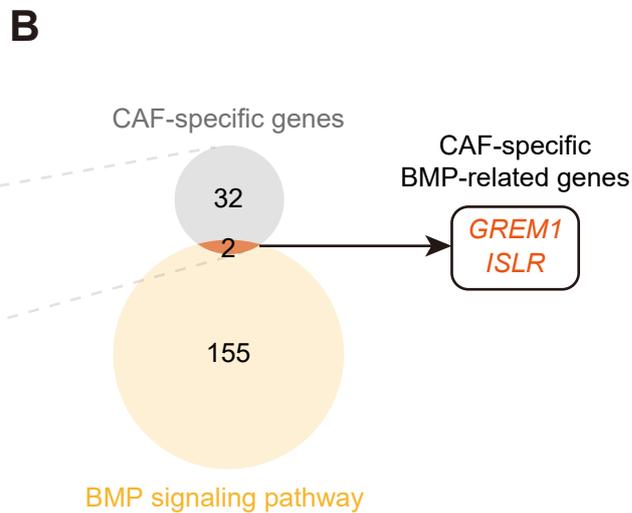
To our knowledge, our study is the first to use AAV8 to target hepatocytes to treat hepatic metastasis<sup>229,236</sup>. Clinical trials have demonstrated that *in vivo* gene delivery by AAV holds great promise in patients with non-neoplastic diseases such as inherited genetic diseases and degenerative neuromuscular disorders<sup>229</sup>. Excitingly, recent clinical trials have shown that hepatocyte-directed gene transfer of coagulation factors by AAV substantially improved clinical symptoms in patients with hemophilia<sup>229,230</sup>. In light of new human liver-tropic bioengineered AAVs<sup>232</sup>, our work suggests that AAV-mediated hepatocyte-directed therapy could, in the future, serve as a novel and well-tolerated cancer therapy. Furthermore, our findings pave the way for AAV-mediated delivery of payloads to modulate not only BMP signaling, but any other number of the relevant biological hallmarks of cancer<sup>8</sup>.

In conclusion, our data show that stromal BMP signaling, inhibited by *GREM1* and promoted by *ISLR*, is biologically relevant in CRC growth, spread, and survival. By targeting the upstream determinants of mesenchymal expression, such as TGF $\beta$  and FOXL1, or by targeting the downstream drivers of BMP signaling, such as *GREM1* and *ISLR*, one may identify new approaches to prevent and to treat cancer.

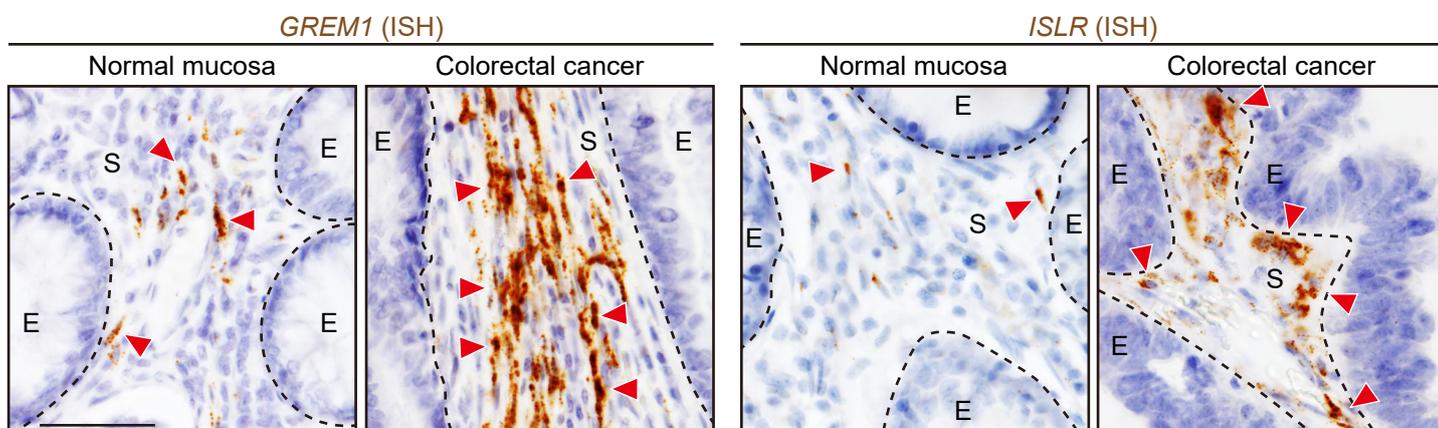
**A** FACS-purified cells from human primary CRC tissues



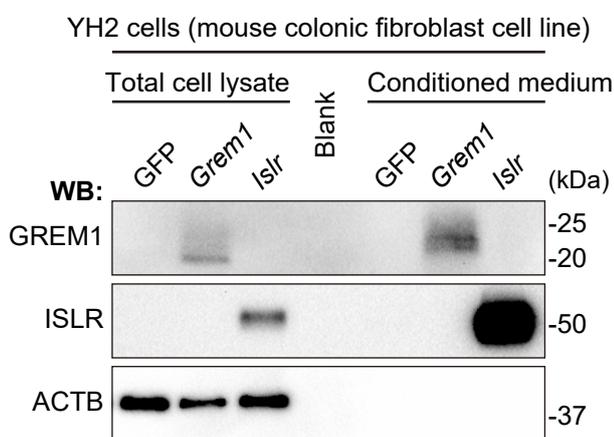
Top 150 differentially upregulated transcripts in each group (GSE39396)



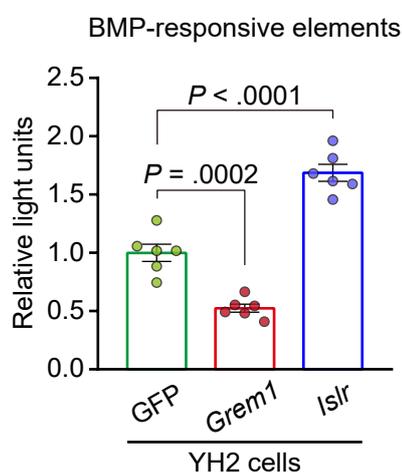
**C**



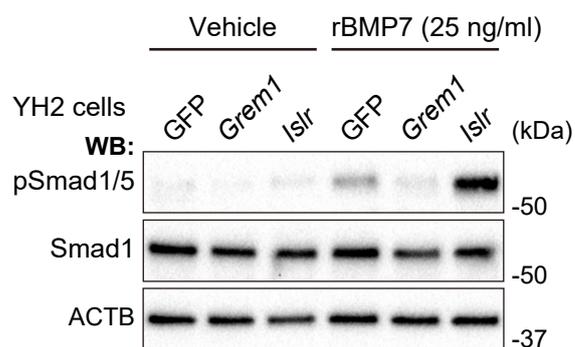
**D**



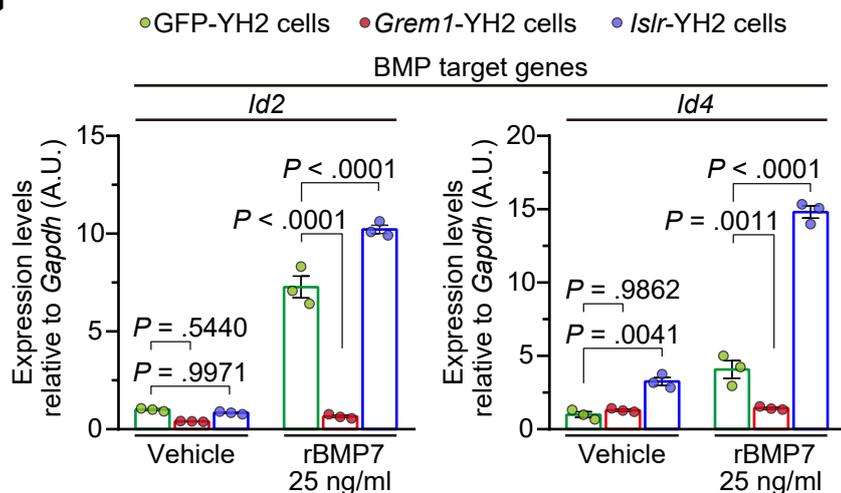
**E**



**F**



**G**



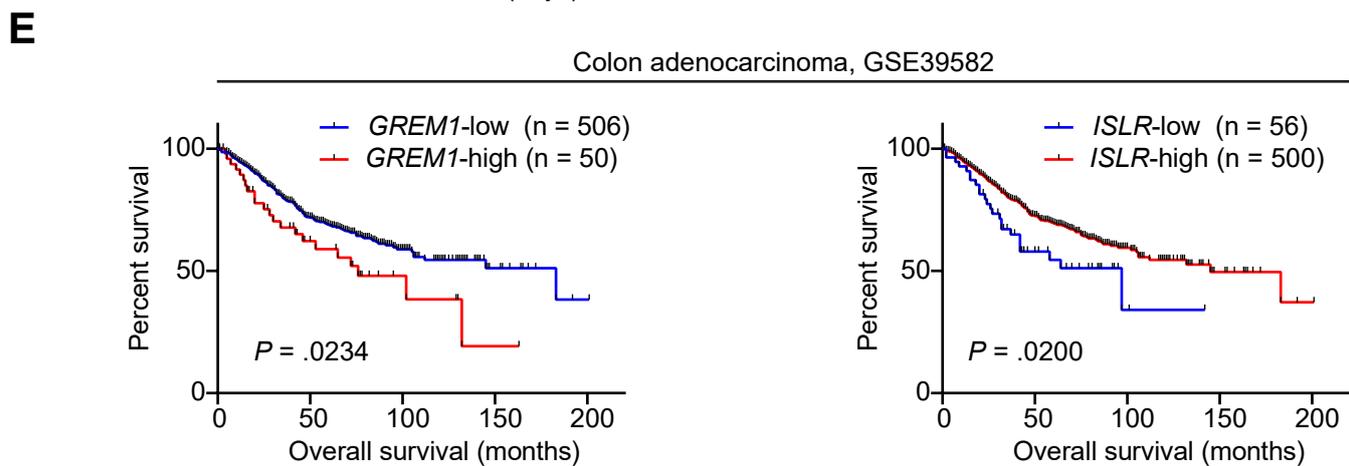
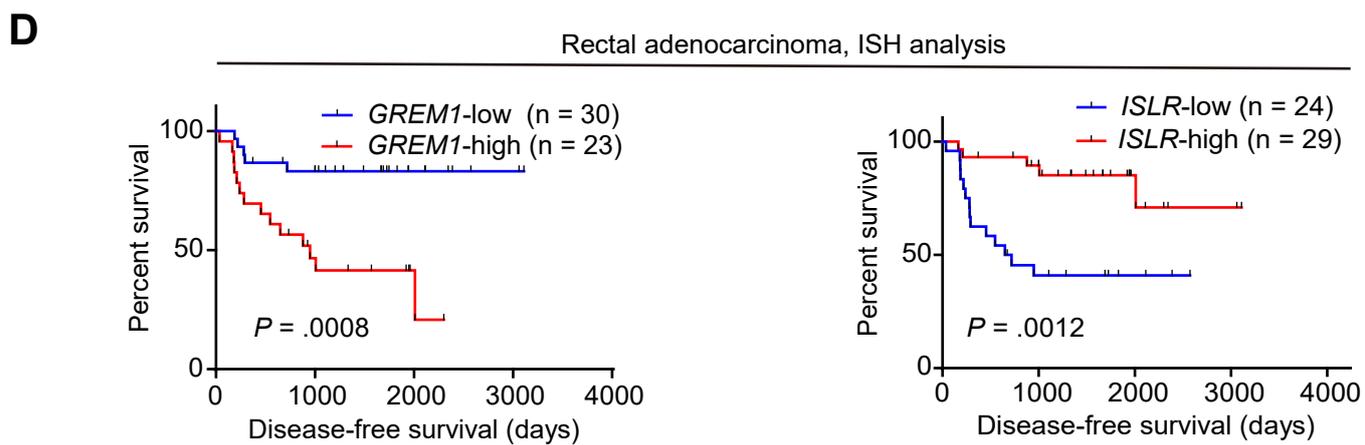
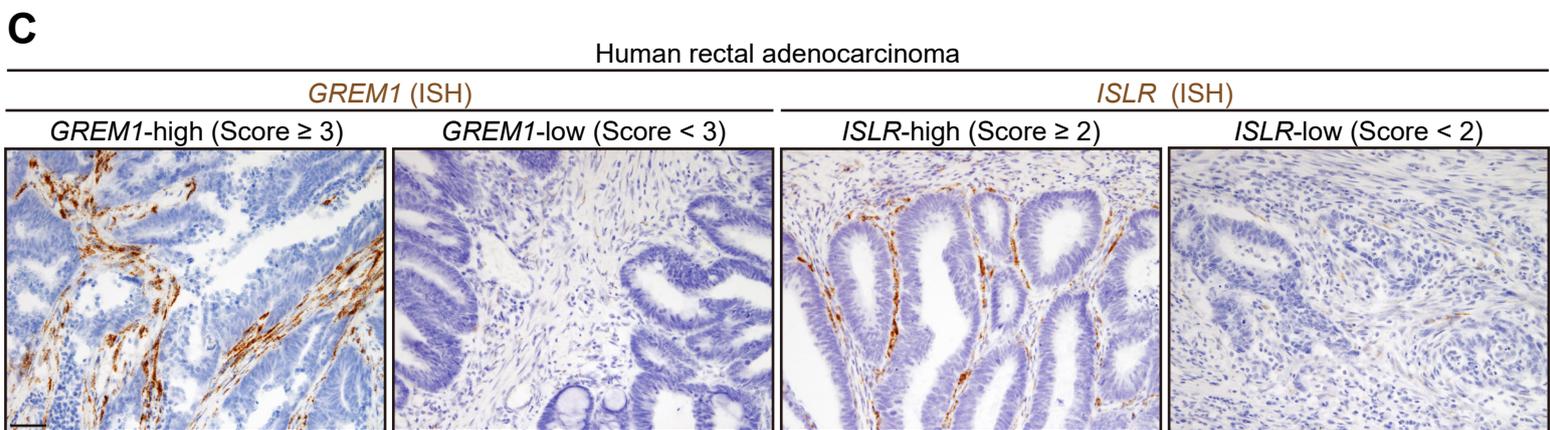
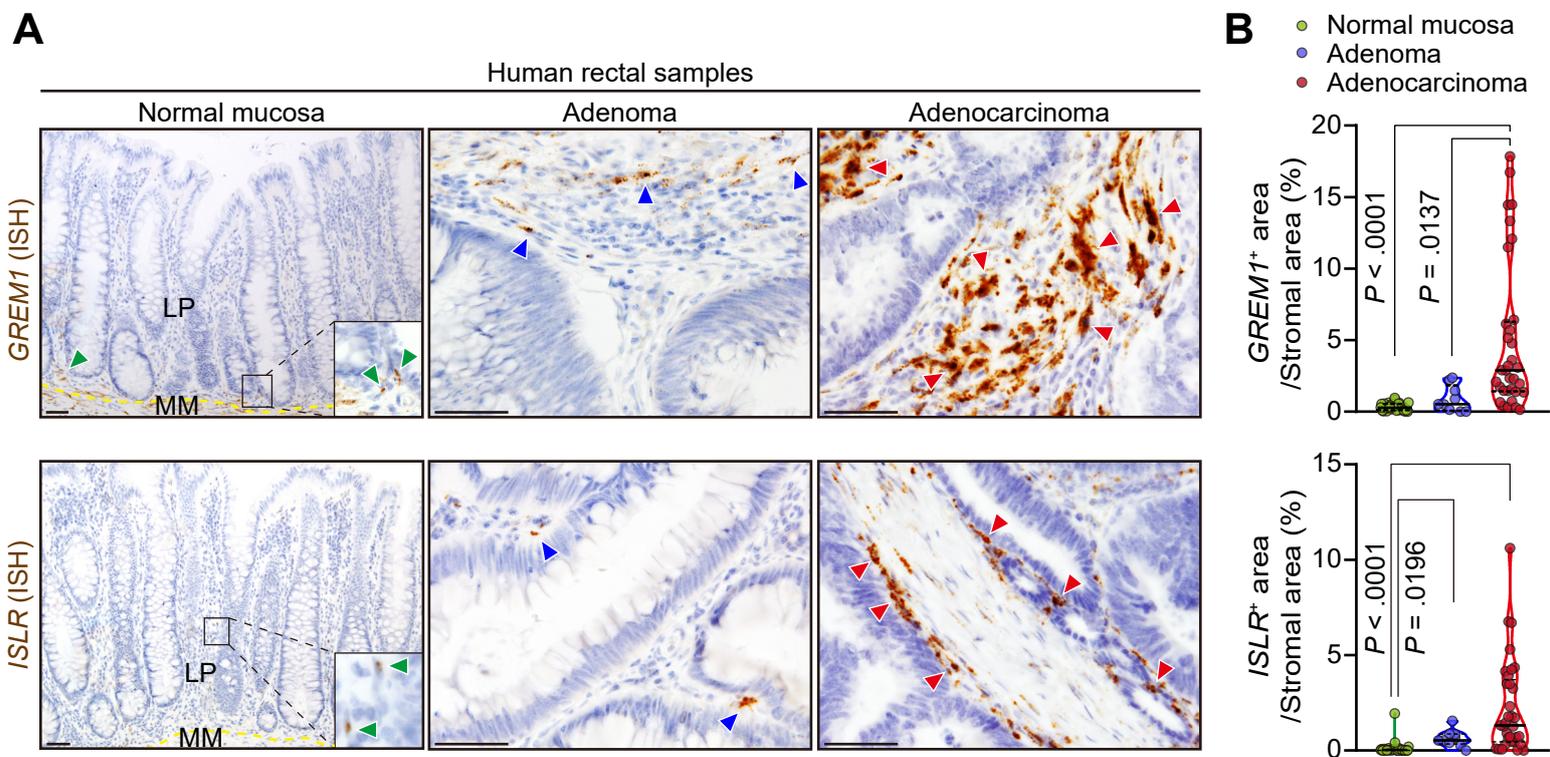
**Figure 1: Identification of *GREM1* and *ISLR* as a BMP antagonist and potentiator, respectively, specifically expressed by CRC CAFs.**

**(A, B)** Analysis of expression microarray data from FACS-purified cells from human primary CRC tissues. **(A)** Venn diagram depicting the overlap of top 150 differentially upregulated transcripts in the three groups as indicated. **(B)** Venn diagram showing the overlap of 34 CAF-specific genes and 157 BMP-related genes identified by Gene ontology of the BMP signaling pathway (GO: 0030509) and Hara *et al.*, *Circ Res*, 2019<sup>213</sup>.

**(C)** *In situ* hybridization (ISH) for *GREM1* and *ISLR* in the human normal colorectal mucosa and CRC. Dotted lines indicate the borders between epithelial cells (E) and the stroma (S). Red arrowheads denote *GREM1* or *ISLR* expression. Scale bar, 50  $\mu$ m.

**(D-G)** Lentivirus-mediated overexpression of *Grem1* and *Islr* in a mouse colonic fibroblast cell line, YH2 cells, represses and augments BMP signaling, respectively. **(D)** Western blotting (WB) showing *Grem1* and *Islr* overexpression in the total cell lysates and conditioned medium. **(E)** Luciferase assays of BMP-responsive elements. n = 6. **(F, G)** YH2 cells were stimulated with recombinant BMP7, followed by WB **(F)** and quantitative reverse-transcription PCR (qRT-PCR). n = 3. **(G)**. A.U., arbitrary unit.

Mean  $\pm$  s.e.m.. One-way ANOVA (E) or two-way ANOVA (G) with post-hoc Tukey's multiple comparisons.



**Figure 2: *GREM1* and *ISLR* expression levels are associated with poor and favorable prognosis in patients with CRC, respectively.**

**(A, B)** ISH for *GREM1* and *ISLR* using human rectal samples. **(A)** Representative images. Yellow dotted lines indicate the borders between the lamina propria (LP) and muscularis mucosa (MM). Green, blue, and red arrowheads denote *GREM1* or *ISLR* expression in the normal mucosa, adenoma, and adenocarcinoma, respectively. **(B)** Violin plots depicting *GREM1* and *ISLR* ISH signal<sup>+</sup> areas in the stroma. 3 high-power fields (HPFs; 400x)/patient; 11 (normal mucosa), 3 (adenoma), 11 patients (adenocarcinoma).

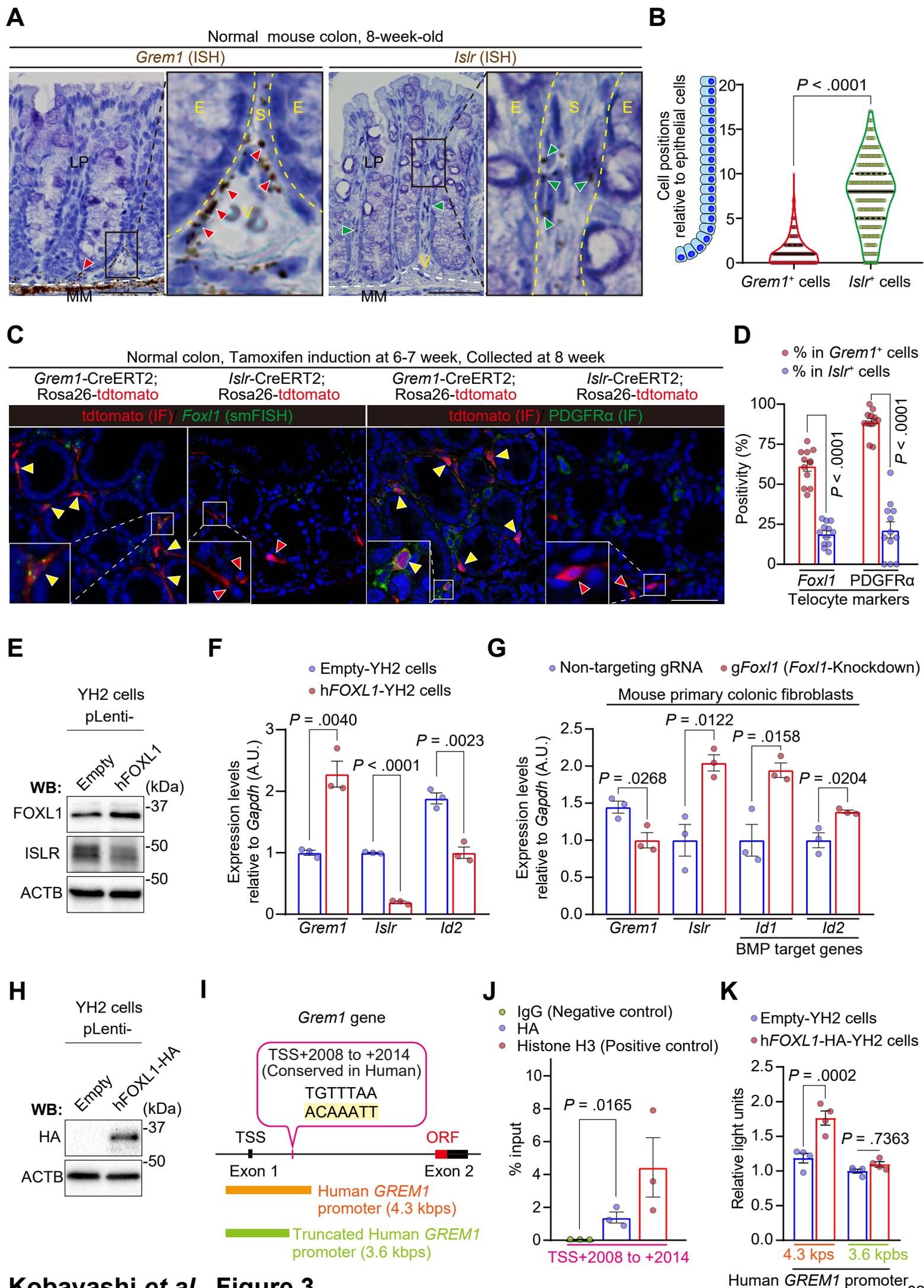
Solid black lines, median; Dotted black lines, quartiles.

**(C, D)** ISH analysis of 53 human primary rectal cancer surgical samples. **(C)** Representative images. Cases with Score  $\geq 3$  and Score  $\geq 2$  were defined as *GREM1*-high and *ISLR*-high, respectively. **(D)** Kaplan-Meier survival curves.

**(E)** Kaplan-Meier survival curves in expression microarray data from 556 primary colon cancer patients.

Kruskal-Wallis test followed by Dunn's post-hoc multiple comparisons (B) and Log-rank test (D and E).

Scale bars, 50  $\mu\text{m}$ .



**Figure 3: *Grem1* and *Islr* identify distinct subpopulations of intestinal fibroblasts in the normal mouse colon and are differentially regulated by FOXL1.**

**(A, B)** ISH for *Grem1* and *Islr* in the adult normal mouse colon. **(A)** Representative images. Red and green arrowheads denote *Grem1*<sup>+</sup> cells and *Islr*<sup>+</sup> cells, respectively. Yellow dotted lines delineate the boundaries between epithelial cells (E) and stromal cells (S). White dotted lines indicate the borders between the lamina propria (LP) and muscularis mucosa (MM). V, blood vessels. **(B)** Violin plots depicting the positions of mesenchymal cells expressing *Grem1* or *Islr* relative to the adjacent epithelial position. 344 *Grem1*<sup>+</sup> cells and 512 *Islr*<sup>+</sup> cells from 20 well-oriented crypts/mouse, 4 mice each. Black solid lines, median; Black dotted lines, quartiles.

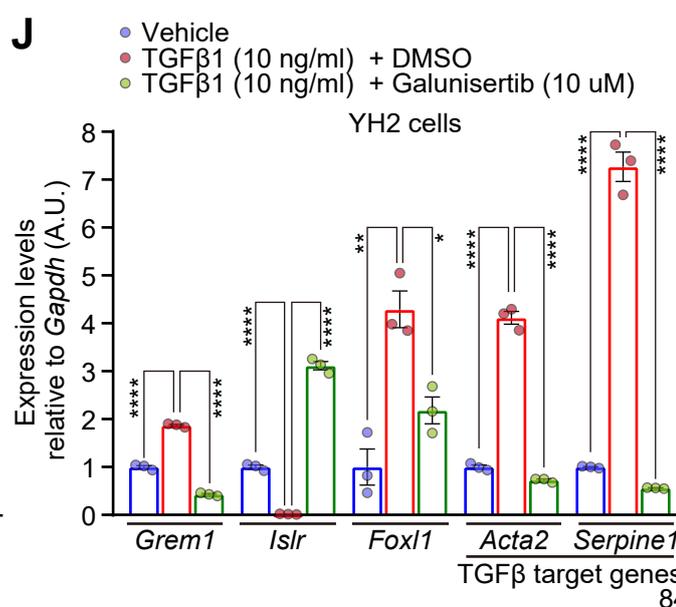
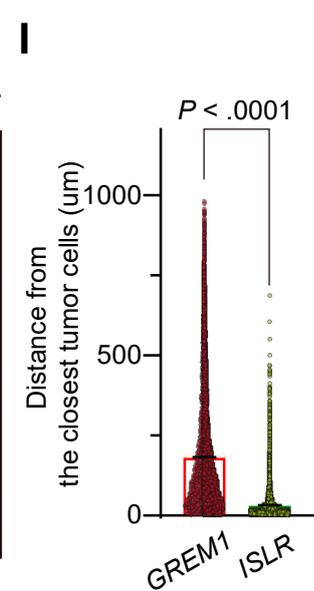
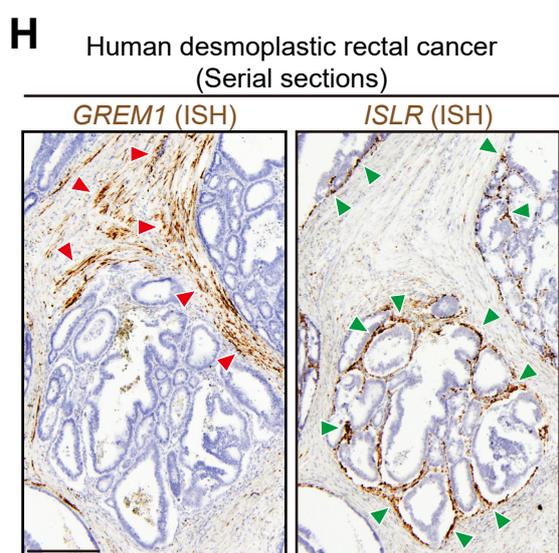
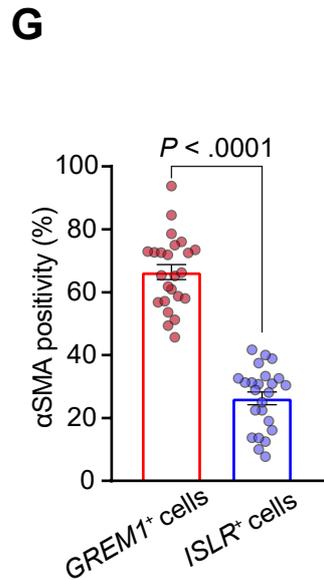
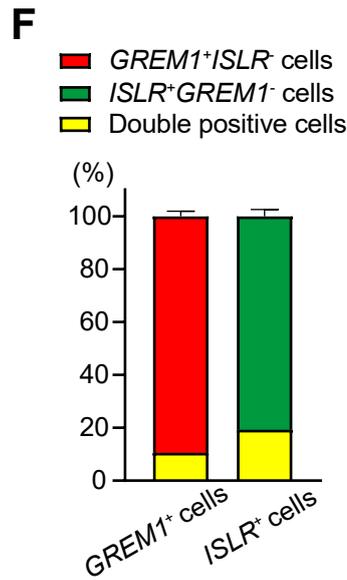
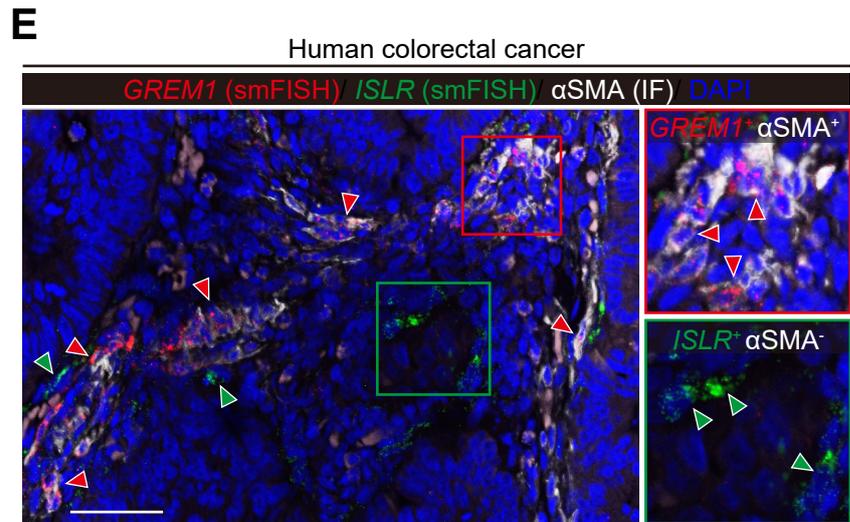
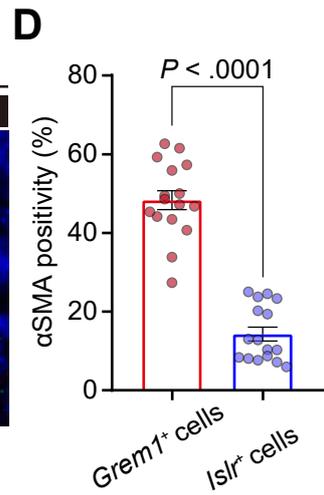
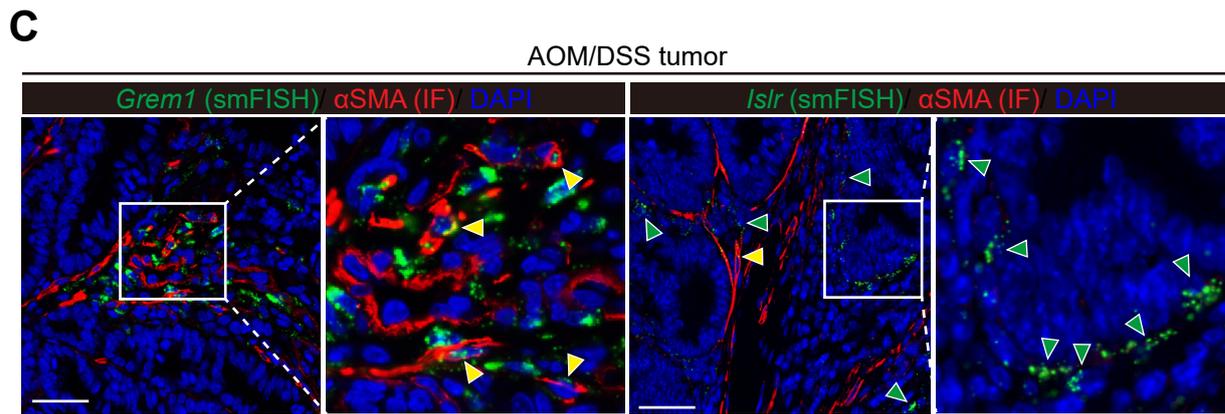
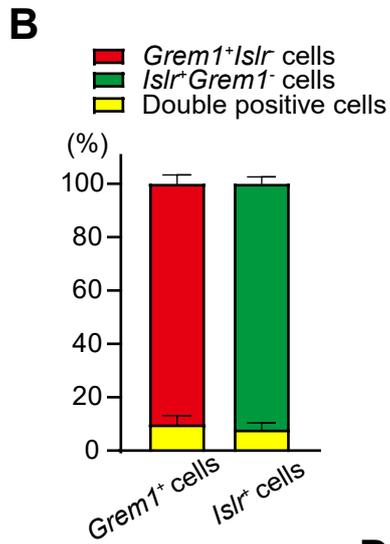
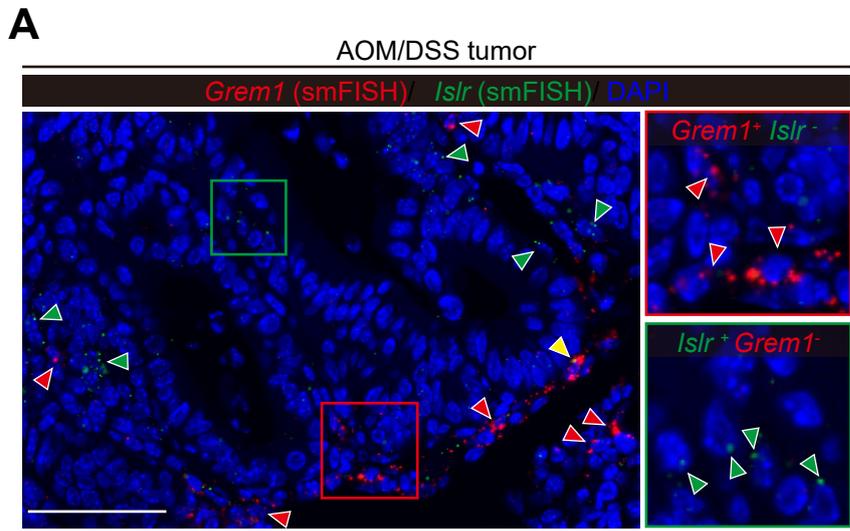
**(C, D)** Single-molecule fluorescent ISH (smFISH) for *Foxl1* and immunofluorescence (IF) for PDGFR $\alpha$  in *Grem1*-CreERT2 mice and *Islr*-CreERT2 mice. **(C)** Representative pictures. Yellow arrowheads indicate double-positive cells (*Grem1*<sup>+</sup>*Foxl1*<sup>+</sup> cells or *Grem1*<sup>+</sup>PDGFR $\alpha$ <sup>+</sup> cells). Red arrowheads denote *Islr*-single-positive cells. **(D)** *Foxl1* positivity and PDGFR $\alpha$  positivity in the *Grem1*<sup>+</sup> cells and *Islr*<sup>+</sup> cells. 4 HPFs (400x)/mouse, 3 mice each.

**(E, F)** Lentivirus-mediated human FOXL1 (hFOXL1) overexpression in YH2 cells induces *Grem1* upregulation and decreases *Islr* expression. **(E)** WB. **(F)** qRT-PCR (n = 3).

**(G)** CRISPR/Cas9-mediated knockdown of *Foxl1* reduces *Grem1* expression while upregulating *Islr* expression in primary mouse colonic fibroblasts as assessed by qRT-PCR (n = 3 mice each).

**(H-K)** FOXL1 interacts with a *Grem1* intron region. **(H)** WB showing hFOXL1-HA overexpression in YH2 cells. **(I)** Schematic representation of a FOXL1-binding site in the mouse *Grem1* intron (highlighted with yellow) and corresponding human *GREM1* promoter regions used in luciferase assays. TSS, transcriptional start site; ORF, open reading frame. **(J)** CHIP-qPCR in YH2 cells (n = 3). **(K)** Luciferase assays of a human *GREM1* promoter (4.3 kbps) and a truncated Human *GREM1* promoter (3.6 kbps) that lacks the FOXL1-binding site (n = 4).

Mean  $\pm$  s.e.m.. Mann-Whitney U-test (B), two-tailed unpaired Student's t-test (D, F, G, and J) and two-way ANOVA with Tukey's post-hoc multiple comparisons (K). Scale bars, 50  $\mu$ m.



**Figure 4: *GREM1*<sup>+</sup> CAFs are myofibroblastic CAFs, which are distinct from *ISLR*<sup>+</sup> CAFs in human and mouse CRC.**

**(A, B)** Dual smFISH for *Grem1* and *Islr* in an AOM/DSS mouse model of CRC. **(A)** Representative pictures. Red, green, and yellow arrowheads denote *Grem1*<sup>+</sup>*Islr*<sup>-</sup> CAFs, *Grem1*<sup>-</sup>*Islr*<sup>+</sup>, and *Grem1*<sup>+</sup>*Islr*<sup>+</sup> CAFs, respectively. **(B)** Semi-quantification of the ratio of double-positive (*Grem1*<sup>+</sup>*Islr*<sup>+</sup>) cells in *Grem1*<sup>+</sup> cells and *Islr*<sup>+</sup> cells. 4 HPFs (400x)/mouse, 4 mice.

**(C, D)** *Grem1* smFISH or *Islr* smFISH followed by  $\alpha$ SMA IF in AOM/DSS tumors. **(C)** Representative pictures. Yellow and green arrowheads indicate double-positive cells (*Grem1*<sup>+</sup> $\alpha$ SMA<sup>+</sup> cells or *Islr*<sup>+</sup> $\alpha$ SMA<sup>+</sup> cells) and *Islr*<sup>+</sup> $\alpha$ SMA<sup>-</sup> cells, respectively. **(D)**  $\alpha$ SMA positivity in *Grem1*<sup>+</sup> cells and *Islr*<sup>+</sup> cells. 4 HPFs/mouse, 4 mice each.

**(E-G)** Dual smFISH for *GREM1* and *ISLR* followed by  $\alpha$ SMA IF in human CRC. **(E)** Representative pictures. Red and green arrowheads denote *GREM1*<sup>+</sup> $\alpha$ SMA<sup>+</sup> cells and *ISLR*<sup>+</sup> $\alpha$ SMA<sup>-</sup> cells, respectively. **(F)** Semi-quantification of the ratio of double-positive (*GREM1*<sup>+</sup>*ISLR*<sup>+</sup>) cells in *GREM1*<sup>+</sup> cells and *ISLR*<sup>+</sup> cells. **(G)**  $\alpha$ SMA positivity in *GREM1*<sup>+</sup> cells and *ISLR*<sup>+</sup> cells.

4-6 HPFs/patient, 5 patients.

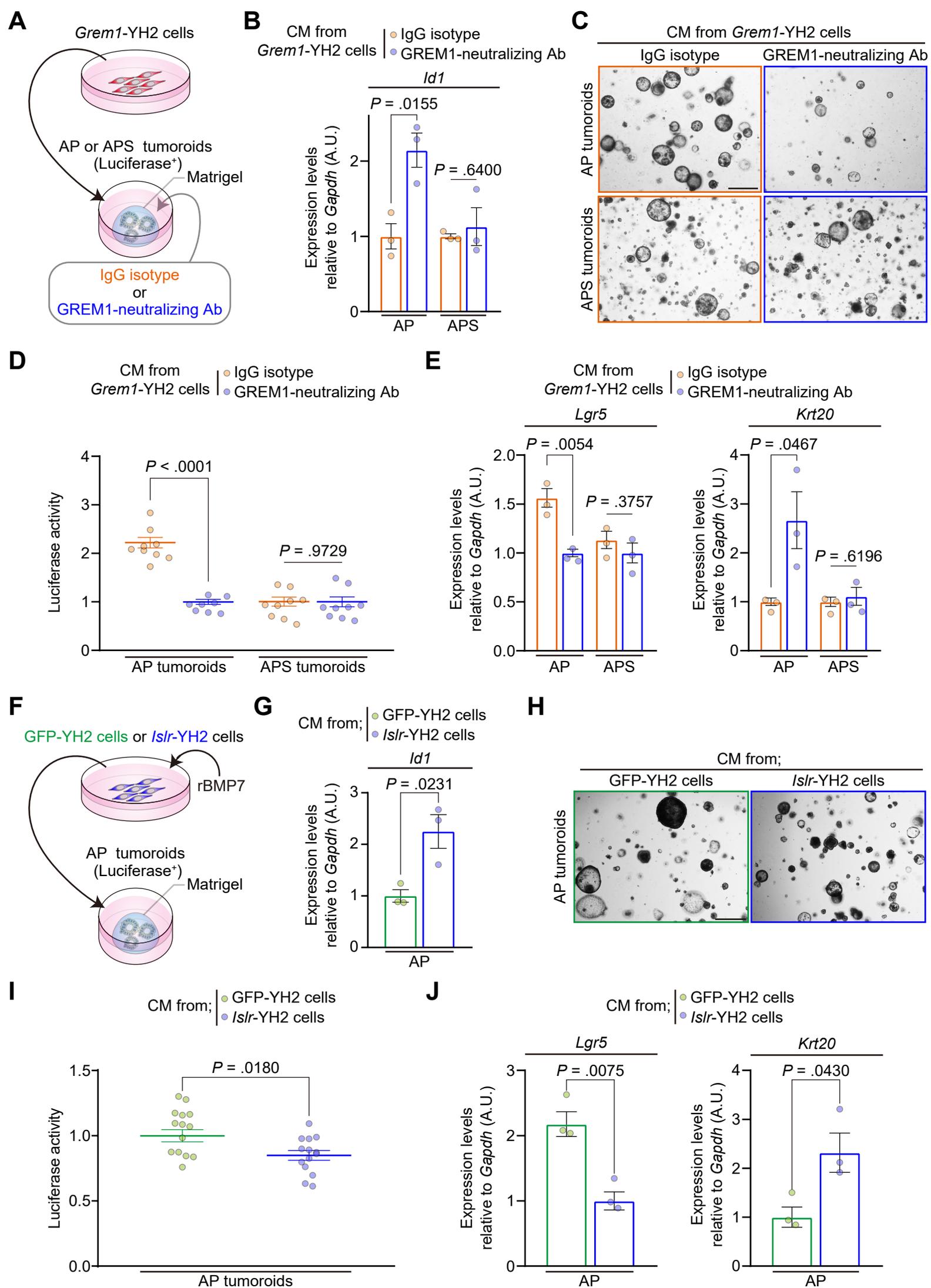
**(H, I)** *GREM1*<sup>+</sup> CAFs are spatially distinct from *ISLR*<sup>+</sup> CAFs in human desmoplastic rectal cancer.

**(H)** Representative pictures of *GREM1* and *ISLR* ISH on human desmoplastic rectal cancer samples. Red and green arrowheads denote *GREM1* and *ISLR* expression, respectively. **(I)** Quantification of the minimum distance between *GREM1* or *ISLR* ISH signals and the closest tumor cells. n = 38396 (*GREM1*) and 18028 DAB<sup>+</sup> signals (*ISLR*) from 3 low-power fields (100x)/patient, 7 patients each.

**(J)** YH2 cells were stimulated with a vehicle, recombinant TGF $\beta$ 1, or recombinant TGF $\beta$ 1 + Galunisertib for 24 hours, followed by qRT-PCR (n = 3).

Mean  $\pm$  s.e.m.. Mann-Whitney U-test (D, G, and I) and one-way ANOVA with Tukey's post-hoc multiple comparisons (J). The boxed areas are magnified in the adjacent panels (A, C, and E).

Scale bars, 50  $\mu$ m (A, C, and E) and 250  $\mu$ m (H). \*\*\*\*, P < 0.0001; \*\*, P = 0.0013; \*, P = 0.0122.



**Figure 5: A GREM1-neutralizing antibody or conditioned medium from *Islr*-overexpressing intestinal fibroblasts restrains CRC tumoroid growth and promotes *Lgr5*<sup>+</sup> stem cell differentiation via increased BMP signaling in tumoroids.**

**(A)** Experimental schematic depicting conditioned medium (CM) transfer from *Grem1*-overexpressing YH2 cells to AP (*Apc*<sup>Δ/Δ</sup> and *Trp53*<sup>Δ/Δ</sup>) tumoroids or APS (*Apc*<sup>Δ/Δ</sup>, *Trp53*<sup>Δ/Δ</sup>, and *Smad4*<sup>Δ/Δ</sup>) tumoroids. Either an IgG isotype or a GREM1-neutralizing antibody was added to the tumoroids.

**(B)** qRT-PCR for *Id1* in AP and APS tumoroids (n = 3).

**(C)** Representative pictures of AP tumoroids and APS tumoroids.

**(D)** Luciferase signals from AP tumoroids and APS tumoroids (n ≥ 8).

**(E)** qRT-PCR for *Lgr5* and *Krt20* in AP and APS tumoroids (n = 3).

**(F)** Experimental schematic depicting CM transfer from *Islr*-overexpressing YH2 cells to AP tumoroids. CM was collected from *Islr*- or GFP-overexpressing YH2 cells incubated with 10 ng/ml of recombinant BMP7 (rBMP7).

**(G)** qRT-PCR for *Id1* in AP tumoroids (n = 3).

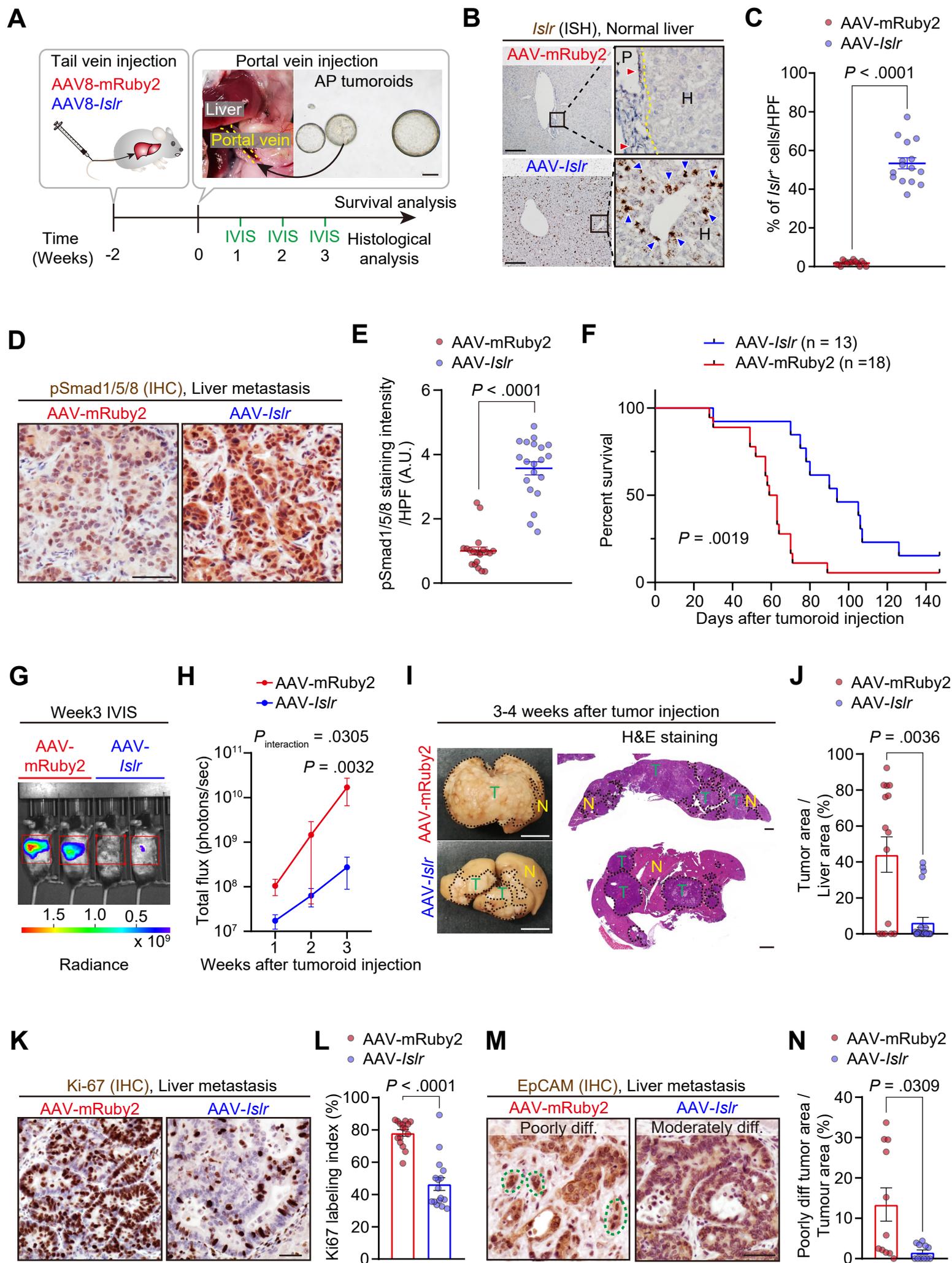
**(H)** Representative pictures of AP tumoroids.

**(I)** Luciferase signals from AP tumoroids (n = 14).

**(J)** qRT-PCR for *Lgr5* and *Krt20* in AP tumoroids (n = 3).

Scale bars, 500 μm. Mean ± s.e.m.. Two-tailed unpaired Student's t-test (B, D, E, G, I, and J).

Note that data normalization was performed within the AP and APS tumoroid groups separately (**B, D, and E**).



**Figure 6: Adeno-associated virus 8-mediated *Islr* overexpression in hepatocytes augments BMP signaling and retards CRC hepatic metastasis growth.**

**(A)** Experimental scheme. Yellow dotted lines outline the portal vein.

**(B, C)** ISH for *Islr* in the liver two weeks after tail vein injection of AAV8-*Islr* or AAV8-mRuby2. **(B)**

Representative images. Red and blue arrowheads denote the endogenous expression of *Islr* in fibroblastic cells in the portal area and ectopic overexpression of *Islr* in hepatocytes, respectively.

The yellow dotted line indicates the border between the portal area (P) and hepatocytes (H) **(C)**

Semi-quantification. 5 HPFs (400x)/mouse, 3 mice each.

**(D, E)** Immunohistochemistry (IHC) for pSmad1/5/8 in liver metastases. **(D)** Representative pictures.

**(E)** Quantification of 3,3'-Diaminobenzidine (DAB) intensity. 5 HPFs/mouse, 4 mice each.

A.U., arbitrary unit.

**(F)** Kaplan-Meier survival curves.

**(G, H)** Luciferase signals from AP tumoroids were assessed by an *in vivo* imaging system (IVIS).

**(G)** Representative images. **(H)** Growth kinetics. Signals within red rectangles in (F) were quantified.

n = 5 (AAV-mRuby2) and 8 (AAV-*Islr*) mice.

**(I, J)** **(I)** Representative macroscopic pictures and hematoxylin and eosin (H&E)-stained sections of liver metastases. Dotted lines indicate borders between tumors (T) and the normal liver (N). **(J)**

Quantification of tumor areas using H&E stained sections. 3 liver pieces/mouse, 5 (AAV-mRuby2)

and 8 (AAV-*Islr*) mice.

**(K, L)** IHC for Ki-67. **(K)** Representative pictures. **(L)** Ki-67 positivity in total epithelial cells.

4 HPFs/mouse, 4 mice each.

**(M, N)** Evaluation of tumor cell differentiation status. **(M)** Representative pictures of IHC for EpCAM.

Green dotted lines indicate tumor budding. diff, differentiated. **(N)** The ratio of poorly differentiated

tumor areas in the total tumor areas. 11 liver pieces each group from 5 (AAV-mRuby2) and 8 (AAV-*Islr*) mice.

Scale bars represent 200  $\mu$ m (A, B), 1 cm (macroscopic pictures in I), 1 mm (H&E staining in I), and

50  $\mu$ m (D, K, and M).

Mean  $\pm$  s.e.m.. Mann-Whitney U-test (C, E, J, L, and N), Log-rank test (F), and two-way repeated-measures ANOVA with post-hoc Sidak's multiple comparison test at Week 3 (H).

## **Acknowledgments**

We thank Kaori Ushida, Kozo Uchiyama, and Riku Takeuchi (Nagoya University, JAPAN) for technical assistance, Vector and Genome Engineering Facility (VGEF), Children's Medical Research Institute (CMRI) (NSW, AUSTRALIA) for producing recombinant AAV vectors, Dr. Andrew Kueh and Dr. Marco Herold (The Walter and Eliza Hall Institute of Medical Research, Victoria, AUSTRALIA) for amplicon sequencing of genetically engineered tumoroids. The Rspo-2 expression plasmid and a mouse colonic fibroblast cell line, YH2, were kind gifts from Professor Antony Burgess (Walter and Eliza Hall Institute of Medical Research, Australia). L-Wnt3a cells were a kind gift from Professor Hans Clevers (Hubrecht Institute, The Netherlands).

# **The balance of stromal BMP signaling mediated by *GREM1* and *ISLR* drives colorectal carcinogenesis.**

Kobayashi *et al.*

## **Supplementary Materials**

1. Supplementary Materials and Methods
2. Supplementary Figures 1-24 and Figure Legends
3. Supplementary Tables 1-5

## Supplementary Materials and Methods

### Human and animal Ethics

Human CRC samples were obtained at the time of surgery from patients who had provided informed consent. This study was approved by the Ethics Committee of Nagoya University Graduate School of Medicine (2017-0127). All animal protocols were approved by the Animal Ethics committees of SAHMRI and Nagoya University Graduate School of Medicine (SAM322, SAM189, and 31434).

### Antibodies

The following antibodies were used in this study. Goat polyclonal anti-GREM1 antibody (1:1000 for Western blotting; WB, AF956, R&D systems), Rabbit polyclonal anti-ISLR antibody (1:1000 for WB, HPA050811, Atlas Antibodies), Mouse monoclonal anti-ACTB (1:5000 for WB, sc-47778, Santa Cruz), Rabbit polyclonal anti-RFP (1:500 for IF, 600-401-379, Rockland), Rabbit monoclonal anti-Phospho-Smad1/5 (1:1000 for WB, clone 41D10, #9516, Cell signaling Technology), Rabbit polyclonal anti-Smad1 (1:1000 for WB, Cell Signaling Technology, #9743), Goat polyclonal anti-mouse PDGFRA (1:50 for immunofluorescence; IF, AF1062, R&D systems), Rabbit polyclonal anti-HA antibody (1:7500 for WB, ab9110, Abcam), Rabbit polyclonal anti-FOXL1 (1:1000 for WB, ABC937, Merck Millipore), Mouse monoclonal anti- $\alpha$ SMA (1:500 for IHC, clone 1A4, Dako), Rabbit monoclonal anti-Smad4 (1:5000 for WB, ab40759, Abcam), Rabbit monoclonal anti-Ki67 (1:200 for IHC, ab16667, Abcam), Rabbit polyclonal anti-EpCAM (1:150 for IHC, abcam71916, Abcam), Sheep polyclonal anti-FAP (1:100 for immunofluorescence, AF3715, R&D systems), and Rabbit polyclonal anti-Phospho-Smad1/5/8 (1:100 for IHC, AB3848-I, Merck Millipore). Mouse monoclonal GREM1-neutralizing antibody (Ab7326) and mouse monoclonal IgG1 isotype (Ab101.4) were provided by UCB Pharma.

### **RNA *in situ* hybridization (ISH)**

All ISH analyses were performed on formalin-fixed and paraffin-embedded human and mouse tissue samples using RNAscope technology (RNAscope 2.5 HD Detection Kit or RNAscope Multiplex Fluorescent Reagent Kit v2, Advanced Cell Diagnostics; ACD) following the manufacturer's instructions. Briefly, tissue sections were baked in a dry oven (HybEZ II Hybridization System, ACD) at 60°C for 1 h and deparaffinized, followed by incubation with Pretreat 1, 2, and 3 (ACD). Slides were incubated with relevant probes for 1 h at 40°C, followed by successive incubations with Amp1 to 6 reagents. Staining was visualized with DAB or TSA Plus Cyanine 5 (NEL745001KT, PerkinElmer). For survival analysis with human CRC samples, ISH samples were quantified by a clinical pathologist in a blinded manner, following a semi-quantitative scoring method recommended by the manufacturer (ACD), where each signal was evaluated as 0 (no staining or < 1 dot/cell), 1 (1-3 dots/cell), 2 (4-9 dots/cell with no or very few dot clusters), 3 (10-15 dots/cell and < 10% dots are in clusters) or 4 (> 15 dots/cell and > 10% dots are in clusters). For other ISH analyses, cells with Score 0 were defined as “negative,” and cells with Score 1 or more were defined as “positive.” In normal colorectal samples, only lamina propria areas were quantified (Figure 2B; Figure 3B and D),

### **Probes used for *in situ* hybridization (ISH)**

*In situ* hybridization (RNAscope) probes used in the study were human *PPIB* (a positive control probe, [NM\\_000942.4](#), region 139 - 989, catalogue number 313901), bacterial *DapB* (a negative control probe, [EF191515](#), regions 414 - 862, catalogue number 310043), human *ISLR* (NM\_005545.3, region 275 - 1322, catalogue number 455481 or 455481-C2), mouse *Islr* (NM\_012043.4, region 763 - 1690, catalogue number 450041 or NM\_012043.4, region 277 - 2225, catalogue number 453321-C2), human *GREM1* (NM\_013372.6, region 175 - 1472, catalogue number 312831 or 312831-C2), mouse *Grem1* (NM\_011824.4, region 398 - 1359, catalogue number 314741), human *FOXL1* ([NM\\_005250.2](#), region 1610 – 2981, catalogue number 558081), and mouse *Foxl1* (NM\_008024.2, region 954 - 1931, catalogue number 407401).

### Luciferase reporter assay

DNA fragments of 4.3 kbp (-1524 to +2841; +1 is the TSS of exon 1 of human *GREM1*) spanning human *GREM1* promoter and intron regions were generated by custom gene synthesis by Gene Universal (New Jersey, US) and inserted into the upstream of the luciferase (NanoLuc) reporter gene in the pNL2.1 vector (Promega). This region contained a putative hFOXL1-binding site at TSS+2103 to +2109 (corresponding to TSS+2008 to +2014 in mouse *Grem1*), as analyzed by the JASPAR database<sup>237</sup>. Sequences from TSS+2103 to +2841, which contains the putative hFOXL1-binding site, was deleted from the *GREM1* promoter (4.3 kbps)-Nanoluc plasmid, resulting in the generation of the truncated *GREM1* promoter (3.6 kbps)-Nanoluc vector.

3.5 kbp human *ISLR* promoter and intron regions (-2321 to +1179; +1 is the TSS of exon 1 of human *ISLR*) were inserted upstream of the luciferase (NanoLuc) reporter gene in the vector pNL2.1 (Promega), as described previously<sup>224</sup>.

The pNL2.1 vectors containing the *GREM1* or *ISLR* promoter regions were co-transfected with pEF1a (Human elongation factor-1 alpha promoter)-Firefly into YH2 cells, using Lipofectamine 2000 (11668019, Thermo Fisher). Cell lysates were collected 48 hours after transfection using a passive lysis buffer (E1941, Promega). Luminescence values of cell lysates were measured using a Nano-Glo Dual-Luciferase Reporter Assay System (N1610, Promega) and a GloMax microplate reader (GM3000, Promega) following the manufacturer's instruction. The Nanoluc luminescence levels were normalized to the Firefly luminescence levels.

To evaluate BMP signaling activation in GFP-, *Islr*-, and *Grem1*-overexpressing YH2 cells, YH2 cells were co-transfected with pGL3 BMP-responsive element-Firefly (Addgene; Plasmid #45126) and pTK-Renilla (E2241, Promega). Cells were grown in DMEM containing 1% FBS, and cell lysates were collected 48 hours after transfection using passive lysis buffer (E1941, Promega). Luminescence values of cell lysates were measured using a Dual-Luciferase Reporter Assay System (E1960, Promega) and a GloMax microplate reader (GM3000, Promega). The Firefly luminescence levels were normalized to the Renilla luminescence levels. Each assay was performed with at least two technical replicates.

### **Promoter analyses and chromatin immunoprecipitation (CHIP)**

Promoter analyses were performed using the JASPAR database<sup>237</sup>. Promoter analyses of the genomic area around the *Grem1* transcriptional start site (TSS) for putative FOXL1-binding sites identified a binding motif for FOXL1 at TSS + 2008 to +2014 bps, an intronic region of the mouse *Grem1* gene, which was conserved in human.

Chromatin was prepared from three independent replicates of *FOXL1*- Hemagglutinin (HA)-overexpressing YH2 cells using a truChIP Chromatin Shearing Kit with Formaldehyde (520154, Covaris) according to the manufacturer's instruction. Chromatin shearing was performed using a Covaris M220 focused ultrasonicator. CHIP assay was performed using EpiQuik Chromatin Immunoprecipitation Kit (P-2002, Epigentek) and IgG isotype (ab171870, Abcam), anti-HA antibody (ab9110, Abcam), anti-Histone H3 antibody (ab1791, Abcam) as per manufacturer's protocol. Prepared DNA samples were analyzed by quantitative polymerase chain reaction (qPCR) using the following primers. Mouse *Grem1* promoter (Forward: 5'-gcaccgttgattaaggctc-3', Reverse: 5'-tgaagatcataggaaagctgtgaag-3')

### **Histopathological examination**

All histopathological examination was performed using formalin-fixed and paraffin-embedded tissues. Paraffin blocks were cut into 4-5  $\mu\text{m}$  sections. Hematoxylin and eosin (H&E) staining and Picro-Sirius red staining were performed by Histology Services (the University of Adelaide) or Sept Sapie (Tokyo, Japan).

### **Immunohistochemistry (IHC) and immunofluorescence (IF)**

Formalin-fixed and paraffin-embedded tissue sections were deparaffinized with xylene and rehydrated with phosphate-buffered saline (PBS), followed by antigen retrieval by boiling samples in antigen retrieval buffer (pH 6 or pH 9, H-3300; Vector Laboratories, S1699; Dako, or S2367; Dako) for 30 min. Inactivation of endogenous peroxidase was performed with 0.5%  $\text{H}_2\text{O}_2$  in methanol for 15 min, followed by washing with PBS. Then, sections were treated with blocking buffer (X0909, Dako) for 30 min, incubated with the indicated primary antibodies overnight at 4°C and washed with

PBS. Sections were incubated with horseradish peroxidase (HRP)-polymer secondary antibody (ab214879 or ab214880, both from Abcam) for 30 min, followed by signal detection with diaminobenzidine (DAB) solution (K3468, DAKO).

For IF studies, deparaffinized sections were treated with blocking buffer (X0909, Dako) for 30 min, incubated with the indicated primary antibodies overnight at 4°C and washed with PBS. Sections were then incubated with Alexa Fluor 488/594/647-conjugated secondary antibodies (All from Thermo Fisher Scientific) for 1 h at room temperature. The sections were then mounted with ProLong Gold antifade reagent containing 4'6-diamidino-2-phenylindole (DAPI; Thermo Fisher Scientific), and fluorescence was examined using a confocal laser-scanning microscope (TCS SP8 MP, Leica) or an inverse immunofluorescence microscope BZ-X710 (Keyence) with optical sectioning.

### **Combined single-molecule fluorescent *in situ* hybridization (smFISH) and immunofluorescence (IF)**

Combined smFISH and IF were implemented by first performing smFISH, followed by IF. After smFISH, the sections were blocked with blocking buffer (X0909, Dako) and then incubated with a primary antibody overnight at 4 °C. The sections were washed in 1x T-PBS 3 times and then incubated with Alexa Fluor 555-conjugated secondary antibody (Thermo Fisher Scientific) for 60 min at room temperature. The sections were then mounted with ProLong Gold antifade reagent containing 4'6-diamidino-2-phenylindole (DAPI; Thermo Fisher Scientific). Fluorescence was examined using an inverse immunofluorescence microscope BZ-X710 (Keyence, Japan) with optical sectioning.

### **Western blot analysis**

Cells were lysed in a lysis buffer (78501, Thermo Fisher) supplemented with cOmplete Protease Inhibitor (Roche) and PhosSTOP Phosphatase Inhibitor cocktails (4906845001, Roche). Lysates were clarified by centrifugation at 12,000 × g for 10 min at 4 °C. Then, sodium dodecyl sulfate (SDS) sample buffer (10 mM Tris-HCl, 2% SDS, 2 mM EDTA, 0.02% bromophenol blue, 6% glycerol; pH 6.8) was added. To prepare conditioned medium and total cell lysates for western blotting, YH2

cells were incubated for 72 hours in protein-free and FBS-free medium (Freestyle 293 expression medium; 12338018, Gibco). The medium was collected, centrifuged at 400 g for 5 minutes, and filtered through a 0.45 µm filter (16533, Sartorius) to remove cell debris. The medium was concentrated 10-fold using Amicon Ultra 3KDa centrifugal filters (UFC500396, Millipore).

The separation was performed by SDS-polyacrylamide gel electrophoresis using a precast gel (4568094, Bio-rad). Proteins were then transferred to Polyvinylidene Difluoride (PVDF) membranes (1620177, Bio-rad) using a semi-dry transfer system (1703940, Bio-rad). The membranes were blocked in 5% milk in PBS containing 0.05% Tween 20, and then incubated with primary antibodies. Proteins were detected by horseradish peroxidase (HRP)-conjugated secondary antibodies (NA9310V, GE Healthcare; N934VS, GE Healthcare; ab97120, Abcam), followed by signal development using an HRP substrate (WBLUR0500, Millipore). The blots were imaged using ChemiDoc MP (Bio-Rad) and quantified using the Image Lab software (Version 6.0.1, Bio-rad).

## Plasmids

The cloning of mouse *Islr* cDNA was described previously<sup>238</sup>. A DNA fragment for the codon-optimized mouse *Grem1* gene was generated using a custom gene synthesis service by Gene Universal (New Jersey, USA). These genes were subcloned into bidirectional pLenti-EF1a (Human elongation factor-1 alpha promoter)-MCS (multiple cloning site)-PGK-Puro vectors, resulting in the generation of pLenti-EF1a-*Islr*-PGK-Puro and pLenti-EF1a-*Grem1*-PGK-Puro.

To overexpress FOXL1 using a lentiviral expression system, DNA fragments for mouse codon-optimized human *FOXL1* gene with either no epitope tags or 3 x Hemagglutinin (HA) tags at the amino-terminus were generated using custom gene synthesis service by Gene Universal. These genes were subcloned into unidirectional pLenti-EF1a-GFP-p2A-Puro-EF1a-MCS vectors, resulting in the generation of pLenti-EF1a-GFP-p2A-Puro-EF1a-*FOXL1* and pLenti-EF1a-GFP-p2A-Puro-EF1a-*FOXL1*-3 x HA.

To enable CRISPR/Cas9-mediated mouse *Foxl1* knockdown, the following mouse non-targeting gRNA (guide RNA) sequence or gRNA sequence targeting mouse *Foxl1* was subcloned into pLentiCRISPR v2 eSpCas9 plasmid by GenScript (New Jersey, USA).

Non-targeting gRNA: 5'- AAAAAGTCCGCGATTACGTC -3'

*gFoxl1*: 5'-GGGCTGTACACGTACAACAG -3'

To overexpress *Islr* using adeno-associated virus (AAV) expression system, mouse *Islr* cDNA was subcloned into a self-complementary AAV plasmid (VPK-430; Cell Biolabs), and a pAAV-CMV-*Islr* vector was generated. As a control, a self-complementary AAV plasmid expressing a red fluorescence protein, mRuby2 (pAAV-CAG-mRuby2; Addgene Plasmid #99123) was used.

### **Lentivirus production and transduction**

293T cells were co-transfected with psPAX2 (Addgene; plasmid #12260) and pMD2.G (Addgene; plasmid #12259), and a lentivirus vector plasmid. At 48 and 72 h after transfection, viral supernatants were harvested, filtered through a 0.45- $\mu$ m filter, and concentrated using Amicon Ultra Centrifugal Filters (Merck Millipore; UFC910024). Concentrated lentivirus particles were used for transduction. 48 h after transduction, positively transduced cells were selected with 2-4  $\mu$ g/mL puromycin if the lentivirus vectors contain a puromycin resistance gene.

### **Retrovirus production and transduction**

293T cells were co-transfected with the pEQ ecotropic-packaging vector<sup>239</sup> and a retrovirus vector plasmid. At 48 and 72 h after transfection, viral supernatants were harvested, filtered through a 0.45- $\mu$ m filter, and concentrated using ultracentrifuge (Optima XPN, Beckman Coulter). Concentrated retrovirus particles were used for transduction.

### **Cell Culture**

A mouse colonic fibroblast cell line, YH2 cells<sup>240</sup>, was generously provided by Professor Tony Burgess (The Walter and Eliza Hall Institute of Medical Research). Normal mouse colonic organoids were isolated from a Rosa26-Cas9 mouse (JAX Stock Number 024858; C57BL/6 x 129 genetic background) housed under pathogen-free conditions in the SAHMRI Bioresources facility. YH2 cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Sigma) supplemented with 10% fetal bovine serum (FBS; Gibco).

All cell lines used were routinely screened for Mycoplasma contamination by MycoAlert Mycoplasma Detection Kit (LT07-118, Lonza).

In Figure 1F, G, and Supplementary Figure 5, 6, YH2 cells were serum-starved in DMEM containing 1% FBS for 48 hours before stimulation with recombinant human BMP7 (PHC9541, Thermo Fisher) or BMP2 (cyt-261, Prospec) at the concentration indicated in figure legends. Proteins and RNAs were collected 30 minutes and 24 hours, respectively, after the addition of the recombinant BMP7 or BMP2. In Figure 4J, YH2 cells were stimulated with a vehicle, 10 ng/ml of recombinant human TGF-beta 1 (240-B-002, R&D systems) + DMSO, or 10 ng/ml of recombinant human TGF-beta 1 + 10 uM of Galunisertib (ADV465749242; AChemBlock), in DMEM containing 1% FBS for 24 hours. In Supplementary Figure 18A and B, recombinant human BMP2 (cyt-261, Prospec), recombinant human GREM1 (provided by UCB pharma), GREM1-neutralizing antibody (Ab7326), and mouse monoclonal IgG1 isotype (Ab101.4) were used for the experiment.

### **Isolation of primary mouse colonic fibroblasts**

Isolation and culture of primary mouse colonic fibroblasts was performed using the protocol described elsewhere<sup>241</sup> with modifications. Briefly, colons were harvested from 8-16-week-old C57BL/6J mice and thoroughly washed with cold PBS. The colon tissues were incubated in PBS supplemented with 3 mM EDTA and 0.05 mM DTT for 60 minutes at room temperature to remove epithelial cells. After washing the colon three times with PBS, the colons were digested in 1 mg of collagenase type IV (17104-019, Gibco), 1 mg of dispase (17105-041, Gibco), and 2000 units of DNase I (D4527, Sigma) in 15 ml of RPMI-5 media (described below) at 37 °C for 30 minutes. The supernatant was filtered through a 70-µm cell strainer, and the cell suspension was plated on T75 flasks. Cells were cultured in RPMI-5 media; Roswell Park Memorial Institute media (R8756, Sigma) supplemented with 5% FBS (Gibco), 2 mM L-Glutamine, 1% Penicillin/Streptomycin, 10 mM HEPES, 1 mM sodium pyruvate, and 0.05 mM 2-mercaptoethanol<sup>241</sup>. At passage 2, cells were transduced with lentivirus encoding Cas9 protein and non-targeting guide RNA (gRNA) or gRNA targeting *Foxl1*. RNA and protein collection was performed at passage 4.

## Organoid culture and organoid genome editing

The basal culture medium for mouse colon organoids was Advanced Dulbecco's modified Eagle medium/F12 (Life Technologies) supplemented with 1x gentamicin/antimycotic/antibiotic (Life Technologies), 10mM HEPES, 2mM GlutaMAX, 1 x B27 (Life Technologies), 1 x N2 (Life Technologies). The following niche factors were used: 50 ng/ml mouse recombinant EGF (Peprotech), 100 ng/ml mouse recombinant noggin (Peprotech), 20% R-spondin-2 conditioned medium, 50% Wnt-3A conditioned medium. Organoids were plated in 50 ul growth factor-reduced Matrigel (356231, Corning) on a 24-well dish, and 500 ul of the medium was added to organoids

Guide RNAs (gRNAs) specific for each target gene were either previously published or designed de novo using the CRISPR design Tool<sup>242</sup>. *Apc* and *p53* gRNA oligos were cloned into pLentiGuide-CMV-dtomato (Modified from; Addgene Plasmid #17452). *Smad4* gRNA oligos were cloned into plentiGuide-Puro (Addgene Plasmid #52963).

gRNA sequences

*Apc*: 5'- GGAAGCCTTGTGGGACATGG -3'

*Trp53*: 5'- GTGTAATAGCTCCTGCATGG -3' <sup>243</sup>

*Smad4*: 5'- CAAAAGCGATCTCCTCCCGA -3' <sup>244</sup>

Normal colon organoids isolated from a Rosa26-Cas9 mouse (JAX Stock Number 024858)<sup>243</sup> were transduced with lentivirus expressing sg*Apc* and sg*Trp53*. 3 days later, media was changed to ADMEM (no Wnt, no Rspo, + Nutlin-3a; SML0580 Sigma) to enrich for correctly targeted clones (Supplementary Figure 16). Following the confirmation of correctly targeted monoclonal lines by amplicon sequence, retrovirus expressing Firefly (Retro-SFG-NES-HSV1-tk-GFP-Luciferase)<sup>245</sup> was transduced to the AP (*Apc*<sup>Δ/Δ</sup>, *Trp53*<sup>Δ/Δ</sup>) tumoroid line. Then, a monoclonal line that showed high luciferase signals was selected by evaluating the luciferase signals by a luminometer. The luciferase-expressing AP monoclonal tumoroid was further transduced with lentivirus expressing sg*Smad4* and a puromycin resistance gene. 3 days later, media was changed to ADMEM (no Wnt, no Rspo, no Noggin, + Puromycin + Nutlin-3a; SML0580 Sigma) to enrich for correctly targeted clones, followed by handpicking monoclonal lines.

Handpicked monoclonal organoid lines were screened for loss of function insertions/deletions using amplicon sequence using the following primers, as described<sup>246</sup>. Overhang sequences are underlined.

*Apc* forward: 5'- CTGAGACTTGACATCGCAGCTTAATTCAGGCAAATCCTAAGAGAG -3'

*Apc* reverse: 5'- GTGACCTATGAACTCAGGAGTCGGTCTGTTTGCCATGAGATTCC -3'

*Trp53* forward: 5'- GTGACCTATGAACTCAGGAGCTAGTGAGGTAGGGAGCGACTTC -3'

*Trp53* reverse: 5'- CTGAGACTTGACATCGCAGCCCAAAGAGCGTTGGGCATGTG -3'

*Smad4* forward: 5'- CTGAGACTTGACATCGCAGCCTGGTGCTCCATTGCTTACT -3'

*Smad4* reverse: 5'- GTGACCTATGAACTCAGGAGTCACTTAATTCCTCGATATTTAAGCTC -3'

To detect a large deletion in *Smad4* in APS tumoroids, PCR was performed using the following primers. *Smad4* forward, 5'- TTGTGTCAGCTCAGAGTGGGTC -3', *Smad4* reverse: 5'- GCAAACCACACGACGATGC -3'. After gel electrophoresis and purification of PCR products, Sanger sequencing for the PCR products was performed to examine mutations.

### Conditioned medium (CM) experiments

In Figure 5A-E and Supplementary Figure 17A-D, monolayer GFP-overexpressing or *Grem1*-overexpressing YH2 cells were cultured for 72 hours in Dulbecco's Modified Eagle Medium (DMEM; Gibco) containing 1% fetal bovine serum (FBS; Gibco), 1% l-glutamine, and 1% Penicillin/Streptomycin. In Figure 5F-J, monolayer GFP-overexpressing or *Islr*-overexpressing YH2 cells were cultured for 72 hours in DMEM containing 1% FBS, 1% l-glutamine, and 1% Penicillin/Streptomycin with 10 ng/ml recombinant human BMP7 (PHC9541, Thermo Fisher). The medium was collected and filtered through a 0.45 µm filter (16533, Sartorius) to remove cell debris. 30 minutes after trypsinized tumoroid fragments equivalent to 2500 cells were plated in 50 µl growth factor-reduced Matrigel (356231, Corning) on a 24-well dish, 500 µl of the CM was added to tumoroids. The medium was changed to new CM every three days. In Figure 5A-E, a mouse GREM1-neutralizing antibody (Ab7326; 100 µg/ml) or a mouse IgG1 isotype (Ab101.4; 100 µg/ml)

was added to each well every time CM was added to the tumoroids throughout the course of experiments.

Organoid area measurements, luciferase assays, and qRT-PCR were performed 8 days and 12 days after plating AP tumoroids and APS tumoroids, respectively, in experiments using *Grem1*-overexpressing YH2 cells. In experiments using *Isl1*-overexpressing YH2 cells, organoid area measurements, luciferase assays, and qRT-PCR were performed 7 days after plating AP tumoroids

### **Luciferase assays for Firefly-expressing tumoroids**

For luciferase assays using Firefly-expressing tumoroids (AP tumoroids and APS tumoroids), following the removal of medium, tumoroids and Matrigel were lysed with 200 ul of 2 x passive lysis buffer (E1941, Promega). Firefly luminescence values of cell lysates were measured using Luciferase Assay Reagent II (E1960, Promega) and a GloMax microplate reader (GM3000, Promega) following the manufacturer's instruction.

### **Tumoroid area quantification**

The images of each well (one picture/independent replicate) were captured using an inverted microscope (IX53, Olympus; a 2 x objective lens). Each Organoid was quantified using the ImageJ software<sup>247</sup> to outline an organoid shape and measure an organoid area.

### **Quantitative PCR**

Total RNA was extracted using TRIzol reagent (Invitrogen) and an RNeasy Mini Kit (Qiagen). Purified RNA samples were reverse-transcribed using Transcriptor Universal cDNA Master (Roche) according to the manufacturer's instruction, followed by the dilution of cDNA at 1:5. Quantitative reverse-transcription PCR (qRT-PCR) of the complementary DNAs (cDNAs) was performed with KAPA PROBE or SYBER FAST qPCR Master Mix (KAPABiosystems) and was run on a QuantStudio 7 Flex Real-Time PCR System (Thermo Scientific). The data were analyzed using the  $2^{-\Delta\Delta Ct}$  method and normalized to *Gapdh* expression levels.

Probes or primers (generated by IDT) used in this study are as follows. Primers were designed using the Primer-BLAST online program<sup>248</sup>.

**Probes:**

Mouse *Grem1* (Mm.PT.58.11631114), Mouse *Islr* (Mm.PT.58.12037488), Mouse *Id1* (Mm.PT.58.6622645.g), Mouse *Id2* (Mm.PT.58.13116812.g), Mouse *Id3* (Mm.PT.58.29482466.g), Mouse *Id4* (Mm.PT.58.6851535), Mouse *Acta2* (Mm.PT.58.16320644), and Mouse *Serpine1* (Mm.PT.58.6413525), Mouse *Gapdh* (Mm.PT.39a.1), Mouse *Lgr5* (Mm.PT.58.12492947), Mouse *Krt20* (Mm.PT.58.43092140).

**Primers:**

Mouse *Grem1* (Forward: 5'- GCTCTCCTTCGTCTTCCTC-3' , Reverse: 5'- AGTGTATGCGGTGCGATTC-3'), Mouse *Islr* (Forward: 5'-TGCGAGCAATCCAGTCCTTAGATG-3' , Reverse: 5'-AGCCCAACAAAGCAGGCACAG-3'), Mouse *Foxl1* (Forward: 5'- GTCGCTCAACGAGTGCTTCG-3' , Reverse: 5'-TGCGCCGATAATTGCCGTTC-3'), and Mouse *Gapdh* (Forward: 5'-CCTCGTCCCGTAGACAAAATG-3' , Reverse: 5'- TGTAGTTGAGGTCAATGAAGGG-3')

**Analyses of publicly available cDNA gene expression microarray and RNA-seq datasets**

For GSE39396<sup>103</sup>, microarray data were downloaded using GEOquery R package version 2.54.1. Differential gene expression was analyzed as described in the limma software manual<sup>249</sup>. Briefly, a linear modeling approach was employed by the function *lmFit*, followed by the empirical Bayes estimation through functions *eBayes* and *topTable*. Differentially upregulated genes were defined as genes that showed log<sub>2</sub> fold change greater than or equal to 2 and false discovery rate less than 0.05. The top 150 differentially upregulated gene probes in FAP<sup>+</sup>CAFs, in each group (FAP<sup>+</sup> CAFs vs. EpCAM<sup>+</sup> cancer cells, FAP<sup>+</sup> CAFs vs. CD31<sup>+</sup> endothelial cells, and FAP<sup>+</sup> CAFs vs. CD45<sup>+</sup> immune cells), were selected based on log<sub>2</sub> fold change. Volcano plots were generated using the *ggplot2* package (version 3.2.1). A human gene list in a Gene Ontology (BMP signaling pathway; GO0030509) was downloaded from AmiGO 2 (<http://amigo.geneontology.org/amigo/search/ontology>).

Expression microarray data (GSE17538, 39582, and 41258) and RNA-seq data (GSE94072) were obtained from Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>). For survival analyses, gene expression levels for the following probes were used. 218469\_PM\_s\_at for *GREM1* and 207191\_PM\_s\_at for *ISLR*. Patient information in GSE41258<sup>250</sup>, GSE39582<sup>251</sup>, and GSE17538<sup>252</sup> were also obtained from GEO. False discovery rate (FDR)-adjusted P-values were calculated by GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r>). In GSE41258, normal colon samples, polyp (adenoma) samples, primary colon tumor samples, and liver metastasis samples included in a previous analysis<sup>250</sup> were analyzed by choosing samples with a label of “Included in analysis; Yes” on the GEO website. Optimal cutoff levels for survival analyses were determined using the X-tile software<sup>253</sup>. Patients with survival time zero (Events at the time of patient enrolment) were not included in survival analysis by a Prism 8 software (Graphpad; [https://www.graphpad.com/guides/prism/8/statistics/stat\\_qa\\_survival\\_analysis.htm](https://www.graphpad.com/guides/prism/8/statistics/stat_qa_survival_analysis.htm)). In GSE94072, fragments per kilobase of exon per million fragments (FPKM) data were directly downloaded from the GEO website (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94072>).

### **The analyses of *GREM1* and *ISLR* expression levels in the Consensus Molecular Subtypes (CMS) of CRC.**

For TCGA data, the CMS labels were downloaded from the supplementary materials of previous publication<sup>31,254</sup>. Expression values and sample IDs were downloaded from the GDC data portal (<https://portal.gdc.cancer.gov/>). Expression Z-scores for tumor samples were calculated based on the standard deviation and mean value for the normal colorectal tissues. Violin plots were then generated using the ggplot2 package (version 3.2.1).

The CMS classifier<sup>31</sup> was applied to the downloaded GSE39582 expression microarray data to classify the colon cancer patients into consensus molecular subtypes (CMSs). This allowed us to classify 566 colon cancer samples into 70 CMS1 tumors, 177 CMS2 tumors, 97 CMS3 tumors, 107 CMS4 tumors, and 115 unclassified tumors.

### **Analyses of publicly available single-cell RNA-seq data**

ScRNA-seq data (GSE132465)<sup>226</sup> were analyzed using BBrowser2 (Version 2.6.4; BioTuring, US). Unique molecular identifier (UMI) counts available on the GEO website (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) were used for normalization by the software. Processed scRNA-seq data (GSE81861)<sup>255</sup> were directly downloaded from the GEO website, and Log<sub>2</sub>(FPKM + 1) values were used to plot *GREM1* and *ISLR* expression.

In Supplementary Figure 4A, fibroblasts include myofibroblasts, stromal 1, stromal 2, and stromal 3 fibroblasts<sup>226</sup>. Endothelial cells include tip-like endothelial cells, stalk-like endothelial cells, proliferative endothelial cells, and lymphatic endothelial cells<sup>226</sup>. In Supplementary Figure 4A and B, immune cells comprise T cells, B cells, myeloid cells, and mast cells<sup>226,255</sup>. In Spearman correlation analyses, given our focus on *GREM1*- or *ISLR*-expressing CAFs, 1156 CRC CAFs that show detectable *GREM1* or *ISLR* transcripts (log normalized *GREM1* expression > 0 or log normalized *ISLR* expression > 0) were analyzed among all CRC CAFs (1501 cells; myofibroblasts, Stromal1, Stromal 2, and Stromal 3 fibroblasts from CRC tissues). Log normalized expression values in the 1156 CRC CAFs were transformed to Z-scores in the Spearman correlation analyses.

### **Tamoxifen administration to mice**

*Grem1*-CreERT2 (JAX Stock No. 027039); Rosa26-LSL (LoxP-stop-LoxP)-tdtomato (JAX Stock No. 007909) and *Islr*-CreERT2<sup>224</sup>; Rosa26-LSL-tdtomato were subjected to tamoxifen administration. 6mg of tamoxifen (T5648, Sigma) dissolved in peanut oil was administered by oral gavage 4 times every other day to 6-week-old adult mice to sufficiently induce CreERT2-mediated recombination, followed by harvest of the 8-week-old mice. The colons and small intestines were removed, opened longitudinally, and fixed in 10% formalin overnight at room temperature. Tissues were subsequently dehydrated and paraffin-embedded. 4 μm sections were used for histological analysis (ISH/IF and IF).

### **Mouse genotyping**

Genomic DNA extracted from mouse ear or tails was used for PCR genotyping. The sequences of the primers were as follows:

*Islr*-Wild-Type (WT) forward, 5'-ACACACGACCTTGGCAAGTCCCAGC-3'; *Islr*-WT reverse, 5'-GTCTGCAATCTGGAAGCCATACTTCTCC-3';

*Islr*-CreERT2 forward, 5'-ACACACGACCTTGGCAAGTCCCAGC-3'; *Islr*-CreERT2 reverse, 5'-CGATCCCTGAACATGTCCATCAGG-3';

*Grem1*-CreERT2 forward, 5'-TTAATCCATATTGGCAGAACGAAAACG-3'; *Grem1*-CreERT2 reverse, 5'-CAGGCTAAGTGCCTTCTCTACA-3'.

Rosa26-LSL-tdtomato forward, 5'-AGATCCACCAGGCCCTGAA-3'; Rosa26-LSL-tdtomato reverse, 5'-GTCTTGAAGTCCACCAGGTAG-3';

Rosa26-Cas9 forward, 5'-CGTCGTCCTTGAAGAAGATGGT-3'; Rosa26-Cas9 reverse, 5'-CACATGAAGCAGCACGACTT-3';

Rosa26-WT forward, 5'-TTCCCTCGTGATCTGCAACTC-3'; Rosa26-WT reverse, 5'-CTTTAAGCCTGCCCAGAAGACT-3'

### **An azoxymethane (AOM)/ dextran sulfate sodium (DSS) mouse model of CRC**

A colitis-associated AOM/DSS mouse model of CRC was generated as described<sup>256</sup>. Briefly, AOM (Sigma, A5486; 10mg/kg) was administered by intraperitoneal injection to C57BL/6J mice obtained from Charles River Laboratories Japan. One week later, the mice commenced the first of three cycles of DSS to induce inflammation (MP Biomedicals, 160110; Molecular Weight 36,000-50,000 Da). Each cycle was comprised of 5 days 3% DSS (weight/volume) then 16 days of normal drinking water. 10 weeks after AOM injection, mice were euthanized, and colons were removed, opened longitudinally, and fixed in 10% formalin overnight at room temperature. Tissues were subsequently dehydrated and paraffin-embedded. 4 µm sections were used for histological analysis.

### **Mouse models of CRC liver metastasis**

Tumoroids were isolated from Matrigel and dissociated to single cells using TrypLE. The cells were filtered through 40-µm meshes to remove cell clumps and suspended in cold PBS containing 10 µM Y-27632 (In Vitro Technologies). 5.0 x 10<sup>5</sup> cells in 100 µl were injected directly into the portal vein using a 33G syringe.

Male and female Rosa26-Cas9 mice (JAX Stock Number 024858; C57BL/6 x 129 genetic background; 6 to 24-week-old), housed under pathogen-free conditions in the SAHMRI Bioresources facility, were used for portal vein injection.

In all animal experiments, mice were allocated randomly to different treatment groups. Sample sizes were determined based on pilot experiments and sample availability.

### **AAV packaging**

AAV8 preparation was performed by Vector & Genome Engineering Facility, Children's Medical Research Institute (NSW, AUSTRALIA). The AAV vectors used in this study were packaged using a standard transient transfection as described previously<sup>257</sup>. In short, pAAV transfer vector, pAd5 helper plasmid<sup>258</sup>, and AAV-helper plasmid encoding rep2 and cap8 (pR2C8) were transfected using PEI (Polysciences; Cat No. 239662) into HEK293 cell. Assembled vector particles were purified using iodixanol- density gradients as described<sup>259</sup>. Vector genomes were quantified using real-time quantitative PCR (RT-qPCR) as previously described<sup>260</sup>.

### **Intravenous AAV8 injection**

Male and female Rosa26-Cas9 mice (JAX Stock Number 024858; C57BL/6 x 129 genetic background; 6 to 24-week-old) were administered intravenously via the tail vein, with a 150- $\mu$ l injection of  $1.0 \times 10^{11}$  viral genomes of AAV8-mRuby2 or AAV8-*Isr* per mouse. A 2-week interval after AAV injection was allowed for protein expression to occur before portal vein injection of tumoroids. In Figure, 6D, E, and I-N, AAV8-treated mice were harvested for histopathological analyses 3-4 weeks after tumor injection, before prominent necrosis complicated histological assessment.

### **Subcutaneous injection of the GREM1-neutralizing antibody**

Subcutaneous injection of the GREM1-neutralizing antibody or IgG isotype was commenced 1 week after tumoroid injection. A 30 mg/kg dose of the GREM1-neutralizing antibody (Ab7326) or IgG

isotype (Ab101.4) was subcutaneously administered to a mouse twice a week until endpoint criteria were met.

### ***In vivo* imaging system (IVIS)**

Liver metastasis tumor growth kinetics was assessed by an *in vivo* imaging system (IVIS) by using a Xenogen IVIS Spectrum Imaging System (Perkin Elmer Inc) 10 minutes after intraperitoneal injection of 150 mg/kg of D-Luciferin (L-8220, Biosynth Carbosynth). Luciferase activity was quantified using Living Image software (Perkin Elmer Inc).

### **A humane endpoint in animal experiments**

In survival analyses, to determine the humane endpoint in animal experiments, a clinical record score was used. Scores were obtained by one point being given for the presence of each of the following observations: weight loss > 15%; hunched/ruffled coat; dehydrated; absence of movement; or facial grimace. Once a score of 3 was reached, the mice were euthanased.

### **Collagen gel contraction assay**

Collagen gels were prepared by mixing YH2 cells with PureCol bovine type I collagen (Advanced Biomatrix), 1× PBS, 1 M NaOH (final concentration: 1.0 mg/mL PureCol). The mixture containing  $3.0 \times 10^5$  cells/mL was seeded in 24-well cell culture plates and allowed to polymerize for 20 min. DMEM supplemented with 10% FBS and 1% penicillin/streptomycin was added, and gels were released from the wells. The gel area was measured with ImageJ (National Institutes of Health). Representative images were processed using Adobe Photoshop CC (Adobe).

### **Human CRC samples**

All ISH was performed using surgery samples of CRC patients who were diagnosed with adenocarcinoma of the colon or rectum at Nagoya University Hospital (Nagoya, Japan). This study was conducted in accordance with the Helsinki Declaration for Human Research and approved by

the Ethics Committee of Nagoya University Graduate School of Medicine (approval number: 2017-0127).

### **Quantitative image analysis**

IHC or ISH images were processed into separate channels representing nuclei staining (Hematoxylin) and IHC or ISH staining (DAB) using a color deconvolution function in a Fiji software (ImageJ; National Institutes of Health). Then, binary images were generated by intensity thresholds, and DAB<sup>+</sup> areas were calculated by ImageJ. To evaluate pSmad1/5/8 staining intensity, following an application of the lower threshold, the DAB staining intensity per high power field was calculated by ImageJ.

In Figure 2B, (1) normal rectal mucosa adjacent to adenocarcinoma, (2) adenoma in carcinoma, and (3) adenocarcinoma were randomly selected from ISH samples, and were used for quantification. Stromal areas (total areas excluding epithelial cells) visualized by hematoxylin counterstaining were outlined and measured by a clinical pathologist, using Adobe Photoshop CC (Adobe). In the normal mucosa, only lamina propria regions were quantified.

Ki-67 labeling index was calculated as the percentage of Ki-67<sup>+</sup> cells in the total epithelial cells as visualized by hematoxylin counterstaining.

The MRI fibrosis tool ([http://dev.mri.cnrs.fr/projects/imagej-macros/wiki/Fibrosis\\_Tool](http://dev.mri.cnrs.fr/projects/imagej-macros/wiki/Fibrosis_Tool)) was used to quantify the percentage of fibrosis areas (Picro-Sirius red<sup>+</sup> areas) at default settings.

### **Distance analysis using ISH samples**

In Figure 4H and I, serial sections of human desmoplastic rectal cancer samples, in which the stroma accounted for more than 40% of the total tumor area, were evaluated. The minimum edge-to-edge distance between ISH signals and the closest tumor glands was calculated by DiAna plugin<sup>261</sup> for ImageJ. Briefly, tumor glands in ISH images were outlined by a clinical pathologist. Then, to identify ISH signals (DAB<sup>+</sup> signals), the ISH images were subjected to color deconvolution before applying intensity thresholding and a watershed algorithm in a Fiji software. ISH signals were segmented using a minimum size threshold by DiAna plugin. Then, the smallest edge-to-edge distance between each ISH signal and the closest tumor gland was computed by DiAna plugin.

### **Evaluation of tumor differentiation and tumor budding**

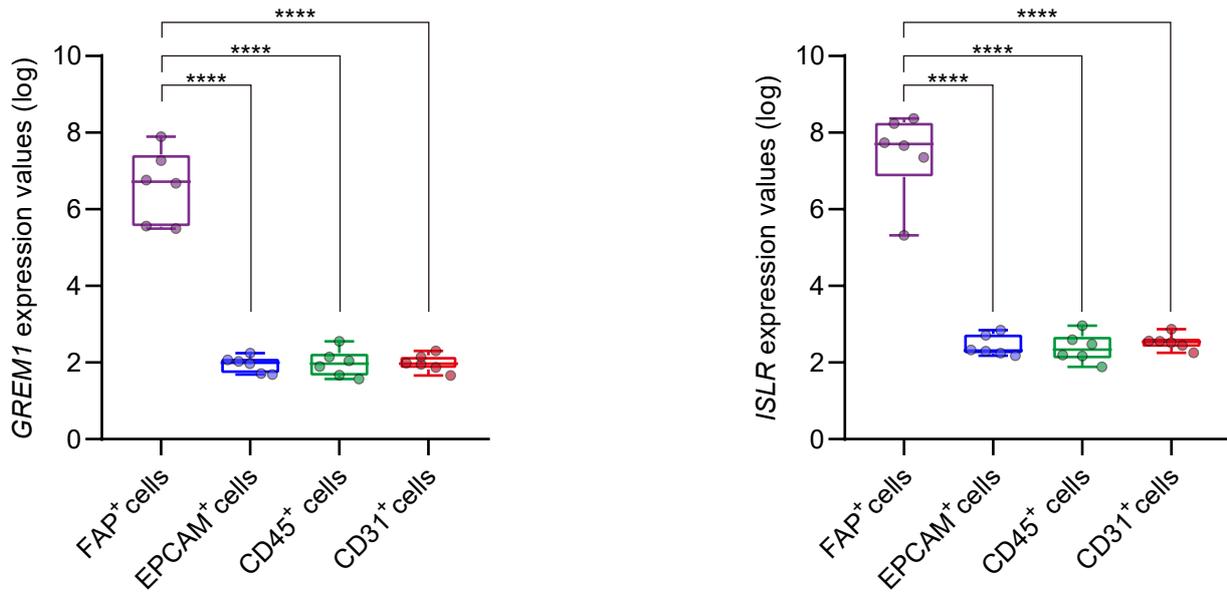
Evaluation of tumor differentiation and tumor budding was performed by a clinical pathologist. Poorly differentiated adenocarcinomas were defined as cancers that exhibit glandular differentiation in 5-50% of the tumor, as described<sup>262</sup>.

Tumor budding was defined as a single tumor cell or a cell cluster of up to 4 tumor cells, as described<sup>263</sup>.

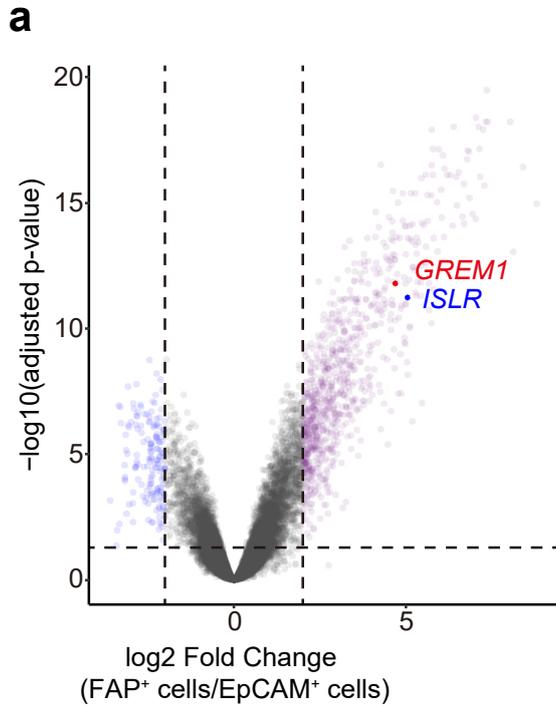
### **Data and materials availability**

Data that support the findings of this study are available from the corresponding authors upon reasonable request. Noncommercially available materials described in this study may be obtained with a material transfer agreement (MTA). Requests for materials should be addressed to S.W.

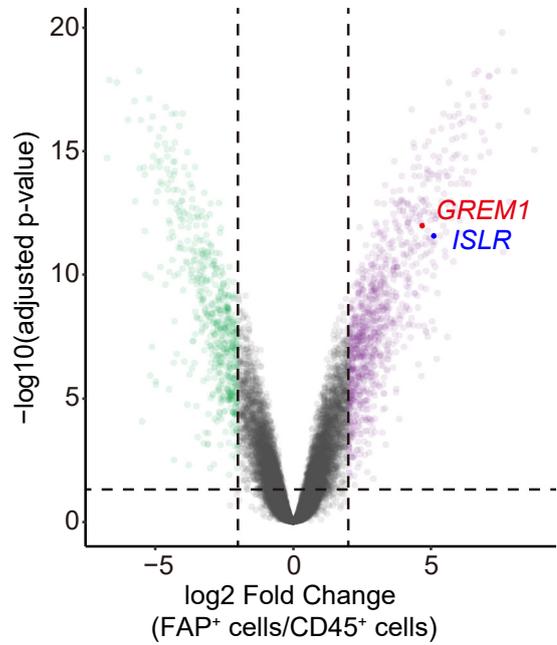
**A** Expression microarray (GSE39396)  
FACS-purified cells from human CRC tissues



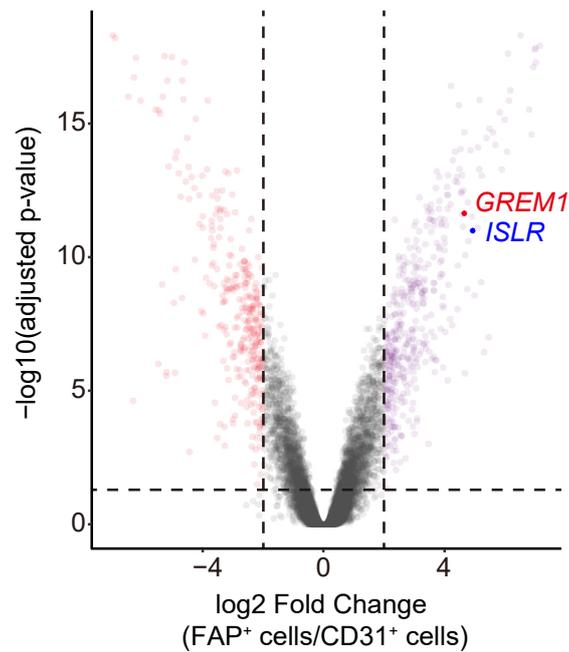
**B**



**b**



**c**



GSE39396  
(Calon et al, Cancer Cell, 2012)

**Supplementary Figure 1: Analyses of a human CRC expression microarray dataset reveal that *GREM1* and *ISLR* are highly expressed in FAP<sup>+</sup> CAFs compared to EpCAM<sup>+</sup> epithelial cells, CD45<sup>+</sup> immune cells, and CD31<sup>+</sup> endothelial cells.**

**(A)** *GREM1* and *ISLR* expression levels in a cDNA (complementary DNA) expression microarray from fluorescence-activated cell sorting (FACS)-purified cells from human primary CRC tissues (GSE39396).

n = 6 patients each. One-way ANOVA followed by Tukey's post-hoc multiple comparisons.

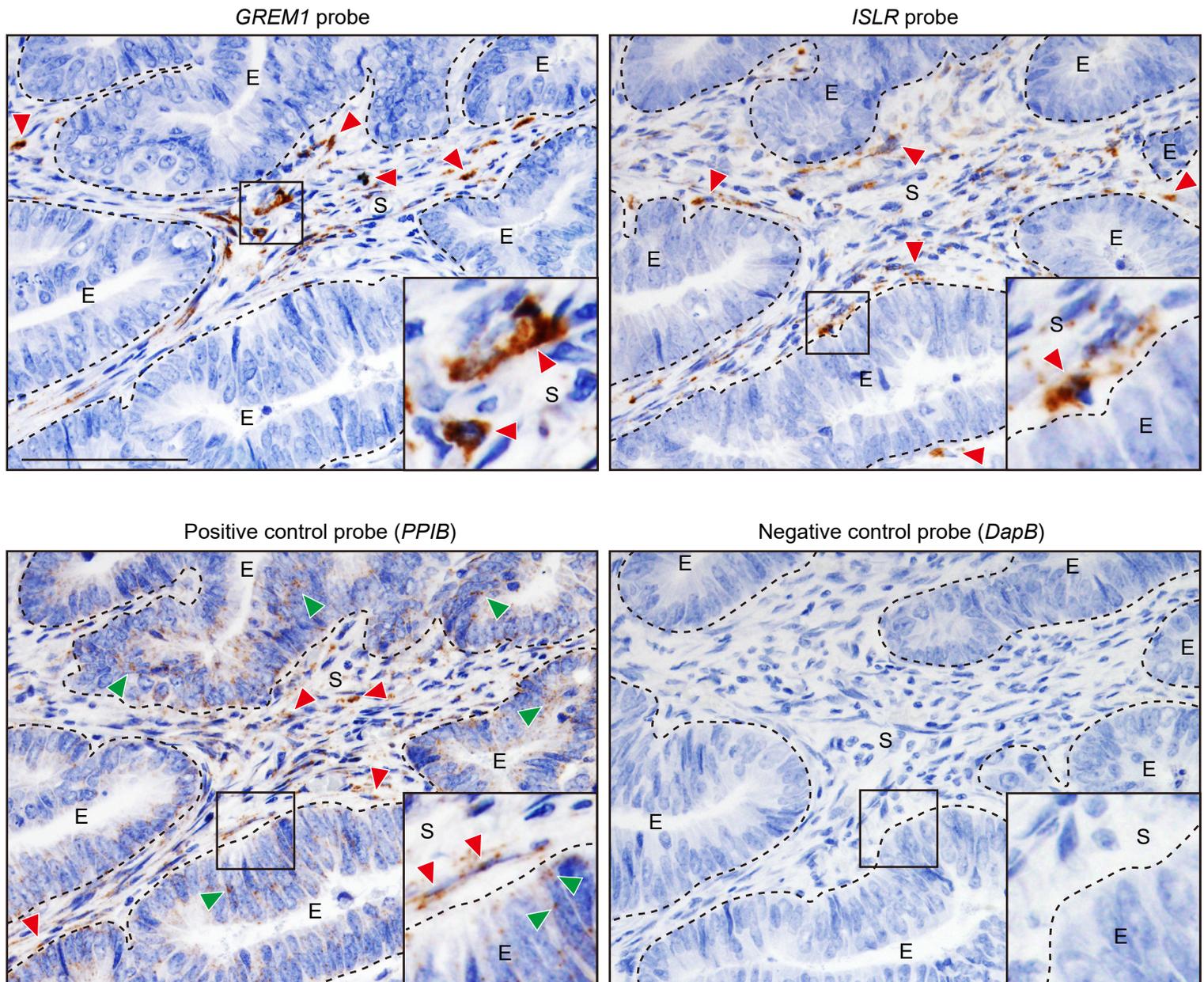
\*\*\*\*, P < 0.0001. Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.

**(B)** Volcano plots showing differentially expressed transcripts between FAP<sup>+</sup> CAFs and EpCAM<sup>+</sup> epithelial cells **(a)**, between FAP<sup>+</sup> CAFs and CD45<sup>+</sup> immune cells **(b)** and between FAP<sup>+</sup>CAFs and CD31<sup>+</sup> endothelial cells **(c)** (GSE39396).

n = 6 patients each. The adjusted P-value cutoffs and the log<sub>2</sub> fold change cutoffs for differentially expressed genes were 0.05 and 2, respectively.

Semi-transparent purple, blue, green, and red dots denote differentially expressed transcripts, which are upregulated in FAP<sup>+</sup> CAFs, EpCAM<sup>+</sup> epithelial cells, CD45<sup>+</sup> immune cells, and CD31<sup>+</sup> endothelial cells, respectively. Gray dots represent transcripts that are not differentially expressed.

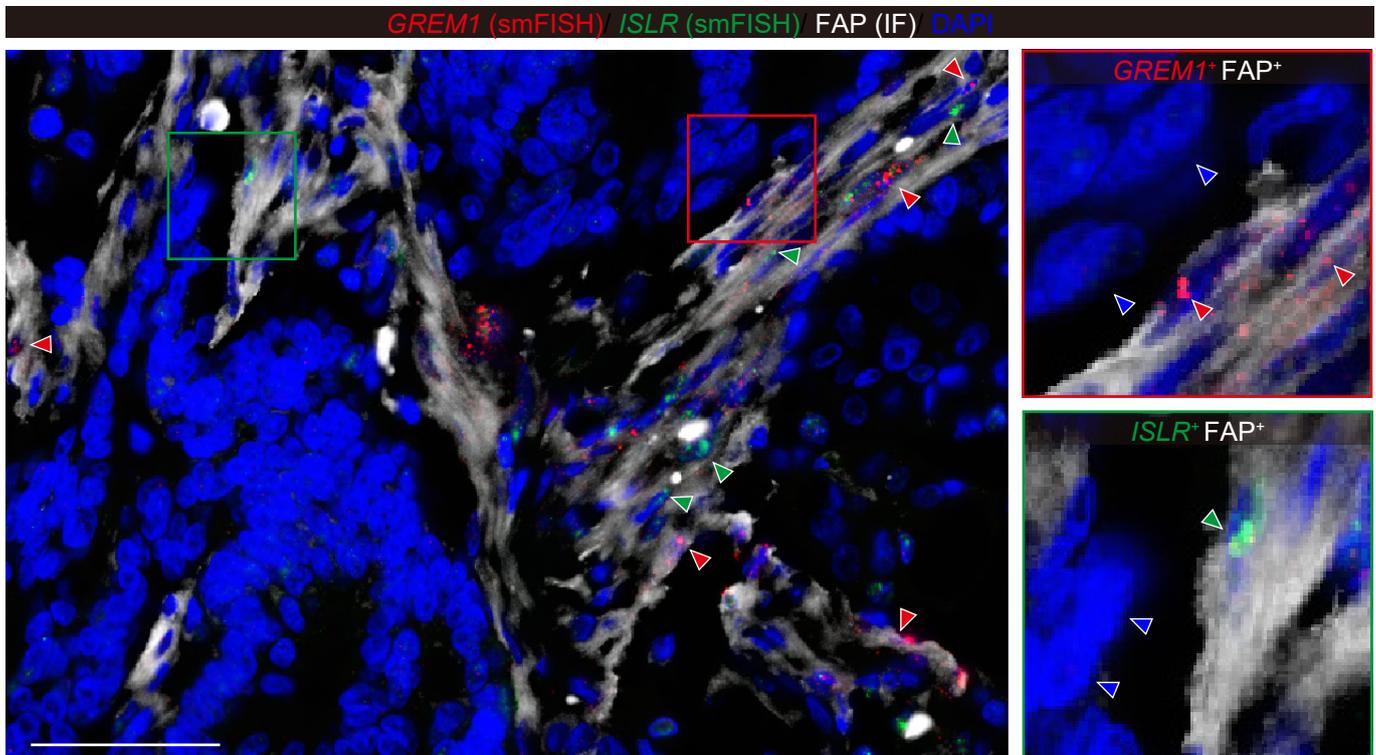
*GREM1* and *ISLR* gene probes used for the analyses are as follows; 218468\_PM\_s\_at (*GREM1*) and 207191\_PM\_s\_at (*ISLR*).



**Supplementary Figure 2: Validation of *in situ* hybridization staining using positive and negative control probes.**

Representative *in situ* hybridization (ISH) pictures using serial sections from human colorectal cancer. ISH for *GREM1* and *ISLR* shows 3,3'-Diaminobenzidine (DAB)<sup>+</sup> staining specifically in the stroma, whereas positive control probe (*PPIB*, a ubiquitously expressed gene) staining is observed both in the epithelial cells and stroma. No apparent background staining is seen in ISH for a negative control probe that targets the bacterial *DapB* gene.

Dotted lines indicate the borders between epithelial cells (E) and the stroma (S). Red and green arrowheads denote DAB<sup>+</sup> staining in the stroma and epithelial cells, respectively. The boxed areas are magnified in the insets. Scale bar, 100  $\mu$ m.

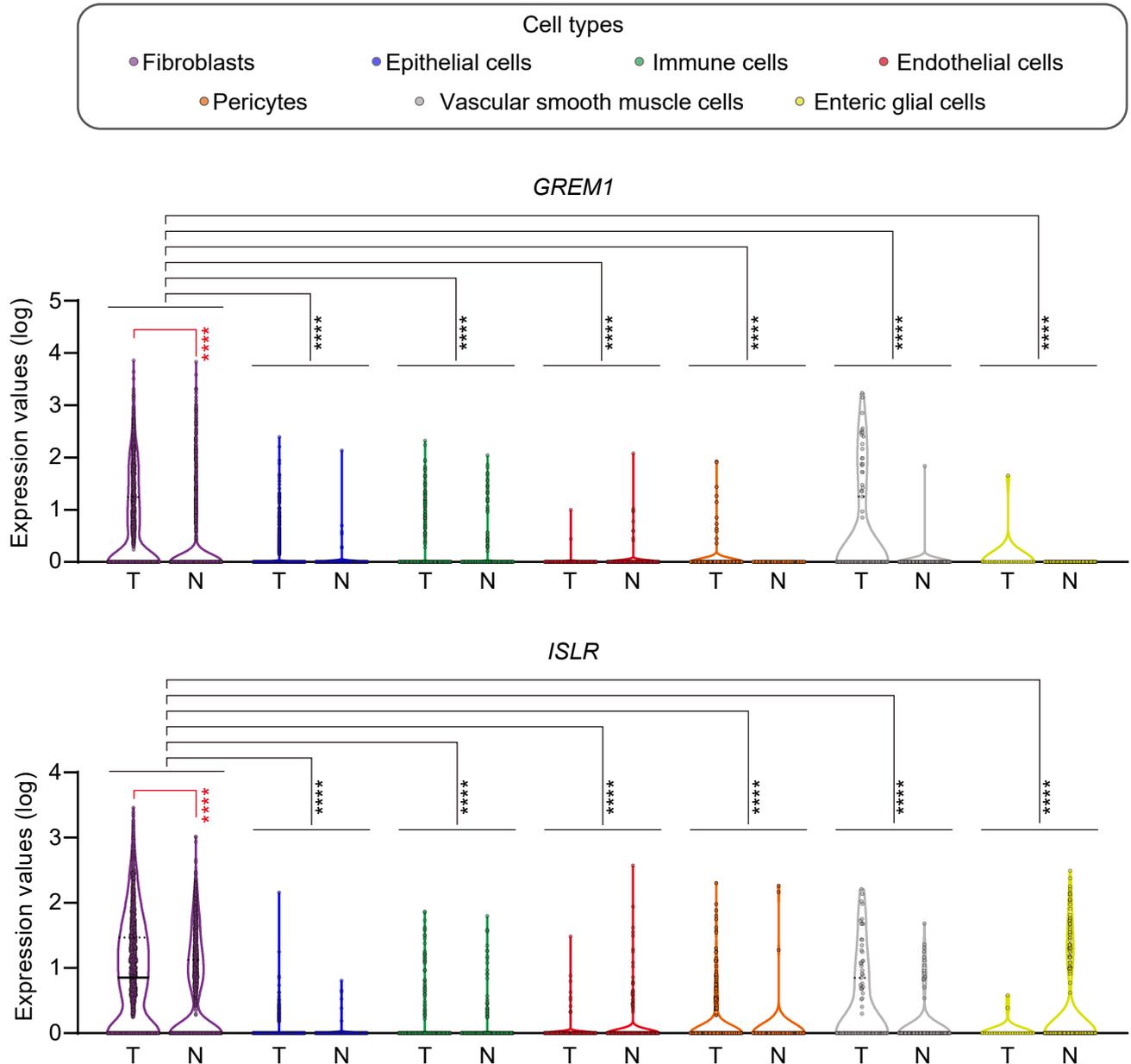


**Supplementary Figure 3: *GREM1* and *ISLR* transcripts are detected in *FAP*<sup>+</sup> CAFs in single-molecule fluorescent *in situ* hybridization on human CRC sections.**

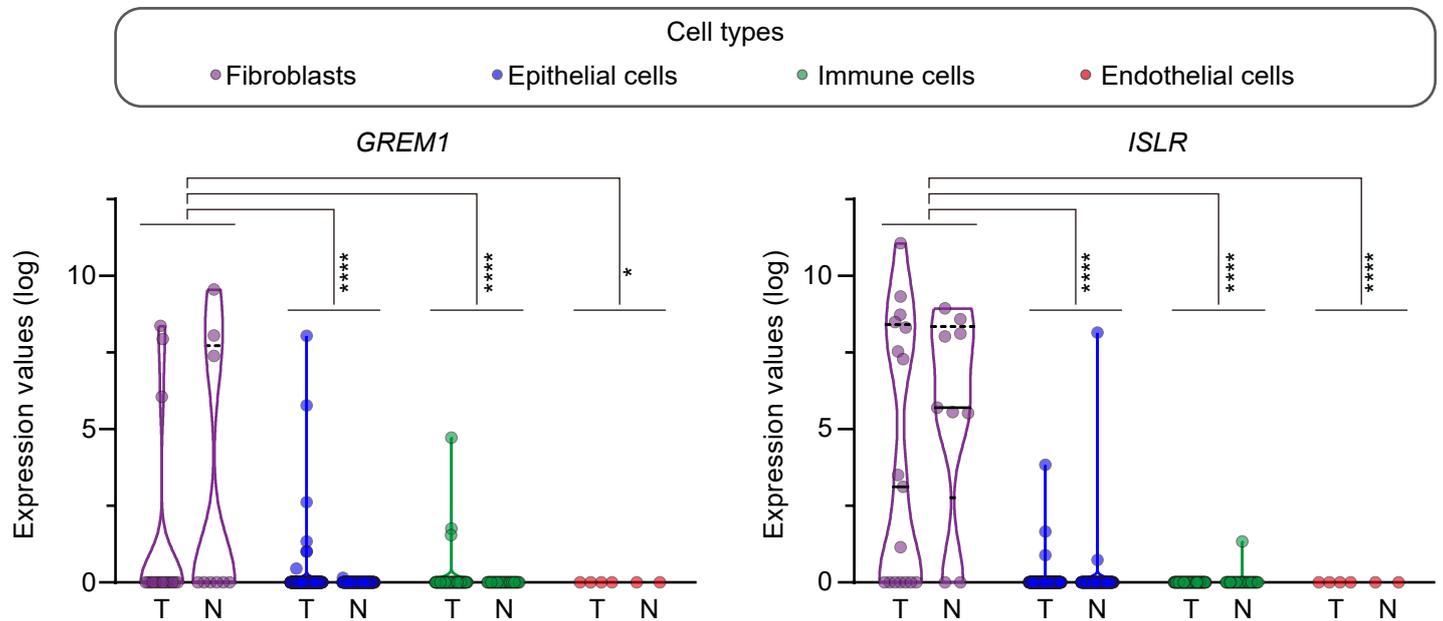
Representative picture of *GREM1* and *ISLR* dual single-molecule fluorescent hybridization (smFISH) followed by *FAP* immunofluorescence (IF), using human CRC tissue sections. Red and green arrowheads denote *GREM1*<sup>+</sup>*FAP*<sup>+</sup> cells and *ISLR*<sup>+</sup>*FAP*<sup>+</sup> cells, respectively. Blue arrowheads indicate *FAP*<sup>-</sup> cells that do not express *GREM1* or *ISLR*. n = 3 patients. Scale bar, 50 μm.

**A**

ScRNA-seq from human primary colorectal cancer and normal mucosa (GSE132465)

(Lee *et al.*, *Nat Genet*, 2020)**B**

ScRNA-seq from human primary colorectal cancer and normal mucosa (GSE81861)

(Li *et al.*, *Nat Genet*, 2017) 16

**Supplementary Figure 4: *GREM1* and *ISLR* are highly expressed in fibroblasts in single-cell RNA-seq data from human normal colorectal mucosa and primary CRC.**

**(A)** Violin plots depicting *GREM1* and *ISLR* expression levels in single-cell RNA-seq (scRNA-seq) of unsorted cells from human primary CRC tissues and normal colorectal tissues (GSE132465)<sup>226</sup>. In fibroblasts, *GREM1* and *ISLR* transcripts are differentially upregulated in tumors compared with normal samples (Wilcoxon tests, red asterisks).

n = 1501, 1961, 17469, 1070, 27080, 12137, 768, 739, 353, 92, 91, 123, 23, and 282 cells (left to right). Tumor samples from 23 patients. Normal samples from 10 of the patients.

T, Tumor samples; N, Normal samples.

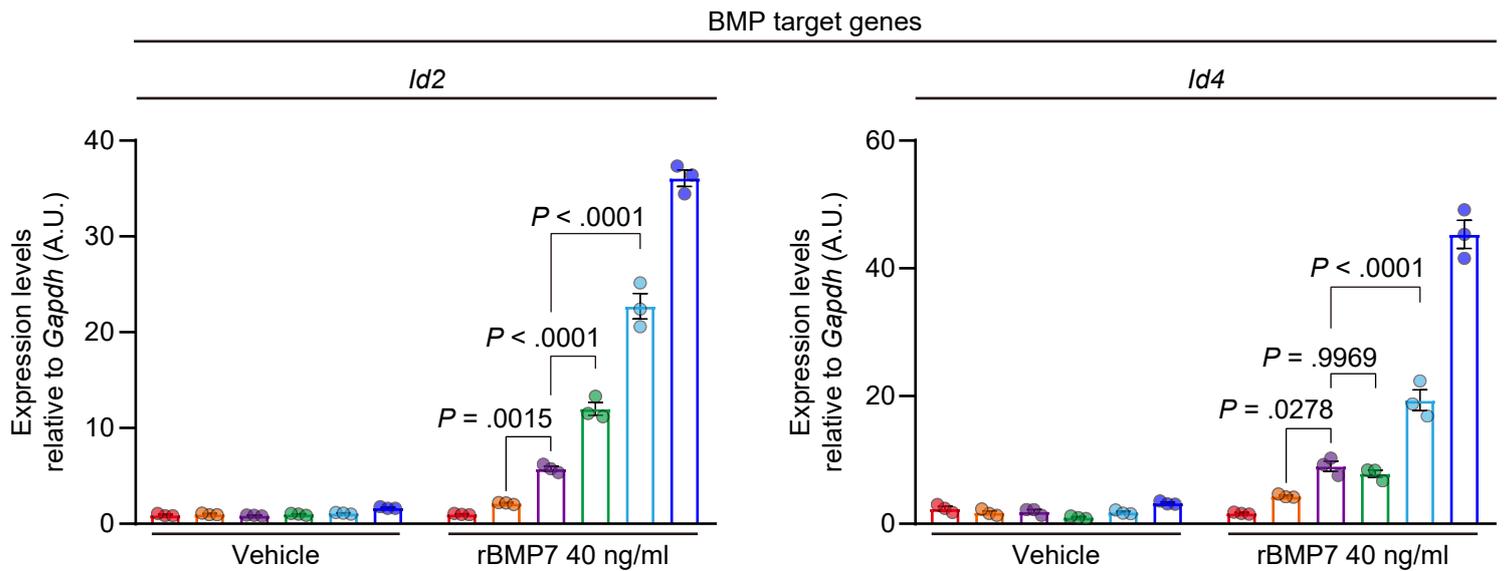
**(B)** Violin plots showing *GREM1* and *ISLR* expression levels in scRNA-seq of unsorted cells from human primary CRC tissues and normal colorectal tissues (GSE81861)<sup>255</sup>.

n = 17, 9, 272, 160, 71, 42, 4, and 2 cells (left to right). Tumor samples from 11 patients. Normal samples from 7 of the patients.

Kruskal-Wallis tests followed by Dunn's post-hoc multiple comparisons were used to compare *GREM1* or *ISLR* expression levels in fibroblasts from tumor and normal samples with those in other cell populations (black asterisks).

\*\*\*\*,  $P < 0.0001$ ; \*,  $P = 0.0132$ . Solid black lines, median; Dotted black lines, quartiles.

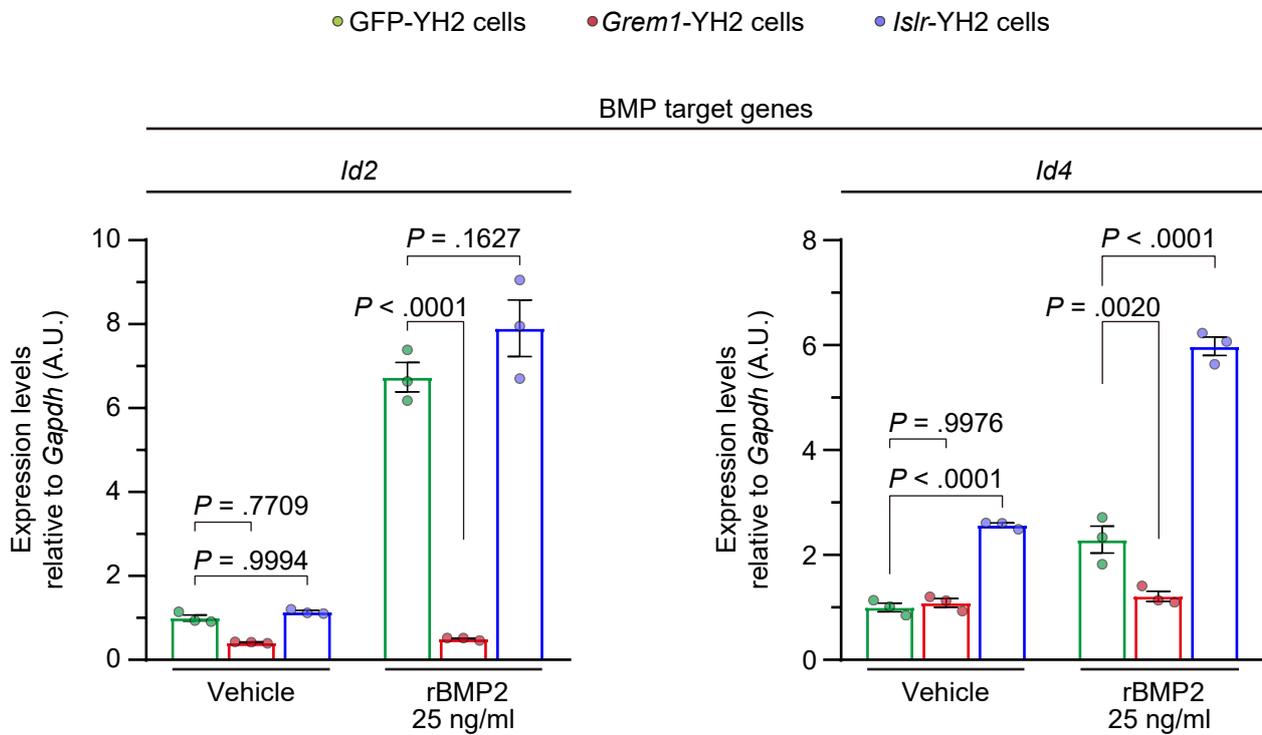
- *Grem1*-YH2    ● *Grem1*-YH2 + GFP-YH2    ● *Grem1*-YH2 + *Islr*-YH2
- GFP-YH2    ● GFP-YH2 + *Islr*-YH2    ● *Islr*-YH2



**Supplementary Figure 5: GREM1 and ISLR overexpression counteract each other's effect on BMP7-*Id2/4* signaling *in vitro*.**

Quantitative reverse-transcription PCR (qRT-PCR) for BMP target genes, *Id2* and *Id4*.

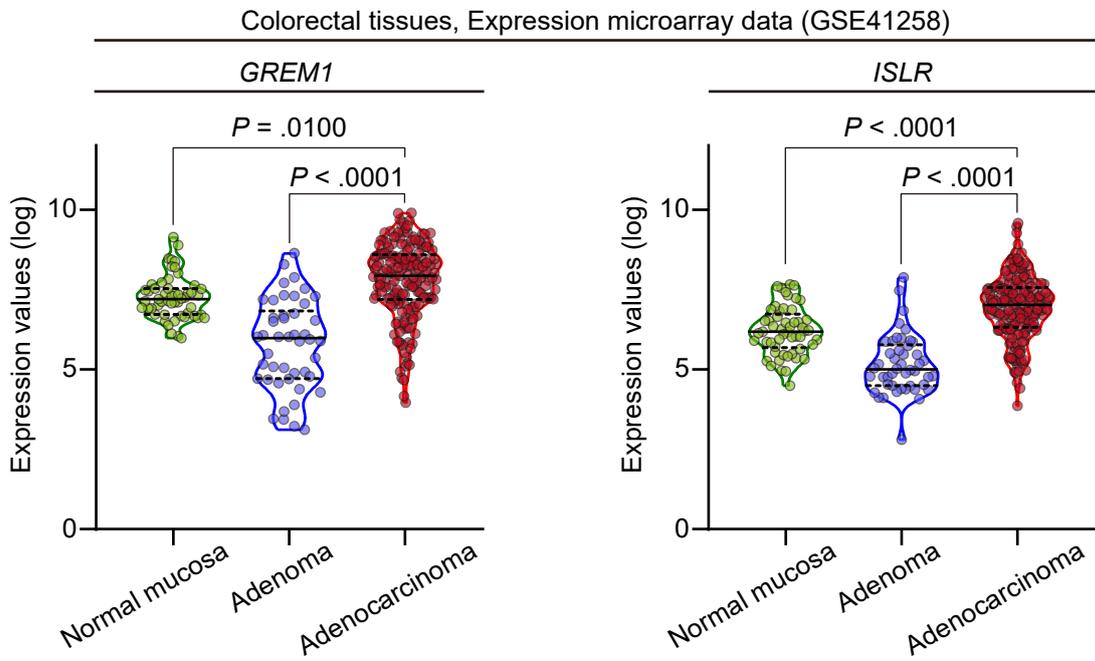
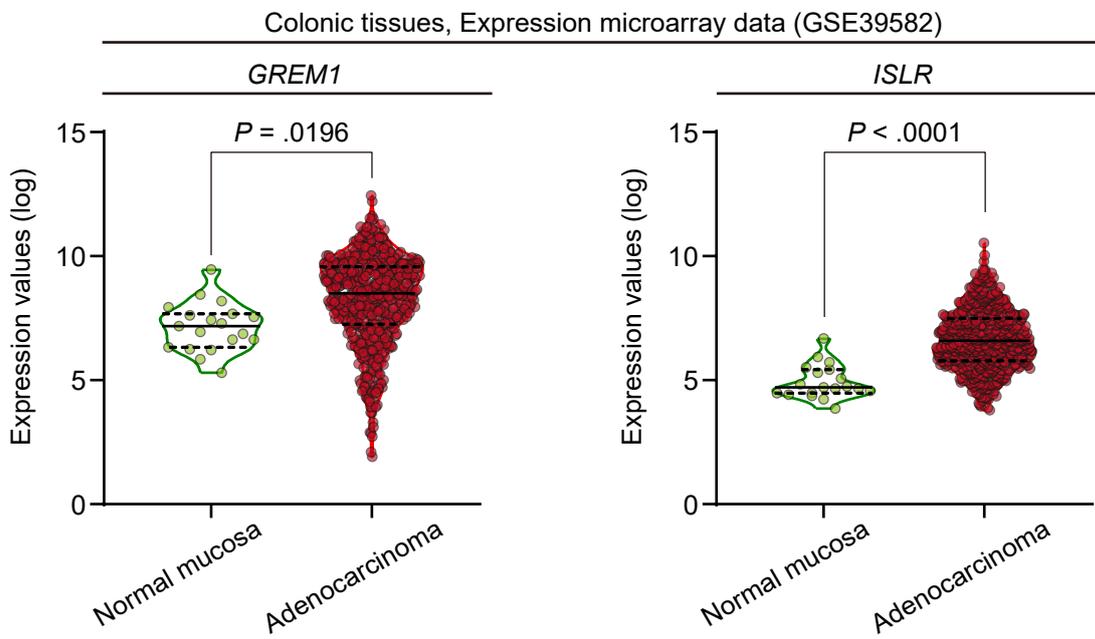
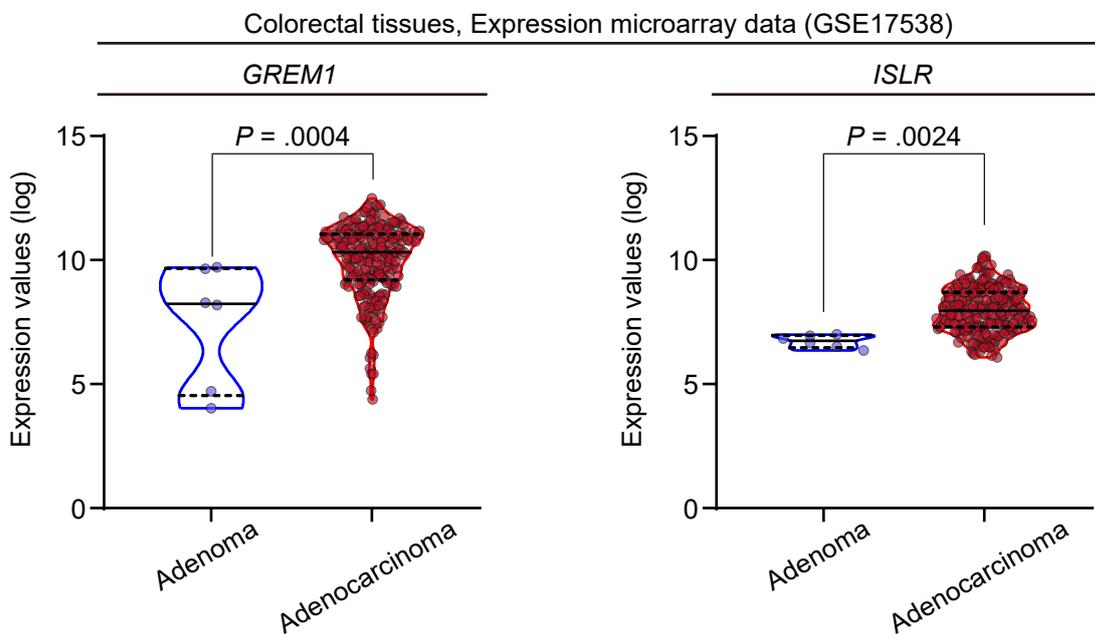
YH2 cells that overexpress GREM1, GFP, or ISLR were admixed at a 1:1 ratio and stimulated with 40 ng/ml of recombinant BMP7 for 24 hours (orange, purple, and aqua). *Grem1*-YH2 cells, GFP-YH2 cells, or *Islr*-YH2 cells alone were used as controls (red, green, and blue, respectively). The same number of cells (a total of  $5.0 \times 10^4$  cells/well) were seeded in a 6-well plate in each condition. n = 3 each. A.U., arbitrary unit. Mean  $\pm$  s.e.m.. Two-way ANOVA with post-hoc Tukey's multiple comparisons.



**Supplementary Figure 6: BMP2-induced increase in *Id4* is prevented by GREM1 and augmented by ISLR in YH2 cells.**

qRT-PCR for *Id2* and *Id4* in YH2 cells. GFP-, *Grem1*-, or *Islr*-overexpressing YH2 cells were treated with 25 ng/ml of recombinant BMP2 for 24 hours. Note that GREM1 prevented the BMP2-induced increase in *Id2* whereas ISLR did not significantly promote the BMP2-mediated increase in *Id2* in YH2 cells.

n = 3 each. A.U., arbitrary unit. Mean  $\pm$  s.e.m.. Two-way ANOVA with post-hoc Tukey's multiple comparisons.

**A****B****C**

**Supplementary Figure 7: *GREM1* and *ISLR* expression levels are upregulated during human CRC carcinogenesis.**

**(A)** Violin plots depicting *GREM1* and *ISLR* expression in cDNA expression microarray data from human normal colorectal mucosa, colorectal adenoma, and primary colorectal adenocarcinoma (GSE41258)<sup>250</sup>.

n = 53 (normal mucosa), 46 (adenoma) and 182 patients (adenocarcinoma).

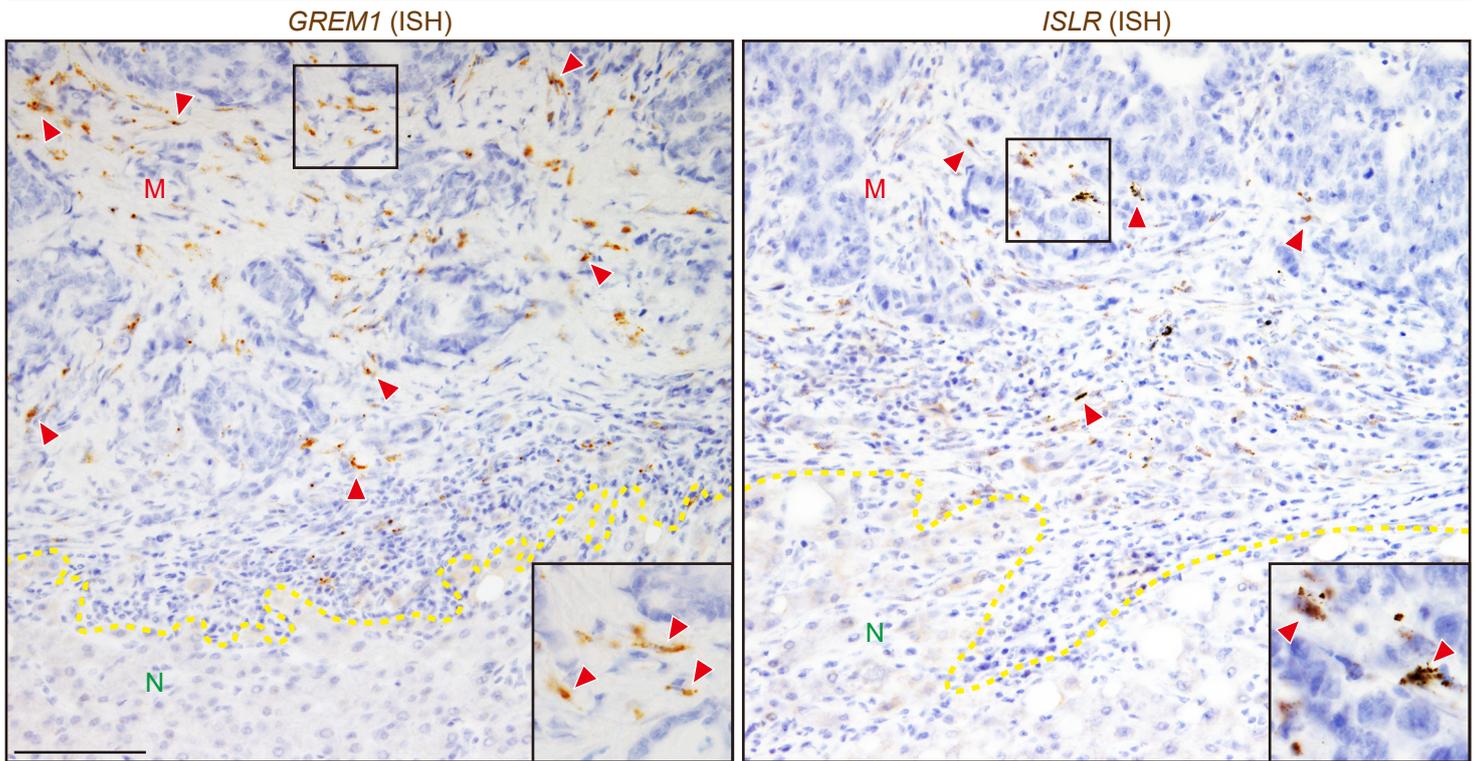
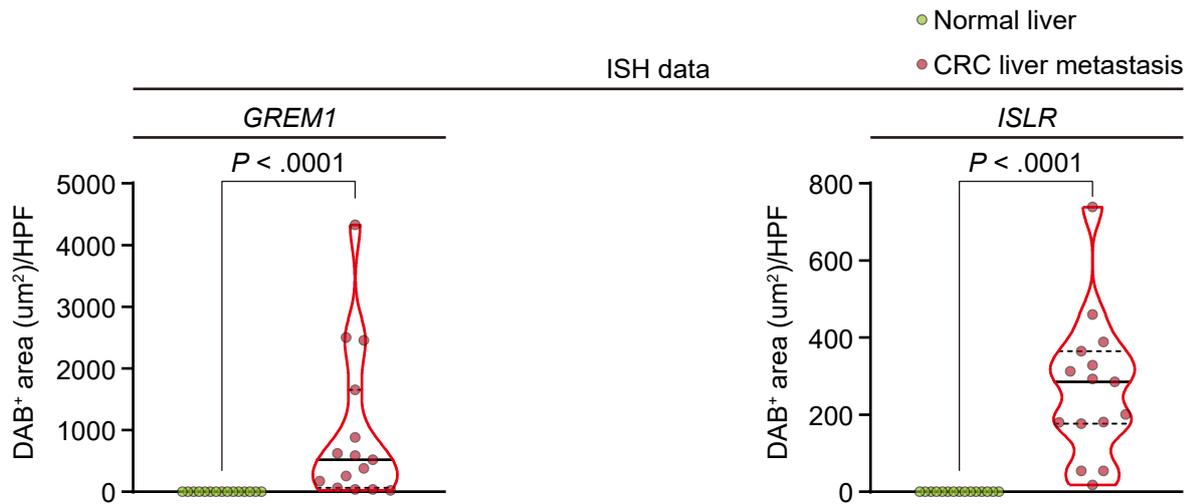
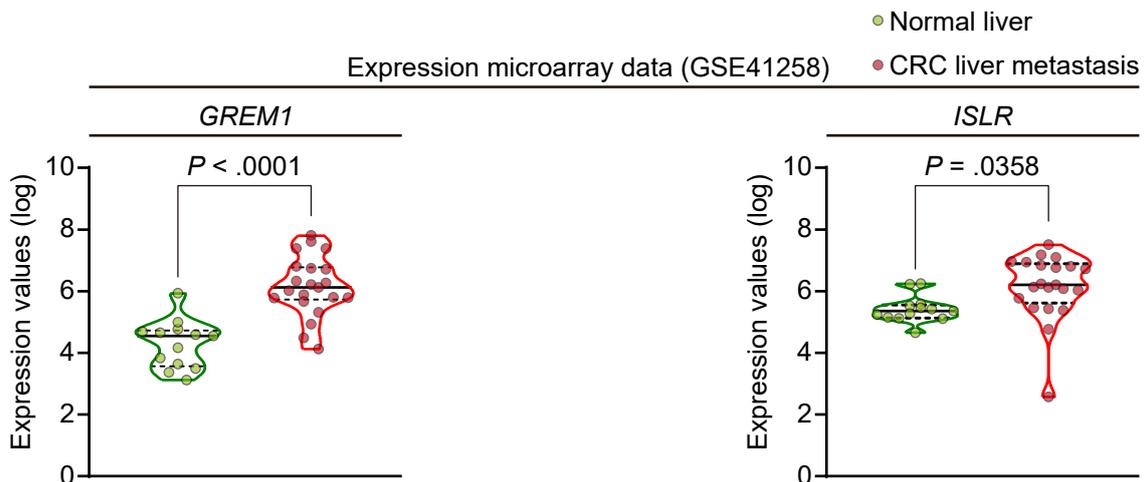
**(B)** Violin plots depicting *GREM1* and *ISLR* expression in cDNA expression microarray data from human normal colon mucosa and primary colon adenocarcinoma (GSE39582)<sup>251</sup>. n = 19 (normal colon mucosa) and 566 patients (primary colon adenocarcinoma).

**(C)** Violin plots depicting *GREM1* and *ISLR* expression in cDNA expression microarray data from human colorectal adenoma and primary colorectal adenocarcinoma (GSE17538)<sup>252</sup>. n = 6 (colorectal adenoma) and 232 patients (primary colorectal adenocarcinoma).

One-way ANOVA followed by Tukey's post-hoc multiple comparisons (A). False discovery rate (FDR)-adjusted P-values are shown (B and C). Solid black lines, median; Dotted black lines, quartiles.

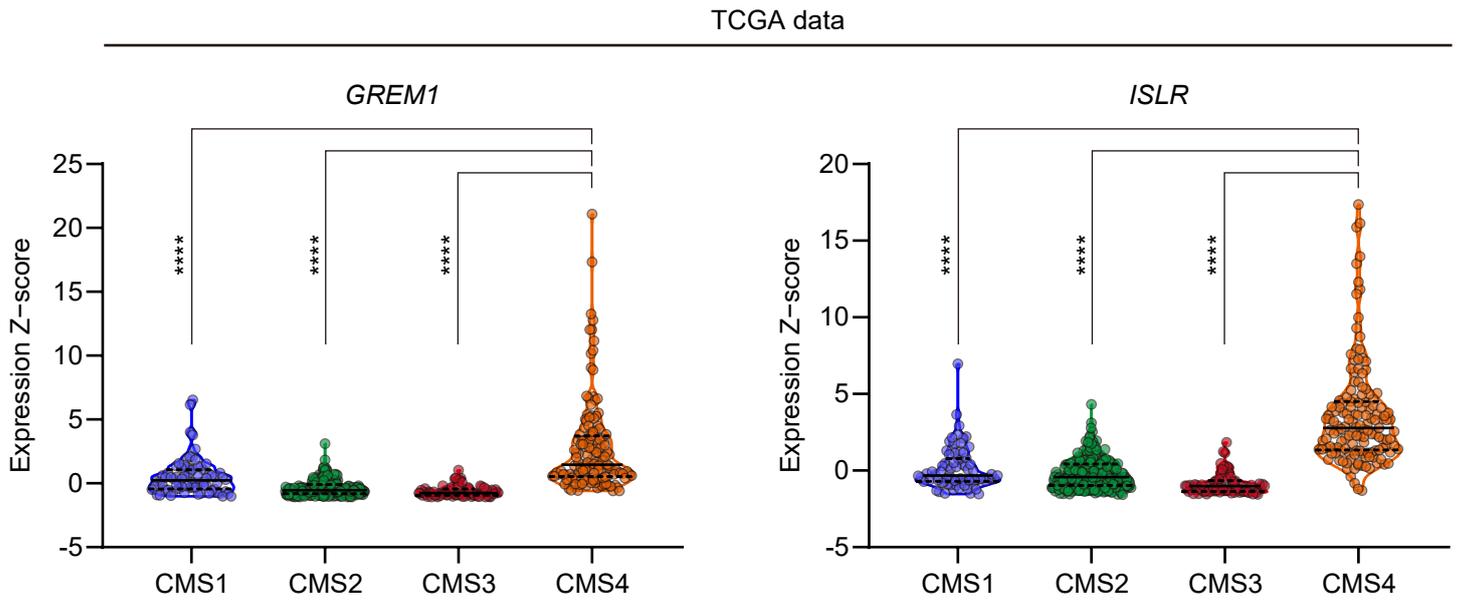
**A**

## Hepatic metastasis of human colorectal cancer

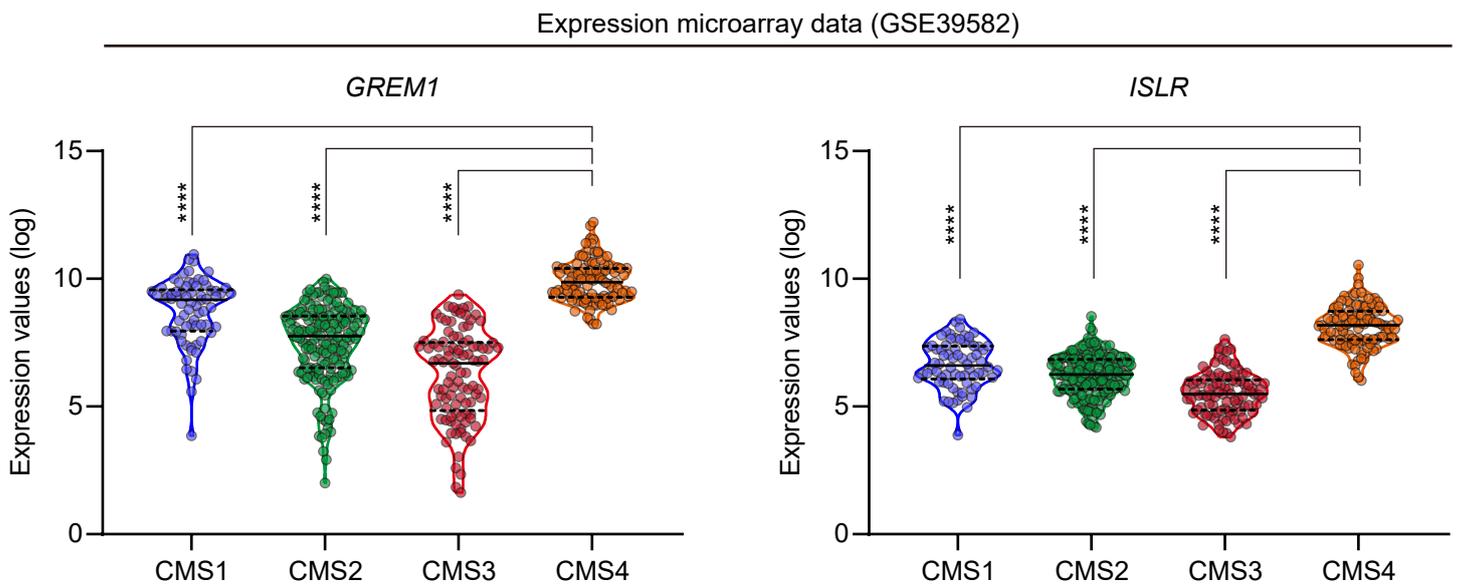
**B****C**

**Supplementary Figure 8: *GREM1* and *ISLR* are expressed in the stroma of human CRC hepatic metastases.**

**(A, B)** ISH for *GREM1* and *ISLR* in the hepatic metastases of human CRC. **(A)** Representative images. *GREM1* and *ISLR* are expressed by fibroblastic cells in the stroma of the liver metastases. Dotted yellow lines indicate the borders between the normal liver (N) and CRC liver metastases (M). Red arrowheads denote *GREM1* or *ISLR* expression. The boxed areas are magnified in the insets. Scale bars, 100  $\mu$ m. **(B)** Violin plots showing quantification of 3,3'-Diaminobenzidine (DAB)<sup>+</sup> areas by ImageJ. 5 HPFs/patient, 3 patients each. Solid black lines, median; Dotted black lines, quartiles. **(C)** Violin plots depicting *GREM1* and *ISLR* expression in cDNA expression microarray data from human normal liver and liver metastasis of CRC (GSE41258). n = 13 (normal liver) and 21 patients (CRC liver metastasis). Solid black lines, median; Dotted black lines, quartiles. Mann-Whitney U-test (B). False discovery rate (FDR)-adjusted P-values are shown in (C).

**A**

(The Cancer Genome Atlas Network, *Nature*, 2012  
; Guinney *et al.*, *Nat Med*, 2015)

**B**

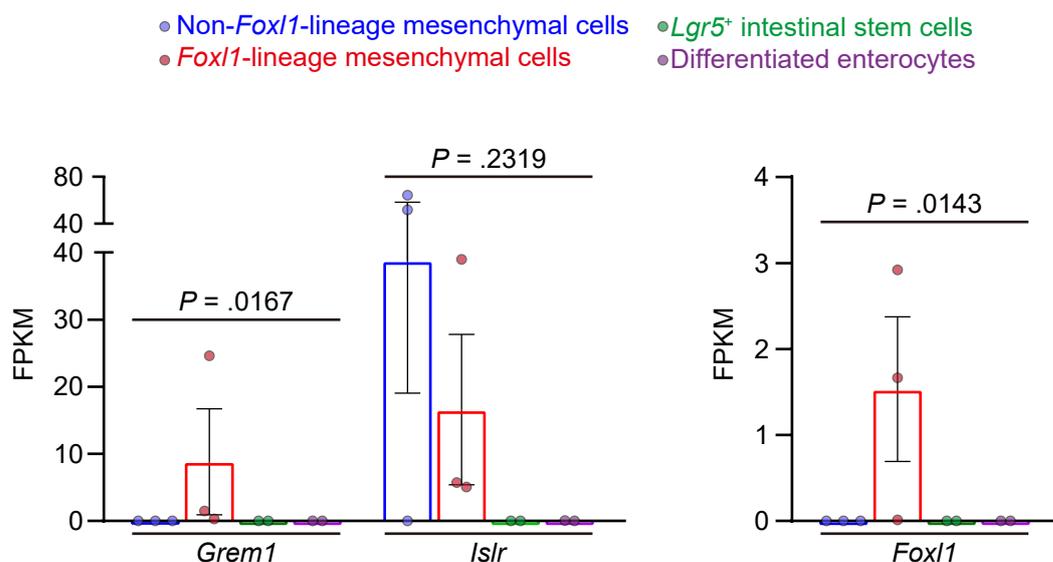
(Marisa *et al.*, *PLoS Med*, 2013)

**Supplementary Figure 9: *GREM1* and *ISLR* are highly expressed in consensus molecular subtype 4 CRC.**

**(A)** Violin plots showing expression levels of *GREM1* and *ISLR* in four consensus molecular subtypes (CMSs)<sup>31</sup>. Primary colon and rectal adenocarcinoma samples in The Cancer Genome Atlas (TCGA) data were analyzed. n = 76 (CMS1), 220 (CMS2), 72 (CMS3), and 143 patients (CMS4).

**(B)** Violin plots depicting *GREM1* and *ISLR* expression levels in four CMSs. Primary colon adenocarcinoma samples were analyzed using cDNA expression microarray data (GSE39582)<sup>251</sup>. n = 70 (CMS1), 177 (CMS2), 97 (CMS3), and 107 patients (CMS4).

Kruskal-Wallis test followed by Dunn's post-hoc multiple comparisons. Solid black lines, median; Dotted black lines, quartiles. \*\*\*\*,  $P < 0.0001$ .



(GSE94072; Shoshkes-Carmel *et al.*, *Nature*, 2018)

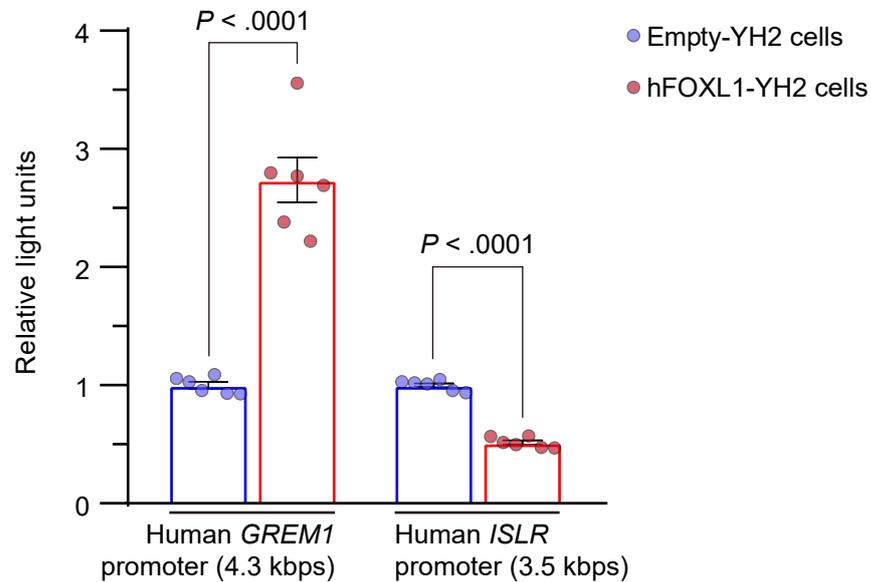
**Supplementary Figure 10: *Grem1* is highly expressed in *Foxl1*-lineage telocytes in the mouse small intestine.**

Expression levels of *Grem1*, *Islr*, and *Foxl1* in RNA-sequence data from FACS-purified non-*Foxl1*-lineage mesenchymal cells, *Foxl1*-lineage mesenchymal cells (sorted from *Fox1*-Cre; Rosa26-YFP or *Fox1*-Cre; Rosa26-mT/mG), *Lgr5*<sup>+</sup> intestinal stem cells (sorted by *Lgr5*-eGFP<sup>high</sup> expression), and differentiated enterocytes (GSE94072).

Note that fragments per kilobase of exon per million fragments (FPKM) for *Grem1* and *Foxl1* were zero in all three biological replicates of non-*Foxl1* lineage mesenchymal cells.

n = 3 (non-*Foxl1*-lineage mesenchymal cells), 3 (*Foxl1*-lineage mesenchymal cells), 2 (*Lgr5*<sup>+</sup> intestinal stem cells) and 2 (Differentiated enterocytes) mice.

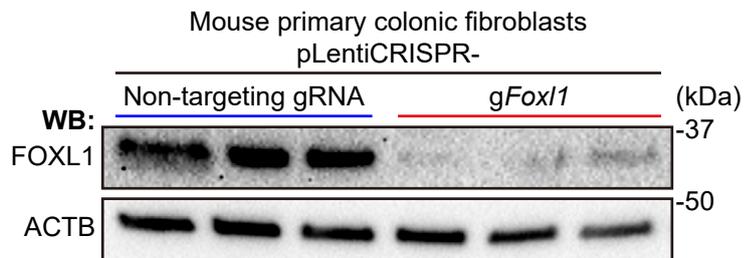
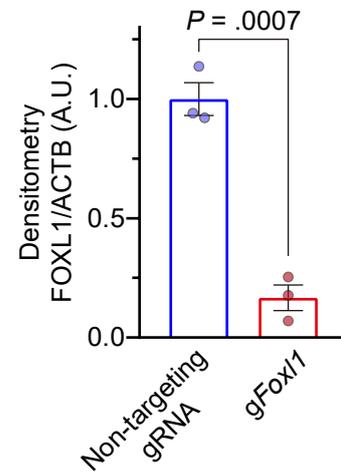
All data are represented as mean ± s.e.m.. Statistical analyses were performed using the Kruskal-Wallis test.



**Supplementary Figure 11: Lentivirus-mediated FOXL1 overexpression increases *GREM1*-promoter activity while decreasing *ISLR*-promoter activity.**

Luciferase assays of Human *GREM1*- or *ISLR*-promoter regions using human FOXL1-overexpressing YH2 cells and control empty YH2 cells.

n = 6 each. Data are represented as mean  $\pm$  s.e.m.. Statistical analysis was performed using two-tailed unpaired Student's t-test.

**A****B**

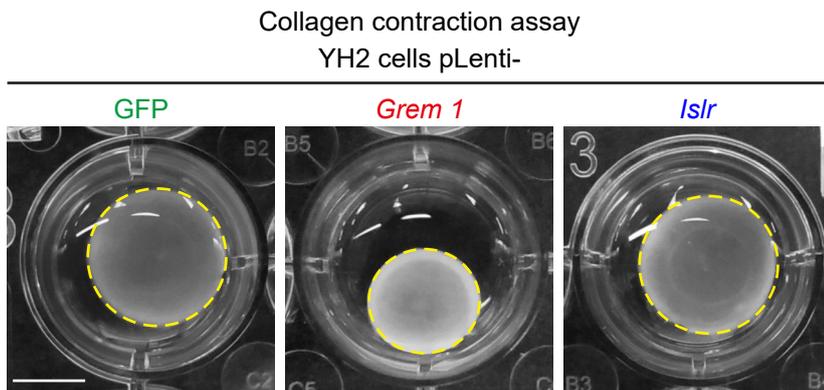
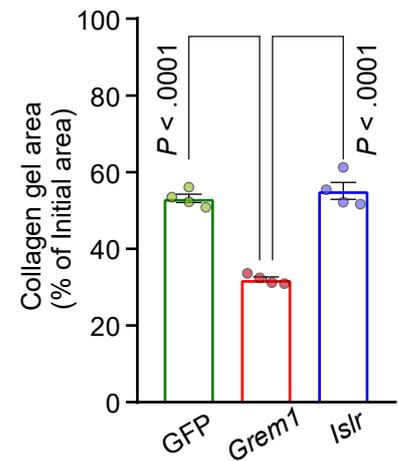
**Supplementary Figure 12: Validation of CRISPR/Cas9-mediated FOXL1-knockdown in mouse primary colonic fibroblasts by Western blot.**

**(A)** Western blotting (WB) showing FOXL1-Knockdown by lentiviral transduction.

CRISPR, clustered regularly interspaced short palindromic repeats; gRNA, guide RNA.

**(B)** Densitometry of Western blotting (A) showing decreased FOXL1 protein expression by lentiviral transduction. A.U., arbitrary unit.

$n = 3$  mice each. Data are represented as mean  $\pm$  s.e.m.. Statistical analysis was performed using two-tailed unpaired Student's t-test.

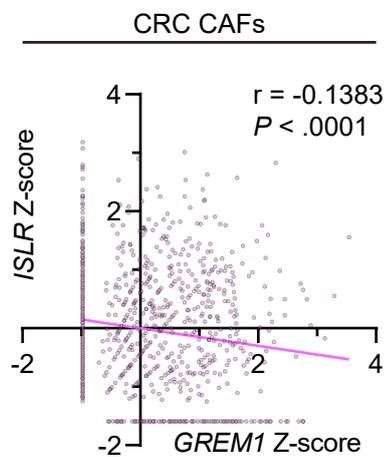
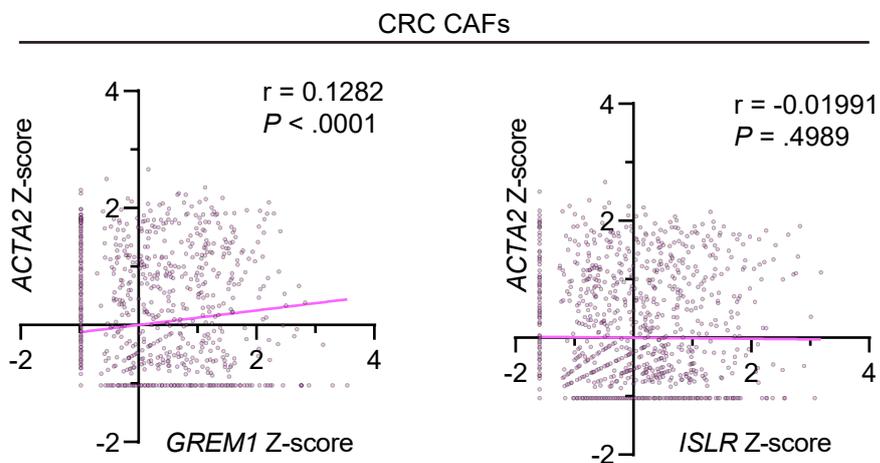
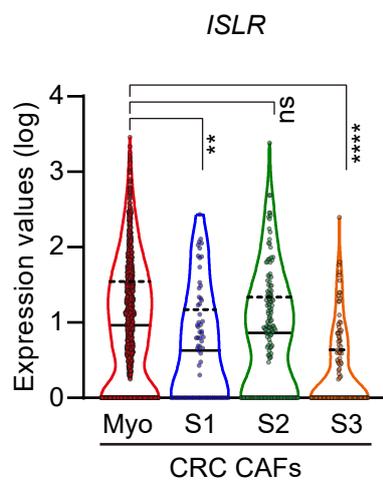
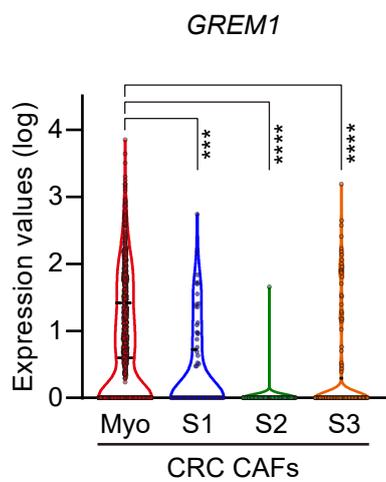
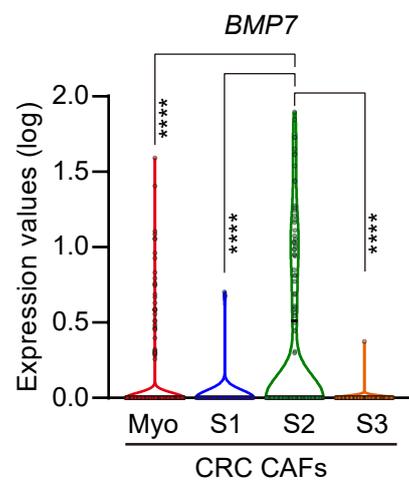
**A****B**

**Supplementary Figure 13: Increased collagen gel contraction in *Grem1*-overexpressing YH2 cells.**

**(A, B)** Collagen gel contraction assay using GFP-, *Grem1*- and *Islr*-overexpressing YH2 cells.

**(A)** Representative pictures. Collagen gel areas are outlined by dotted yellow lines. Scale bars, 5 mm. **(B)** Quantification of collagen gel areas.

Collagen gel areas were evaluated 12 hours after seeding cells.  $n = 4$  each. Data are represented as mean  $\pm$  s.e.m.. Statistical analysis was performed using one-way ANOVA with Tukey's post-hoc multiple comparisons.

**A****B****C****D**

(Lee et al., Nat Genet, 2020)

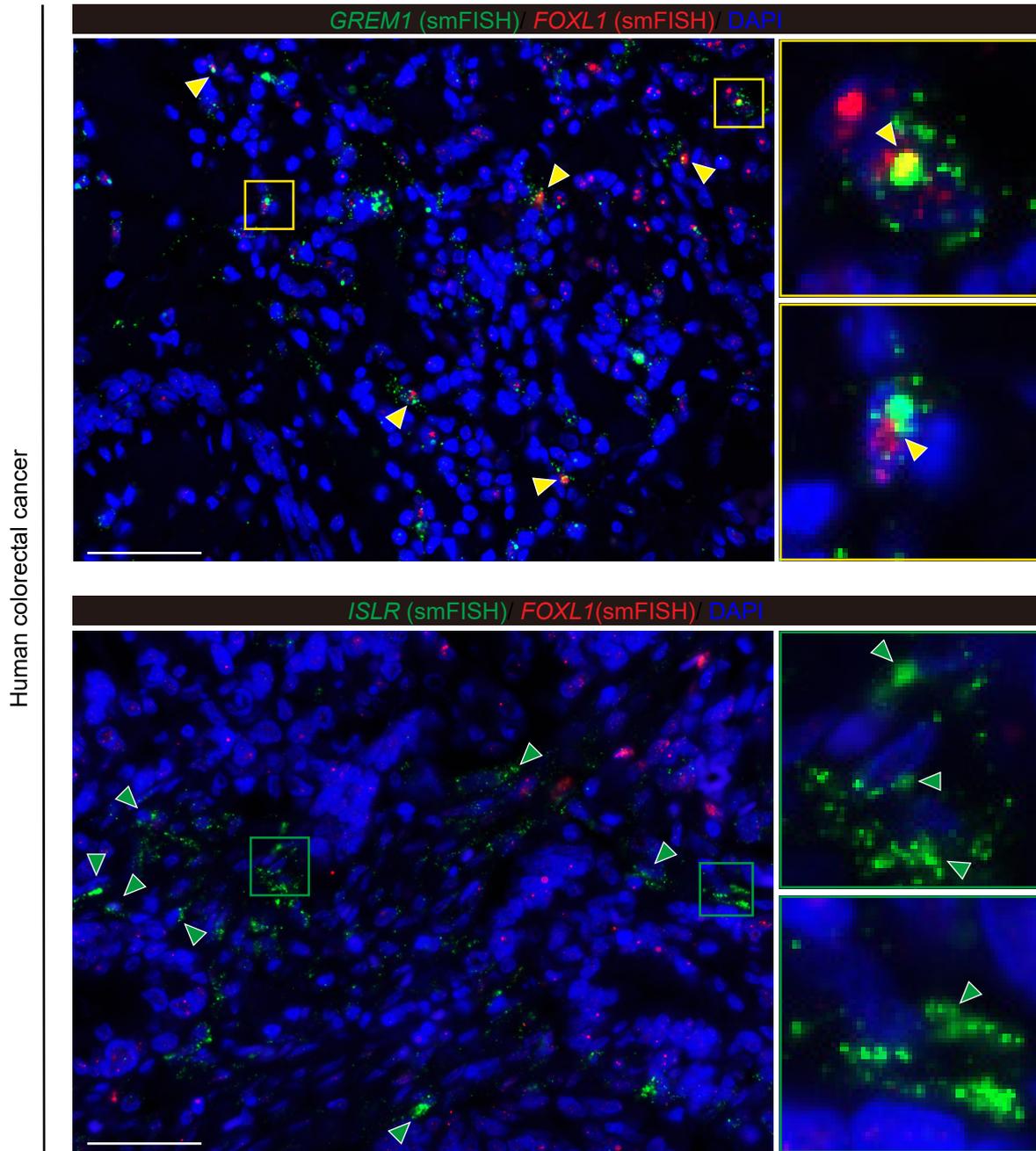
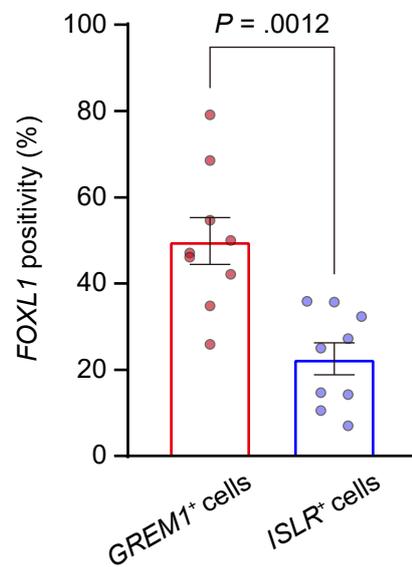
**Supplementary Figure 14: ScRNA-seq from human CRC tissues reveals that *GREM1* identifies a myofibroblastic CAF subpopulation that is distinct from *ISLR*<sup>+</sup> CAFs.**

**(A, B)** Spearman correlation analyses of gene expression levels in scRNA-seq data from human CRC tissues (GSE132465)<sup>226</sup>. **(A)** *GREM1* transcript levels are inversely correlated with *ISLR* expression in human CRC CAFs. **(B)** *GREM1*, but not *ISLR*, transcripts are positively correlated with *ACTA2* transcripts. In a total of 1501 CAFs, 1156 human CRC CAFs that express *GREM1* or *ISLR* transcripts were analyzed. Spearman r and p-values are reported in the figure. Solid pink line, linear regression.

**(C, D)** Violin plots depicting *GREM1* and *ISLR* expression (C) and BMP7 expression (D) in four different CRC CAF subclusters<sup>226</sup>. Myo, Myofibroblasts; S1, Stromal 1 fibroblasts; S2, Stromal 2 fibroblasts; S3, Stromal 3 fibroblasts. n = 1146 (Myofibroblasts), 73 (S1 fibroblasts), 158 (S2 fibroblasts), and 124 cells (S3 fibroblasts).

Kruskal-Wallis test followed by Dunn's post-hoc multiple comparisons.

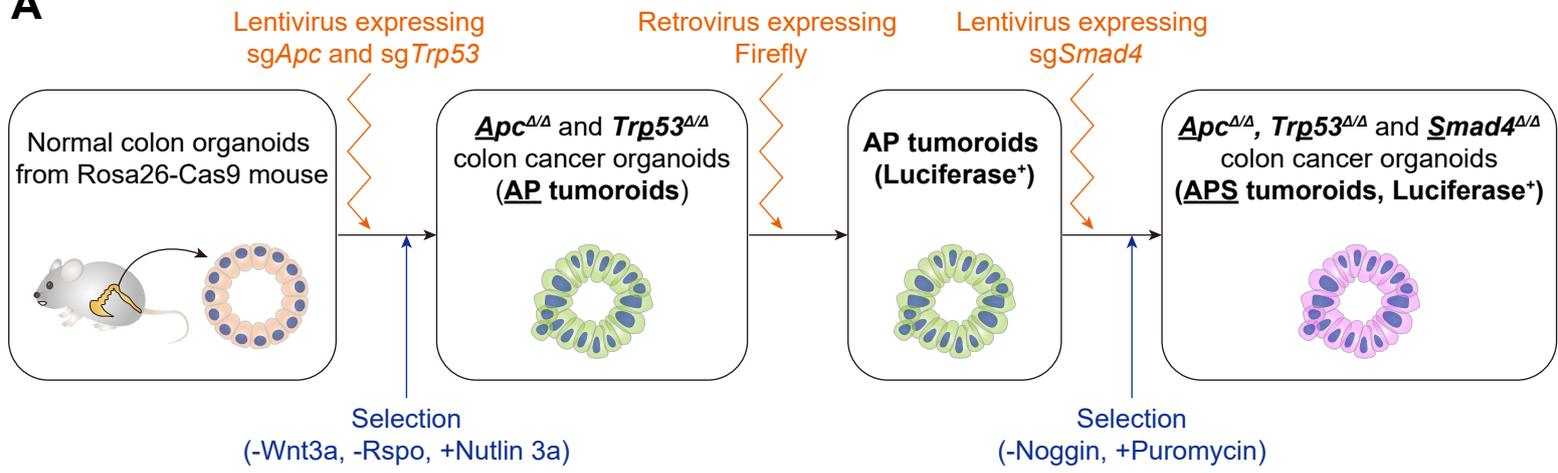
\*\*\*\*, P < 0.0001; \*\*\*, P = 0.0005; \*\*, P = 0.0098; ns, P = 0.1375. Solid black lines, median; Dotted black lines, quartiles.

**A****B**

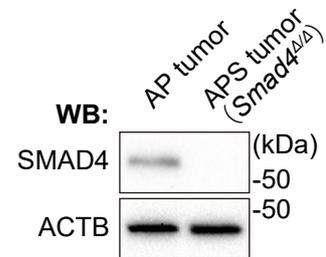
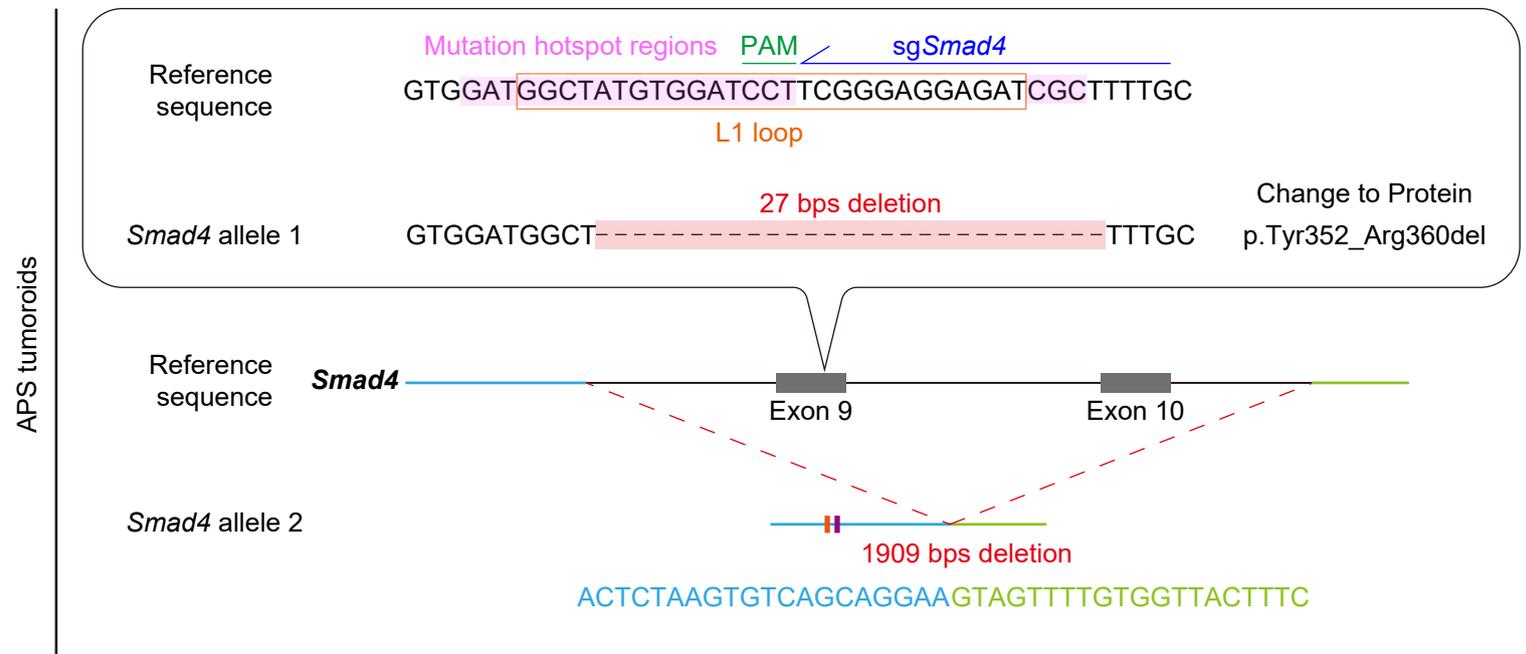
**Supplementary Figure 15: *GREM1*<sup>+</sup> CAFs shows higher positivity for *FOXL1* in human CRC than *ISLR*<sup>+</sup> CAFs.**

**(A, B)** Dual smFISH for *GREM1/FOXL1* (upper panel in A) and *ISLR/FOXL1* (lower panel in A) in human CRC samples. **(A)** Representative pictures. Yellow and green arrowheads denote *GREM1*<sup>+</sup>*FOXL1*<sup>+</sup> cells and *ISLR*<sup>+</sup>*FOXL1*<sup>-</sup> cells, respectively. The boxed areas are magnified in the adjacent panels. **(B)** Semi-quantification of *FOXL1* positivity in *GREM1*<sup>+</sup> cells and *ISLR*<sup>+</sup> cells. 3 HPFs/patient, 3 patients.

Mean  $\pm$  s.e.m.. Mann-Whitney U-test. Scale bars, 50  $\mu$ m.

**A****B**

| <i>Apc</i>            |  |                    |
|-----------------------|--|--------------------|
| Reference sequence    | GACCAGGAAGCCTTGTGGGACA TGGGGGCA                      | Change to Protein  |
| <i>Apc</i> allele 1   | GACCAGGAAGCCTTGTGGGACA <b>1bp insertion</b> TGGGGGCA | p.Met717AsnfsTer37 |
| <i>Apc</i> allele 2   | GACCAGGAAGCCTTGT <b>5bp deletion</b> TGGGGGCA        | p.Trp715TyrfsTer37 |
| <i>Trp53</i>          |  |                    |
| Reference sequence    | TACATGTGTAATAGCTCCTGCA TGGGGGGC                      | Change to Protein  |
| <i>Trp53</i> allele 1 | TACATGTGTAATAGCTCCTGCA <b>2bp deletion</b> GGGGGC    | p.Met240ArgfsTer27 |
| <i>Trp53</i> allele 2 | TACATGTGTAATAGCTCCTGCA <b>1bp insertion</b> TGGGGGGC | p.Met240AsnfsTer28 |

**C****D**

**Supplementary Figure 16: Generation of *Apc*<sup>ΔΔ</sup>, *Trp53*<sup>ΔΔ</sup> organoids (AP tumoroids) and *Smad4*-mutant AP tumoroids (APS tumoroids)**

**(A)** Schematic illustration of the CRISPR/Cas9-mediated sequential engineering of AP tumoroids and APS tumoroids from normal colon organoids isolated from a Rosa26-Cas9 mouse. Correctly targeted clones were enriched by selection medium as indicated. Monoclonal lines were handpicked after each transduction step. sg, single guide RNA.

**(B)** *Apc* and *Trp53* DNA sequence verification of biallelic insertion/deletion mutations, which result in prematurely truncated proteins.

fs, frameshift mutation; Ter, translation termination (stop) codon; PAM, Protospacer adjacent motif.

**(C)** Decreased SMAD4 protein expression in APS tumoroids was validated by Western blotting (WB).

**(D)** *Smad4* DNA sequence verification of biallelic mutations, resulting in disruption of the sequence of L1 loop protein, which is involved in the formation of SMAD protein complex formation<sup>264</sup>.

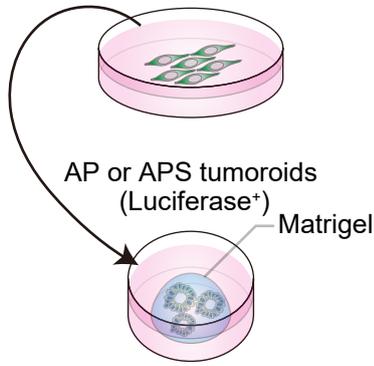
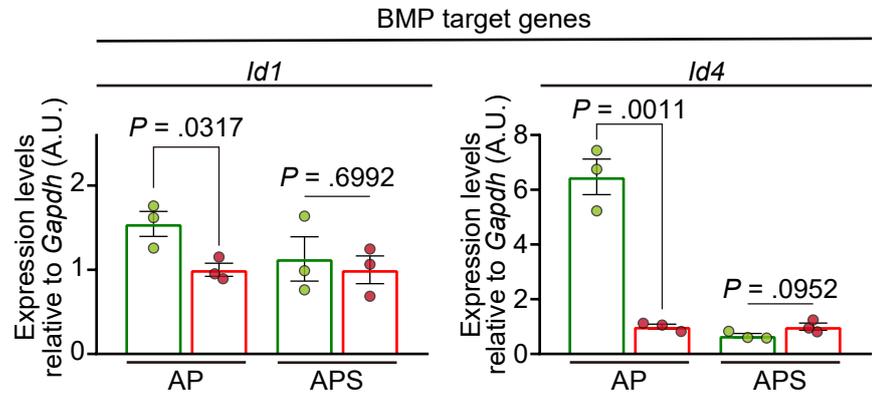
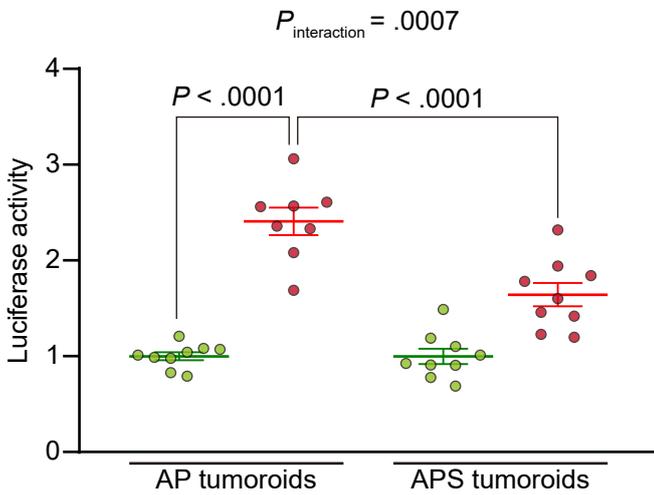
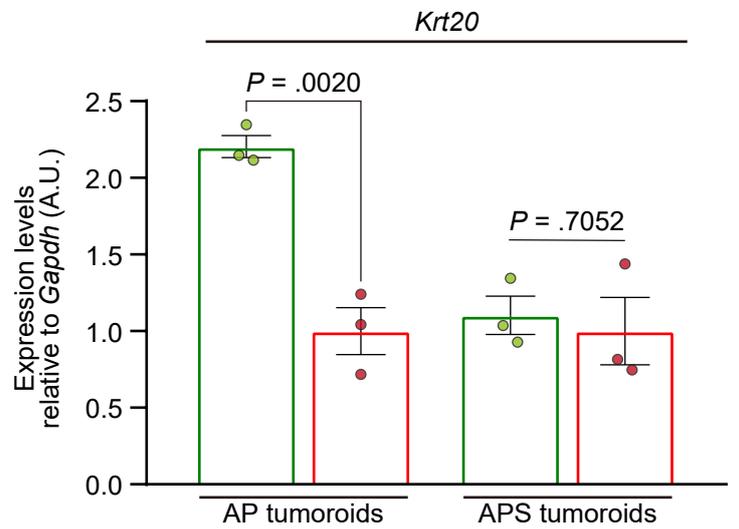
sg*Smad4* was designed to target the genomic sequence corresponding to *SMAD4* mutation hotspot regions in human CRC<sup>264</sup>, which were mapped to the L1 loop and its adjacent region<sup>264</sup>.

To simplify graphical presentation, the following mutations were not shown in the schematics in (D).

*Smad4* allele 1, c.1005A>G (Silent mutation in Exon 9); *Smad4* allele2, c.953-831A>T (Mutation in an intron; Orange bar) and c.953-805\_953-804insAAAAAATAAATAAAAAAATAAAT (Mutation in an intron; Magenta bar).

Changes to proteins are reported according to the Human Genome Variation Society nomenclature.

Reference sequences are as follows. *Apc*, NM\_001360980.1; *Trp53*, NM\_011640.3; *Smad4*, NM\_008540.3.

**A**GFP-YH2 cells or *Grem1*-YH2 cells**B**CM from; ● GFP-YH2 cells  
● *Grem1*-YH2 cells**C**CM from; ● GFP-YH2 cells  
● *Grem1*-YH2 cells**D**CM from; ● GFP-YH2 cells  
● *Grem1*-YH2 cells

**Supplementary Figure 17: Conditioned medium transfer from *Grem1*-overexpressing YH2 cells decreases BMP signaling and accelerates organoid growth in AP tumoroids, but not in APS tumoroids.**

**(A)** Experimental schematic depicting conditioned medium transfer from *Grem1*-overexpressing YH2 cells or GFP-overexpressing YH2 cells to AP (*Apc*<sup>ΔΔ</sup>, *Trp53*<sup>ΔΔ</sup>) tumoroids or APS (*Apc*<sup>ΔΔ</sup>, *Trp53*<sup>ΔΔ</sup>, *Smad4*<sup>ΔΔ</sup>) tumoroids.

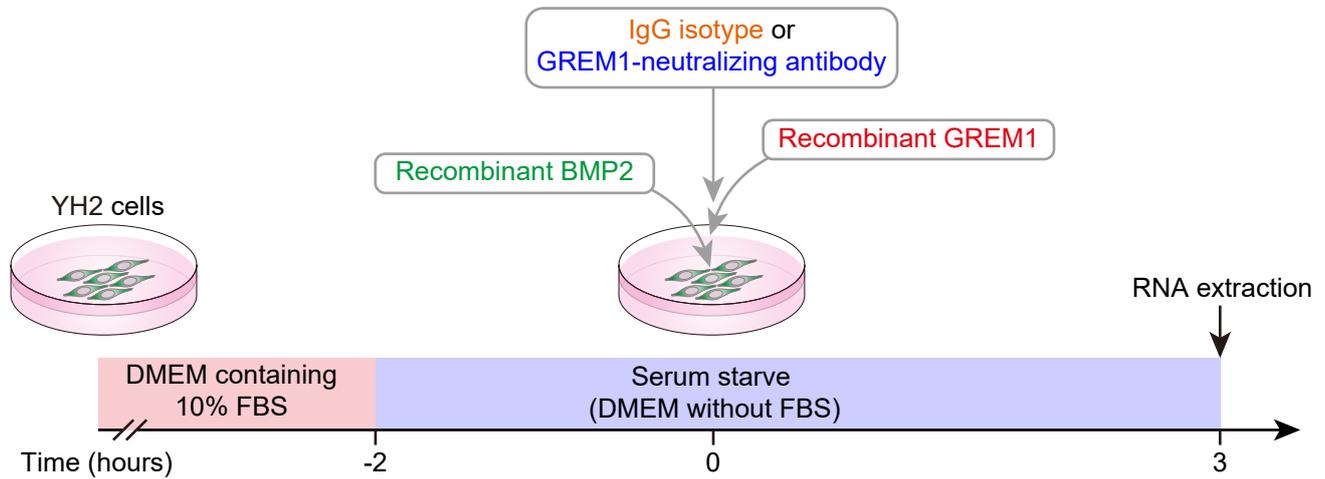
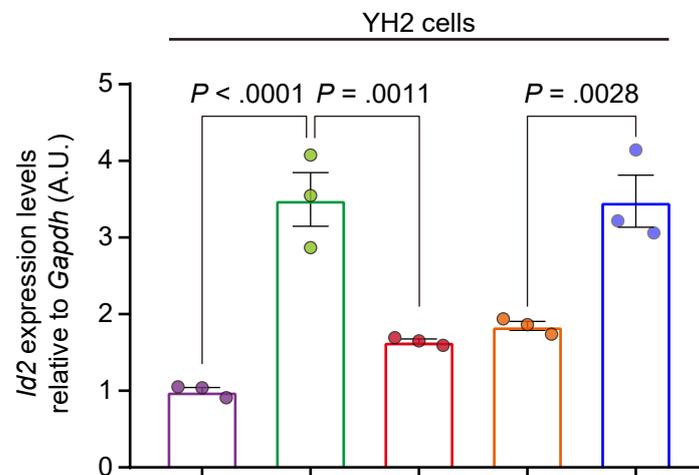
**(B)** qRT-PCR for BMP target genes in AP and APS tumoroids (n = 3). A.U., Arbitrary Units.

**(C)** Luciferase signals from AP tumoroids and APS tumoroids (n ≥ 8).

**(D)** qRT-PCR for *Krt20* in AP tumoroids and APS tumoroids (n = 3).

Note that data normalization was performed within the AP and APS tumoroid groups separately (**B-D**).

Statistical analyses were performed using two-tailed unpaired Student's t-test (B and D) and two-way ANOVA with Tukey's post-hoc multiple comparisons (C). Data are represented as mean ± s.e.m..

**A****B**

|  |   |   |   |   |   |
|--|---|---|---|---|---|
| Recombinant BMP2 (100 ng/ml)           | - | + | + | + | + |
| Recombinant GREM1 (200 ng/ml)          | - | - | + | + | + |
| Isotype IgG (20 ug/ml)                 | - | - | - | + | - |
| GREM1-neutralizing antibody (20 ug/ml) | - | - | - | - | + |

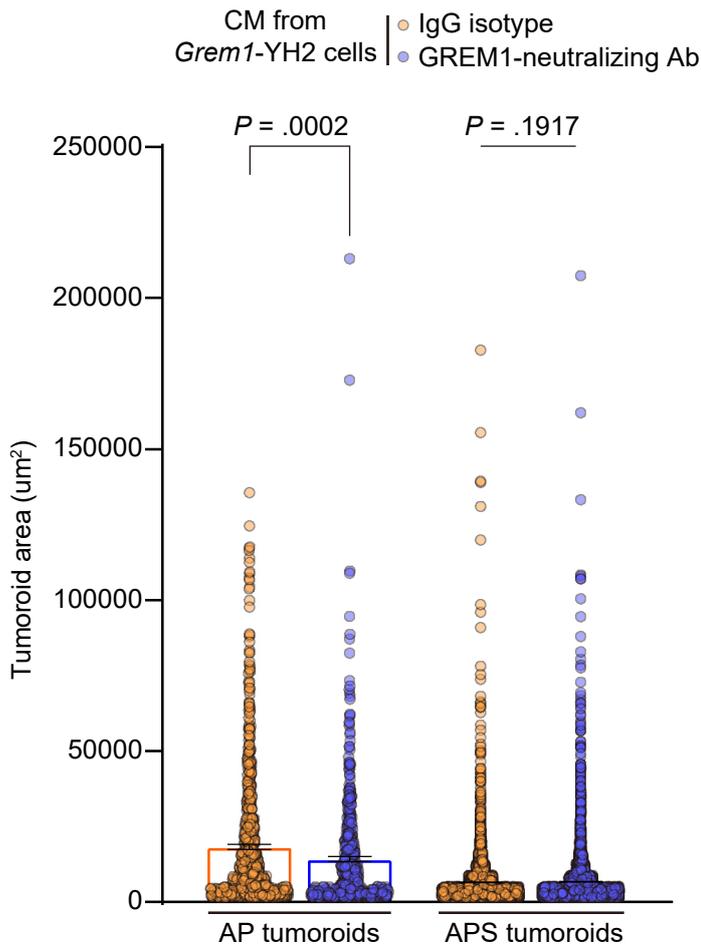
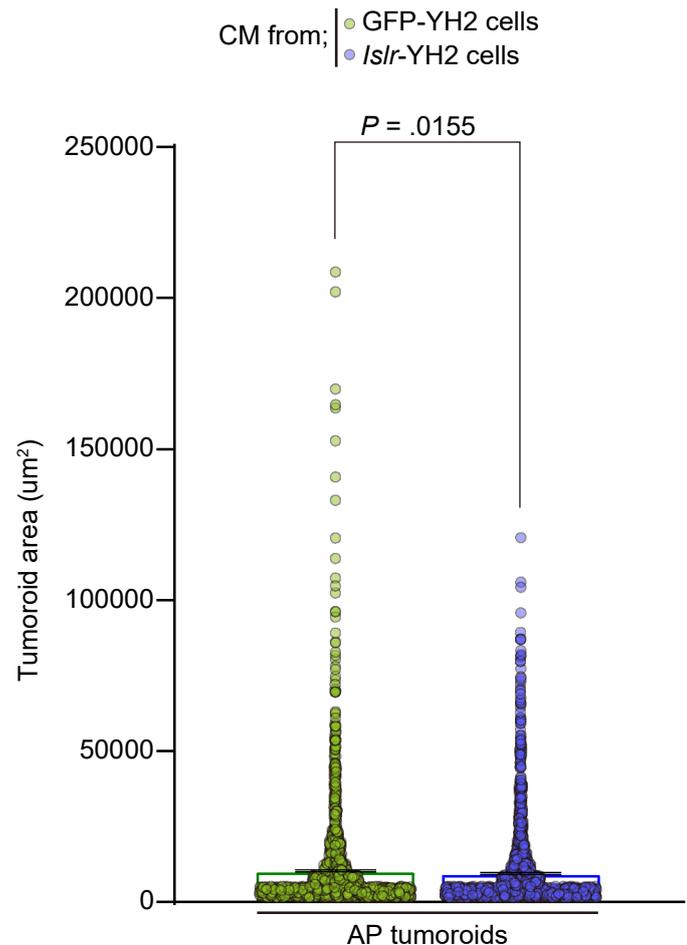
**Supplementary Figure 18: Treatment with a GREM1-neutralizing antibody blocks GREM1-mediated inhibition of BMP2 signaling *in vitro*.**

**(A, B)** Validation of the GREM1-neutralizing antibody using YH2 cells. **(A)** Experimental schematic.

YH2 cells were serum-starved for 2 hours before treating with the indicated proteins for 3 hours.

**(B)** qRT-PCR of *Id2* in YH2 cells was performed to confirm the effect of the GREM1-neutralizing antibody in neutralizing recombinant GREM1 and restoring BMP signaling. n = 3 each.

Data are represented as mean  $\pm$  s.e.m.. Statistical analyses were performed using one-way ANOVA with Tukey's post-hoc multiple comparisons (B).

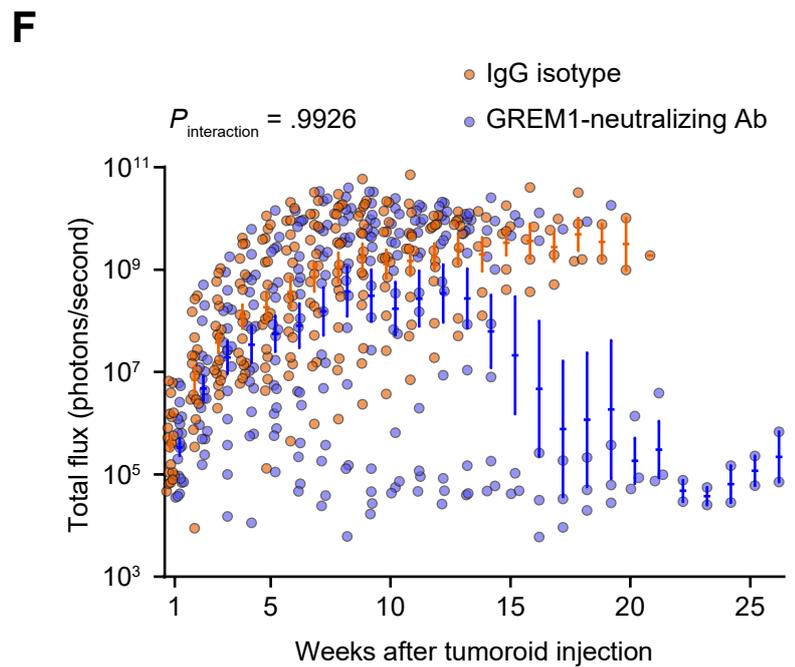
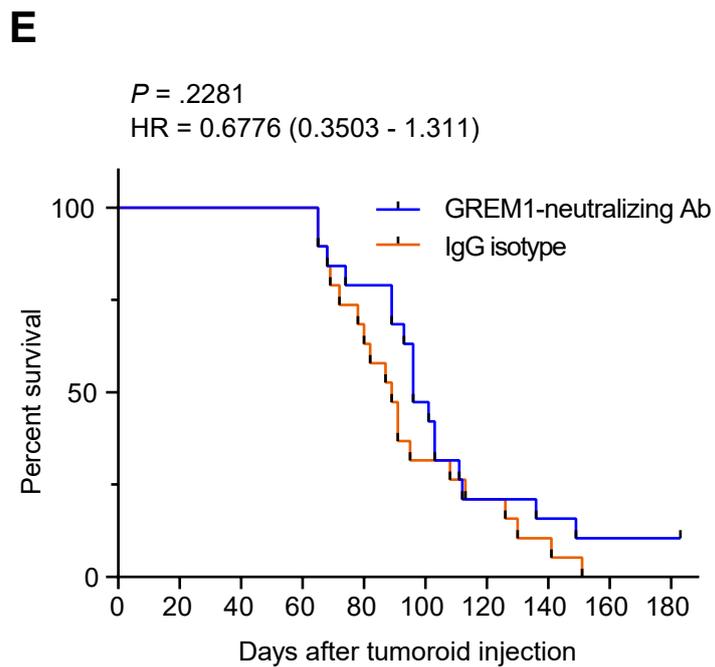
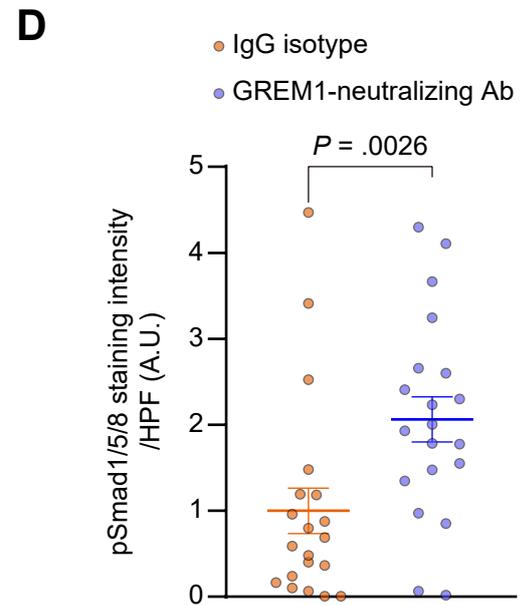
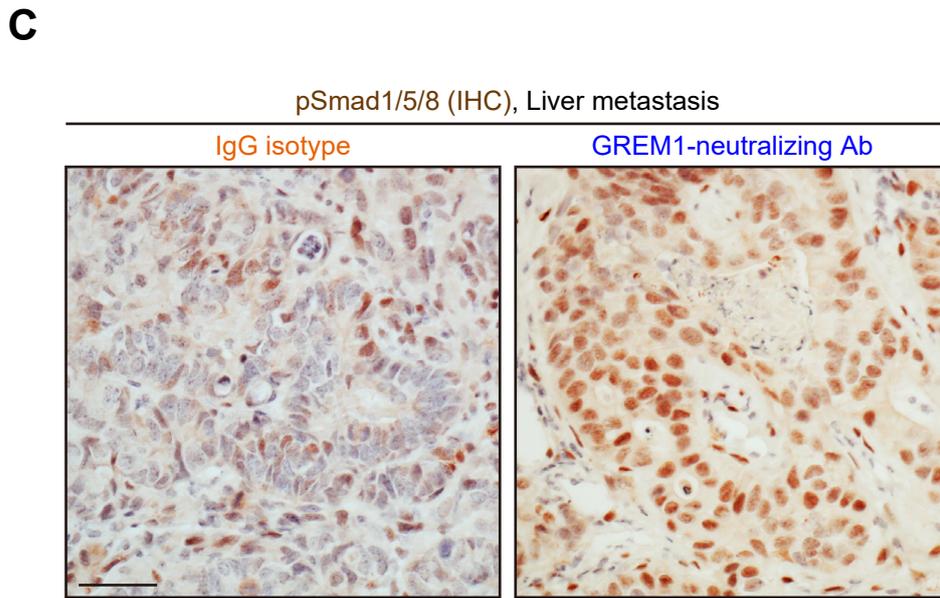
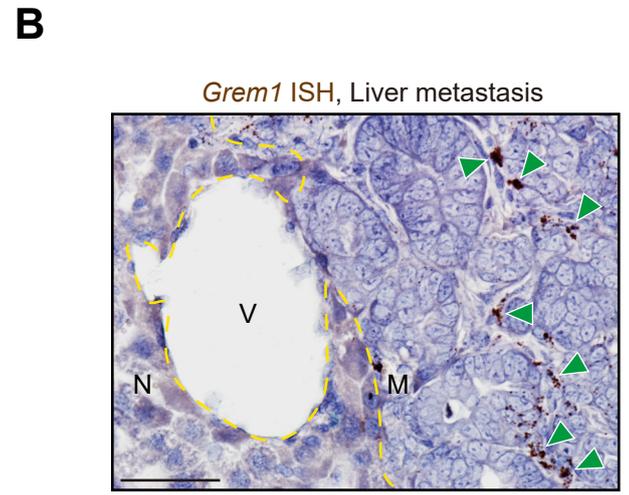
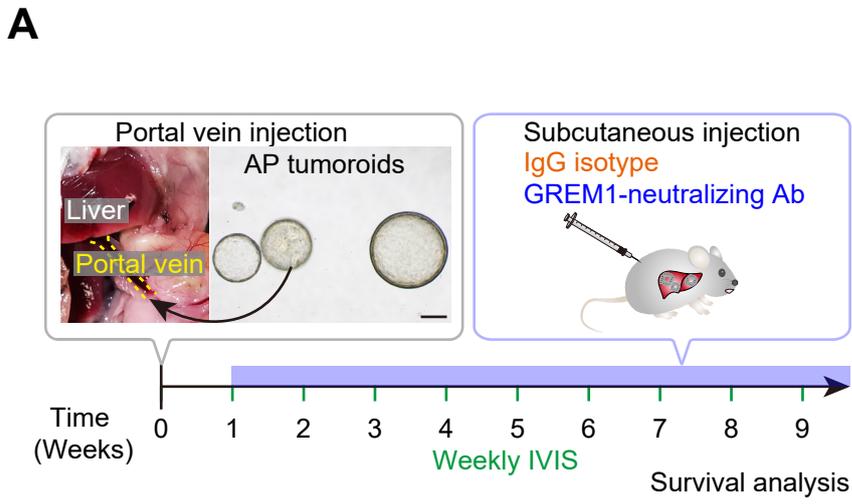
**A****B**

**Supplementary Figure 19: Treatment with a GREM1-neutralizing antibody or conditioned medium from *Islr*-overexpressing intestinal fibroblasts attenuates tumoroid growth as assessed by tumoroid area.**

**(A)** Treatment with a GREM1-neutralizing antibody reduces AP tumoroid areas.  $n = 790, 522, 3571,$  and 3494 tumoroids (left to right) from 6 (AP tumoroids) and 9 independent replicates (APS tumoroids). Note that AP tumoroid and APS tumoroids were evaluated 8 and 12 days after being plated. In this experimental condition, APS tumoroids generated smaller tumoroid areas than AP tumoroids when treated with an IgG isotype ( $P < 0.0001$ , Mann-Whitney U-test). CM, Conditioned Medium.

**(B)** Conditioned medium transfer from *Islr*-overexpressing YH2 cells decreases AP tumoroid areas.  $n = 1873$  (GFP-YH2 cells) and 1713 tumoroids (*Islr*-YH2 cells) from 9 independent replicates. The AP tumoroids were evaluated 7 days after being plated.

Tumoroid areas were assessed by Image J. Mean  $\pm$  s.e.m., Mann-Whitney U-test (A and B).



**Supplementary Figure 20: A GREM1-neutralizing antibody-treated group showed enhanced BMP signaling and a trend toward prolonged survival in a mouse model of CRC liver metastasis.**

**(A)** Experimental scheme for portal vein injection of AP tumoroids and subsequent GREM1-neutralizing antibody administration. The GREM1-neutralizing antibody was subcutaneously administered 1 week after portal vein injection of AP tumoroids until mice reach humane endpoints. Yellow dotted lines outline the portal vein. Scale bar, 200  $\mu\text{m}$ .

**(B)** ISH for *Grem1* in the mouse CRC liver metastasis. *Grem1* was expressed by the fibroblastic cells in the liver metastasis mesenchyme (M), but not in the normal liver (N). Yellow dotted lines demarcate the normal liver (N) and CRC liver metastasis (M). Green arrowheads indicate *Grem1* expression. V, blood vessel. n = 3 mice. Scale bar, 50  $\mu\text{m}$ .

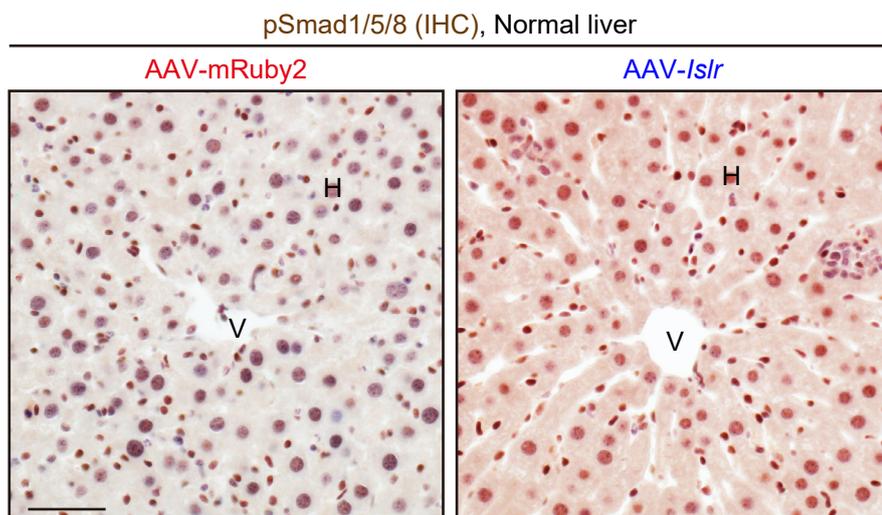
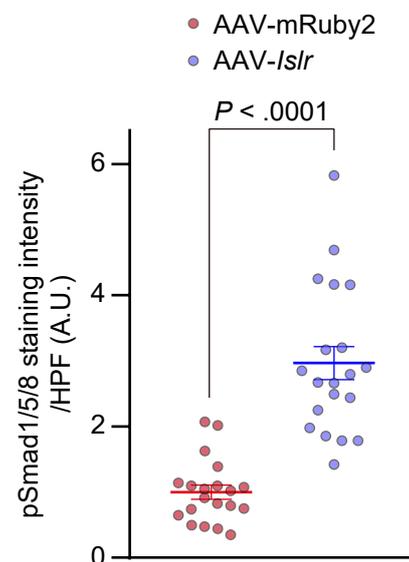
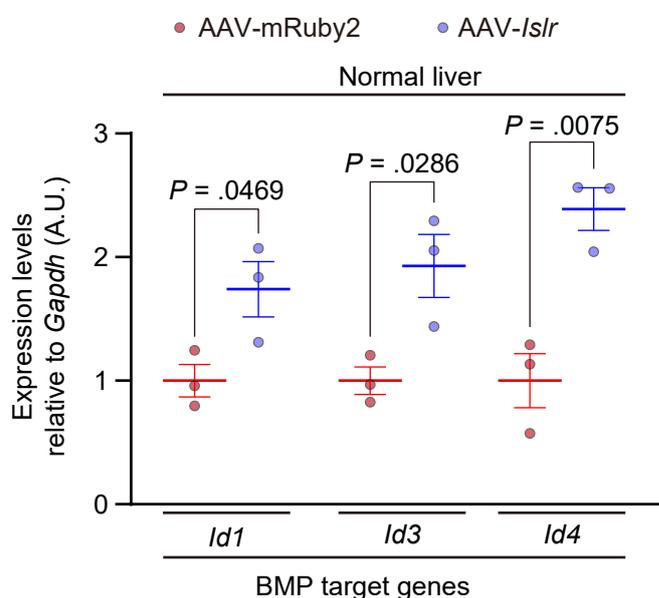
**(C, D)** Immunohistochemistry (IHC) for pSmad1/5/8 in the liver metastases developed in IgG isotype-treated or GREM1-neutralizing antibody-treated mice. **(C)** Representative pictures. Scale bar, 50  $\mu\text{m}$ . **(D)** Quantification of 3,3'-Diaminobenzidine (DAB) intensity by ImageJ. 4 HPFs/mouse, 5 mice each. A.U., arbitrary unit.

**(E)** Kaplan-Meier survival curves. n = 19 (IgG isotype) and 19 mice (GREM1-neutralizing antibody). Hazard ratio (HR) and 95% confidence interval of the hazard ratio are shown (Log-rank method).

**(F)** Growth kinetics of CRC hepatic metastases was assessed during the survival analysis in (E). AP tumoroid-derived luciferase signals were evaluated using an *in vivo* imaging system (IVIS).

Mean  $\pm$  s.e.m.. Mann-Whitney U-test (D), Log-rank test (E), and two-way repeated-measures ANOVA (F). In (F), in order to exclude the effect of loss of samples during the experiment, statistical analysis was performed using IVIS signal values before Week 10, when the first mouse reached a humane endpoint.

Note that the same pictures were used in Figure 6A and Supplementary Figure 20A in order to show the procedure of portal vein injection.

**A****B****C**

**Supplementary Figure 21: AAV-*Islr* treatment augments BMP signaling in the normal mouse liver.**

(A, B) Immunohistochemistry (IHC) for pSmad1/5/8 in the normal liver from AAV-*Islr*-treated or AAV-mRuby2-treated mice. (A) Representative pictures. Scale bar, 50  $\mu$ m. V, central vein; H, hepatocytes. (B) Quantification of 3,3'-Diaminobenzidine (DAB) intensity by ImageJ. 5 HPFs/mouse, 4 mice each. A.U., arbitrary unit.

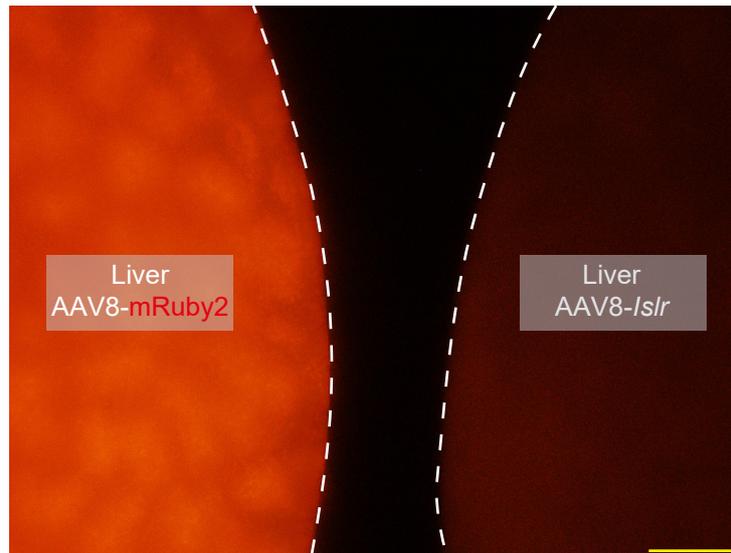
(C) qRT-PCR for BMP target genes using normal liver tissue lysates. n = 3 mice each.

Mice were harvested 5-6 weeks after tail vein injection (3-4 weeks after tumor injection) (A-C).

Mean  $\pm$  s.e.m.. Mann-Whitney U-test (B) and two-tailed unpaired Student's t-test (C).

**A**

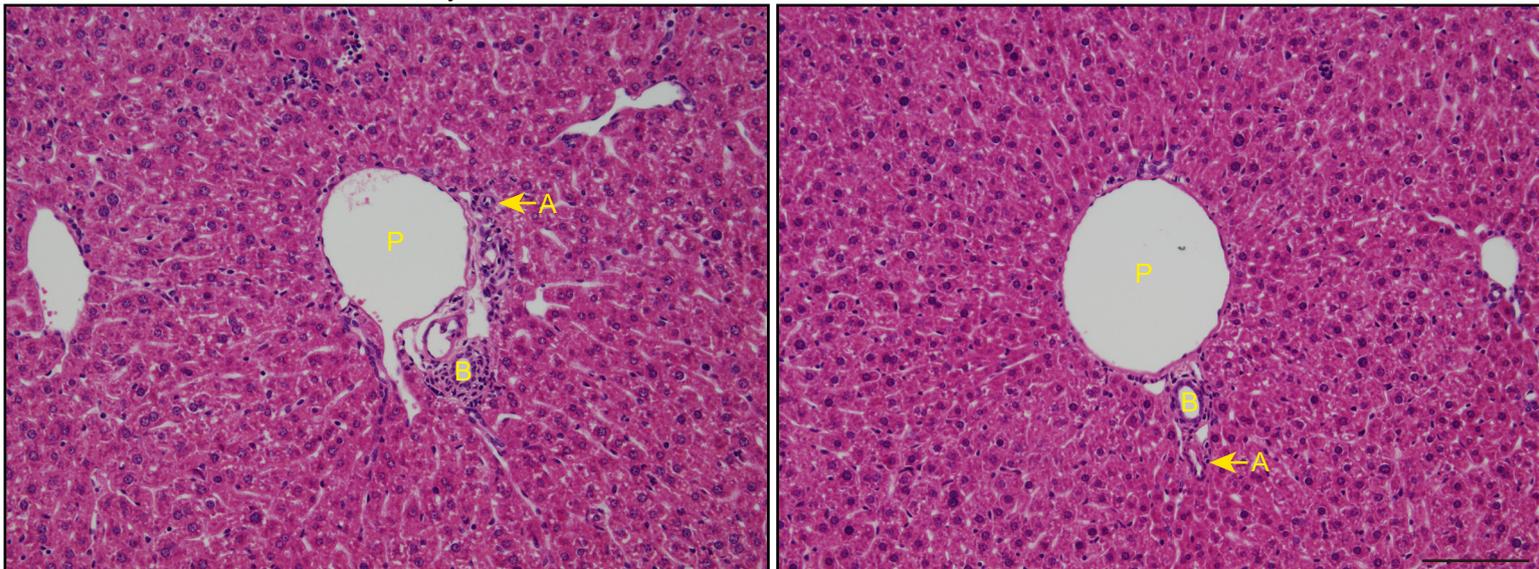
Fluorescence microscopy, Liver  
2 weeks after tail vein injection of AAV8

**B**

H&E staining, Liver, 2 weeks after tail vein injection

AAV8-mRuby2

AAV8-Is/r



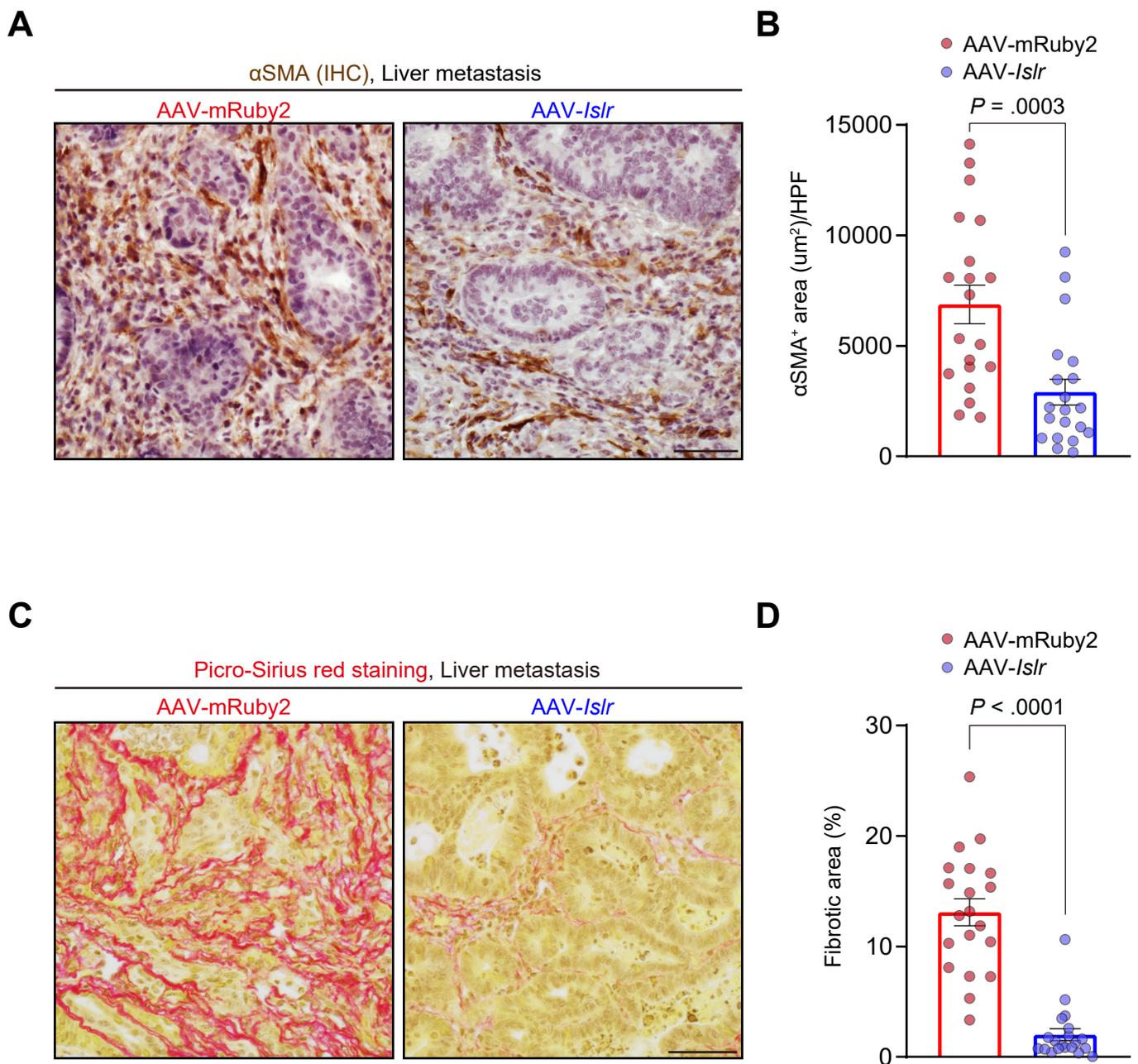
**Supplementary Figure 22: No liver injury was apparent in mice administered AAV8-Is/r via histological analysis.**

**(A)** Fluorescent microscopy image showing expression of a red fluorescence protein (mRuby2) in the liver of mice harvested two weeks after tail vein injection of AAV8-mRuby2 and AAV8-Is/r.

White dotted lines indicate the edges of the livers. Yellow scale bar, 500  $\mu$ m.

**(B)** Hematoxylin and eosin (H&E) staining of the liver section from mice harvested two weeks after tail vein injection of AAV8-mRuby2 and AAV8-Is/r. n = 3 mice each.

A, hepatic artery (arrows); B, bile duct; P, portal vein. Scale bar, 100  $\mu$ m.



**Supplementary Figure 23: Mice administered AAV8-*Islr* show reduced fibrosis in CRC liver metastases.**

(A, B) Immunohistochemistry (IHC) for  $\alpha$ SMA using CRC liver metastases in mice administered AAV8-*Islr* or AAV8-mRuby2. (A) Representative pictures. (B) Quantification of 3,3'-Diaminobenzidine (DAB)<sup>+</sup> areas by ImageJ. 5 HPFs/mouse, 4 mice each.

(C, D) Picro-Sirius red staining for collagen using the CRC liver metastasis sections. (C) Representative pictures. (D) Quantification of Picro-Sirius-red-positive areas by ImageJ. 5 HPFs/mouse, 4 mice each.

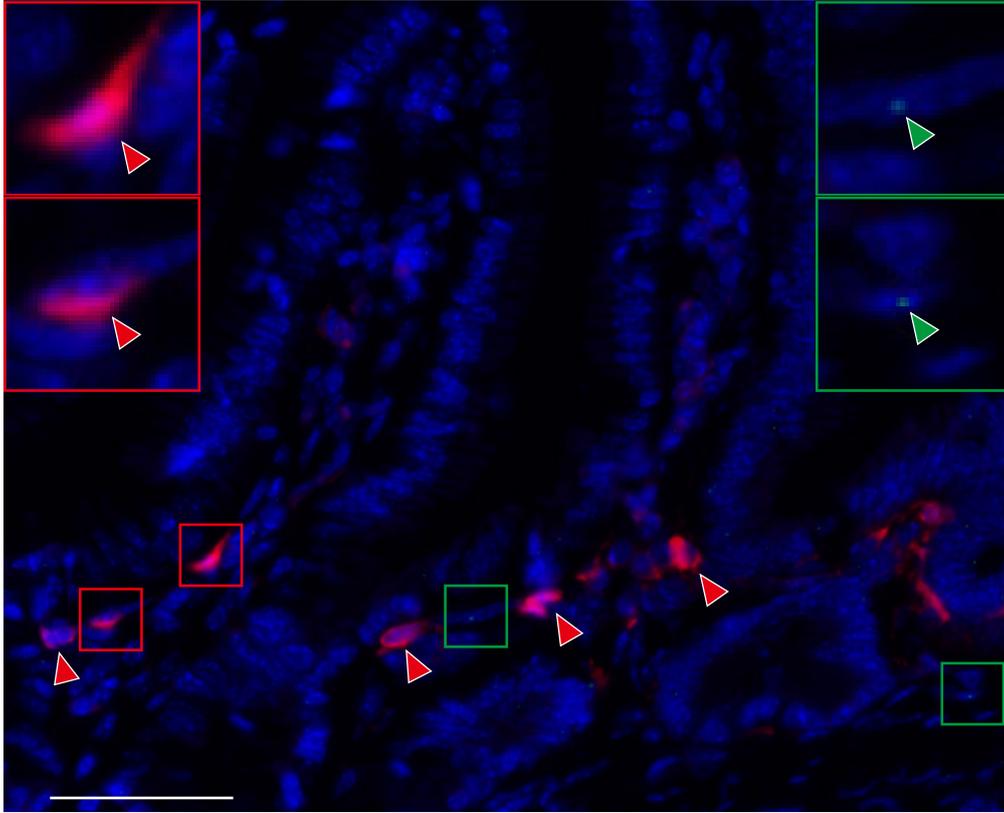
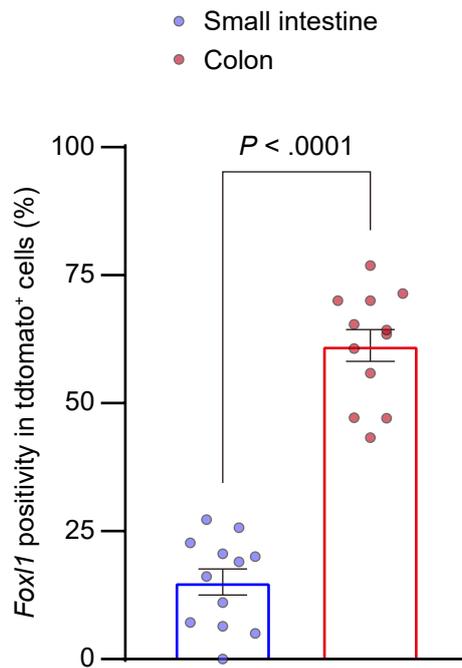
Mean  $\pm$  s.e.m.. Mann-Whitney U-test (B and D). Scale bars, 50  $\mu$ m.

**A**

Normal small intestine, Tamoxifen induction at 6-7 week, Collected at 8 week

*Grem1*-CreERT2; *Rosa26*-*tdtomato*

*tdtomato* (IF) *Foxl1* (smFISH)

**B**

**Supplementary Figure 24: *Grem1*<sup>+</sup> cells in the small intestine show lower *Foxl1* positivity than those in the colon.**

**(A, B)** Single-molecule fluorescent ISH (smFISH) for *Foxl1* and immunofluorescence (IF) for tdTomato in the small intestine from *Grem1*-CreERT2; *Rosa26*-tdTomato mice. **(A)** Representative picture. Red and green arrowheads denote *Grem1*<sup>+</sup>*Foxl1*<sup>-</sup> cells and *Grem1*<sup>-</sup>*Foxl1*<sup>+</sup> cells, respectively. The boxed areas are magnified in the insets. **(B)** *Foxl1* positivity in the *Grem1*<sup>+</sup> cells in the small intestine (blue). For statistical comparison, the result with the colon from Figure 3D is shown on the right (red). 4 HPFs (400x)/mouse, 3 mice each. Two-tailed unpaired Student's t-test. Scale bars, 50  $\mu$ m.

## Supplementary Table 1

List of 34 genes upregulated specifically in CRC CAFs

---

|                 |                 |               |               |
|-----------------|-----------------|---------------|---------------|
| <i>ADAMTS2</i>  | <i>COL1A1</i>   | <i>CTSK</i>   | <i>NOTCH3</i> |
| <i>AEBP1</i>    | <i>COL1A2</i>   | <i>CXCL14</i> | <i>PALLD</i>  |
| <i>ANTXR1</i>   | <i>COL3A1</i>   | <i>CYGB</i>   | <i>POSTN</i>  |
| <i>C11orf96</i> | <i>COL5A1</i>   | <i>DCN</i>    | <i>SDC2</i>   |
| <i>C1R</i>      | <i>COL5A2</i>   | <i>GREM1</i>  | <i>TAGLN</i>  |
| <i>C1S</i>      | <i>COL6A1</i>   | <i>ISLR</i>   | <i>THBS2</i>  |
| <i>C3</i>       | <i>COL6A3</i>   | <i>LUM</i>    | <i>VCAN</i>   |
| <i>CDR1</i>     | <i>COL8A1</i>   | <i>MEG3</i>   |               |
| <i>COL12A1</i>  | <i>CRISPLD2</i> | <i>MXRA5</i>  |               |

---

(GSE39396)

This list of 34 genes (corresponding to 63 transcripts in the microarray data) represents an intersection of the Venn diagram in **Figure 1A**.

*GREM1* and *ISLR* are highlighted in red.

## Supplementary Table 2

Clinicopathological characteristics of rectal cancer patients analyzed by *GREM1* and *ISLR* ISH.

|   | All        | <i>GREM1</i><br>low | <i>GREM1</i><br>high | P values     | <i>ISLR</i><br>low | <i>ISLR</i><br>high | P values     |
|---|------------|---------------------|----------------------|--------------|--------------------|---------------------|--------------|
| Number of patients                          | 53         | 30                  | 23                   |              | 24                 | 29                  |              |
| Median age at diagnosis (range), years      | 64 (28-89) | 65 (38-88)          | 64 (28-89)           |              | 66 (36-89)         | 63 (28-82)          |              |
| Median disease-free survival months (range) | 40 (1-104) | 53 (6-104)          | 29 (1-77)            |              | 22 (1-86)          | 52 (6-104)          |              |
| Age, n (%)                                  |            |                     |                      | 0.775        |                    |                     | 0.780        |
| < 60  | 19 (35.8%) | 10 (33.3%)          | 9 (39.1%)            |              | 8 (33.3%)          | 11 (37.9%)          |              |
| ≥ 60  | 34 (64.2%) | 20 (66.7%)          | 14 (60.9%)           |              | 16 (66.7%)         | 18 (62.1%)          |              |
| Sex, n (%)                                  |            |                     |                      | 1.000        |                    |                     | 0.762        |
| Male  | 38 (71.7%) | 21 (70.0%)          | 17 (73.9%)           |              | 18 (75.0%)         | 20 (69.0%)          |              |
| Female                                      | 15 (28.3%) | 9 (30.0%)           | 6 (26.1%)            |              | 6 (25.0%)          | 9 (31.0%)           |              |
| pT status, n (%)                            |            |                     |                      | 0.512        |                    |                     | 0.284        |
| pCR   | 1 (1.9%)   | 1 (3.3%)            | 0 (0.0%)             |              | 1 (4.2%)           | 0 (0.0%)            |              |
| Tis   | 1 (1.9%)   | 1 (3.3%)            | 0 (0.0%)             |              | 0 (0.0%)           | 1 (3.4%)            |              |
| T1  | 3 (5.7%)   | 2 (6.7%)            | 1 (4.3%)             |              | 1 (4.2%)           | 2 (6.9%)            |              |
| T2  | 13 (24.5%) | 8 (26.7%)           | 5 (21.7%)            |              | 4 (16.7%)          | 9 (31.0%)           |              |
| T3  | 27 (50.9%) | 12 (40.0%)          | 15 (65.2%)           |              | 12 (50.0%)         | 15 (51.7%)          |              |
| T4  | 8 (15.1%)  | 6 (20.0%)           | 2 (8.7%)             |              | 6 (25.0%)          | 2 (6.9%)            |              |
| pN status, n (%)                            |            |                     |                      | <b>0.031</b> |                    |                     | <b>0.010</b> |
| N0  | 38 (71.7%) | 21 (70.0%)          | 17 (73.9%)           |              | 13 (54.2%)         | 25 (86.2%)          |              |
| N1  | 9 (17.0%)  | 7 (23.3%)           | 2 (8.7%)             |              | 8 (33.3%)          | 1 (3.4%)            |              |
| N2  | 4 (7.5%)   | 0 (0.0%)            | 4 (17.4%)            |              | 2 (8.3%)           | 2 (6.9%)            |              |
| N3  | 2 (3.8%)   | 2 (6.7%)            | 0 (0.0%)             |              | 1 (4.2%)           | 1 (3.4%)            |              |
| pM status, n (%)                            |            |                     |                      | 1.000        |                    |                     | 0.453        |
| M0  | 52 (98.1%) | 29 (96.7%)          | 23 (100.0%)          |              | 23 (95.8%)         | 29 (100.0%)         |              |
| M1  | 1 (1.9%)   | 1 (3.3%)            | 0 (0.0%)             |              | 1 (4.2%)           | 0 (0.0%)            |              |
| pTNM stage, n (%)                           |            |                     |                      | 0.978        |                    |                     | <b>0.004</b> |
| pCR   | 1 (1.9%)   | 1 (3.3%)            | 0 (0.0%)             |              | 1 (4.2%)           | 0 (0.0%)            |              |
| Stage 0 (Tis)                               | 1 (1.9%)   | 1 (3.3%)            | 0 (0.0%)             |              | 0 (0.0%)           | 1 (3.4%)            |              |
| Stage I                                     | 12 (22.6%) | 7 (23.3%)           | 5 (21.7%)            |              | 1 (4.2%)           | 11 (37.9%)          |              |
| Stage II                                    | 24 (45.3%) | 12 (40.0%)          | 12 (52.2%)           |              | 11 (45.8%)         | 13 (44.8%)          |              |
| Stage III                                   | 14 (26.4%) | 8 (26.7%)           | 6 (26.1%)            |              | 10 (41.7%)         | 4 (13.8%)           |              |
| Stage IV                                    | 1 (1.9%)   | 1 (3.3%)            | 0 (0.0%)             |              | 1 (4.2%)           | 0 (0.0%)            |              |
| Differentiation, n (%)                      |            |                     |                      | 1.000        |                    |                     | <b>0.035</b> |
| Well  | 6 (11.3%)  | 4 (13.3%)           | 2 (8.7%)             |              | 0 (0.0%)           | 6 (20.7%)           |              |
| Moderate                                    | 41 (77.4%) | 23 (76.7%)          | 18 (78.3%)           |              | 20 (83.3%)         | 21 (72.4%)          |              |
| Poor  | 4 (7.5%)   | 2 (6.7%)            | 2 (8.7%)             |              | 2 (8.3%)           | 2 (6.9%)            |              |
| Mucinous                                    | 2 (3.8%)   | 1 (3.3%)            | 1 (4.3%)             |              | 2 (8.3%)           | 0 (0.0%)            |              |
| Neoadjuvant treatment, n (%)                |            |                     |                      | 0.266        |                    |                     | <b>0.029</b> |
| No  | 24 (45.3%) | 16 (53.3%)          | 8 (34.8%)            |              | 15 (62.5%)         | 9 (31.0%)           |              |
| Chemotherapy                                | 29 (54.7%) | 14 (46.7%)          | 15 (65.2%)           |              | 9 (37.5%)          | 20 (69.0%)          |              |
| <i>ISLR</i> expression, n (%)               |            |                     |                      | 0.579        |                    |                     |              |
| Low   | 24 (45.3%) | 15 (50%)            | 9 (39.1%)            |              |                    |                     |              |
| High  | 29 (54.7%) | 15 (50%)            | 14 (60.9%)           |              |                    |                     |              |

P value less than 0.05 is shown in bold characters (Fisher's exact test).

### Supplementary Table 3

Univariate and multivariate Cox regression analysis of *GREM1* and *ISLR* expression levels and disease-free survival of rectal cancer patients

| Variable               | Disease-free survival |               |                       |               |
|------------------------|-----------------------|---------------|-----------------------|---------------|
|                        | Univariate analysis   |               | Multivariate analysis |               |
|                        | HR (95% CI)           | P Value       | HR (95% CI)           | P Value       |
| <i>GREM1</i>           |                       | <b>0.0025</b> |                       | <b>0.0001</b> |
| Low                    | 1.000                 |               | 1.000                 |               |
| High                   | 4.882 (1.748-13.631)  |               | 8.74 (2.864-26.675)   |               |
| <i>ISLR</i>            |                       | <b>0.0032</b> |                       | <b>0.0013</b> |
| Low                    | 1.000                 |               | 1.000                 |               |
| High                   | 0.213 (0.076-0.596)   |               | 0.124 (0.034-0.444)   |               |
| Age                    |                       | 0.9768        |                       |               |
| < 60                   | 1.000                 |               |                       |               |
| ≥ 60                   | 0.986 (0.388-2.507)   |               |                       |               |
| Sex                    |                       | 0.6791        |                       |               |
| Female                 | 1.000                 |               |                       |               |
| Male                   | 0.815 (0.310-2.146)   |               |                       |               |
| pTNM Stage             |                       | <b>0.0401</b> |                       | 0.8158        |
| pCR and Stage 0, I, II | 1.000                 |               | 1.000                 |               |
| Stage III,IV           | 2.609 (1.044-6.520)   |               | 0.869 (0.267-2.833)   |               |
| Differentiation        |                       | <b>0.0210</b> |                       | 0.1391        |
| Well and Moderate      | 1.000                 |               | 1.000                 |               |
| Poor and Mucinous      | 3.770 (1.221-11.637)  |               | 2.742 (0.72-10.435)   |               |
| Neoadjuvant treatment  |                       | 0.7506        |                       |               |
| No                     | 1.000                 |               |                       |               |
| Chemotherapy           | 0.863 (0.348-2.138)   |               |                       |               |

P values less than 0.05 are shown in bold characters.

## Supplementary Table 4

Univariate and multivariate Cox regression analysis of *GREM1* and *ISLR* expression levels and overall survival of colon cancer patients.

| Variable         | Overall survival    |               |                       |               |
|------------------|---------------------|---------------|-----------------------|---------------|
|                  | Univariate analysis |               | Multivariate analysis |               |
|                  | HR (95% CI)         | P Value       | HR (95% CI)           | P Value       |
| <i>GREM1</i>     |                     | <b>0.0190</b> |                       | <b>0.0132</b> |
| Low (n = 506)    | 1.000               |               | 1.000                 |               |
| High (n = 49)    | 1.704 (1.092-2.660) |               | 1.768 (1.127-2.775)   |               |
| <i>ISLR</i>      |                     | <b>0.0222</b> |                       | <b>0.0132</b> |
| Low (n = 56)     | 1.000               |               | 1.000                 |               |
| High (n = 499)   | 0.606 (0.394-0.931) |               | 0.577 (0.373-0.891)   |               |
| Age              |                     | 0.0758        |                       | 0.1226        |
| < 60 (n = 150)   | 1.000               |               | 1.000                 |               |
| ≥ 60 (n = 405)   | 1.369 (0.968-1.937) |               | 1.315 (0.929-1.862)   |               |
| Sex              |                     | 0.0589        |                       | 0.0648        |
| Female (n = 249) | 1.000               |               | 1.000                 |               |
| Male (n = 306)   | 1.327 (0.989-1.779) |               | 1.319 (0.983-1.768)   |               |

(GSE39582)

P values less than 0.05 are shown in bold characters.

Note that one patient whose age at diagnosis was unknown (Sample ID: GSM972293) was excluded from the univariate and multivariate Cox regression analysis.

## Supplementary Table 5

No overlap between *GREM1*<sup>high</sup> patients and *ISLR*<sup>low</sup> patients in the colon cancer patient dataset (GSE39582).

|                   | <i>ISLR</i> low | <i>ISLR</i> high | Total        |
|-------------------|-----------------|------------------|--------------|
| <i>GREM1</i> low  | 56 (11.1%)      | 450 (88.9%)      | 506 (100.0%) |
| <i>GREM1</i> high | 0 (0.0%)        | 50 (100.0%)      | 50 (100.0%)  |
| Total             | 56              | 500              | 556          |

P = 0.005863 (Fisher's exact test)

Number of cases and the percentage within *GREM1*-low (upper row) or *GREM1*-high cases (lower row) are shown.

## **Chapter 3: The origin and contribution of cancer-associated fibroblasts in colorectal carcinogenesis.**

Having elucidated the functional dichotomy in colorectal CAFs, I next addressed the question of CAF origins. Contrasting with the emerging understanding in CAF heterogeneity, the origins of CAFs are poorly understood<sup>3</sup>. Here, using genetic fate-mapping mouse models, this study identified intestinal Leptin-receptor (*Lepr*)-lineage cells as a major contributor to colorectal CAFs. Through RNA-sequencing of CAFs and immunohistochemistry, I found that these *Lepr*-lineage cells express melanoma cell adhesion molecule (*Mcam*), a colorectal stroma-specific gene. I explored the clinical significance of MCAM expression in human CRC samples. Finally, the biological role of stromal *Mcam* in CRC progression was investigated using *Mcam*-knockout mice.

## Statement of Authorship

|                     |   |
|---------------------|---|
| Title of Paper      | The origin and contribution of cancer-associated fibroblasts in colorectal carcinogenesis   |
| Publication Status  | Submitted work written in a manuscript style  |
| Publication Details | Kobayashi, H., Gieniec, K. A., Lannagan, T. R. M., Wang, T., Asai, N., Mizutani, Y., Iida, T., Ando, R., Sakai, A., Suzuki, N., Ichinose, M., Wright, J. A., Vrbanac, L., Ng, J. Q., Goyne, J., Lawrence, M. J., Sammour, T., Hayakawa, Y., Klebe, S., Shin, A. E., Asfaha, S., Bettington, M., Rieder, F., Arpaia, N., Danino, T., Butler, L. M., Burt, A. D., Leedham, S. J., Rustgi, A. K., Mukherjee, S., Takahashi, M., Wang, T. C., Enomoto, A., Woods, S. L., Worthley, D. L. (Unpublished). The origin and contribution of the tumor stroma in colorectal carcinogenesis. |

## Principal Author

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of Principal Author  | Hiroki Kobayashi   |      |         |
| Contribution to the Paper | Conceived and designed the study. Performed histopathological analyses. Performed experiments with <i>Mcam</i> -KO mice. Performed statistical analyses, Interpreted data. Wrote the manuscript. Primary author.   |      |         |
| Overall Percentage (%)    | 60%  |      |         |
| Certification:            | This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. |      |         |
| Signature                 |  | Date | 6/11/20 |

## Co-author contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all the co-author contributions is equal to 100% less the candidate's stated contribution.

Note that multiple authors have given written permission for my primary supervisor (Dr. Daniel Worthley) to sign on their behalf.

|                           |   |      |          |
|---------------------------|---|------|----------|
| Name of co-author         | Krystyna Gieniec  |      |          |
| Contribution to the paper | Discussed manuscript content, performed data interpretation. Performed histopathological analyses. Modified a tumor organoid line. Primary co-author. |      |          |
| Signature                 |   | Date | 30/10/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Tamsin Lannagan  |      |          |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, extracted samples for RNA-sequencing, developed the AOM/DSS colorectal cancer mouse model in the laboratory, performed animal experiments using <i>Lepr-Cre</i> , <i>Grem1-CreERT2</i> and <i>Acta2-RFP</i> (bone marrow transplant) mice, performed histopathological analyses. Primary co-author. |      |          |
| Signature                 |  | Date | 17/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Tongtong Wang  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, extracted samples for RNA-sequencing, analysed RNA-sequencing and microarray datasets, performed statistical analyses, assisted in animal monitoring. |      |         |
| Signature                 |  | Date | 6/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Naoya Asai  |      |         |
| Contribution to the paper | Supervised development of work, performed animal experiments using <i>Islr-CreERT2</i> mice, helped in data interpretation. |      |         |
| Signature                 |   | Date | 9/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Yasuyuki Mizutani                                |      |          |
| Contribution to the paper | Performed smFISH, helped in data interpretation. |      |          |
| Signature                 |  | Date | 17/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Tadashi Iida                                     |      |          |
| Contribution to the paper | Performed smFISH, helped in data interpretation. |      |          |
| Signature                 |  | Date | 18/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Ryota Ando                                       |      |          |
| Contribution to the paper | Performed smFISH, helped in data interpretation. |      |          |
| Signature                 |  | Date | 17/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Akihiro Sakai                                    |      |          |
| Contribution to the paper | Performed smFISH, helped in data interpretation. |      |          |
| Signature                 |  | Date | 19/02/21 |

|                           |  |  |  |
|---------------------------|--|--|--|
| Name of co-author         | Nobumi Suzuki  |  |  |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |  |  |

|           |  |      |         |
|-----------|--|------|---------|
| Signature |  | Date | 9/11/20 |
|-----------|--|------|---------|

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Mari Ichinose  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Josephine Wright   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 6/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Laura Vrbanac  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 6/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Jia Ng   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 6/11/20 |

|                           |  |  |  |
|---------------------------|--|--|--|
| Name of co-author         | Jarrad Goyne   |  |  |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |  |  |

|           |  |      |          |
|-----------|--|------|----------|
| Signature |  | Date | 19/02/21 |
|-----------|--|------|----------|

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Matthew Lawrence   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |          |
|---------------------------|--|------|----------|
| Name of co-author         | Tarik Sammour  |      |          |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |          |
| Signature                 |  | Date | 10/11/20 |

|                           |  |      |            |
|---------------------------|--|------|------------|
| Name of co-author         | Yoku Hayakawa  |      |            |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |            |
| Signature                 |  | Date | 11/18/2020 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Sonja Klebe   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, collected and provided human colorectal cancer clinical data and samples for the MCAM tissue microarray. |      |         |
| Signature                 |   | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Alice Eunjung Shin   |      |         |
| Contribution to the paper | Performed animal experiments using <i>Krt19-Cre</i> mice, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Samuel Asfaha   |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content. |      |         |
| Signature                 |   | Date | 9/11/20 |

|                           |  |      |            |
|---------------------------|--|------|------------|
| Name of co-author         | Mark Bettington  |      |            |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, collected and provided human colorectal cancer samples of different grades. |      |            |
| Signature                 |  | Date | 18.11.2020 |

|                           |  |      |            |
|---------------------------|--|------|------------|
| Name of co-author         | Florian Rieder   |      |            |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |            |
| Signature                 |  | Date | 19.02.2021 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Nicholas Arpaia  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 1/12/20 |

|                   |            |  |  |
|-------------------|------------|--|--|
| Name of co-author | Tal Danino |  |  |
|-------------------|------------|--|--|

|                           |  |      |          |
|---------------------------|--|------|----------|
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |          |
| Signature                 |  | Date | 17/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Lisa Butler   |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content. |      |         |
| Signature                 |   | Date | 6/11/20 |

|                           |   |      |         |
|---------------------------|---|------|---------|
| Name of co-author         | Alastair Burt   |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content. |      |         |
| Signature                 |   | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Simon Leedham  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation, provided <i>Apc</i> <sup>Min/+</sup> samples. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Anil Rustgi  |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 1/12/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Siddhartha Mukherjee   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Masahide Takahashi   |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

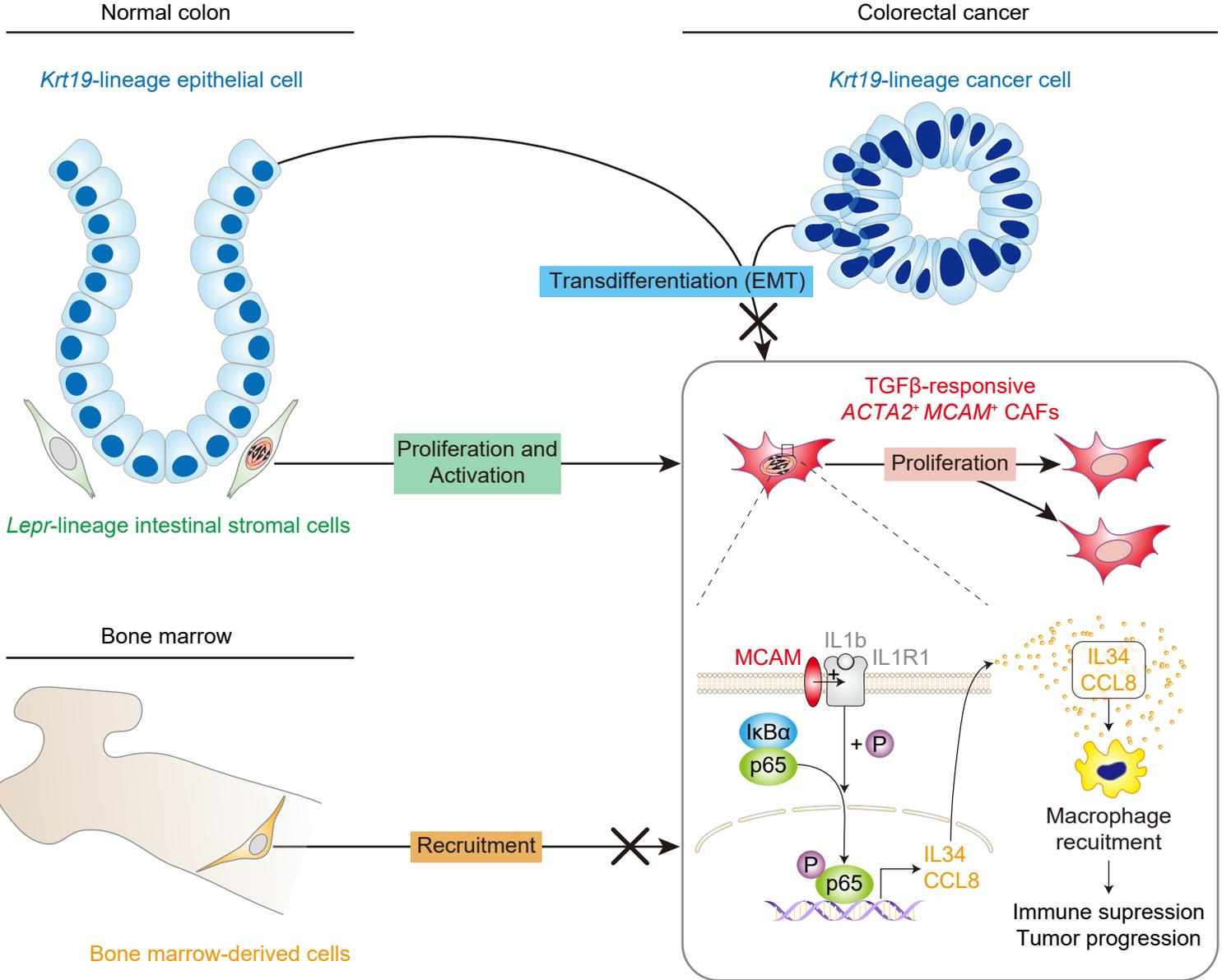
|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Timothy Wang   |      |         |
| Contribution to the paper | Discussed manuscript content, helped in data interpretation. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Atsushi Enomoto  |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation and manuscript evaluation, wrote the manuscript. Corresponding author. |      |         |
| Signature                 |  | Date | 9/11/20 |

|                           |  |  |  |
|---------------------------|--|--|--|
| Name of co-author         | Susan Woods  |  |  |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation and manuscript evaluation, wrote the manuscript. Corresponding author. |  |  |

|           |  |      |         |
|-----------|--|------|---------|
| Signature |  | Date | 6/11/20 |
|-----------|--|------|---------|

|                           |  |      |         |
|---------------------------|--|------|---------|
| Name of co-author         | Daniel Worthley  |      |         |
| Contribution to the paper | Supervised development of work, discussed manuscript content, helped in data interpretation and manuscript evaluation, wrote the manuscript. Corresponding author. |      |         |
| Signature                 |  | Date | 6/11/20 |



**Stromagenesis during CRC progression**



# The origin and contribution of cancer-associated fibroblasts in colorectal carcinogenesis

Short title: Colorectal cancer-associated fibroblasts

Hiroki Kobayashi<sup>1,2,3,4\*</sup>, Krystyna A. Gieniec<sup>1,2\*</sup>, Tamsin RM. Lannagan<sup>1,2\*</sup>, Tongtong Wang<sup>1,2</sup>, Naoya Asai<sup>5</sup>, Yasuyuki Mizutani<sup>3,6</sup>, Tadashi Iida<sup>3,6</sup>, Ryota Ando<sup>3</sup>, Akihiro Sakai<sup>3</sup>, Nobumi Suzuki<sup>1,2,7</sup>, Mari Ichinose<sup>1,2</sup>, Josephine A Wright<sup>2</sup>, Laura Vrbanac<sup>1,2</sup>, Jia Q Ng<sup>1,2</sup>, Jarrad Goynes<sup>1,2</sup>, Matthew J. Lawrence<sup>8</sup>, Tarik Sammour<sup>1,2,8</sup>, Yoku Hayakawa<sup>7</sup>, Sonja Klebe<sup>9</sup>, Alice E. Shin<sup>10</sup>, Samuel Asfaha<sup>11</sup>, Mark L. Bettington<sup>12,13,14</sup>, Florian Rieder<sup>15,16</sup>, Nicholas Arpaia<sup>17,18</sup>, Tal Danino<sup>18, 19</sup>, Lisa M. Butler<sup>1,2</sup>, Alastair D. Burt<sup>1,20</sup>, Simon J. Leedham<sup>21</sup>, Anil K. Rustgi<sup>18</sup>, Siddhartha Mukherjee<sup>22</sup>, Masahide Takahashi<sup>3,4,23</sup>, Timothy C. Wang<sup>22</sup>, Atsushi Enomoto<sup>3,25</sup>, Susan L. Woods<sup>1,2,25,26</sup>, and Daniel L. Worthley<sup>2,24,25</sup>

<sup>1</sup>Adelaide Medical School, University of Adelaide, Adelaide, SA, 5000, Australia

<sup>2</sup>South Australian Health and Medical Research Institute (SAHMRI), Adelaide, SA, 5000, Australia

<sup>3</sup>Department of Pathology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, 466-8550, Japan

<sup>4</sup>Division of Molecular Pathology, Center for Neurological Disease and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Aichi, 466-8550, Japan

<sup>5</sup>Department of Molecular Pathology, Graduate School of Medicine, Fujita Health University, Toyoake, Aichi, 470-1192, Japan

<sup>6</sup>Department of Gastroenterology and Hepatology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, 466-8550, Japan

<sup>7</sup>Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, 113-0033, Japan

<sup>8</sup>Colorectal Unit, Department of Surgery, Royal Adelaide Hospital, Adelaide, SA, 5000, Australia.

<sup>9</sup>Department of Anatomical Pathology, Flinders Medical Centre, Bedford Park, Adelaide, SA, 5001, Australia

<sup>10</sup>Pathology and Laboratory Medicine, Schulich School of Medicine & Dentistry, University of Western Ontario, London, ON, Canada

<sup>11</sup>Department of Medicine, University of Western Ontario, London, ON, Canada

<sup>12</sup>Envoi Specialist Pathologists, Kelvin Grove 4059, Queensland, Australia

<sup>13</sup>Faculty of Medicine, University of Queensland, Herston 4006, Queensland, Australia

<sup>14</sup>QIMR Berghofer Medical Research Institute, Herston 4006, Queensland, Australia

<sup>15</sup>Department of Gastroenterology, Hepatology, and Nutrition, Digestive Diseases and Surgery Institute, Cleveland Clinic Foundation, Cleveland, Ohio, USA

<sup>16</sup>Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, Ohio, USA

<sup>17</sup>Department of Microbiology and Immunology, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA.

<sup>18</sup>Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY, USA.

<sup>19</sup>Department of Biomedical Engineering, Columbia University, New York, NY, USA

<sup>20</sup>Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne NE2 4HH, UK

<sup>21</sup>Intestinal Stem Cell Biology Lab, Wellcome Trust Centre Human Genetics, University of Oxford, Oxford, UK

<sup>22</sup>Department of Medicine and Irving Cancer Research Center, Columbia University, New York, NY, USA

<sup>23</sup>International Center for Cell and Gene Therapy, Fujita Health University, Toyoake, Aichi, 470-1192, Japan

<sup>24</sup>GastroIntestinal Endoscopy, Lutwyche 4030, Queensland, Australia

<sup>25</sup>Co-corresponding authors

<sup>26</sup>Lead contact

\*These authors contributed equally.

## Grant Support

This study was supported by grants from the National Health and Medical Research Council (APP1156391 to D.L.W., S.L.W.) (APP1081852 to D.L.W., APP1140236 to SLW, APP1099283 to DLW,); Cancer Council SA Beat Cancer Project on behalf of its donors and the State Government of South Australia through the Department of Health (MCF0418 to S.L.W., D.L.W., and PRF1117 to L.M.B.); a Grant-in-Aid for Scientific Research (S) (26221304 to M.T.) commissioned by the Ministry of Education, Culture, Sports, Science and Technology of Japan; AMED-CREST (Japan Agency for Medical Research and Development, Core Research for Evolutional Science and Technology;19gm0810007h0104 and 19gm1210008s0101 to A.E.); the Project for Cancer Research and Therapeutic Evolution (P-CREATE) from AMED (19cm0106332h0002 to A.E.); Japan Society for the Promotion of Science Overseas Challenge Program for Young Researchers (to H.K.), Takeda Science Foundation Fellowship (to H.K.), Greaton International Ph.D. Scholarship (to H.K.), Lions Medical Research Foundation Scholarship (to K.G.).

## Abbreviations

ANOVA, analysis of variance; AOM, azoxymethane; CAFs, cancer-associated fibroblasts; CMS, consensus molecular subtype; CRC, colorectal cancer; DSS, dextran sodium sulfate; FACS, fluorescence-activated cell sorting; IF, immunofluorescence; IHC, immunohistochemistry; qRT-PCR, quantitative reverse-transcription polymerase chain reaction; scRNA-seq, single-cell RNA-sequencing; smFISH, single-molecule fluorescent *in situ* hybridization; TGF $\beta$ , transforming growth factor  $\beta$

## Correspondence

D.L.W.; [dan@gastros.com.au](mailto:dan@gastros.com.au), Group Leader, Gut Health, SAHMRI

Address: SAHMRI (5 South), North Terrace, Adelaide SA 5000, AUSTRALIA, 08-8128-4386

S.L.W.; [susan.woods@adelaide.edu.au](mailto:susan.woods@adelaide.edu.au), Group Leader, Gut Cancer, the University of Adelaide,

Address: SAHMRI (5 South), North Terrace, Adelaide SA 5000, AUSTRALIA, 08-8128-4386

A.E.; [enomoto@iar.nagoya-u.ac.jp](mailto:enomoto@iar.nagoya-u.ac.jp), Professor, Department of Pathology, Nagoya University School of Medicine

Address: 65 Tsurumai-Cho, Showa-Ku, Nagoya, 466-8550, JAPAN, 05-2744-2093

### **Disclosures**

F.R. is a consultant to or on the advisory board of Agomab, Allergan, AbbVie, Boehringer-Ingelheim, Celgene/BMS, CDISC, Cowen, Genentech, Gilead, Gossamer, Guidepoint, Helmsley, Index Pharma, Janssen, Koutif, Mestag, Metacrine, Morphic, Origo, Pfizer, Pliant, Prometheus Biosciences, Receptos, RedX, Roche, Samsung, Surrozen, Takeda, Techlab, Theravance, Thetis, and UCB. The remaining authors disclose no conflicts.

**Transcript Profiling:** GSE162508

**Writing Assistance:** N/A

## **Author Contributions**

H.K., K.A.G., T.R.L., S.A., A.E., M.T., T.C.W., S.L.W., and D.L.W. conceived and designed the study. H.K., K.A.G., and T.R.L. performed most of the experiments. H.K. and T.W. (Tongtong Wang) performed statistical analyses. T.W. and H.K. analyzed RNA-seq and microarray datasets. Y.M., T.I., R.A., and A.S. performed smFISH. N.A. and A.E.S. performed animal experiments using *Islr-CreERT2* and *Krt19-Cre* mice, respectively. S.K. and M.L.B. collected and provided human CRC clinical data and samples. H.K., K.A.G., and T.R.L. performed histopathological analyses. H.K. and K.A.G. performed experiments with *Mcam-KO* mice. N.A., S.A., A.D.B., M.T., A.E., S.L.W., and D.L.W. supervised the project. H.K., A.E., S.L.W., and D.L.W. wrote the manuscript. All authors contributed substantially to the discussion of content for the article, reviewed and/or edited the manuscript before submission.

## **Word count:**

Abstract: 260/260 words

Main manuscript: 6998/7000 words

(inclusive of main text; references; figure legends)

**Total number of figures and tables:** 6 figures

# Abstract

## Background and aims

Cancer-associated fibroblasts (CAFs) play an important role in colorectal cancer (CRC) progression and predict poor prognosis in CRC patients. However, the cellular origins of CAFs remain unknown, making it challenging to therapeutically target these cells. Here, we aimed to identify the origins and contribution of colorectal CAFs associated with poor prognosis.

## Methods

To elucidate CAF origins, we used a CRC mouse model in 5 different fate-mapping mouse lines with BrdU dosing. RNA-sequencing of fluorescence-activated cell sorting (FACS)-purified CRC CAFs was performed to identify a potential therapeutic target in CAFs. To examine the prognostic significance of the novel stromal target, CRC patient RNA-sequencing data and tissue microarray were used. CRC organoids were injected into the colon of knockout mice to assess the mechanism by which the stromal gene contributes to colorectal tumorigenesis.

## Results

Our lineage-tracing studies revealed that, in CRC, many ACTA2<sup>+</sup> CAFs emerge through proliferation from intestinal pericryptal Leptin receptor (*Lepr*)<sup>+</sup> cells. These *Lepr*-lineage CAFs, in turn, express melanoma cell adhesion molecule (MCAM), a CRC stroma-specific marker we identified using RNA-sequencing. High MCAM expression induced by TGF- $\beta$  was inversely associated with patient survival in human CRC. In mice, stromal *Mcam* knockout attenuated orthotopically injected colorectal tumoroid growth and improved survival through decreased tumor-associated macrophage recruitment. Mechanistically, fibroblast MCAM interacted with interleukin-1 receptor 1 to augment nuclear factor- $\kappa$ B-IL34/CCL8 signaling that promotes macrophage chemotaxis.

## Conclusion

In colorectal carcinogenesis, pericryptal *Lepr*-lineage cells proliferate to generate MCAM<sup>+</sup> CAFs that shape the tumor-promoting immune microenvironment. Preventing the expansion/differentiation of *Lepr*-lineage CAFs or inhibiting MCAM activity could be effective therapeutic approaches for CRC.

## Keywords:

colorectal cancer; tumor microenvironment; alpha-smooth muscle actin ( $\alpha$ SMA); CD146

## Introduction

Colorectal cancer (CRC) is a leading cause of cancer-related death. Cancer-associated fibroblasts (CAFs) are histologically prominent and biologically important in CRC initiation, progression, and metastasis<sup>3</sup>. CAFs contribute to carcinogenesis via secretion of growth factors, cytokines, pro-angiogenic factors, and extracellular matrix<sup>3</sup>. Recent studies using immunophenotyping and single-cell RNA-sequencing (scRNA-seq) have revealed that CAFs contain heterogeneous subpopulations<sup>226,265,266</sup>. It is now apparent that distinct CAF populations have different consequences on cancer growth. Some CAFs promote while others retard cancer growth<sup>6</sup>. The cellular origins of CAFs, whether promoting or retarding, are poorly understood<sup>3</sup>. With respect to the development and consequences of CAFs on CRC growth, there remain at least three unresolved questions. Firstly, are CAFs newly generated cells arising through proliferation, or simply old cells acquiring a new phenotype? Secondly, if any of the CAFs emerge through proliferation, what is their cellular origin? Thirdly, what CAF-derived factors promote cancer progression, and could those be targeted with novel stromal therapies?

Theoretically, CAFs could arise through at least four non-mutually exclusive mechanisms; **proliferation (A)**, **activation (B)**, **transdifferentiation (C)**, and **recruitment (D)**<sup>3</sup>. Although studies using autochthonous mouse models of cancers have indicated that some CAFs undergo proliferation (**A; Proliferation**)<sup>4,54,267</sup>, the relative contribution of proliferating and non-proliferating CAFs to the entire pool remains unclear. Induced by factors such as transforming growth factor  $\beta$  (TGF- $\beta$ )<sup>266,268</sup>, quiescent fibroblasts might undergo phenotypic conversion into activated CAFs, an old cell, but with a new mask (**B; Activation**). Thirdly, several fate-mapping studies have indicated that non-fibroblast lineage cells, such as epithelial cells, could transdifferentiate into CAFs through epithelial-to-mesenchymal transition (EMT) (**C; Transdifferentiation**)<sup>58,59</sup>. Lastly, bone marrow transplantation experiments have indicated that about 20% of ACTA2 (Protein name, alpha-smooth muscle actin;  $\alpha$ SMA)<sup>+</sup> CAFs were recruited from the bone marrow in a mouse model of gastric cancer (**D; Recruitment**)<sup>18</sup>. Human studies have also suggested that bone marrow contribution can be detected in CAFs in several neoplasias including CRC<sup>86</sup>. Others, however, have suggested that

local precursors were a predominant contributor to ACTA2<sup>+</sup> CAFs<sup>84</sup>. Thus, the origin of CAFs remains uncertain. In contrast to fibrosis in organs such as the liver, kidney, and skin, in which the origins of myofibroblasts have been extensively investigated<sup>190</sup>, to our knowledge, no previous CAF studies have comprehensively performed lineage-tracing experiments to track the aforementioned four possible CAF sources.

Leptin receptor (*Lepr*) is a well-established marker for perivascular mesenchymal cells, which support bone marrow hematopoietic stem cell (HSC) maintenance<sup>208</sup>. Previous fate-mapping studies have demonstrated that *Lepr*-expressing cells give rise to bone and adipocytes formed in the adult normal bone marrow<sup>269</sup> as well as myofibroblasts in primary myelofibrosis<sup>270</sup>. However, the significance of *Lepr*-lineage cells in the development of CAFs is unknown.

Similar to *Lepr*, MCAM (melanoma cell adhesion molecule, also known as CD146 or MUC18) is highly expressed by perivascular stromal cells in the bone marrow and suggested to be important in the HSC niche<sup>271</sup>. MCAM is also expressed by endothelial cells, melanoma cells, pericytes, and CAFs<sup>272,273</sup>. MCAM expressed in endothelial and melanoma cells contributes to cancer progression by promoting cancer cell growth, angiogenesis, and metastasis<sup>273-275</sup>. Recently, scRNA-seq analyses have revealed that *MCAM* defines a subset of pericyte-like CAFs that secrete tumor-promoting immunomodulatory cytokines in human cholangiocarcinoma and breast cancer<sup>276,277</sup>. The biological role of MCAM<sup>+</sup> CAFs, however, has been poorly defined in human CRC or organoid models of CRC.

This study, for the first time, comprehensively addresses the cellular origins, dynamics, and consequences of specific CAFs in CRC. Using lineage tracing, we identify intestinal pericryptal *Lepr*-lineage cells as a major source of proliferating CAFs in a mouse model of CRC. Next, by combining fluorescence-activated cell sorting (FACS), RNA-sequencing (RNA-seq), and immunohistochemistry, we show these CAFs express MCAM. We investigate the clinical significance of MCAM expression using RNA-seq data and tissue microarray from human CRC samples. Finally, we uncover the mechanism of stromal MCAM action in CRC using newly generated *Mcam*-null mice and mouse colonoscopy.

## Materials and Methods

### Statistical analysis

Comparison of 2 groups was performed using two-tailed unpaired t-tests or Mann–Whitney U tests. For multiple comparisons, we used analysis of variance (ANOVA) or Kruskal-Wallis test. For survival analyses, Kaplan-Meier survival estimation with a Log-rank (Mantel-cox) test was performed. Statistical analyses were conducted using GraphPad Prism 8.00 (GraphPad) or SPSS Statistics ver. 25 (IBM). P-values of less than 0.05 were considered statistically significant.

For all other Materials and Methods, see **Supplementary Materials**.

## Results

### *Desmoplasia is increased during colorectal carcinogenesis in human and mice.*

To explore whether desmoplasia is increased during colorectal carcinogenesis and identify a suitable mouse model to investigate this, we performed immunohistochemistry for ACTA2, a well-established marker for CAFs, in human normal colorectal mucosa, adenoma, and adenocarcinoma samples. The ratio of ACTA2<sup>+</sup> fibroblasts in the total stromal cells increased from normal to low-grade adenoma to high-grade adenoma, and ultimately adenocarcinoma (**Figure 1A and B**). The elevated ACTA2 expression level during colorectal carcinogenesis was corroborated by an analysis of expression microarray data from human colorectal tissues (**Supplementary Figure 1A**). Analyses of scRNA-seq data from human CRC tissues<sup>226</sup> also demonstrated that ACTA2 expression is increased in CAFs compared with normal fibroblasts, with the highest ACTA2 transcripts observed in pericytes among various CAF subpopulations (**Figure 1C; Supplementary Figure 1B and C**).

Next, we investigated the prognostic significance of ACTA2 expression in The Cancer Genome Atlas (TCGA) data. High ACTA2 expression was inversely associated with overall survival in patients with CRC (**Figure 1D**). High ACTA2 expression, as well as high expression of *FAP*, an activated fibroblast marker<sup>3</sup>, was consistently associated with poor prognosis across multiple

expression datasets from CRC patients (**Supplementary Figure 2**). The highest *ACTA2* expression was observed in the poor-prognosis stroma-rich molecular subtype of CRC (Consensus Molecular Subtype 4; CMS4)<sup>31</sup> (**Figure 1E**).

We then sought to explore whether *ACTA2*<sup>+</sup> fibroblasts are similarly increased in mouse models of CRC. To this end, we performed *ACTA2* immunohistochemistry using tumors from the azoxymethane (AOM)/dextran sulfate sodium (DSS) (**Figure 1F and G**) and *Apc*<sup>Min/+</sup> mouse models. In line with a previous study<sup>165</sup>, *ACTA2* expression was significantly elevated in the stroma of AOM/DSS tumors compared with the adjacent normal mucosa (**Figure 1H and I**). Similarly, small intestinal tumors from *Apc*<sup>Min/+</sup> mice showed an increase in stromal *ACTA2* expression in comparison with the adjacent normal tissue, but to a lesser extent than the AOM/DSS mouse model (**Supplementary Figure 3A and B**). Taken together, these data suggest that *ACTA2*<sup>+</sup> CAF number increases throughout colorectal carcinogenesis in humans, and this is recapitulated in the AOM/DSS mouse model of CRC.

#### ***A subpopulation of CRC CAFs arises through proliferation in human and mice.***

Having confirmed the importance of CAFs across multiple human CRC sample sets and identified a mouse CRC model in which to study CAFs, we addressed the question of whether CAFs emerge through cell division or simply increase *ACTA2* expression in existing cells. Co-staining for *ACTA2* and Ki67 using human colorectal samples revealed that the percentage of *ACTA2* and Ki67 double-positive cells (i.e., proliferating *ACTA2*<sup>+</sup> CAFs) was increased in high-grade adenoma and adenocarcinoma compared to normal colorectal mucosa, with about 10% of *ACTA2*<sup>+</sup> CAFs marked by Ki67 in adenocarcinoma (**Figure 2A and B**). Analysis of scRNA-seq data from human CRC and normal mucosa<sup>226</sup> confirmed that a subcluster of *ACTA2*<sup>+</sup> CAFs expressed *MKI67*, and this co-expressing population was not found in fibroblasts from the normal mucosa (**Supplementary Figure 4A-C**). These data suggest that human CRC CAFs undergo mitosis during malignant progression.

Ki67 only temporarily marks actively cycling cells, so our analysis of proliferation of human CRC CAFs may underestimate CAFs that divided at an earlier time-point. To capture the entire population of CAFs that underwent proliferation throughout carcinogenesis, we took advantage of continuous 5-bromodeoxyuridine (BrdU) labeling in the AOM/DSS mouse model, with BrdU dosing beginning at the onset of observable tumors (**Figure 2C**). The ratio of ACTA2<sup>+</sup> CAFs that incorporated BrdU was significantly elevated in AOM/DSS tumors, compared to the adjacent non-neoplastic colon, with approximately 45% of ACTA2<sup>+</sup> CAFs marked by BrdU, proving that they have divided since the onset of tumorigenesis (**Figure 2D and E**). Contrasting with the BrdU incorporation ratio, the ratio of actively proliferating CAFs (Ki67<sup>+</sup>ACTA2<sup>+</sup> CAFs) in the total pool of ACTA2<sup>+</sup> CAFs was only about 3% and was not significantly different from the ratio of proliferating fibroblasts in the normal mouse colorectal mucosa (**Figure 2D and E**). These data indicate that, in AOM/DSS tumors, the majority of ACTA2<sup>+</sup> CAFs at humane endpoint were in quiescent G0 phase as evaluated by Ki67 negativity, but almost half of the CAFs had undergone cell division and incorporated BrdU at some point during tumorigenesis. Overall, our analyses of mouse CRC suggest that almost half of the ACTA2<sup>+</sup> CAFs emerge through proliferation and that proliferating CAFs are also a feature of human colorectal carcinogenesis.

### ***Lepr-lineage stromal cells are a major contributor to the proliferating CAF population in AOM/DSS CRC.***

Having observed that almost half of the CAFs in CRC developed through cell division, we next sought to establish the cellular origin of these new CAFs by using a lineage-tracing strategy. We selected transgenic mouse lines that **(1)** identified putative colorectal mesenchymal stem-progenitor cells (*Lepr*-Cre; Rosa26-LSL-tdtomato<sup>208</sup>, *Grem1*-CreERT2; Rosa26-LSL-ZsGreen<sup>46</sup> and *Islr*-CreERT2; Rosa26-LSL-tdtomato<sup>6</sup> or **(2)** labeled epithelium (*Krt19*-Cre; Rosa26-*mt/mG*<sup>278</sup>) or **(3)** marked bone marrow-derived cells through a combination of bone marrow from *Acta2*-RFP mouse transplanted into non-RFP recipients (**Figure 3A-E**). These experiments were coupled with BrdU labeling (**Figure 2C**), with tamoxifen administered to the inducible Cre lines at postnatal day 6.

Immunofluorescence for EPCAM, a pan-epithelial cell marker, showed that all *Lepr*-, *Grem1*-, and *Islr*-lineage cells were observed only within the EPCAM<sup>-</sup> stroma, validating their mesenchymal identity (**Figure 3A**). Approximately half of ACTA2<sup>+</sup> CAFs and 75% of proliferating BrdU<sup>+</sup>ACTA2<sup>+</sup> CAFs were *Lepr*-lineage-positive in AOM/DSS tumors, with a smaller proportion of ACTA2<sup>+</sup> CAFs derived from the *Grem1*-lineage and *Islr*-lineage (**Figure 3A and C-E**). *Lepr*-lineage cells also represented a predominant contributor to PDGFRA<sup>+</sup> CAFs in the AOM/DSS model (**Supplementary Figure 5A and B**). Together, these results suggest that *Lepr*-lineage stromal cells are a major source of proliferating ACTA2<sup>+</sup> CAFs.

### ***Neither epithelium nor bone marrow recruitment contributed to ACTA2<sup>+</sup> CAFs in the AOM/DSS mouse model of CRC.***

The epithelium has been shown to generate CAFs in mouse models of pancreatic and breast cancers<sup>58,59</sup>. Therefore, we explored whether colonic epithelial cells could similarly undergo epithelial-mesenchymal transition into colorectal CAFs. For this purpose, we utilized constitutive *Krt19*-Cre; Rosa26-*mt/mG* mice to track the fate of *Krt19*-lineage colonic epithelial cells<sup>278</sup>. All colonic cells with epithelial morphology were marked following reporter recombination by Cre recombinase driven by the *Krt19* promoter<sup>278</sup>. However, no *Krt19*-lineage cells were positive for ACTA2 in either normal colon or AOM/DSS tumors (**Figure 3B and D**). This suggests that, at least in this mouse model of CRC, the epithelium is not a source of ACTA2<sup>+</sup> CAFs.

Next, to assess the contribution of bone marrow-derived cells to the AOM/DSS tumor stroma, we performed bone marrow transplantation experiments using an *Acta2*-RFP (red fluorescent protein) reporter mouse as a donor. Initially, we validated that, in *Acta2*-RFP mice that did not undergo bone marrow transplantation, RFP was expressed by fibroblastic cells in AOM/DSS tumors, confirming that the *Acta2* promoter is active in this CRC mouse model (**Supplementary Figure 6A and B**). To perform bone marrow transplantation from *Acta2*-RFP mice, wild-type recipient mice were subjected to total body irradiation and transplanted with whole bone marrow cells from *Acta2*-RFP donor mice. Then, the mice were treated with AOM/DSS to induce colorectal tumors (**Supplementary Figure 7A**). Quantitative polymerase chain reaction (qPCR) for RFP using genomic DNA isolated from the

bone marrow of the recipient mice confirmed engraftment of RFP<sup>+</sup> cells in the recipient bone marrow (**Supplementary Figure 7B**). Transplanted *Acta2*-RFP<sup>+</sup> cells were also observed in the small intestine of the wild-type recipients, further validating the engraftment (**Supplementary Figure 7C**). However, no bone marrow-transplanted RFP<sup>+</sup> cells were observed in AOM/DSS tumors in wild-type recipient mice (**Figure 3B and D**). This indicates that, at least in this experimental CRC model, CAFs did not arise via recruitment from the bone marrow, but only from local precursors.

Collectively, our data with five distinct genetically engineered mouse models suggest that tissue-resident *Lepr*-lineage stromal cells are a key contributor to the ACTA2<sup>+</sup> CAFs in the AOM/DSS mouse model of CRC.

***Lepr*-lineage intestinal stromal cells undergo proliferation and differentiation into ACTA2<sup>+</sup> CAFs during AOM/DSS carcinogenesis.**

Having identified intestinal *Lepr*-lineage stromal cells as a predominant contributor to the CRC mesenchyme, we next sought to characterize *Lepr*-lineage cells in the normal colon and AOM/DSS tumors. In the normal colonic mucosa, pericryptal *Lepr*-lineage cells were preferentially located near the base of the crypts (**Figure 3F and G**). *Lepr*-lineage stromal cells in AOM/DSS tumors exhibited higher ACTA2 positivity than *Lepr*-lineage stromal cells in the normal mucosa, indicating that *Lepr*-lineage cells underwent phenotypic conversion into ACTA2<sup>+</sup> CAFs during carcinogenesis (**Supplementary Figure 8A and B**). BrdU labeling in AOM/DSS-treated mice revealed that *Lepr*-lineage cells showed higher proliferation in AOM/DSS tumors, in comparison with the adjacent normal mucosa (**Supplementary Figure 8C and D**). Single-molecule fluorescent RNA *in situ* hybridization (smFISH) for *Lepr* revealed that active expression of *Lepr* in *Lepr*-lineage cells was reduced in the AOM/DSS tumor compared with the normal colon (**Supplementary Figure 8E and F**). Together, these findings indicate that intestinal *Lepr*-lineage stromal cells undergo expansion and CAF differentiation at the expense of *Lepr* expression during AOM/DSS colorectal carcinogenesis.

### **Identification of MCAM as a CRC stroma-specific marker that defines a subset of *Lepr*-lineage proliferating CAFs.**

Lower *Lepr* expression in the CRC mesenchyme could potentially make it challenging to therapeutically target *Lepr*-lineage CAFs based on active *Lepr* expression in established cancers. Therefore, we next aimed to identify a stromal factor that is actively expressed in the CRC mesenchyme as a potential therapeutic stromal target to treat CRC.

As a strategy to identify the most biologically relevant stromal targets, we were inspired by the parallels between cancer and developmental biology<sup>279,280</sup>. For example, factors involved in CAF activation, such as TGF- $\beta$ , also play a crucial role in the early development of gastrointestinal organs<sup>3,280</sup>. Therefore, we decided to triangulate the fibroblastic factors that were significantly upregulated in both tumorigenesis and development compared to adult colonic fibroblasts.

We first sorted fibroblasts using a negative selection strategy. Fibroblasts were selected based on their lack of expression of blood cell markers (CD45 and Ter119), an endothelial marker (CD31) and an epithelial marker (EPCAM), from AOM/DSS tumors and developmental colon at postnatal day 14, as well as the normal adult colon (**Figure 4A**). Fibroblast markers such as *Grem1*, *Acta2*, and *Fap* were highly expressed in the FACS-sorted mesenchymal cells (CD45<sup>-</sup>, Ter119<sup>-</sup>, CD31<sup>-</sup>, EPCAM<sup>-</sup>), validating their enrichment for fibroblasts (**Supplementary Figure 9**). RNA-sequencing from the FACS-purified fibroblasts revealed that 342 genes were differentially upregulated in both the AOM/DSS tumors and the early postnatal colons when compared with the normal adult colon fibroblasts (**Figure 4B; Step 1**). Next, we analyzed the prognostic significance of these 342 genes by performing survival analysis using TCGA data, resulting in the selection of 46 genes that were associated with human CRC survival (**Figure 4B; Step 2, Supplementary Table 1**). Next, to focus on stroma-specific targets, rather than genes expressed in multiple cell types, using our RNA-seq data from normal adult colon and AOM/DSS tumors, we selected 18 stroma-specific genes that were upregulated in fibroblasts compared to epithelial cells (**Figure 4B; Step 3, Supplementary Table 1**). Then, to examine for genes expressed at the protein level in human CRC stroma, we interrogated human CRC immunohistochemistry data in the Human Protein Atlas database and selected 6 proteins that were highly expressed in the CRC stroma (**Figure 4B; Step 4, and**

**Supplementary Figure 10).** Finally, our immunohistochemistry data for candidate genes showed that MCAM was the only candidate that was consistently upregulated in the stroma of AOM/DSS tumors and the developmental colon, compared to the normal adult colon (**Figure 4C and D; Supplementary Figure 11**). This systematic approach allowed us to identify MCAM as a colonic mesenchyme-specific gene that is upregulated in both tumorigenesis and development, and inversely associated with human CRC survival.

Next, we explored the stromal MCAM expression in human and mouse colorectal tissues. Analyses of scRNA-seq from human CRC tissues<sup>226</sup> and ulcerative colitis samples revealed that the high *MCAM* expression was observed in pericytes compared with other cell subpopulations such as endothelial cells, epithelial cells, and immune cells (**Supplementary Figure 12A and B**). In AOM/DSS tumors, co-immunofluorescence for CD31, ACTA2, CD45, and EPCAM showed that approximately 45% of MCAM<sup>+</sup> cells expressed a pericyte/CAF marker, ACTA2 (**Supplementary Figure 12C**).

To characterize the cellular sources of MCAM<sup>+</sup> CAFs in CRC, we performed immunofluorescence for MCAM in the three mesenchymal fate-mapping mouse models (*Lepr*-Cre, *Grem1*-CreERT2, and *Islr*-CreERT2 mice). Our data revealed that about 80% of MCAM<sup>+</sup>ACTA2<sup>+</sup> CAFs were derived from the *Lepr*-lineage in AOM/DSS tumors (**Figure 4E; Supplementary Figure 13A-D**). We also co-stained MCAM and BrdU in AOM/DSS-treated mice that were administered BrdU during carcinogenesis. In keeping with previous scRNA-seq data showing that *Mcam* was highly expressed by a proliferative subpopulation of CAFs<sup>267</sup>, more than half of the MCAM<sup>+</sup> cells were positive for BrdU, indicating that the majority of MCAM<sup>+</sup> cells arose through proliferation (**Figure 4F and G**). Collectively, these data indicate that MCAM identifies *Lepr*-lineage proliferating CAFs in mouse colorectal tumors.

***Increased MCAM expression is associated with Consensus Molecular Subtype (CMS) 4 and predicts poor survival in patients with CRC.***

We investigated the clinical significance of MCAM expression in CRC patients. Consistent with the observed upregulation of MCAM during mouse colorectal tumorigenesis, MCAM expression was

increased in the human adenoma-carcinoma sequence (**Figure 5A and B**). Analyses of expression microarray data from human colorectal tissues also showed that *MCAM* transcripts were elevated during colorectal carcinogenesis (**Supplementary Figure 14A and B**). Furthermore, scRNA-seq data from human colorectal tissues<sup>226</sup> demonstrated that, among fibroblast subpopulations, *MCAM* expression was increased in pericytes during carcinogenesis (**Supplementary Figure 14C and D**).

Analyses of the TCGA dataset showed that the highest expression of *MCAM* was observed in poor-prognosis immunosuppressive CMS4 tumors (**Figure 5C**). Given that TGF- $\beta$  signaling activation is a defining characteristic of CMS4 CRC<sup>31</sup>, we reasoned that TGF- $\beta$  might upregulate *MCAM* expression. In keeping with our hypothesis, stimulation of a mouse colonic fibroblast cell line, YH2 cells, with recombinant TGF- $\beta$ 1 enhanced *Mcam* transcript levels as well as a TGF- $\beta$  target gene, *Acta2* (**Figure 5D**). This was rescued by co-treatment with Galunisertib, a specific inhibitor for TGF- $\beta$  receptor 1. In keeping with this, scRNA-seq data<sup>226</sup> showed positive correlations between *MCAM* and *ACTA2* expression in human colorectal CAFs (**Figure 5E and Supplementary Figure 15A**). Analysis of TCGA and expression microarray data also demonstrated that *MCAM* transcripts were positively correlated with expression levels of TGF- $\beta$  target genes such as *ACTA2* and *SERPINE1* (**Supplementary Figure 15B**).

Next, to confirm the clinical association between *MCAM* expression and survival, we performed *MCAM* immunohistochemistry using tissue microarrays from our own independent cohort of 101 CRC patients. Consistent with a previous paper<sup>281</sup>, high *MCAM* expression was an independent prognostic factor for poor overall survival in CRC patients (**Figure 5F and G; Supplementary Tables 2 and 3**). Moreover, analyses of four independent CRC datasets confirmed that high *MCAM* expression was inversely associated with survival (**Supplementary Figure 16**). Taken together, these data indicate that high *MCAM* expression driven, at least in part, by TGF- $\beta$ , predicts poor prognosis in human CRC.

**Genetic deletion of stromal *Mcam* inhibits colorectal tumorigenicity and improves survival via decreased Nuclear Factor- $\kappa$ B-IL34/CCL8-mediated macrophage recruitment.**

Finally, to delineate the mechanism by which MCAM contributes to CRC progression, we generated *Mcam*-knockout mice using CRISPR/Cas9-mediated genome engineering and colonoscopically injected luciferase-expressing *Apc* $\Delta/\Delta$ , *Kras*<sup>G12D/ $\Delta$</sup> , *Trp53* $\Delta/\Delta$  mouse CRC organoids (hereafter termed AKP tumoroids) into the colon of *Mcam*-knockout and wild-type mice (**Figure 6A; Supplementary Figures 17A-C and 18A-D**). In this mouse model, more than half of MCAM<sup>+</sup> cells were ACTA2<sup>+</sup> CAFs (**Supplementary Figure 17B and C**). Consistent with our earlier *MCAM* expression and survival analyses from human CRC, *Mcam*-knockout mice showed prolonged survival after tumoroid injection (**Figure 6B**). *Mcam*-knockout mice also demonstrated reduced tumoroid-derived luciferase signals by in vivo imaging system (IVIS), decreased tumor volumes and colonoscopic tumor scores (**Figure 6C-F; Supplementary Figure 19A and B**).

Immunohistochemistry for various immune cell markers revealed that infiltration of CD68<sup>+</sup> macrophages and CD11b<sup>+</sup> myeloid-derived cells was decreased in tumors from *Mcam*-knockout mice (**Figure 6G and H**). This was accompanied with decreased FOXP3<sup>+</sup> regulatory T cells and increased CD8<sup>+</sup> cytotoxic T cells in *Mcam*-knockout mice (**Supplementary Figure 20A and B**). In our mouse model, we did not observe alterations in vasculature density by *Mcam* knockout as evaluated by CD31 immunohistochemistry (**Supplementary Figure 20C and D**).

Consistent with our mouse immunophenotyping data, gene set enrichment analysis (GSEA) using TCGA data revealed positive enrichment of macrophage/monocyte chemotaxis genes in *MCAM*<sup>high</sup> cancers compared with *MCAM*<sup>low</sup> tumors (**Supplementary Figure 21**). We hypothesized that MCAM<sup>+</sup> CAFs might promote tumor-associated macrophage (TAM) recruitment, contributing to the immunosuppressive tumor microenvironment. To identify macrophage/monocyte chemoattractants secreted by MCAM<sup>+</sup> CAFs, we first performed differential gene expression analysis using scRNA-seq data from human CRC<sup>226</sup> and found that 462 genes were upregulated in *MCAM*<sup>high</sup> CAFs compared to *MCAM*<sup>low</sup> CAFs (**Figure 6I**). Next, using gene ontologies, we examined transcripts encoding cytokines and chemokines involved in macrophage/monocyte chemotaxis. This analysis

identified *IL34* and *CCL8* (also known as monocyte chemoattractant protein-2) as genes with roles in TAM recruitment that are upregulated in *MCAM*<sup>high</sup> CAFs.

Next, to assess whether MCAM could promote IL34 and CCL8 expression, we overexpressed MCAM in YH2 cells by lentiviral transduction and stimulated MCAM-YH2 cells with recombinant interleukin (IL)-1 $\beta$ , which is known to induce IL34 and CCL8 expression in fibroblasts<sup>282,283</sup>. As expected, IL-1 $\beta$ -treated MCAM-YH2 cells showed decreased I $\kappa$ B $\alpha$  expression, increased phosphorylation of nuclear factor- $\kappa$ B (NF- $\kappa$ B; p65), and enhanced luciferase signals from NF- $\kappa$ B-responsive elements, leading to upregulation of *Il34* and *Ccl8* (**Figure 6J-L; Supplementary Figure 22**). These alterations were rescued by co-treatment with IKK16, a selective inhibitor for I $\kappa$ B kinase. We reasoned that MCAM might act as a co-receptor for IL-1 $\beta$  receptor, IL1R1, to potentiate IL-1 $\beta$ -NF- $\kappa$ B-IL34/CCL8 signaling. To this end, we lentivirally transduced YH2 cells with MCAM-hemagglutinin (HA) epitope tag or, as a control, mScarlet-HA, and performed immunoprecipitation with an anti-HA antibody. The co-immunoprecipitation revealed that MCAM interacted with IL1R1 (**Figure 6M**). Reciprocal co-immunoprecipitation of MYC-tagged IL1R1 using an anti-MYC antibody verified the interaction of IL1R1 with MCAM (**Supplementary Figure 23**). In line with our *in vitro* data, tumors from *Mcam*-knockout mice showed decreased NF- $\kappa$ B activation as assessed by phospho-p65 immunohistochemistry (**Figure 6N and O**). In human CRC, TCGA and expression microarray data confirmed that *MCAM* expression was positively correlated with *IL34* and *CCL8*, as well as *CD68* and *ITGAM* (CD11b) expression (**Supplementary Figure 24**). Collectively, our data indicate that MCAM alters the immune microenvironment and accelerates CRC progression, in part, through increased TAM recruitment mediated by IL1R1-NF- $\kappa$ B-IL34/CCL8 signaling.

## Discussion

In this study, we have shown that about 50% of ACTA2<sup>+</sup> CAFs in CRC were generated through proliferation, with the remaining 50% acquired through new or preserved ACTA2 expression in existing fibroblasts (i.e., activation). 75% of the proliferating ACTA2<sup>+</sup> CAFs were derived from intestinal *Lepr*-lineage stromal cells. These *Lepr*<sup>+</sup> pericryptal fibroblasts are also the chief origin of proliferating MCAM<sup>+</sup> CAFs. High stromal MCAM expression is associated with poor clinical outcomes in patients with CRC. Furthermore, transgenic knockout of *Mcam* in the colorectal tumor microenvironment limits tumor growth and improves survival by modifying TAM recruitment and immune landscapes. These data suggest that MCAM, a prominent cell surface protein, could prove to be a valuable novel stromal target in the prevention and treatment of CRC.

Several previous studies have indicated that recruitment from the bone marrow could contribute to CAFs in mouse models of cancers such as gastric and breast cancer<sup>18,284</sup>. In contrast, one paper demonstrated that no *Acta2*-RFP<sup>+</sup> CAFs were detected in intestinal tumors developed in a parabiosis study of an *Apc*<sup>Min/+</sup> with an *Acta2*-RFP mouse<sup>84</sup>. In agreement with this, we found that no ACTA2<sup>+</sup> CAFs were derived from the bone marrow in an AOM/DSS model of CRC. Similarly, we did not find any evidence that epithelial cells contributed to ACTA2<sup>+</sup> CAFs, at least, in our AOM/DSS CRC model, despite several papers indicating EMT as a potential source in pancreatic and breast tumor models<sup>58,59</sup>. It is plausible, and indeed likely, that the origins and contributions of CAFs are context-dependent, depending on cancer stage, site, and cancer genetics. It would be intriguing, in future studies, to investigate whether different precursor cells give rise to functionally distinct CAFs in primary versus metastatic cancer models.

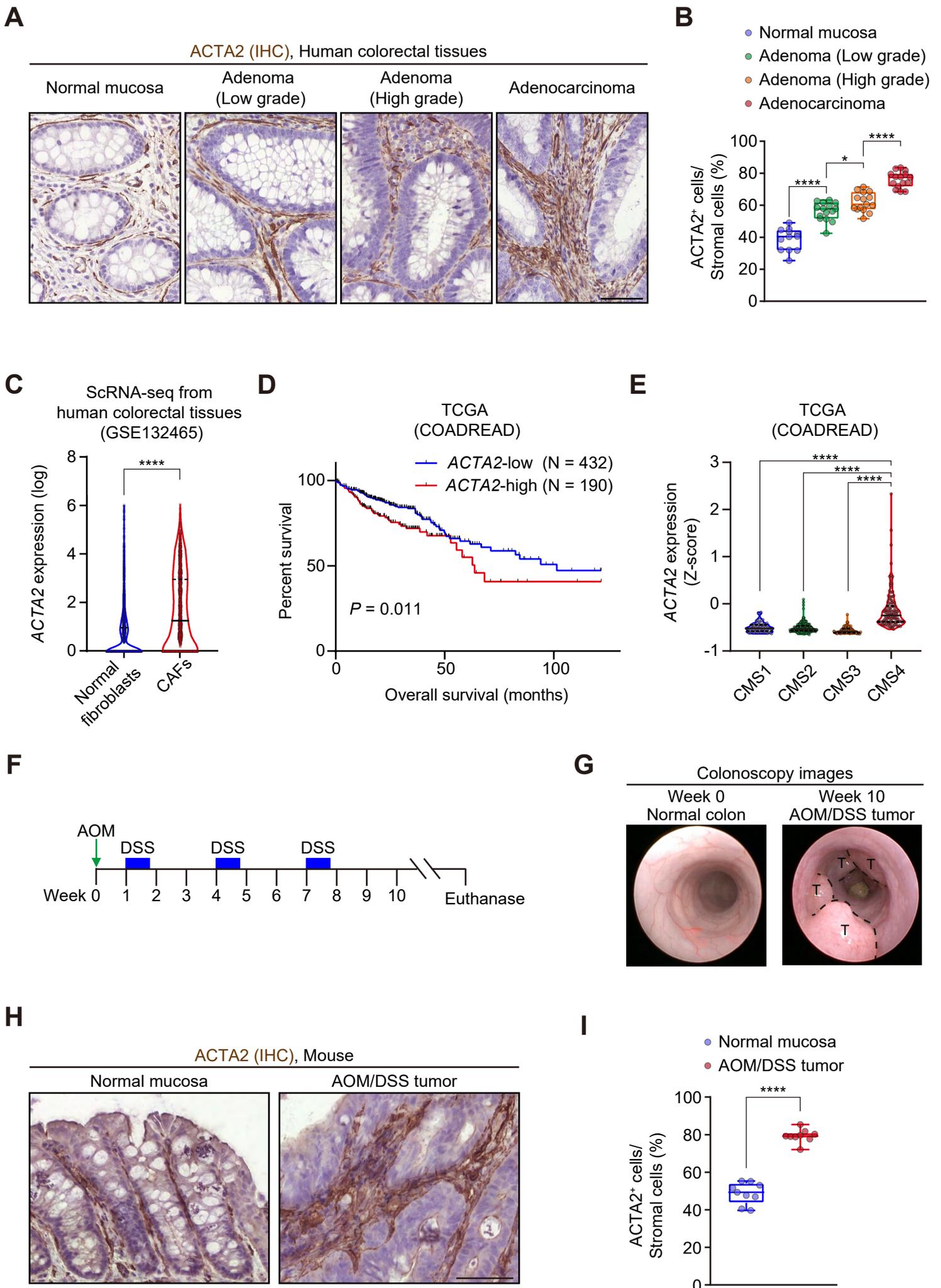
Intestinal normal and neoplastic epithelium develop from stem-progenitor cell hierarchies<sup>285</sup>. Analogous to this, we have previously shown that *Grem1*<sup>+</sup> intestinal reticular stromal cells identify connective tissue stem cells in the normal small intestine<sup>46</sup>. Here, our data indicate that the majority of CRC CAFs, however, arise not from *Grem1*<sup>+</sup> cells, but from intestinal *Lepr*-lineage pericryptal cells. Interestingly, a recent paper found that *Gli1*<sup>+</sup> pancreatic stellate cells could contribute to approximately half of ACTA2<sup>+</sup> CAFs in a mouse model of pancreatic cancer<sup>286</sup>. Further research is

required to determine the hierarchical or overlapping relationship between *Lepr*-lineage and *Gli1*-lineage CAFs in different tissues in health and neoplasia.

One limitation of the present study is that we have not been able to ascertain whether *Lepr*-lineage CAFs display cellular plasticity during cancer development as has been shown to occur in cancer stem cells<sup>285</sup> or whether they undergo an irreversible “lineage-restricted” differentiation. Given that CAFs are considered to exhibit tumor stage-dependent phenotypes<sup>3,287</sup>, it is conceivable that *Lepr*-lineage CAFs could adapt to dynamic phenotypic shifts during colorectal carcinogenesis and co-evolve with epithelial genetic events.

This work also demonstrated that MCAM is an attractive therapeutic target that modifies the immunosuppressive milieu through augmenting NF-κB signaling, key signaling that defines inflammatory phenotypes in CAFs<sup>64,70,268</sup>. Excitingly, MCAM-neutralizing antibodies show promising results in restraining cancer progression in preclinical models, including a model of CRC<sup>274,275</sup>. Future research should focus on investigating whether co-treatment of the MCAM-neutralizing antibody and an immune checkpoint inhibitor could unleash a cytotoxic immune response against immunologically “cold” cancers that are resistant to immunotherapies.

In conclusion, our data show that *Lepr*-lineage intestinal stromal cells, resident at the pericryptal base in the normal colon, proliferate in colorectal carcinogenesis to generate MCAM<sup>+</sup> CAFs. We also show that MCAM is an important factor in sculpting the detrimental immune microenvironment responsible for driving colorectal carcinogenesis and the associated poor patient outcome. In the future, approaches to reduce the expansion of *Lepr*<sup>+</sup> pericryptal cells, prevent their differentiation into MCAM<sup>+</sup> CAFs, and inhibit the activity of MCAM-mediated NF-κB signaling axis in mature CAFs, may all have considerable clinical value in the treatment of colorectal cancer.



Kobayashi *et al.*, Figure 1.

**Figure 1: ACTA2 expression is increased during colorectal carcinogenesis in humans and mice.**

**(A, B)** Immunohistochemistry (IHC) for ACTA2 in human colorectal samples. **(A)** Representative pictures. **(B)** Quantification of ACTA2 positivity in total stromal cells (visualized by hematoxylin counterstaining). 3 high power fields (HPFs, 400x)/patient, 4-5 patients each.

**(C)** Violin plots depict ACTA2 transcripts in normal fibroblasts (n = 2053 cells) and CRC CAFs (n = 1854 cells) assessed by single-cell RNA-sequencing (scRNA-seq) from human colorectal tissues.

**(D)** Kaplan-Meier survival curves in The Cancer Genome Atlas (TCGA) dataset.

**(E)** Violin plots showing ACTA2 expression level in four consensus molecular subtypes (CMS). n = 76 (CMS1), 220 (CMS2), 72 (CMS3), and 143 patients (CMS4).

**(F)** Scheme for the experimental course of azoxymethane (AOM)/dextran sulfate sodium (DSS)-induced colorectal carcinogenesis.

**(G)** Representative endoscopic images of the normal colon mucosa and AOM/DSS tumors. T, tumors.

**(H, I)** Immunohistochemistry for ACTA2 in the normal mucosa and AOM/DSS tumors. **(H)** Representative pictures. **(I)** Quantification of ACTA2 positivity in total stromal cells. 3 HPFs/mouse, 3 mice each.

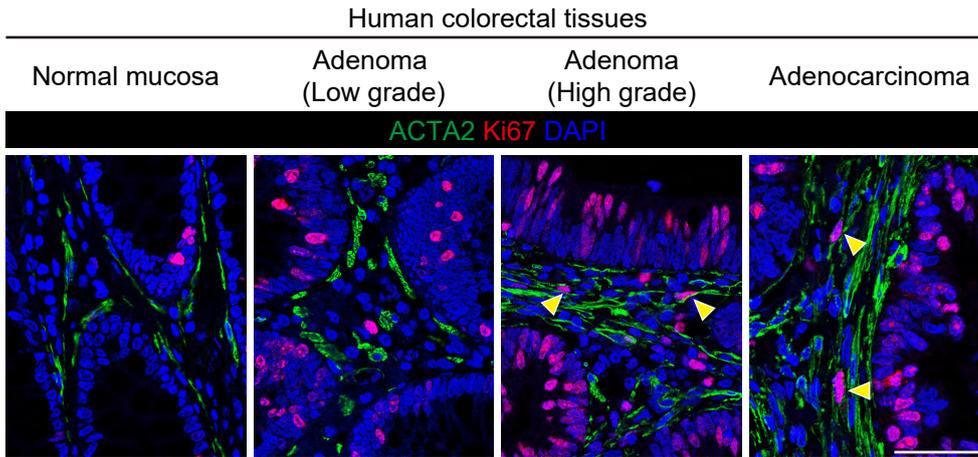
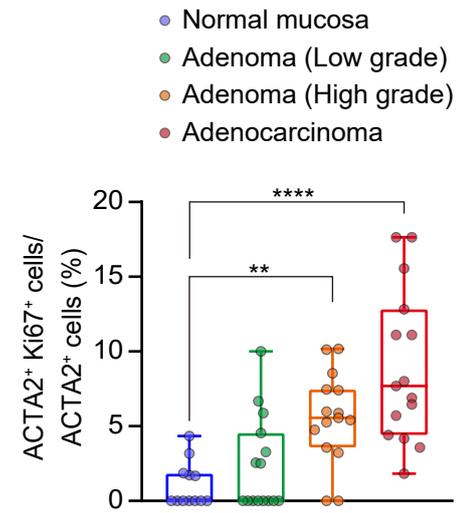
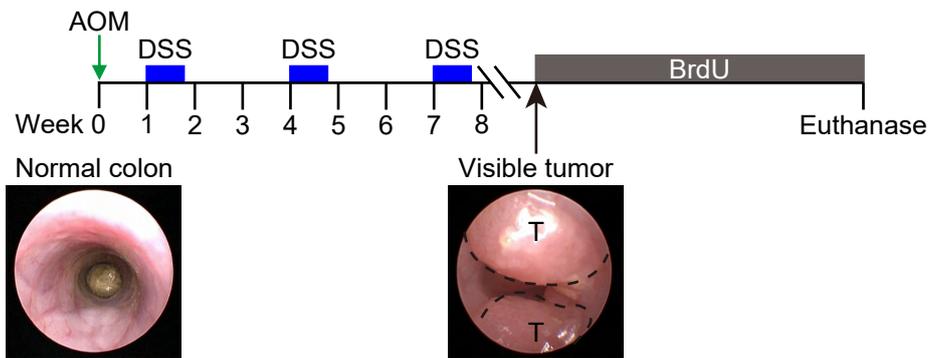
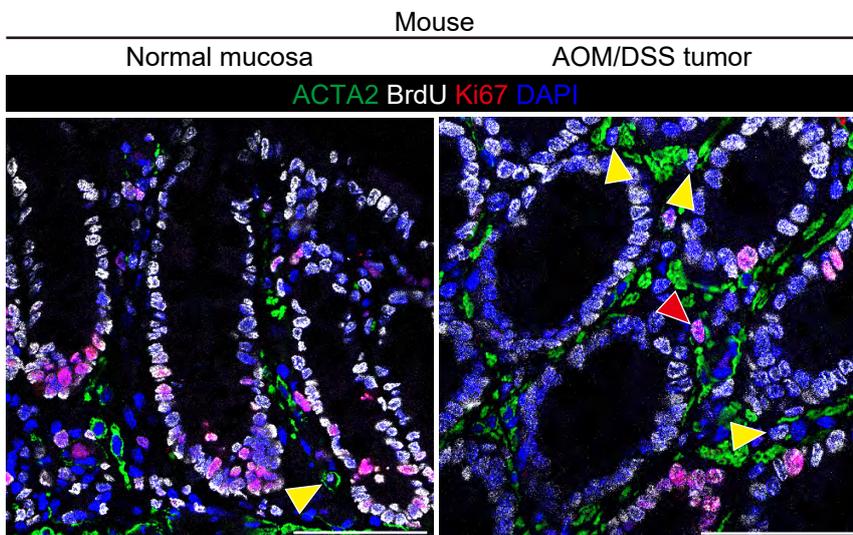
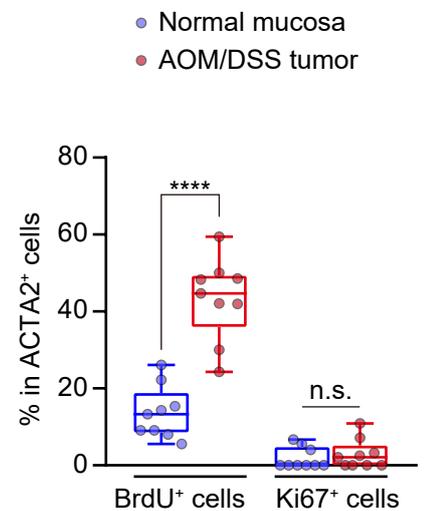
One-way ANOVA followed by Tukey's post-hoc multiple comparison test (B), Wilcoxon rank-sum test (C), Log-rank test (D), Kruskal-Wallis test followed by Dunn's multiple comparisons test (E), and two-tailed unpaired Student's t-test (I).

\*\*\*\*,  $P < 0.0001$ ; \*\*,  $P = 0.00299$ ; \*,  $P = 0.0451$ .

Scale bars, 50  $\mu\text{m}$ .

Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles (B and I).

Solid black lines, median; Dotted black lines, quartiles (C and E).

**A****B****C****D****E**

**Figure 2: A subset of ACTA2<sup>+</sup> CAFs proliferate during colorectal carcinogenesis in humans and mice.**

**(A, B)** Co-immunofluorescence for ACTA2 and Ki67 in human colorectal samples. **(A)**

Representative pictures. Yellow arrowheads denote proliferating CAFs (ACTA2<sup>+</sup>Ki67<sup>+</sup> cells). **(B)**

Quantification of Ki67 positivity in total ACTA2<sup>+</sup> cells. 3 HPFs (400x)/patient, 4-5 patients each.

**(C)** Scheme for the experimental course of AOM/DSS-induced colon carcinogenesis and 5-bromodeoxyuridine (BrdU) administration. Continuous BrdU administration was commenced after a visible tumor was observed via mouse colonoscopy. T, tumor.

**(D, E)** Co-immunofluorescence for ACTA2, BrdU, and Ki67 in the normal colon mucosa and

AOM/DSS tumors. **(D)** Representative images. Yellow and red arrowheads denote

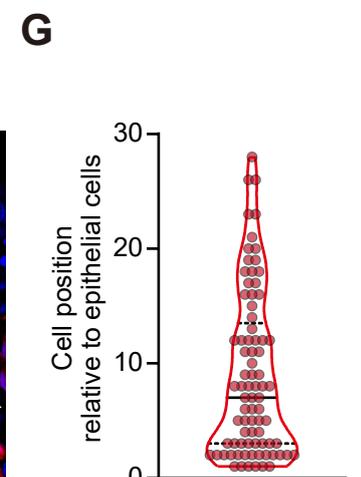
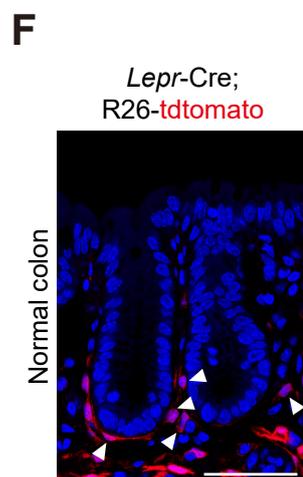
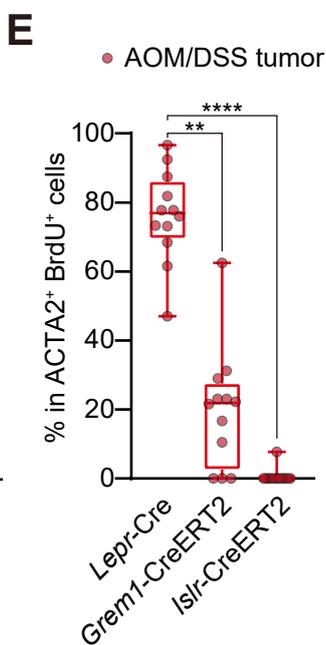
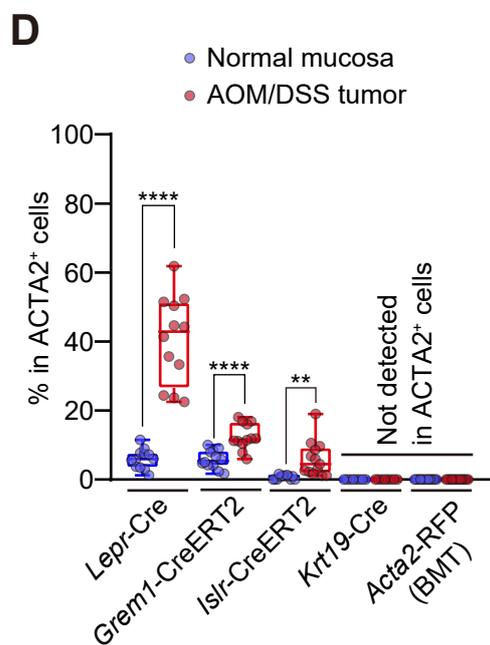
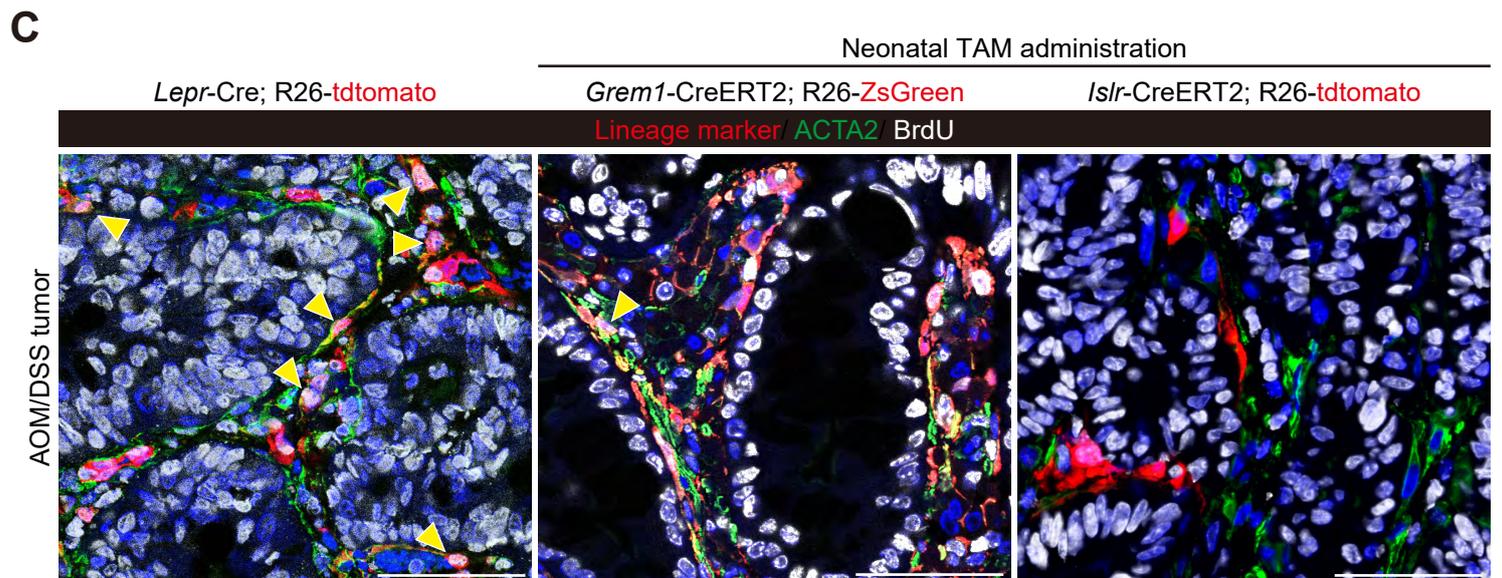
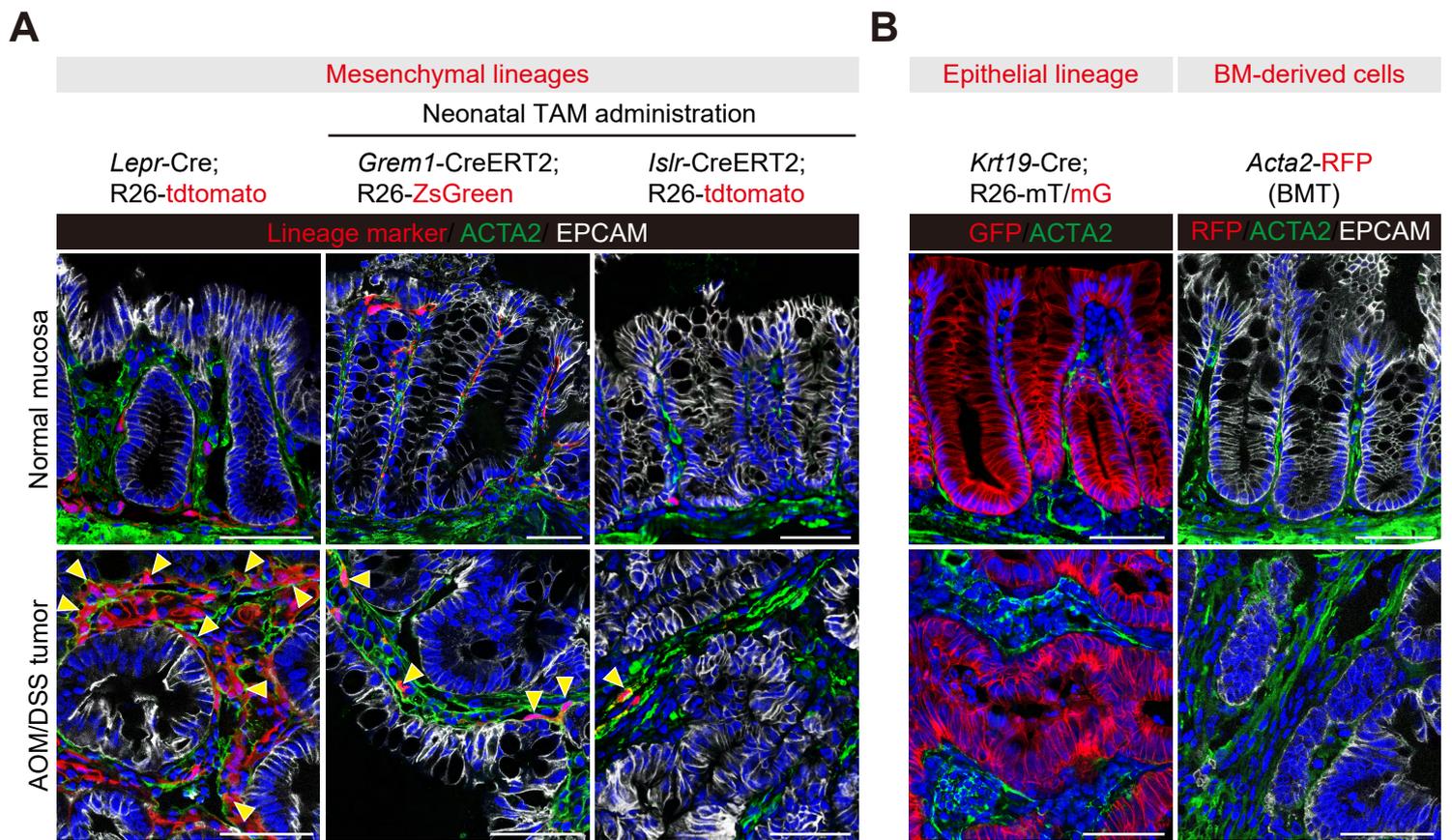
ACTA2<sup>+</sup>BrdU<sup>+</sup>Ki67<sup>-</sup> cells and ACTA2<sup>+</sup>BrdU<sup>+</sup>Ki67<sup>+</sup> cells, respectively. **(E)** Quantification of BrdU positivity (left) and Ki67 positivity (right) in total ACTA2<sup>+</sup> cells.

\*\*\*\*,  $P < 0.0001$ ; \*\*,  $P = 0.0077$ ; n.s.,  $P = 0.5049$ .

Kruskal-Wallis test followed by Dunn's multiple comparisons test (B) and two-tailed unpaired Student's t-test (E)

Scale bars, 50  $\mu$ m.

Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.



**Figure 3: Proliferating ACTA2<sup>+</sup> CAFs derive predominantly from *Lepr*-lineage cells in an AOM/DSS mouse model of CRC.**

**(A)** Immunofluorescence for ACTA2 and EPCAM in the normal colon mucosa and AOM/DSS tumors using fate-mapping mouse models. Yellow arrowheads denote lineage-marker<sup>+</sup>ACTA2<sup>+</sup> cells. See Figure 3D for quantification. R26, Rosa26-loxP-stop-loxP; BM, bone marrow; BMT, bone marrow transplantation; TAM, tamoxifen.

**(B)** Immunofluorescence for ACTA2 in the normal mucosa and AOM/DSS tumors using *Krt19-Cre* mice. (left). Immunofluorescence for ACTA2 and EPCAM in the normal mucosa and AOM/DSS tumor, using a wild-type recipient mouse transplanted with bone marrow cells from an *Acta2-RFP* mouse (right).

**(C)** Immunofluorescence for ACTA2 and BrdU in AOM/DSS tumors using the BrdU-treated fate-mapping mouse models. Yellow arrowheads denote proliferating CAFs that were derived from each cellular lineage (lineage-marker<sup>+</sup>ACTA2<sup>+</sup>BrdU<sup>+</sup> cells). See Figure 3E for quantification.

**(D)** Quantification of the ratio of lineage-marker<sup>+</sup> cells in total ACTA2<sup>+</sup> cells in the normal mucosa and AOM/DSS tumors. 4 HPFs/mouse. 3 mice (*Lepr-Cre*, *Grem1-CreET2*, *Islr-CreERT2*, *Acta2-RFP*) and 2 mice (*Krt19-Cre*).

**(E)** Quantification of the ratio of lineage marker<sup>+</sup> cells in total proliferating CAFs. 4 HPFs/mouse. 3 mice each.

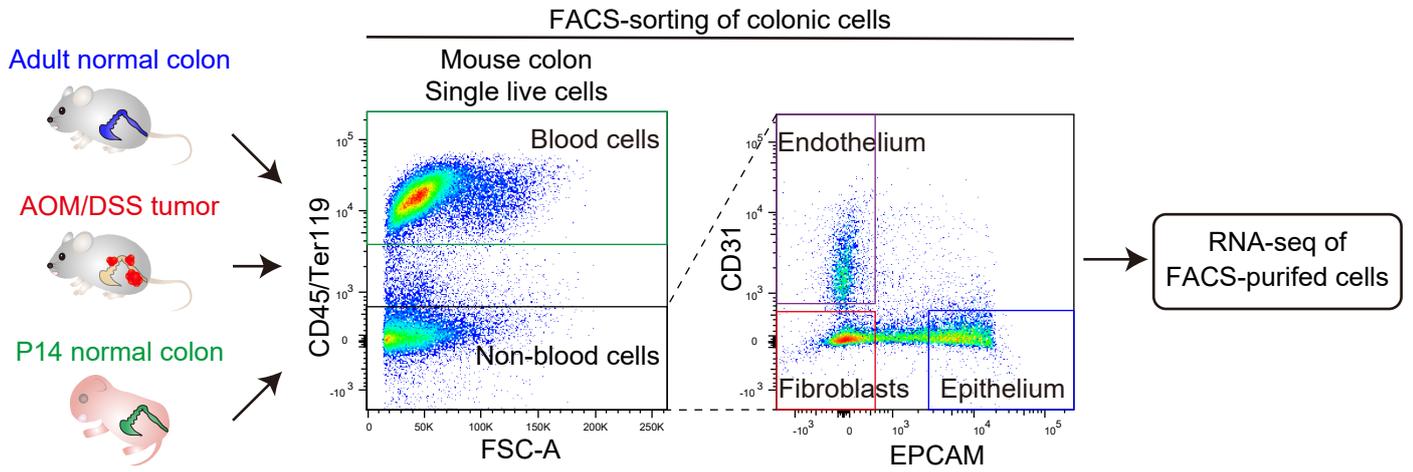
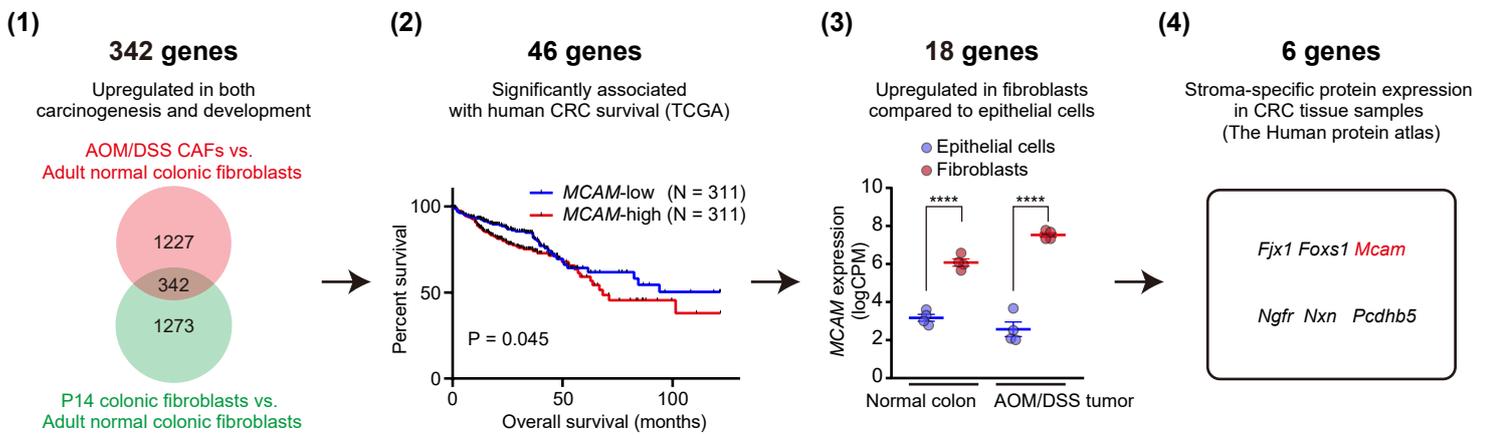
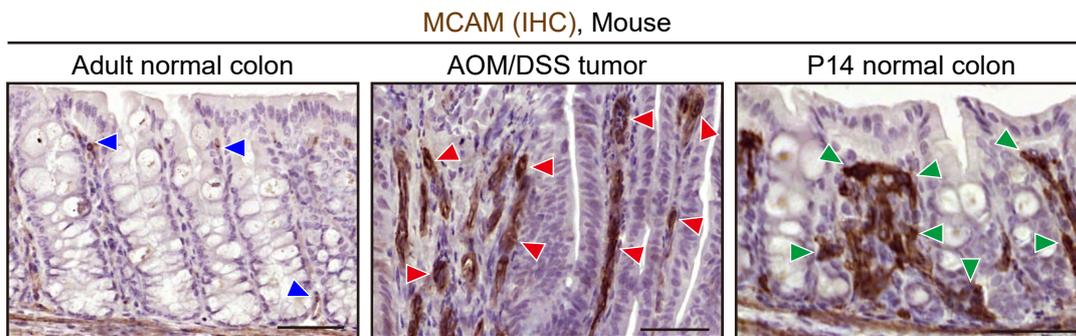
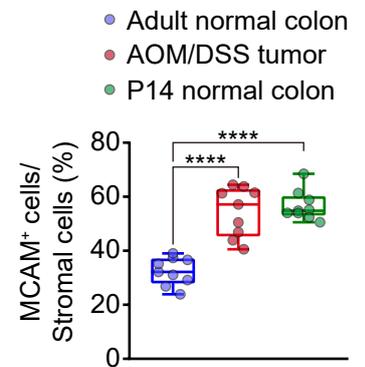
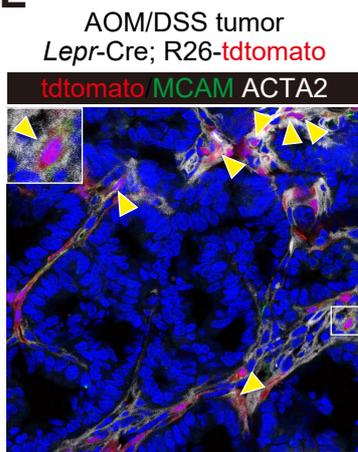
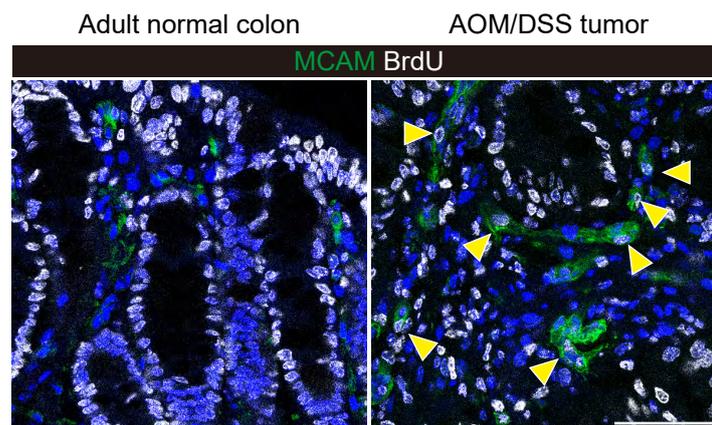
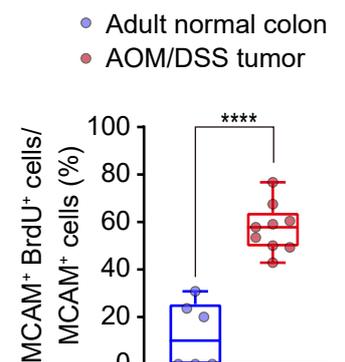
**(F, G)** Cellular positions of *Lepr*-lineage stromal cells in the normal adult mouse colon. **(F)** Representative pictures. White arrowheads denote *Lepr*-lineage tdTomato<sup>+</sup> cells. **(G)** Violin plots showing the positions of pericryptal *Lepr*-lineage stromal cells relative to the adjacent epithelial position. n = 81 *Lepr*-lineage cells from 3 mice. Solid black lines, median; Dotted black lines, quartiles.

Scale bars, 50  $\mu$ m.

Two-tailed unpaired t-test with Welch's correction (D) and Kruskal-Wallis test followed by Dunn's multiple comparisons test (E).

\*\*\*\*, P < 0.0001; \*\*, P = 0.0030 (D); \*\*, P = 0.0043 (E).

Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.

**A****B****C****D****E****F****G**

**Figure 4: Identification of MCAM as a CRC mesenchyme-specific marker that represents a subset of *Lepr*-lineage proliferating cells.**

**(A)** Experimental schematic for isolating colonic fibroblasts from the normal adult colon, AOM/DSS tumors, and postnatal day 14 colon. Gating strategy to isolate CD45<sup>-</sup>Ter119<sup>-</sup>CD31<sup>-</sup>EPCAM<sup>-</sup> fibroblasts by fluorescence-activated cell-sorting (FACS) is shown for one mouse adult normal colon. n = 4 mice each.

**(B)** Strategy to identify a colonic stromal gene upregulated in development and carcinogenesis, which is associated with human CRC survival. **(1)** Venn diagram showing 342 genes upregulated in AOM/DSS tumors and postnatal day 14 colon, compared with the normal adult colon fibroblasts. **(2)** Survival analysis using TCGA dataset. **(3)** Using our RNA-seq data, genes upregulated in EPCAM<sup>-</sup>CD31<sup>-</sup>CD45<sup>-</sup>Ter119<sup>-</sup> fibroblasts compared with EPCAM<sup>+</sup> epithelial cells, both in the normal adult colon and AOM/DSS tumors, were selected. Mean ± s.e.m. **(4)** The Human protein atlas data were used to select genes whose protein expression was restricted to the CRC stroma. *Mcam* is highlighted in red.

**(C, D)** Immunohistochemistry for MCAM. **(C)** Representative images. Blue, red, and green arrowheads denote MCAM expression in the normal adult colon, AOM/DSS tumor, and postnatal day 14 colon, respectively. **(D)** Quantification of the ratio of MCAM<sup>+</sup> cells in total stromal cells (visualized by hematoxylin counterstaining). 3 HPFs/mouse, 3 mice each.

**(E)** Co-Immunofluorescence for MCAM and ACTA2 using AOM/DSS tumors from *Lepr*-Cre; Rosa26-tdtomato mice. Yellow arrowheads denote *Lepr*-lineage MCAM<sup>+</sup> ACTA2<sup>+</sup> CAFs. See Supplementary Figure 13C and D for quantification and separate channel images.

**(F, G)** Co-immunofluorescence for MCAM and BrdU. **(F)** Representative images. Yellow arrowheads denote proliferating MCAM<sup>+</sup> cells. **(G)** Quantification of the ratio of MCAM<sup>+</sup>BrdU<sup>+</sup> cells in total MCAM<sup>+</sup> cells.

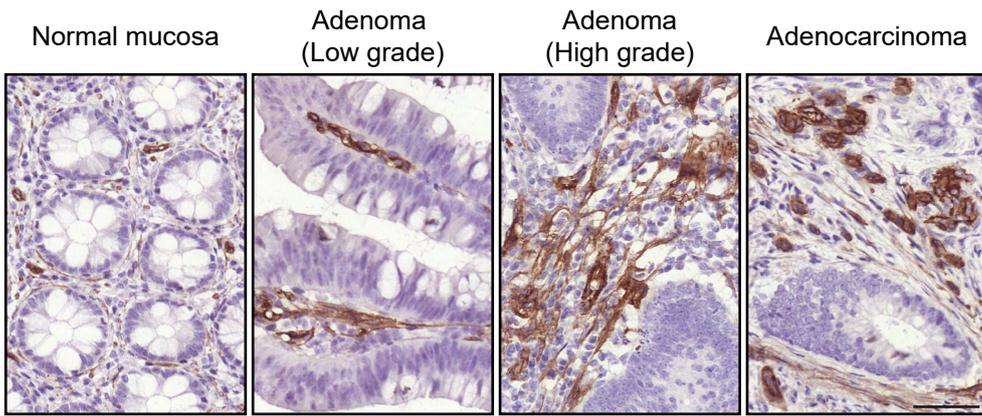
Scale bars, 50 μm.

\*\*\*\*, P < 0.0001.

Log-rank test (B(2)), one-way ANOVA followed by Tukey's post-hoc multiple comparison test (B(3) and D), and two-tailed unpaired Student's t-test (G)

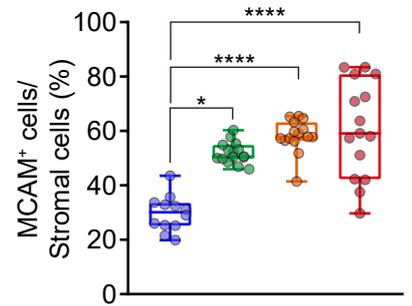
Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.

**A** MCAM (IHC), Human colorectal tissues

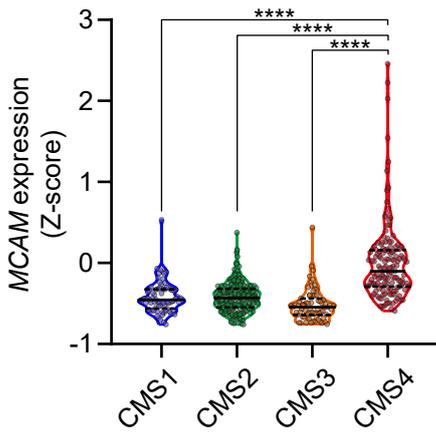


**B**

- Normal mucosa
- Adenoma (Low grade)
- Adenoma (High grade)
- Adenocarcinoma



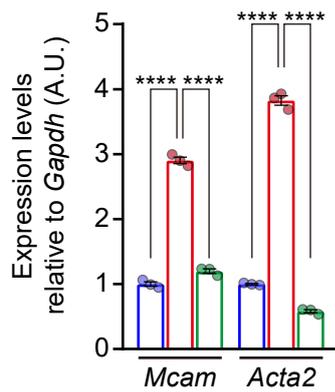
**C** TCGA (COADREAD)



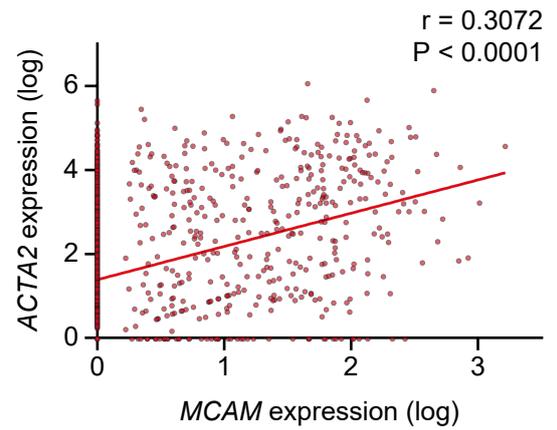
**D**

- Vehicle
- TGFβ1 + DMSO
- TGFβ1 + Galunisertib

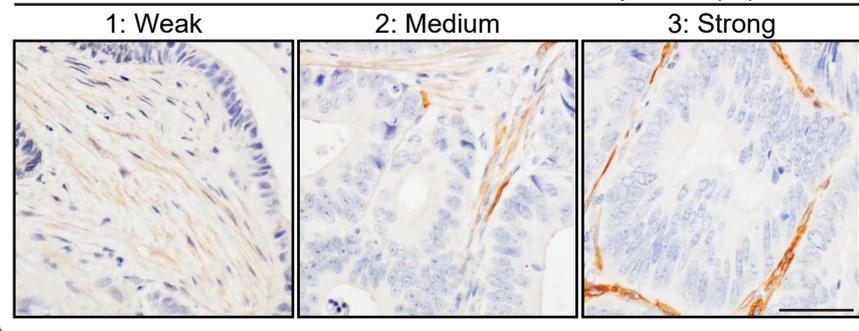
YH2 cells (mouse colonic fibroblasts)



**E** ScRNA-seq data from human CRC tissues (GSE132465) CAFs



**F** Human colorectal cancer MCAM intensity score (IS)

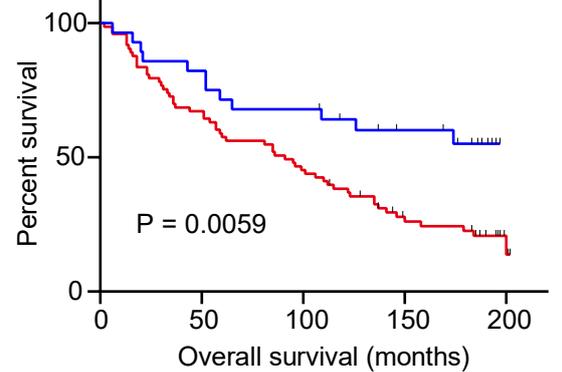


Proportion score (PS)  
 1: <10% of stroma  
 2: 11-20% of stroma  
 3: > 20% of stroma

Total score (TS) = IS + PS  
 MCAM-high: Average TS ≥ 3.3  
 MCAM-low: Average TS < 3.3

**G**

- MCAM-low (N = 28)
- MCAM-high (N = 73)



**Figure 5: High stromal MCAM expression driven, in part, by TGF- $\beta$ , is associated with poor survival in patients with CRC.**

**(A, B)** Immunohistochemistry for MCAM in human colorectal samples. **(A)** Representative pictures.

**(B)** Quantification of MCAM positivity in total stromal cells (visualized by hematoxylin counterstaining). 3 HPFs (400x)/patient, 4-5 patients each.

**(C)** Violin plots showing *MCAM* expression levels in four CMSs.

n = 76 (CMS1), 220 (CMS2), 72 (CMS3), and 143 patients (CMS4). Solid black lines, median; Dotted black lines, quartiles.

**(D)** A mouse colonic fibroblast cell line, YH2, was incubated with vehicle, recombinant TGF $\beta$ 1, or recombinant TGF $\beta$ 1 + TGF $\beta$ 1-receptor inhibitor (Galunisertib) for 24 hours, followed by quantitative reverse-transcription PCR (qRT-PCR). mean  $\pm$  s.e.m. n = 3.

**(E)** ScRNA-seq data show *MCAM* transcript levels are positively correlated with *ACTA2* expression in colorectal CAFs. n = 1854 CAFs. Solid line, linear regression

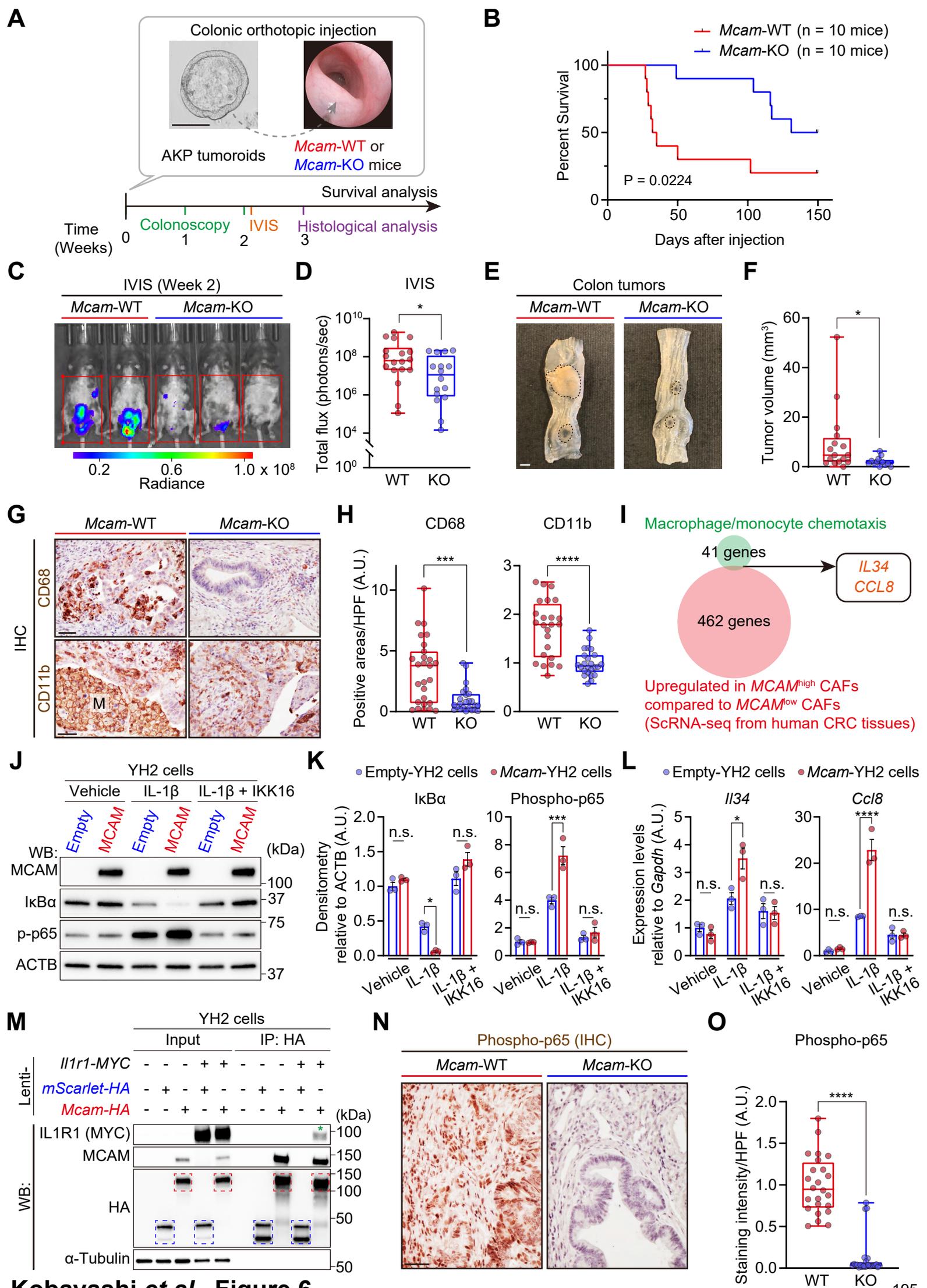
**(F, G)** MCAM immunohistochemistry in a CRC tissue microarray **(F)** Representative images of MCAM immunostaining showing each intensity score. Cases with total scores (the sum of intensity and proportion scores) of  $\geq 3.3$  were defined as MCAM-high cancer. **(G)** Kaplan-Meier survival curves.

Scale bars, 50  $\mu$ m.

\*\*\*\*, P < 0.0001; \*, P = 0.0124.

Kruskal-Wallis test followed by Dunn's multiple comparisons test (B and C), one-way ANOVA followed by Tukey's post-hoc multiple comparison test (D), Spearman correlation (E), and Log-rank test (G).

Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles (B).



**Figure 6: Stromal MCAM promotes CRC progression via IL1R1-p65-IL34/CCL8 signaling-mediated macrophage recruitment.**

**(A)** Experimental scheme showing orthotopic injection of *Apc*<sup>Δ/Δ</sup>, *Kras*<sup>G12D/Δ</sup>, *Trp53*<sup>Δ/Δ</sup> CRC organoids (AKP tumoroids) into the colon. WT, wild type; KO, knockout; IVIS, *in vivo* imaging system.

**(B)** Kaplan-Meier survival curves.

**(C, D)** Luciferase signals from AKP tumoroids were assessed by IVIS. Luciferase values in red boxes were quantified. 18 *Mcam*-WT mice and 16 KO mice.

**(E, F)** Macroscopic evaluation of colon tumors. Mice were harvested 3 weeks after tumoroid injection. **(E)** Representative pictures. Dotted lines indicate tumors. **(F)** Quantification of tumor volumes. 2 injections/mouse, 8 *Mcam*-WT mice and 6 KO mice

**(G, H)** Immunohistochemistry for CD68 and CD11b. **(G)** Representative pictures. M, macrophages as assessed by morphology. **(H)** Quantification of 3,3'-Diaminobenzidine (DAB)-positive areas. 3 HPFs (400x)/tumor, 1-2 tumors/mouse, 5 mice each group. A.U., arbitrary unit.

**(I)** Venn diagram showing the overlap of 41 macrophage/monocyte chemoattractant genes identified by Gene Ontologies and 462 genes upregulated in *MCAM*<sup>high</sup> CAFs compared with *MCAM*<sup>low</sup> CAFs (scRNA-seq data from GSE132465).

**(J, K, L)** Lentivirus-mediated overexpression of MCAM augments IL-1β-p65-*Il34/Ccl8* signaling in YH2 cells. MCAM-overexpressing or empty YH2 cells were stimulated with recombinant IL-1β, followed by Western blotting (WB; **J, K**) and qRT-PCR (**L**). mean ± s.e.m. n = 3 each. p-p65, phosphorylated p65.

**(M)** Immunoprecipitation (IP) for MCAM-hemagglutinin (HA) tag with an anti-HA antibody, followed by western blotting. A green asterisk denotes the interaction of MCAM-HA with IL1R1. An anti-MYC antibody was used to detect IL1R1 protein tagged with MYC. Blue and red dotted boxes indicate mScarlet-HA and MCAM-HA proteins, respectively.

**(N, O)** Immunohistochemistry for phosphorylated p65. **(N)** Representative pictures. **(O)** Quantification of DAB intensity. 3 HPFs/tumor, 1-2 tumors/mouse, 5 mice each group.

Scale bars, 200 μm (A), 2 mm (E), 50 μm (G and N)

All histopathological analyses were performed using mice harvested 3 weeks after tumoroid injection.

Log-rank test (B), two-tailed unpaired t-test with Welch's correction (D), Mann-Whitney U-test (F, H, and O), and two-way ANOVA followed by Tukey's post-hoc multiple comparison test (K and L).

\*\*\*\*,  $P \leq 0.0001$ ; \*\*\*,  $P \leq 0.001$ ; \*,  $P \leq 0.05$ ; n.s.,  $P > 0.05$

Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.

## **Acknowledgments**

We thank Kaori Ushida, Kozo Uchiyama (Nagoya University, JAPAN) for their technical assistance. The mouse colonic fibroblast cell line, YH2, was a kind gift from Professor Antony Burgess (Walter and Eliza Hall Institute of Medical Research, Australia). We acknowledge the facilities and the scientific and technical assistance of the South Australian Genome Editing (SAGE) Facility, the University of Adelaide, and the South Australian Health and Medical Research Institute. SAGE is supported by Phenomics Australia. Phenomics Australia is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) program.

# The origin and contribution of cancer-associated fibroblasts in colorectal carcinogenesis

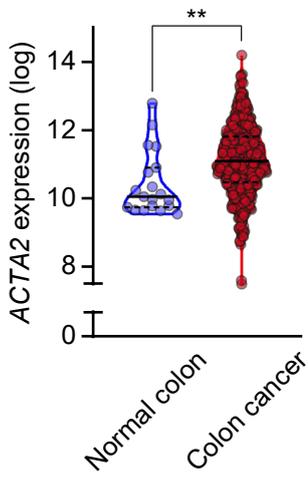
Kobayashi *et al.*

## Supplemental Information

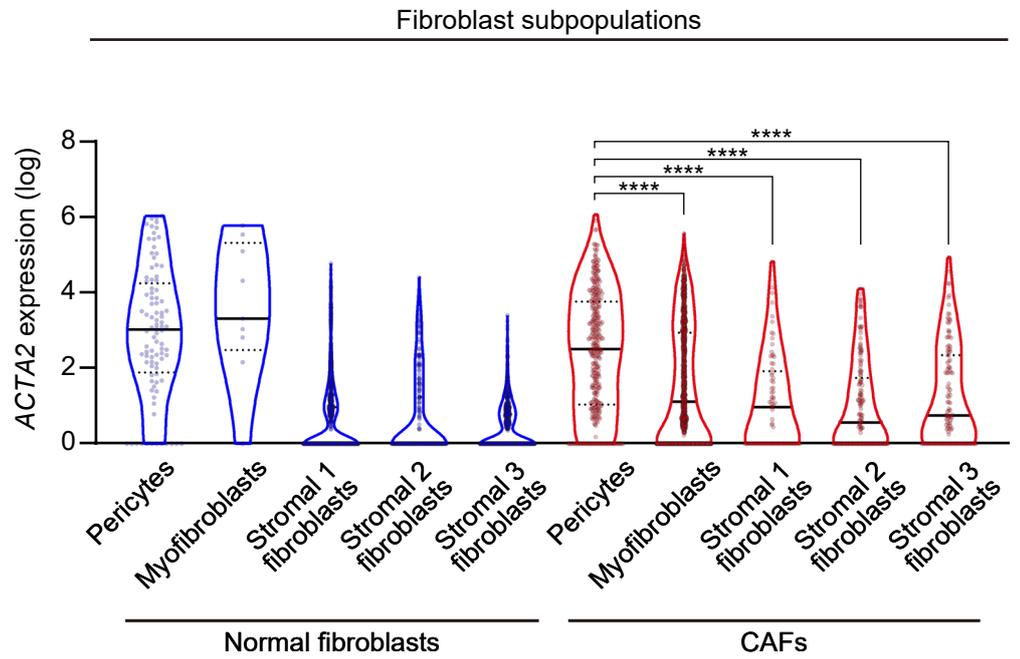
1. Supplementary Figures 1-24 and figure legends
2. Supplementary Table 1-3
3. Supplementary Materials and Methods

**A**

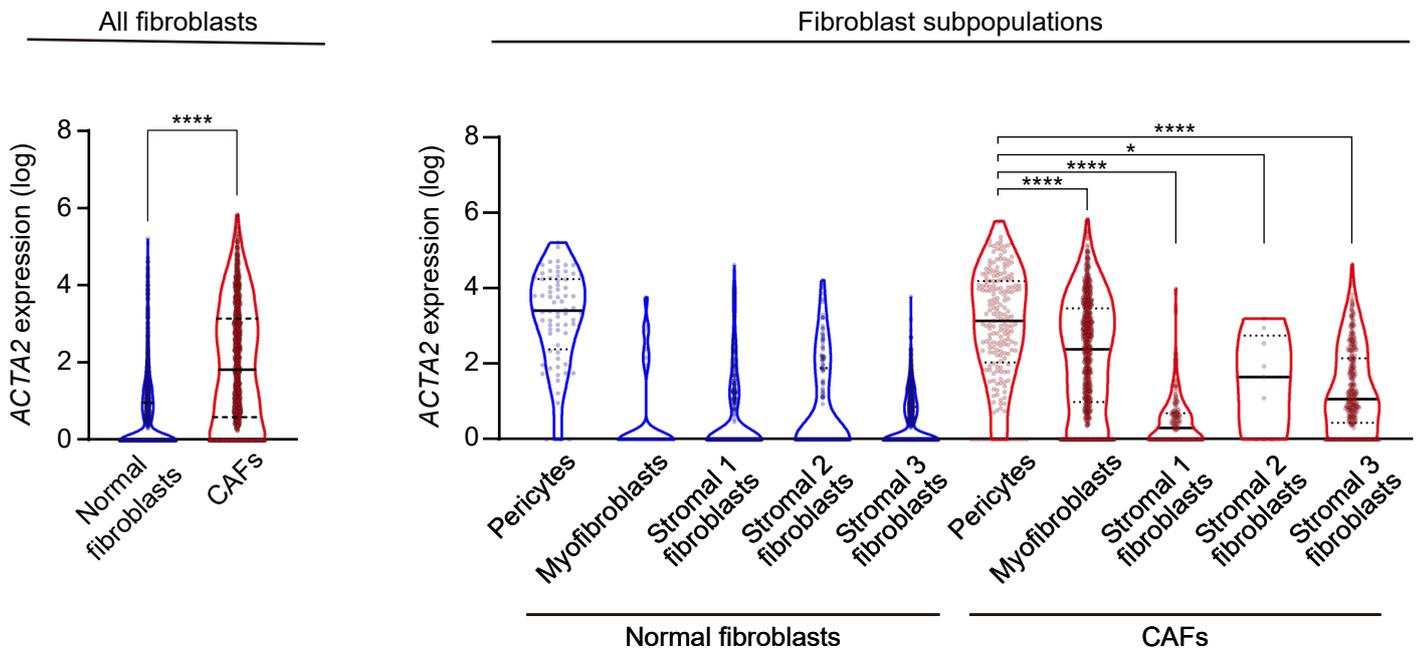
Expression microarray data  
from human colorectal tissues  
(GSE39582)

**B**

ScRNA-seq from human colorectal tissues (GSE132465)

**C**

ScRNA-seq from human colorectal tissues (GSE144735)



**Supplementary Figure 1: ACTA2 expression is increased during human colorectal carcinogenesis, with the highest ACTA2 transcripts observed in pericytes in CAF subpopulations.**

**(A)** Violin plots depicting ACTA2 expression in expression microarray data from human normal colorectal mucosa and primary colorectal adenocarcinoma (GSE39582)<sup>251</sup>. n = 19 (normal colorectal mucosa) and 566 patients (primary colorectal adenocarcinoma).

**(B)** Violin plots show ACTA2 transcript levels in fibroblasts from normal colorectal mucosa and CRC tissues. Single-cell RNA-sequencing (ScRNA-seq) data from human colorectal tissues (GSE132465) were analyzed<sup>226</sup>. Data with all fibroblasts are shown in Figure 1C.

n = 92, 9, 972, 236, 744, 353, 1146, 73, 158, and 124 cells (left to right; normal pericytes to Stromal 3 fibroblasts in tumors) from 23 CRC patients.

**(C)** Violin plots show ACTA2 transcript levels in fibroblasts from normal colorectal mucosa and CRC tissues. ScRNA-seq data from human colorectal tissues (GSE144735) were analyzed<sup>226</sup>.

n = 2429 normal fibroblasts from colorectal mucosa and 1483 CRC CAFs from 6 patients (All fibroblasts)

n = 72, 26, 417, 151, 1763, 201, 778, 159, 9, and 336 cells (Fibroblast subpopulations; left to right; normal pericytes to Stromal 3 fibroblasts in tumors)

In the scRNA-seq datasets (GSE132465 and GSE144735), fibroblasts are defined as pericytes, myofibroblasts, and Stromal 1-3 fibroblasts.

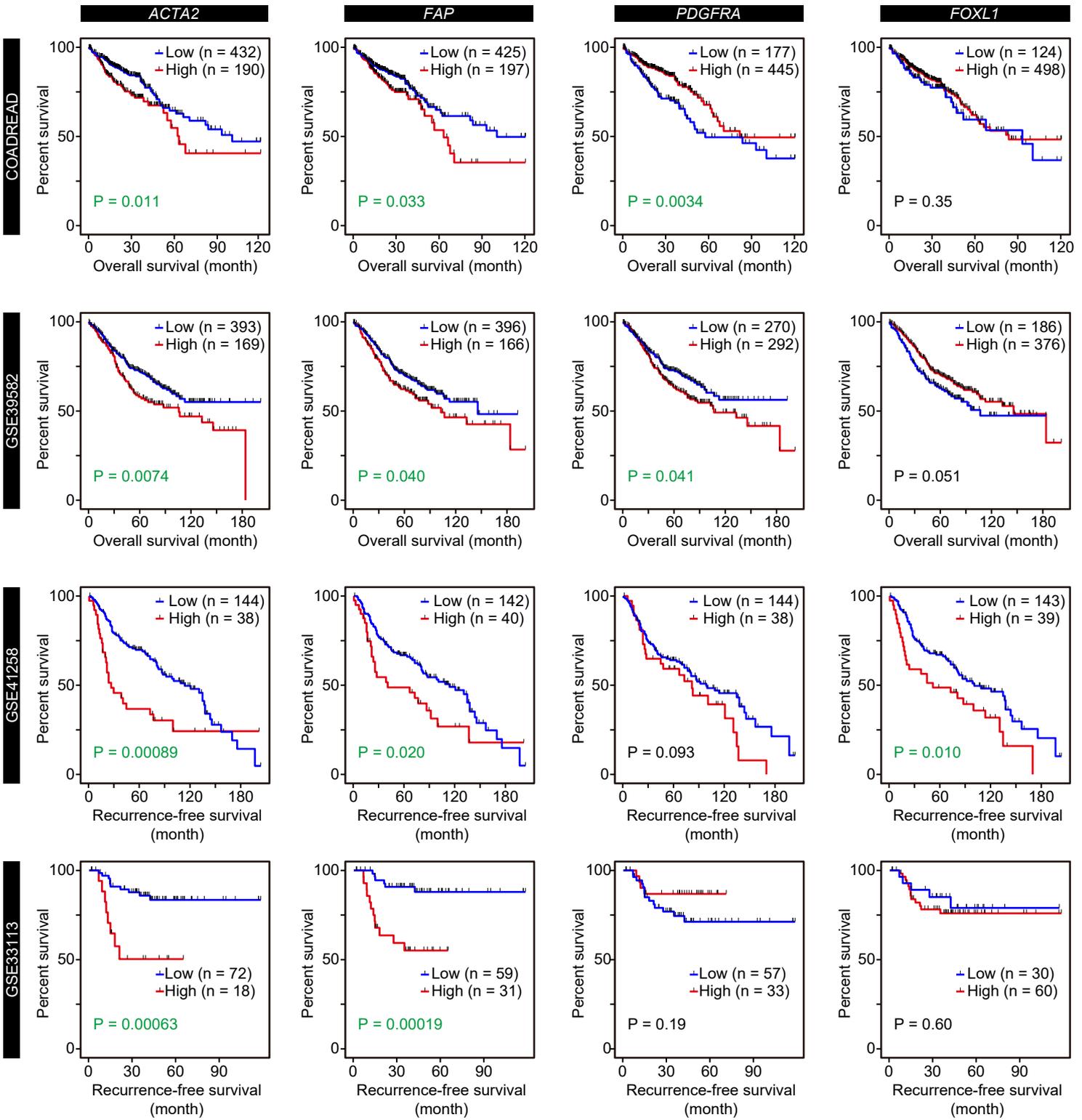
False discovery rate (FDR)-adjusted P-values (A), Kruskal-Wallis test followed by Dunn's post-hoc multiple comparisons (B and right-sided graph in C), and Wilcoxon rank-sum test (left-sided graph in C).

\*\*\*\*, P < 0.0001; \*, P = 0.0452

Solid black lines, median; Dotted black lines, quartiles.

Fibroblast markers

Gene expression datasets from human CRC tissues

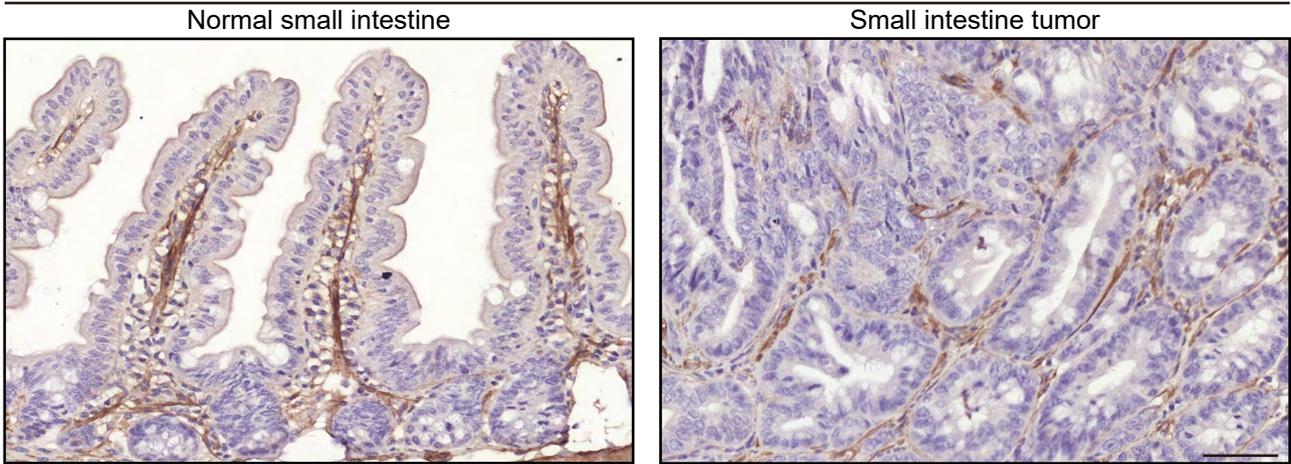
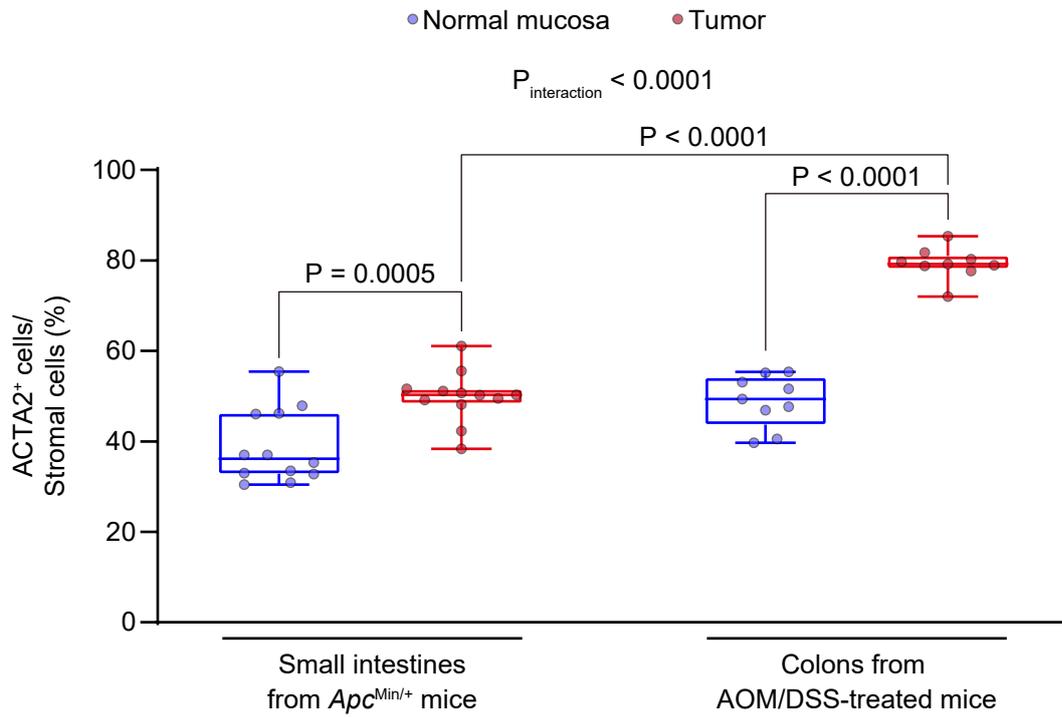


**Supplementary Figure 2: Analyses of multiple human CRC expression datasets reveal that high *ACTA2* and *FAP* expression are associated with poor survival.**

Associations between survival and transcript levels of intestinal fibroblast markers were assessed using 4 independent human CRC expression datasets<sup>250,251,254,288</sup>.

Log-rank test. P-values less than 0.05 are shown in green.

Note that the same survival analysis of the TCGA dataset using *ACTA2* expression is shown in Figure 1D as well.

**A**ACTA2 (IHC), *Apc*<sup>Min/+</sup> mouse**B**

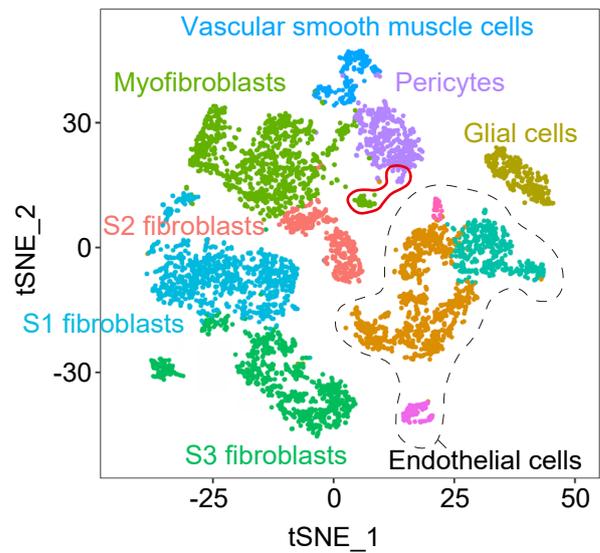
**Supplementary Figure 3: ACTA2 expression is increased in intestinal tumors from *Apc*<sup>Min/+</sup> mice compared with the adjacent normal mucosa, but to a lesser extent than the AOM/DSS mouse model.**

**(A, B)** Immunohistochemistry (IHC) for ACTA2 using the normal small intestine and small intestinal tumors in *Apc*<sup>Min/+</sup> mice. **(A)** Representative pictures. **(B)** Quantification of the ratio of ACTA2<sup>+</sup> cells in total stroma cells (as visualized by hematoxylin counterstaining). 3 HPFs/mouse, 3-4 mice each. Two-way ANOVA followed by Tukey's post-hoc multiple comparison test (B). Note that, for statistical comparisons, the data with AOM/DSS-treated mice (Figure 1H and I) are also shown in Supplementary Figure 3B.

Scale bar, 50µm. Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.

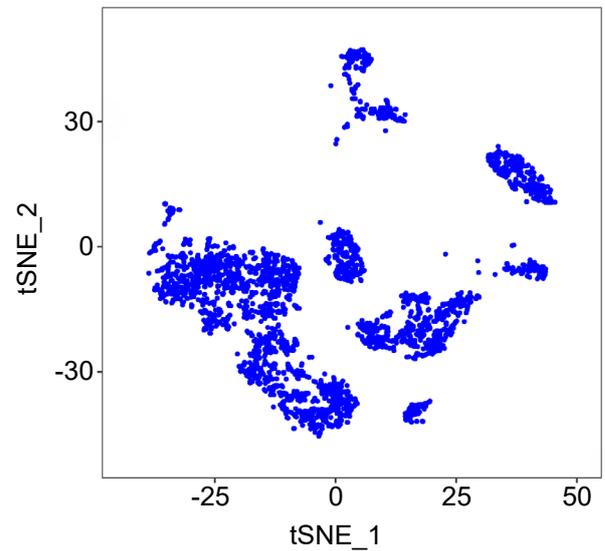
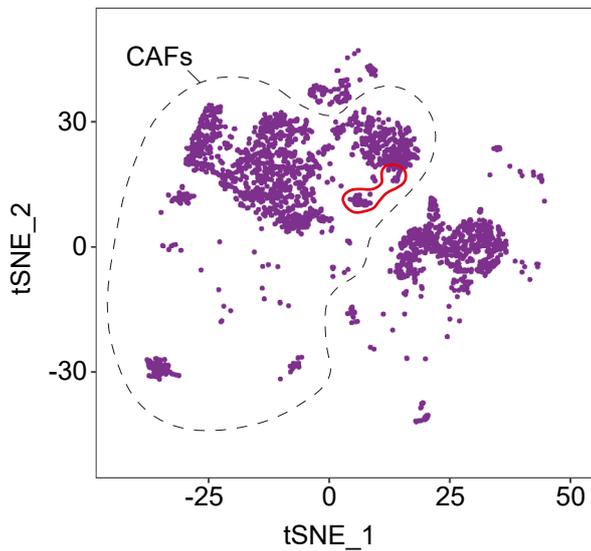
**A**

5933 stromal cells from  
human CRC and normal mucosa tissues

**B**

Stromal cells from CRC tissues

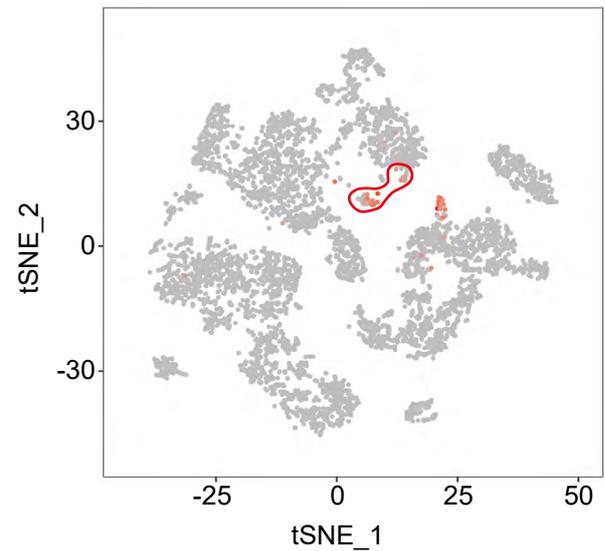
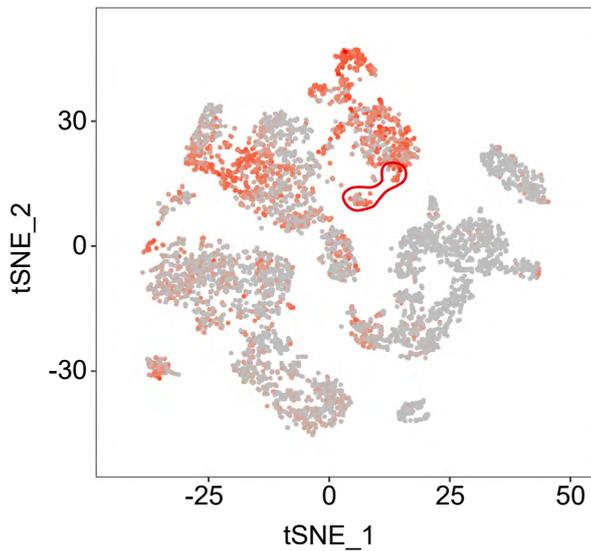
Stromal cells from normal mucosa tissues

**C**

 *ACTA2*<sup>+</sup>*MKI67*<sup>+</sup> CAFs

*ACTA2*

*MKI67*



(GSE132465; Lee *et al.*, *Nat Genet*, 2020)

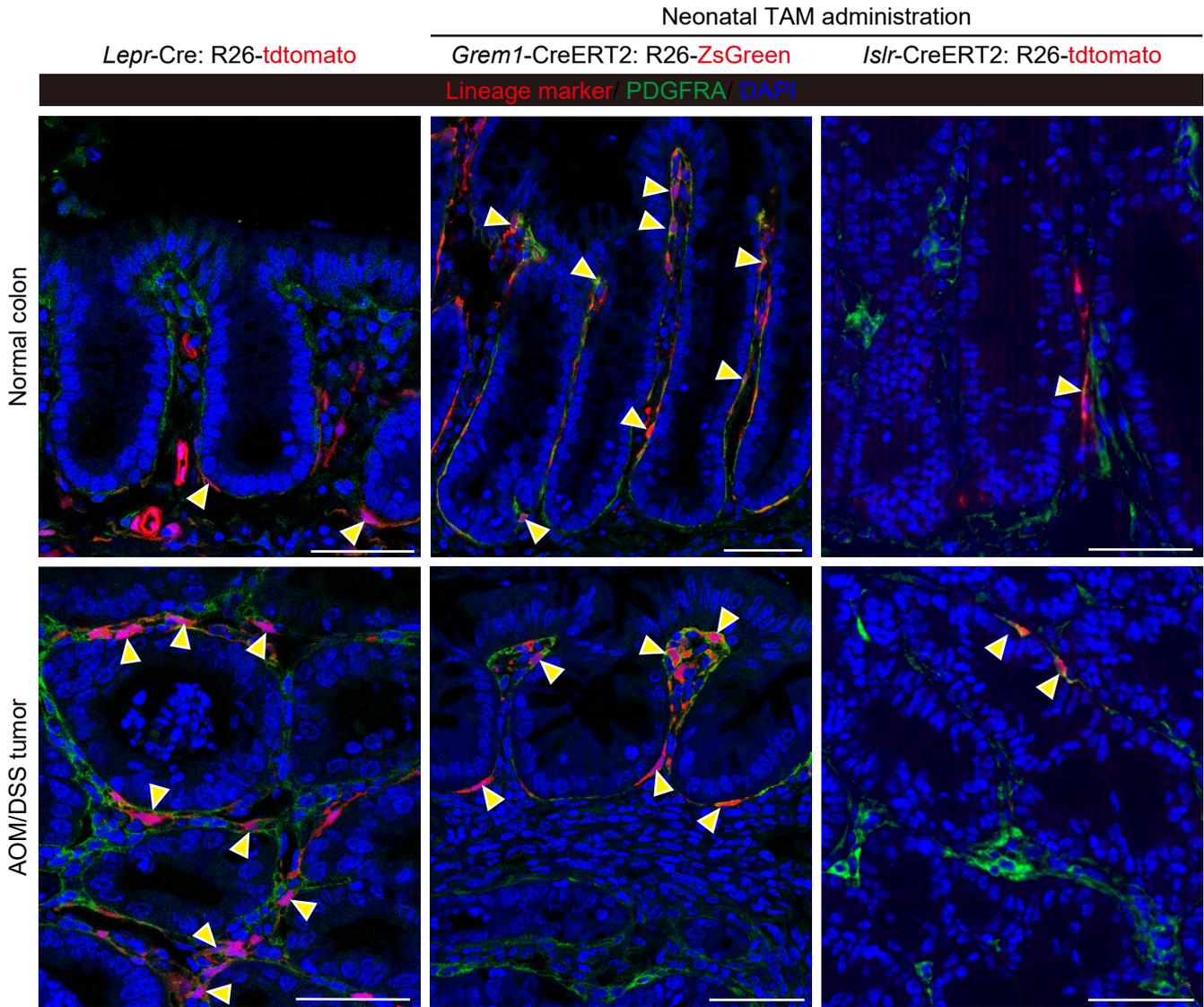
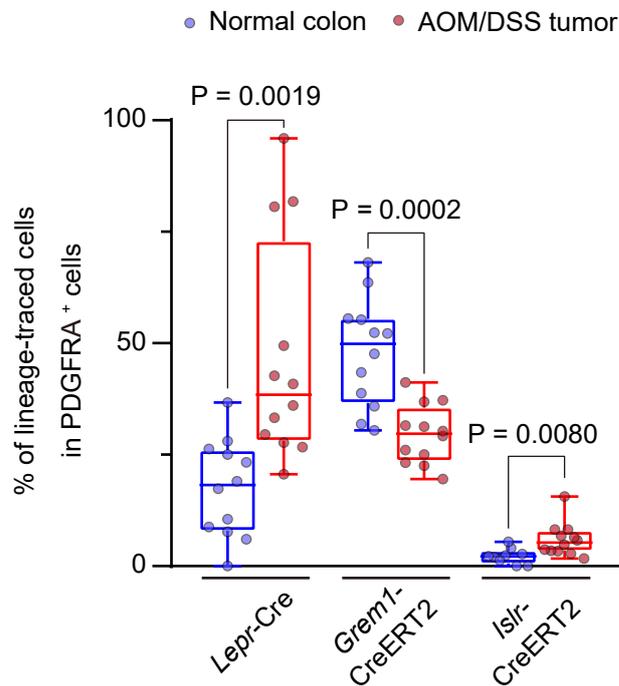
**Supplementary Figure 4: Single-cell RNA-sequencing data from human CRC tissues reveal that a subpopulation of *ACTA2*<sup>+</sup> CAFs express *MKI67*.**

**(A-C)** t-distributed stochastic neighbor embedding (t-SNE) plot showing human colorectal stromal cell subpopulations. 5933 stromal cells from 23 CRC patients were analyzed using single-cell RNA-sequencing (scRNA-seq) from unsorted colorectal cells (GSE132465)<sup>226</sup>.

Cluster annotations **(A)**. Sample origins (CRC tissues or normal mucosa tissues) **(B)**. The intensity of the red represents the expression levels of *ACTA2* (left) or *MKI67* (right) **(C)**.

Red circles denote a proliferating CAF cluster characterized by the co-expression of *ACTA2* and *MKI67*.

CAFs are defined as pericytes, myofibroblasts, and Stromal 1-3 fibroblasts found in CRC tissues.

**A****B**

**Supplementary Figure 5: *Lepr*-lineage cells are a major contributor to PDGFRA<sup>+</sup> CAFs in AOM/DSS tumors.**

**(A, B)** Immunofluorescence for PDGFRA in the normal colon mucosa and AOM/DSS tumor in three genetic fate-mapping mouse models.

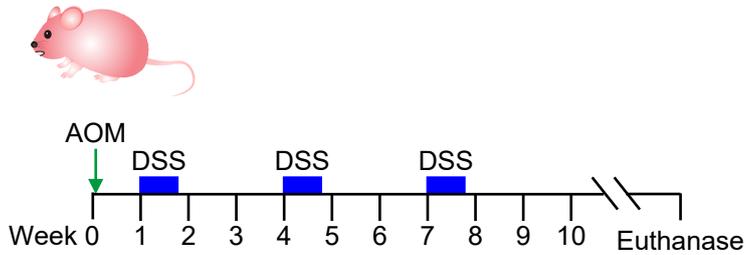
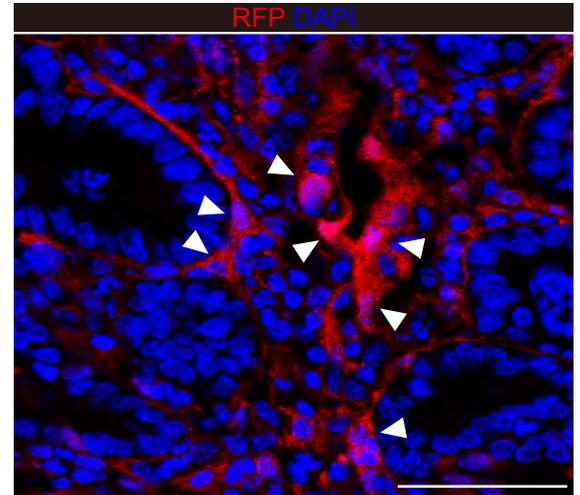
**(A)** Representative pictures. Yellow arrowheads denote lineage-marker (tdtomato or ZsGreen)<sup>+</sup> PDGFRA<sup>+</sup> cells. R26, Rosa26-loxP-stop-loxP.

**(B)** Quantification of the ratio of lineage-marker<sup>+</sup> cells in total PDGFRA cells. 3-4 HPFs/mouse, 3 mice each.

Scale bars, 50  $\mu$ m.

Two-tailed unpaired t-test with Welch's correction (B)

Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.

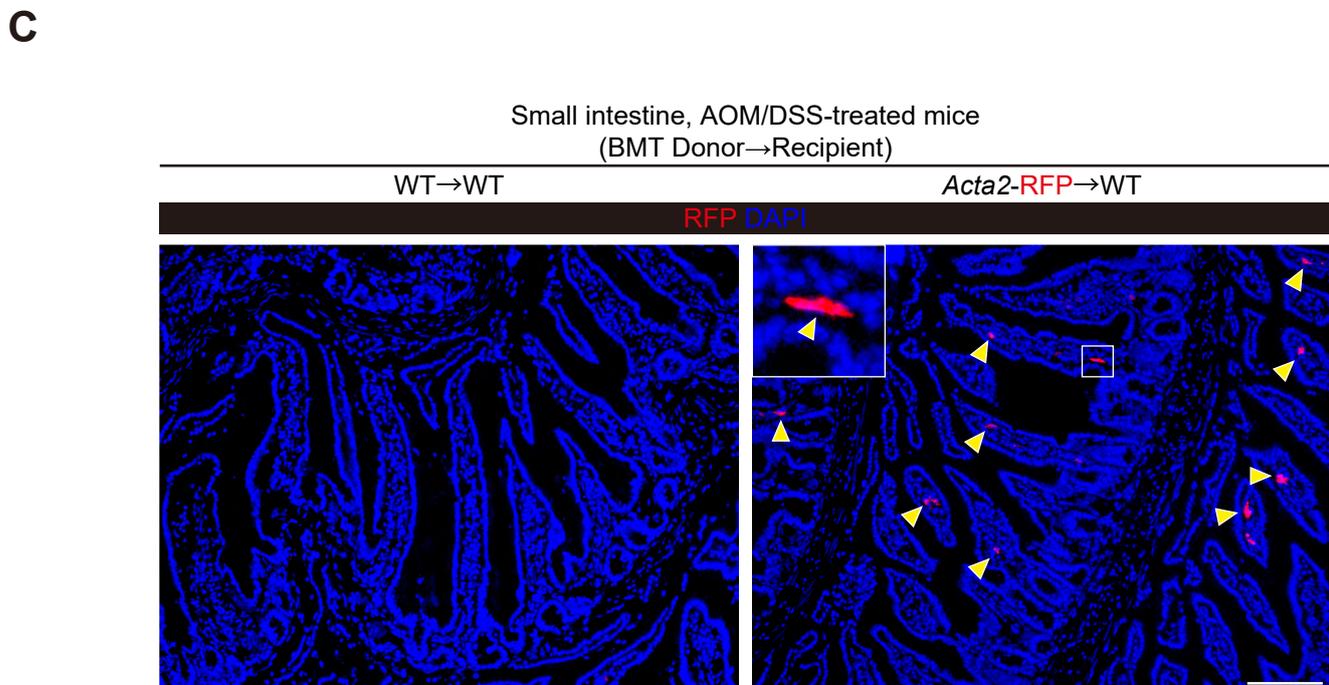
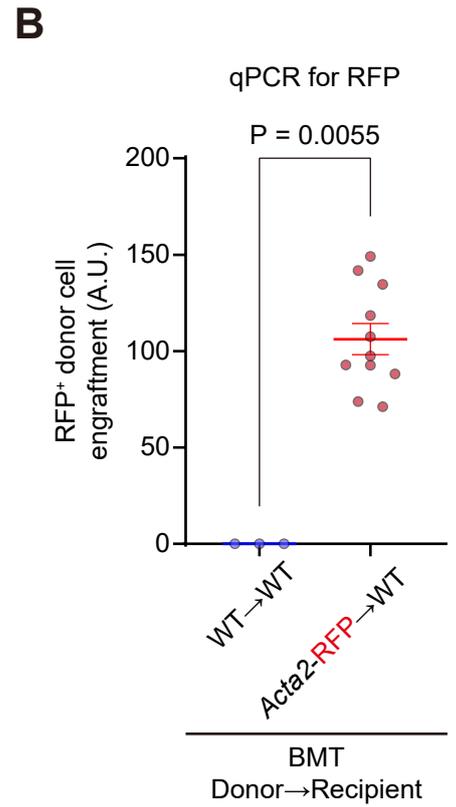
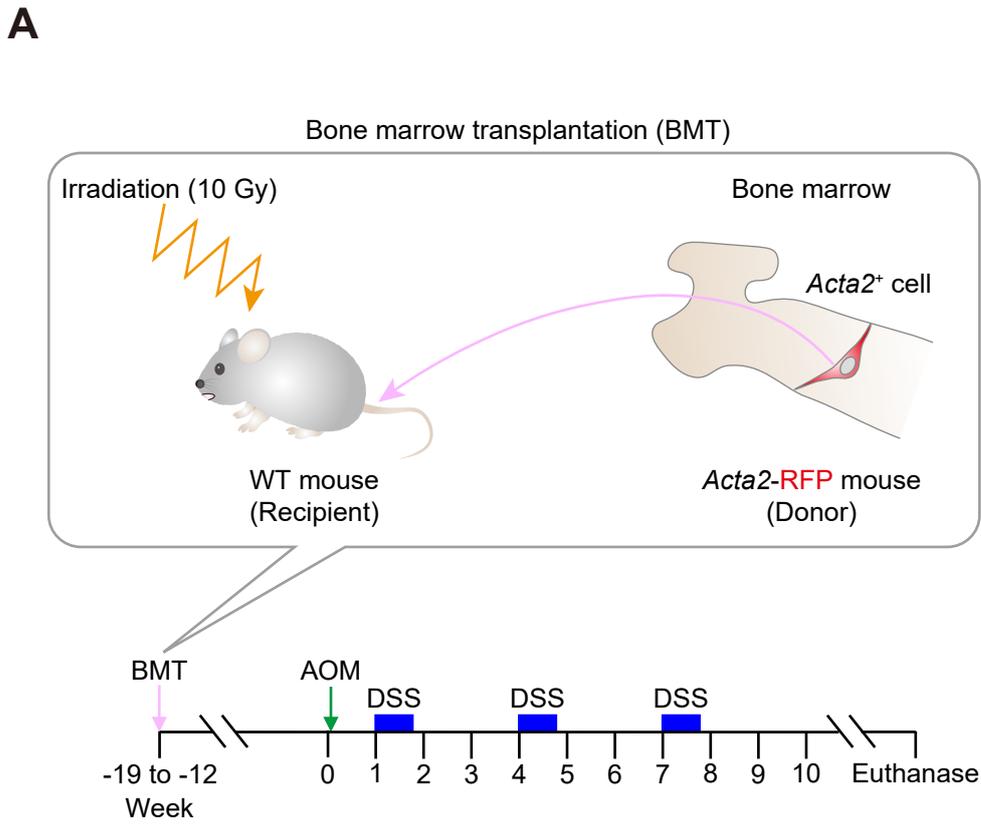
**A***Acta2-RFP* mouse**B**AOM/DSS tumor in an *Acta2-RFP* mouse

**Supplementary Figure 6: RFP reporter protein is expressed in AOM/DSS tumors from *Acta2-RFP* mice.**

**(A)** Scheme for the experimental course of azoxymethane (AOM)/ dextran sulfate sodium (DSS)-induced colorectal carcinogenesis. *Acta2-RFP* mice were treated with AOM/DSS. RFP, Red Fluorescent Protein.

**(B)** Representative picture showing RFP fluorescence in colorectal tumors from AOM/DSS-treated *Acta2-RFP* mice. White arrowheads denote RFP expression. Note that RFP was expressed by spindle-shaped fibroblastic cells in the tumor, suggesting that RFP promoter was active in CAFs in this mouse model.

n ≥ 5 mice. Scale bar, 50 μm.



**Supplementary Figure 7: Validation of bone marrow-derived cell engraftment in recipient wild-type mice.**

**(A)** Experimental schematic depicting the bone marrow transplantation experiment and the following AOM/DSS treatment. Lethally irradiated C57B/6 wild-type (WT) recipient mice were transplanted with bone marrow cells from the *Acta2*-RFP donor mice. AOM administration was commenced 12-19 weeks after bone marrow transplantation (BMT). Gy, Gray.

**(B)** Quantitative PCR (qPCR) for RFP using genomic DNA isolated from the recipient mice. To prepare a negative control for RFP genomic DNA from the bone marrow, wild-type recipient mice were transplanted with bone marrow cells from wild-type mice (WT→WT).

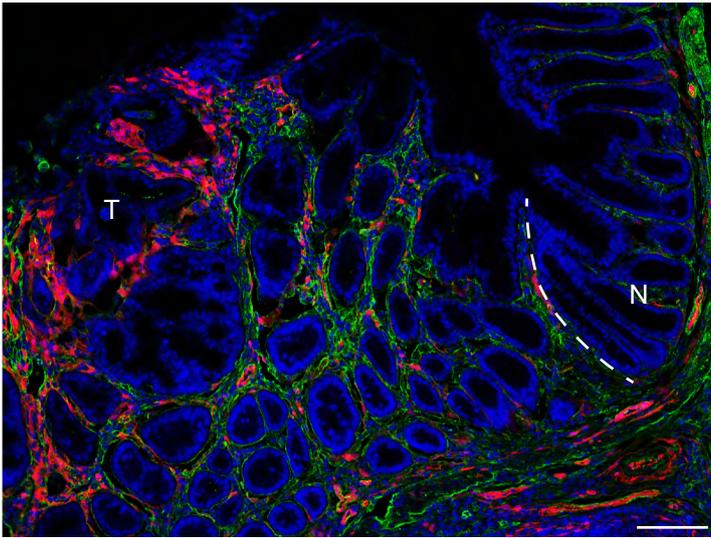
n = 3 (WT→WT) and 11 mice (*Acta2*-RFP→WT). A.U., Arbitrary Unit.

Mean ± s.e.m. Mann-Whitney U-test

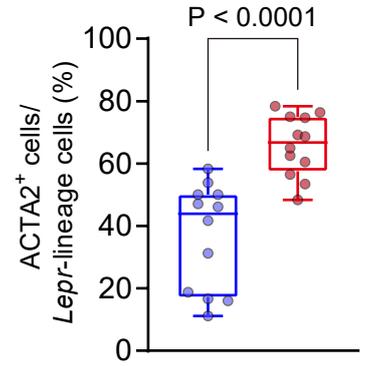
**(C)** Representative fluorescent microscopic images of the small intestines from wild-type mice that underwent bone marrow transplantation. Yellow arrowheads denote bone-marrow transplanted RFP<sup>+</sup> cells engrafted in the small intestine of a wild-type recipient mouse. n = 3 mice each.

**A***Lepr-Cre; Rosa26-tdtomato*, AOM/DSS tumor

tdtomato ACTA2 DAPI

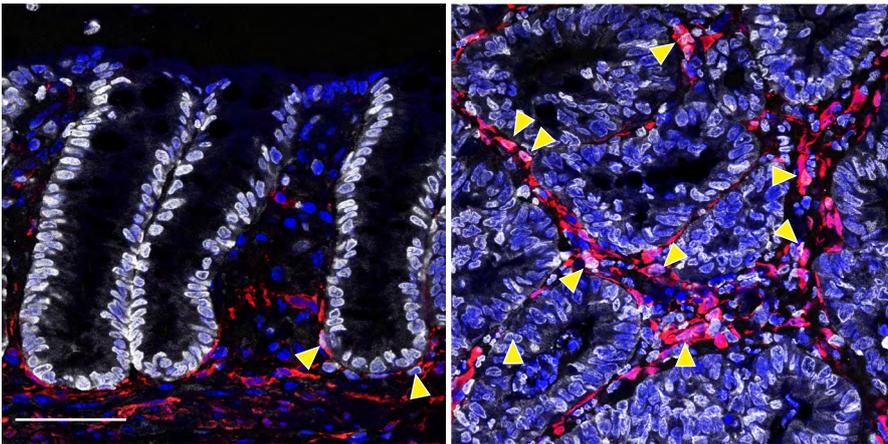
**B**

- Normal mucosa
- AOM/DSS tumor

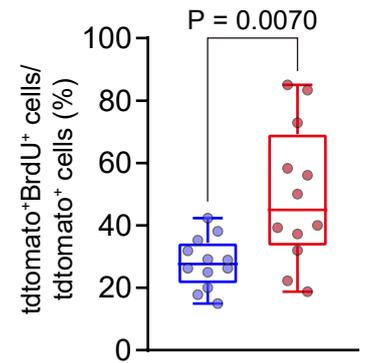
**C***Lepr-Cre; Rosa26-tdtomato*

Normal mucosa AOM/DSS tumor

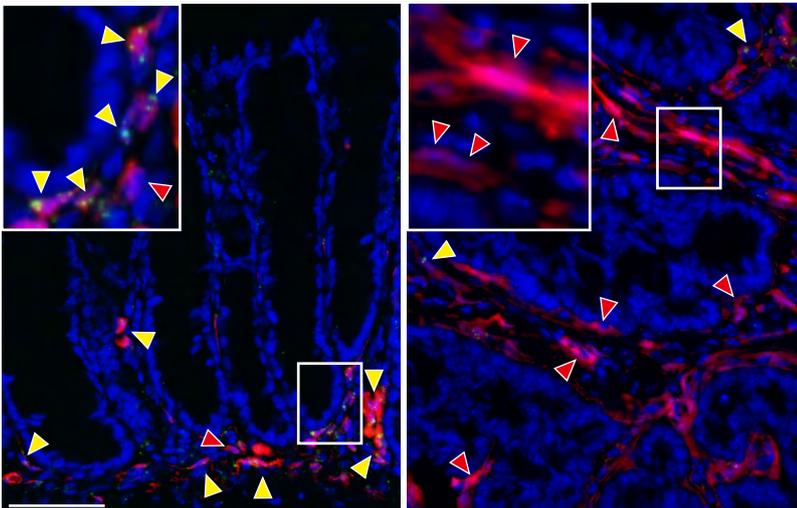
tdtomato BrdU DAPI

**D**

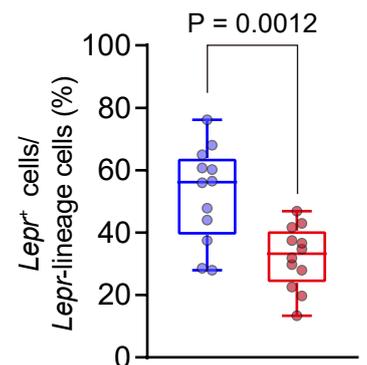
- Normal mucosa
- AOM/DSS tumor

**E***Lepr-Cre; Rosa26-tdtomato*

Normal mucosa AOM/DSS tumor

tdtomato *Lepr* (smFISH) DAPI**F**

- Normal mucosa
- AOM/DSS tumor



**Supplementary Figure 8: *Lepr*-lineage stromal cells undergo proliferation and differentiation into ACTA2<sup>+</sup> CAFs at the expense of *Lepr* expression during colorectal carcinogenesis.**

**(A)** Representative low-power magnification picture of an AOM/DSS tumor in a *Lepr*-Cre; Rosa26-tdtomato mouse immunostained with ACTA2. The number of ACTA2<sup>+</sup> cells and *Lepr*-lineage cells are increased in the tumor (T) compared with the normal colon (N). White dotted lines denote a border between the tumor and normal colon.

**(B)** Quantification of the ratio of ACTA2<sup>+</sup> cells in *Lepr*-lineage<sup>+</sup> cells (tdtomato<sup>+</sup> cells).

4 HPFs/mouse, 3 mice.

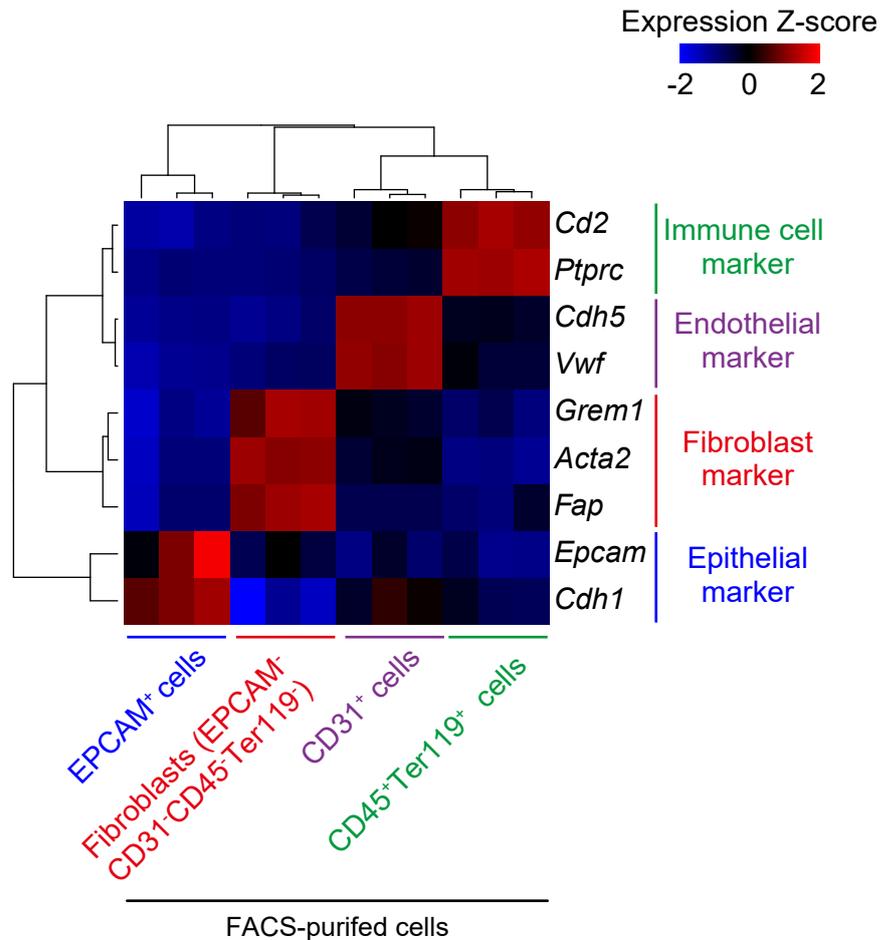
**(C, D)** Immunofluorescence for tdtomato and BrdU using the normal mucosa and AOM/DSS tumors in BrdU-treated *Lepr*-Cre; Rosa26-tdtomato mice. **(C)** Representative pictures. Yellow arrowheads denote proliferating *Lepr*-lineage cells (BrdU<sup>+</sup> tdtomato<sup>+</sup> cells). **(D)** Quantification of proliferating *Lepr*-lineage cells in total *Lepr*-lineage cells. 4 HPFs/mouse, 3 mice.

**(E, F)** Combined *Lepr* single-molecule fluorescent *in situ* hybridization (smFISH) and tdtomato immunofluorescence (IF). Active expression of *Lepr* is decreased in *Lepr*-lineage cells in AOM/DSS tumors, compared with that in the normal colons. **(E)** Representative pictures. Yellow arrowheads denote *Lepr*-lineage cells that actively express *Lepr* transcripts (*Lepr* smFISH<sup>+</sup> tdtomato IF<sup>+</sup> cells). Red arrowheads denote *Lepr*-lineage cells that do not actively express *Lepr* transcripts (*Lepr* smFISH<sup>-</sup> tdtomato IF<sup>+</sup> cells). Boxed areas are magnified in the insets. **(F)** Quantification of the ratio of *Lepr*-expressing cells in *Lepr*-lineage cells. 4 HPFs/mouse, 3 mice.

Two-tailed unpaired t-test with Welch's correction (B, D, and F).

Scale bars, 100  $\mu$ m (A), and 50  $\mu$ m (C and E).

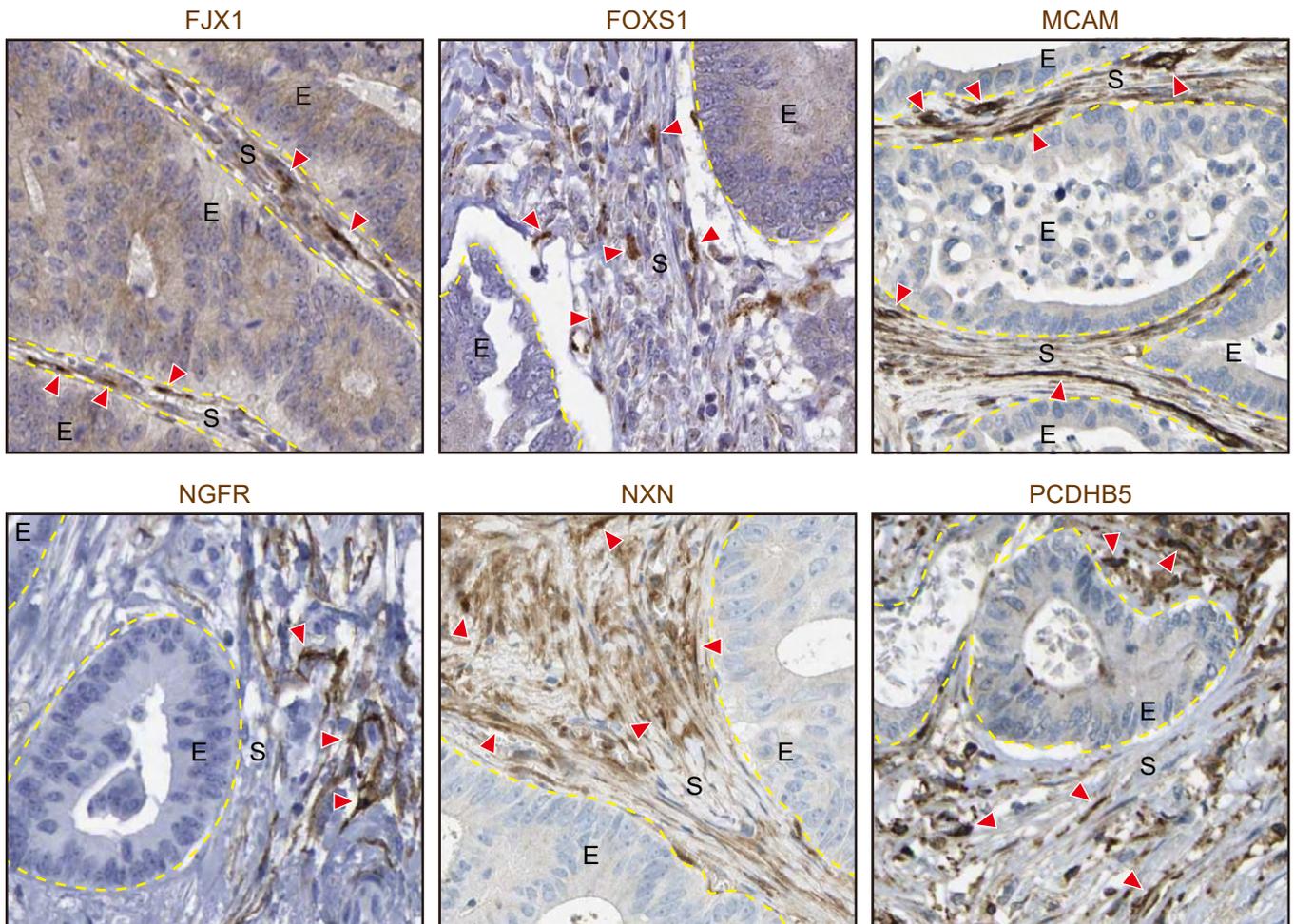
Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.



**Supplementary Figure 9: Fibroblast makers are highly expressed in a FACS-purified CD45<sup>-</sup>, Ter119<sup>-</sup>, CD31<sup>-</sup> and EPCAM<sup>-</sup> colonic mesenchymal cell population.**

Quantitative reverse-transcription polymerase chain reaction (qRT-PCR) for selected marker genes in the fluorescence-activated cell sorting (FACS)-purified normal colonic cell populations was performed, followed by hierarchical clustering. n = 3 mice.

*Ptprc*, the gene name for CD45.



(The Human Protein Atlas)

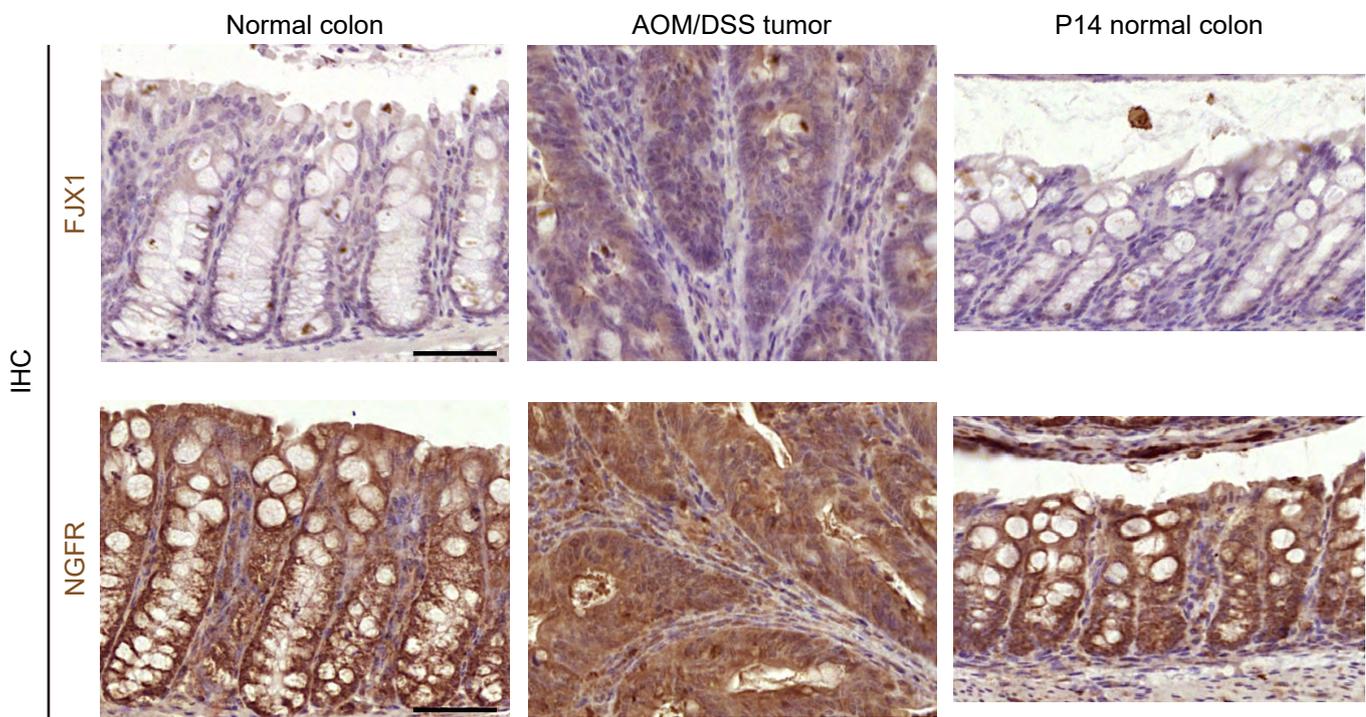
**Supplementary Figure 10: Immunohistochemistry data in the Human Protein Atlas show that six candidate proteins are expressed in the human CRC stroma compared with cancer cells.**

Histopathological analysis was performed for genes selected in Figure 4B (3) to examine whether protein expression was observed in human CRC tissue sections. Immunohistochemistry pictures were obtained from the Human Protein Atlas<sup>289</sup>.

Red arrowheads denote protein expression in the stroma. Yellow dotted lines indicate borders between epithelial cells (E) and the stroma (S).

IHC, immunohistochemistry

Mouse colonic tissues



**Supplementary Figure 11: Immunohistochemistry for several candidate genes using mouse colonic tissues did not confirm their stroma-specific protein expression, except for MCAM.**

Representative immunohistochemistry (IHC) images for several candidate genes listed in Figure 4B(4). Immunohistochemistry data for MCAM is shown in Figure 4C.

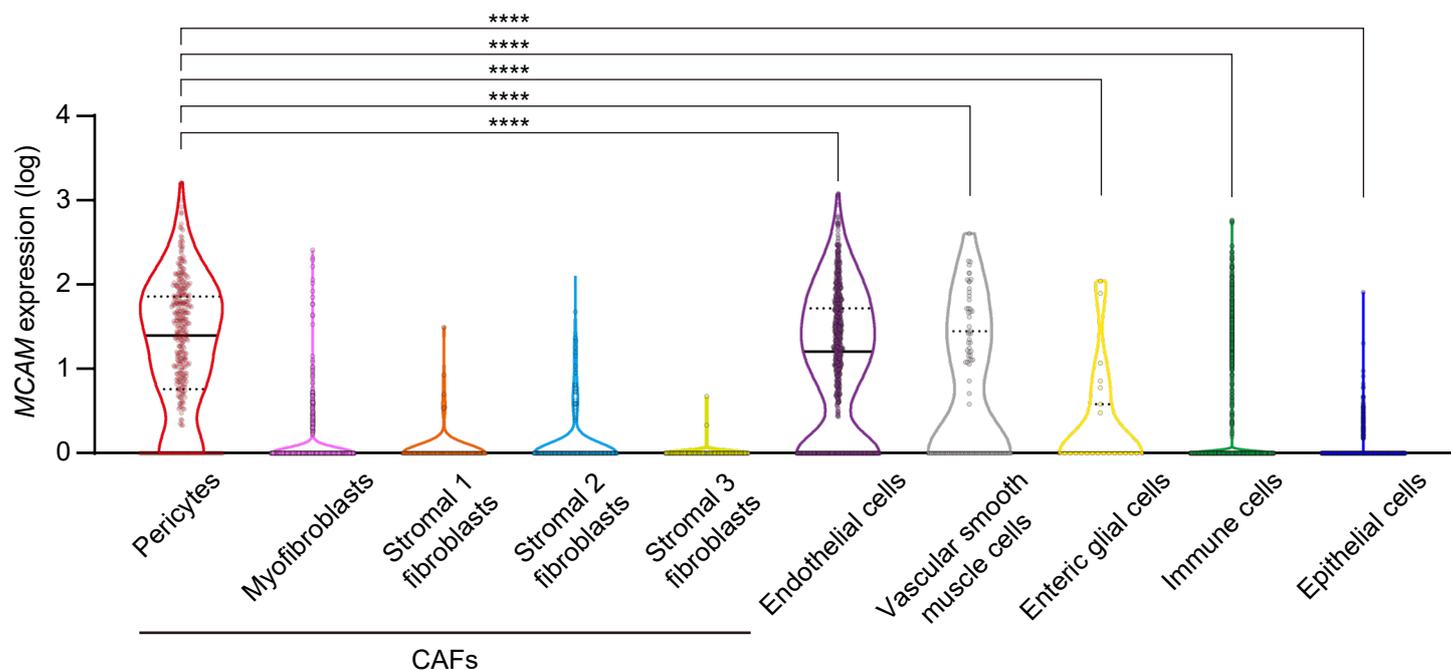
Note that immunohistochemistry for FOXS1, NXN, and PCDHB5 was not performed because of the lack of published immunohistochemistry-grade anti-mouse antibodies that could selectively stain the stroma.

P14, postnatal day 14.

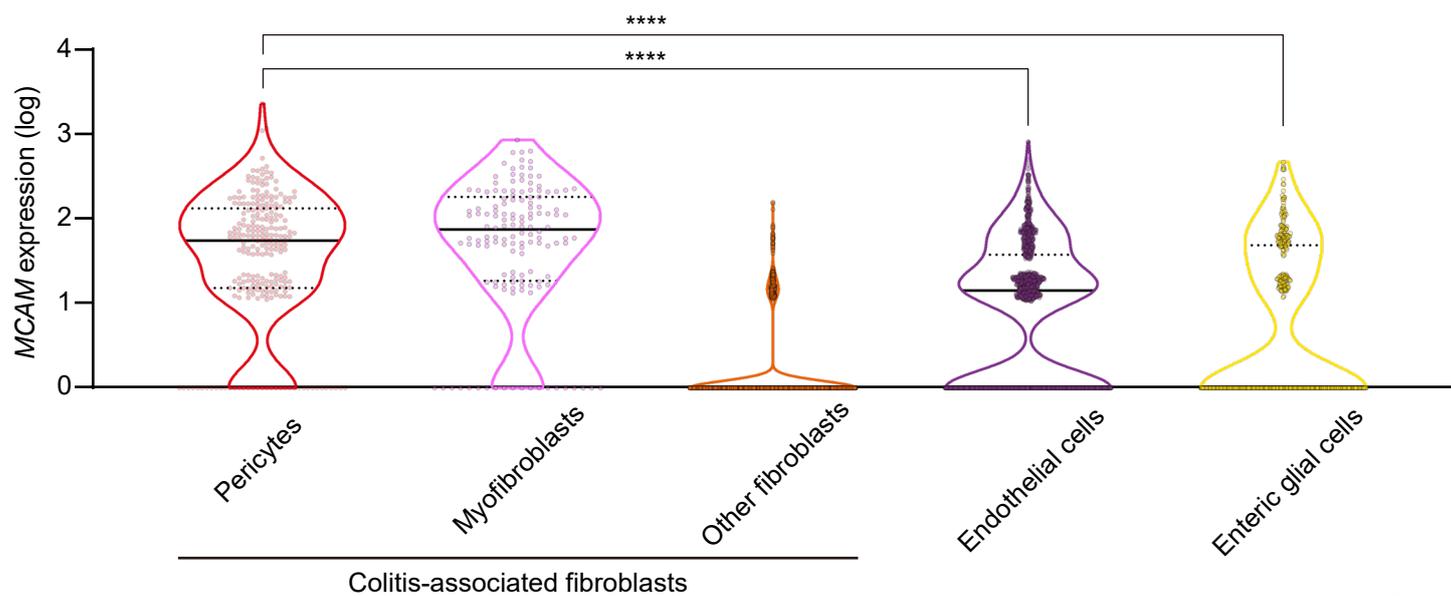
Scale bar, 50  $\mu$ m.

**A**

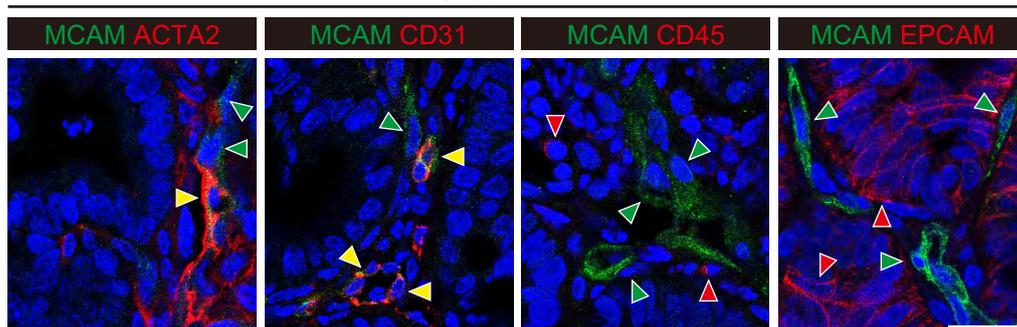
ScRNA-seq from human primary CRC tissues (GSE132465)

(Lee *et al.*, *Nat Genet*, 2020)**B**

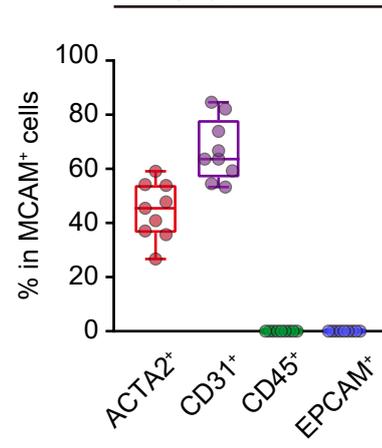
ScRNA-seq from colonic stromal cells from ulcerative colitis patients (GSE114374)

(Kinchen *et al.*, *Cell*, 2018)**C**

AOM/DSS tumor

**D**

AOM/DSS tumor



**Supplementary Figure 12: *MCAM* is expressed by fibroblasts in human CRC, ulcerative colitis samples, and mouse AOM/DSS tumors.**

**(A)** Violin plots depict *MCAM* expression in scRNA-seq data from human CRC tissues (GSE132465)<sup>226</sup>. n = 353, 1146, 73, 158, 124, 768, 91, 23, 27080 and 17469 cells (left to right) from 23 CRC patients. CAFs are defined as pericytes, myofibroblasts, and Stromal 1-3 fibroblasts.

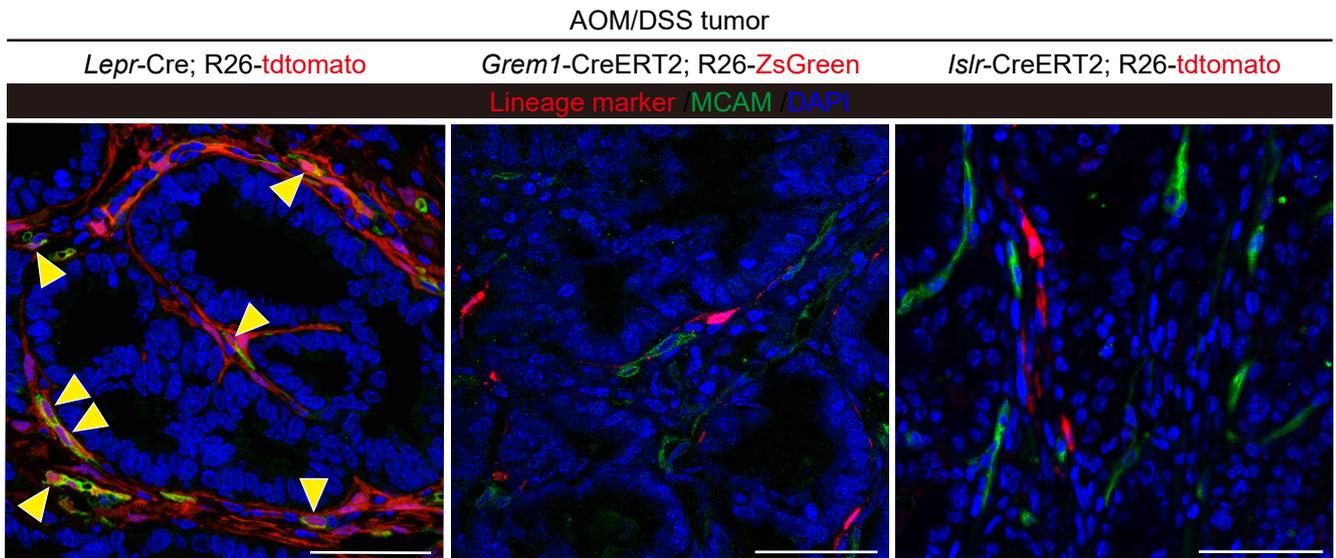
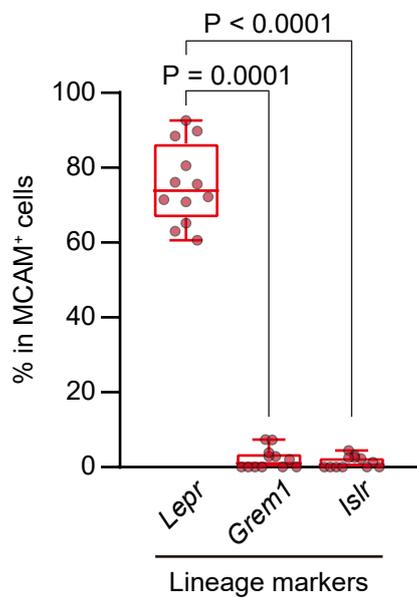
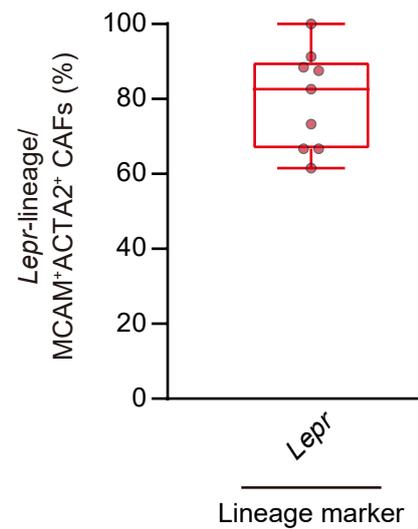
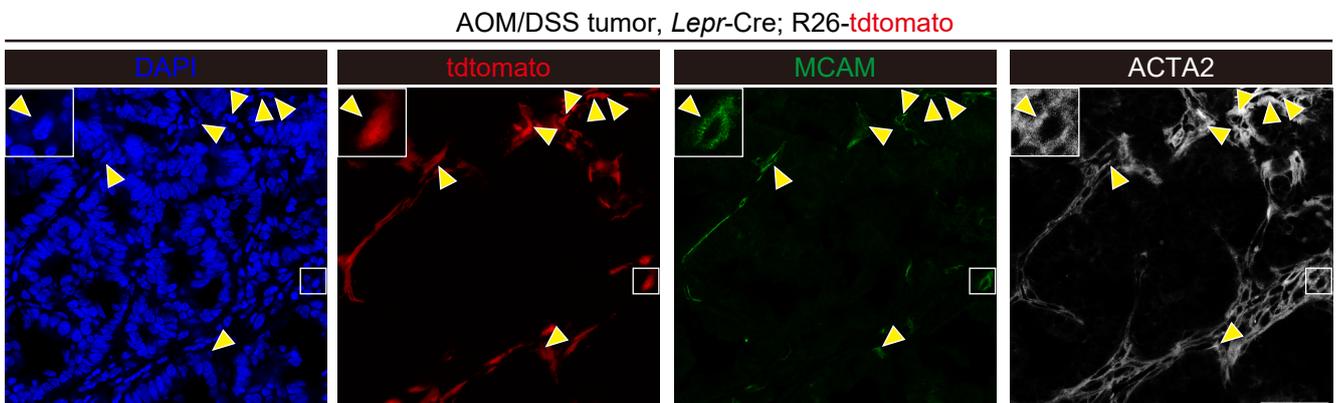
**(B)** Violin plots showing *MCAM* expression in human colonic mesenchymal cells from ulcerative colitis patient samples. ScRNA-seq data from magnetic-activated cell sorting (MACS)-purified colonic mesenchymal cells (EPCAM<sup>-</sup>, CD45<sup>-</sup>, and CD235a<sup>-</sup> cells) from 5 ulcerative colitis patients were analyzed (GSE114374)<sup>290</sup>. n = 243, 142, 2848, 1203, and 384 cells (left to right).

**(C, D)** *MCAM* is expressed by ACTA2<sup>+</sup> CAFs and CD31<sup>+</sup> endothelial cells in AOM/DSS tumors. Co-immunofluorescence for *MCAM* and lineage-marker proteins (ACTA2, CD31, CD45, and EPCAM) was performed. **(C)** Representative images. Yellow, green, and red arrowheads denote double-positive cells, *MCAM* single-positive cells, and lineage-marker single-positive cells, respectively. **(D)** Quantification of *MCAM* positivity in each lineage. 3 HPFs/mouse, 3 mice each. Scale bar, 50  $\mu$ m. Kruskal-Wallis test followed by Dunn's post-hoc multiple comparisons (A and B).

\*\*\*\*, P < 0.0001

Solid black lines, median; Dotted black lines, quartiles (A and B).

Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles (D).

**A****B****C****D**

**Supplementary Figure 13: *Lepr*-lineage intestinal stromal cells, but not *Grem1*-lineage cells or *Islr*-lineage cells, are a major contributor to MCAM<sup>+</sup> CAFs in AOM/DSS tumors.**

**(A, B)** Immunofluorescence for MCAM was performed in AOM/DSS tumors from *Lepr*-Cre; Rosa26-tdtomato mice, *Grem1*-CreERT2; Rosa26-ZsGreen mice, and *Islr*-CreERT2; Rosa26-tdtomato mice.

**(A)** Representative pictures. Yellow arrowheads indicate *Lepr*-lineage MCAM<sup>+</sup> cells. Lineage markers (tdtomato or ZsGreen) are shown in red pseudocolor. **(B)** Quantification of the lineage-marker (tdtomato or ZsGreen)<sup>+</sup> cells in total MCAM<sup>+</sup> cells. 4 HPFs/mouse, 3 mice each.

**(C, D)** Co-immunofluorescence for MCAM and ACTA2 in AOM/DSS tumors from *Lepr*-Cre; Rosa26-tdtomato mice. **(C)** Quantification of the percentage of *Lepr*-lineage cells in MCAM<sup>+</sup>ACTA2<sup>+</sup> CAFs.

**(D)** Representative separate channel images. A merged picture is shown in Figure 4E. Yellow arrowheads denote *Lepr*-lineage MCAM<sup>+</sup> ACTA2<sup>+</sup> CAFs. A boxed area is magnified in the inset.

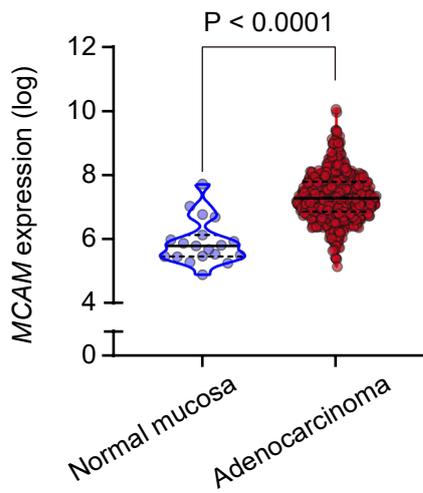
Scale bars, 50  $\mu$ m.

Kruskal-Wallis test followed by Dunn's multiple comparisons test (B).

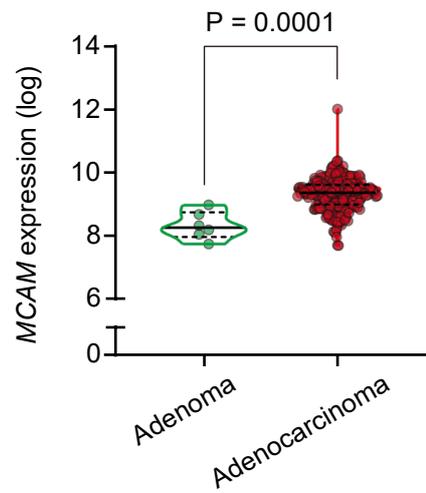
Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.

**A**

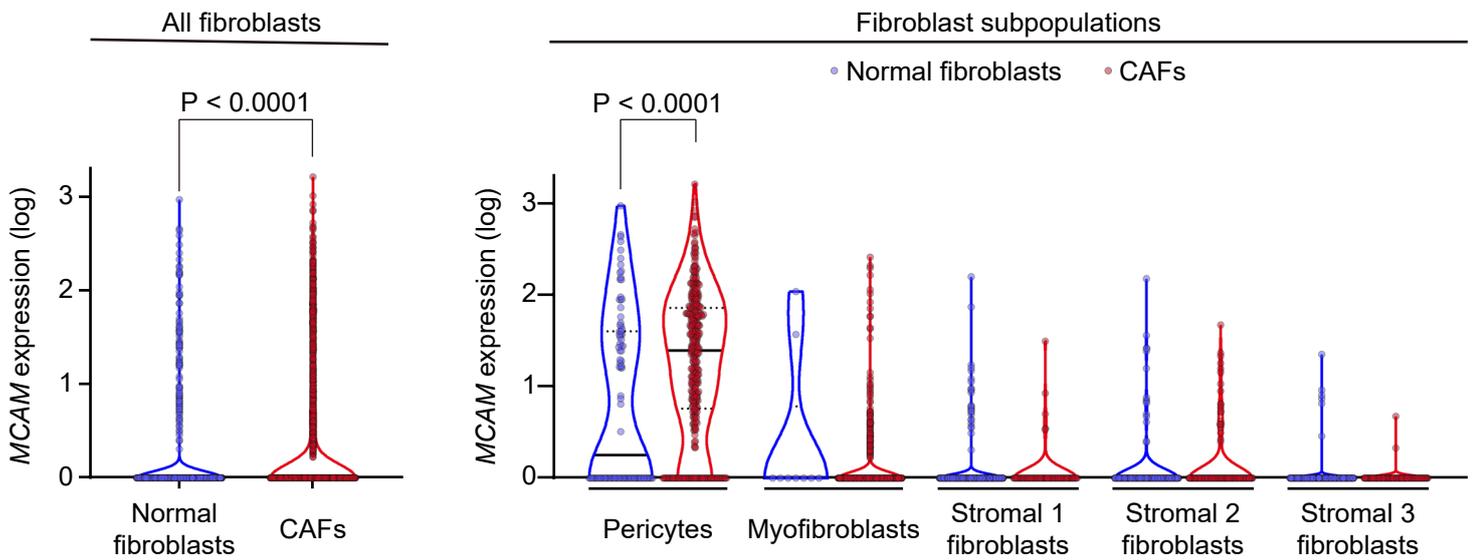
Colorectal tissues,  
Expression microarray data (GSE39582)

**B**

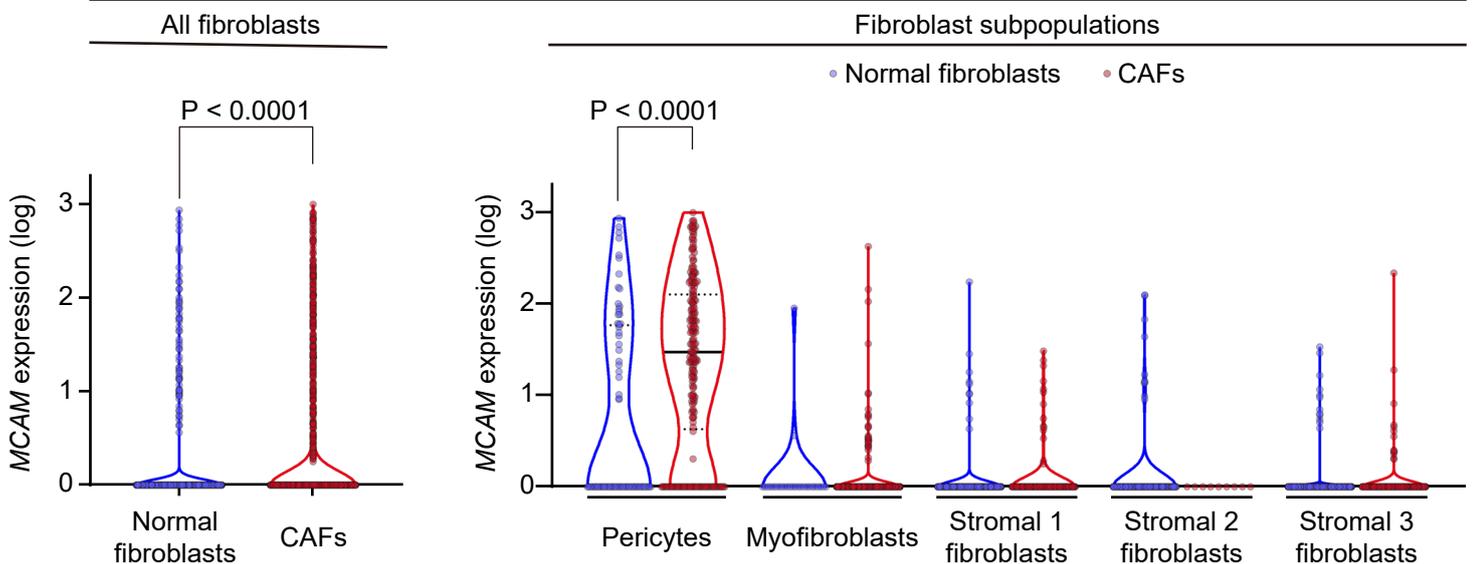
Colorectal tissues,  
Expression microarray data (GSE17538)

**C**

ScRNA-seq from human colorectal tissues (GSE132465)

**D**

ScRNA-seq from human colorectal tissues (GSE144735)



**Supplementary Figure 14: *MCAM* expression is increased in fibroblasts, especially in pericytes, during human CRC carcinogenesis.**

**(A)** Violin plots depicting *MCAM* expression in cDNA expression microarray data from human normal colorectal mucosa and primary colorectal adenocarcinoma (GSE39582)<sup>251</sup>. n = 19 (normal colorectal mucosa) and 566 patients (primary colorectal adenocarcinoma).

**(B)** Violin plots show *MCAM* expression in cDNA expression microarray data from human colorectal adenoma and primary colorectal adenocarcinoma (GSE17538)<sup>252</sup>. n = 6 (colorectal adenoma) and 232 patients (primary colorectal adenocarcinoma).

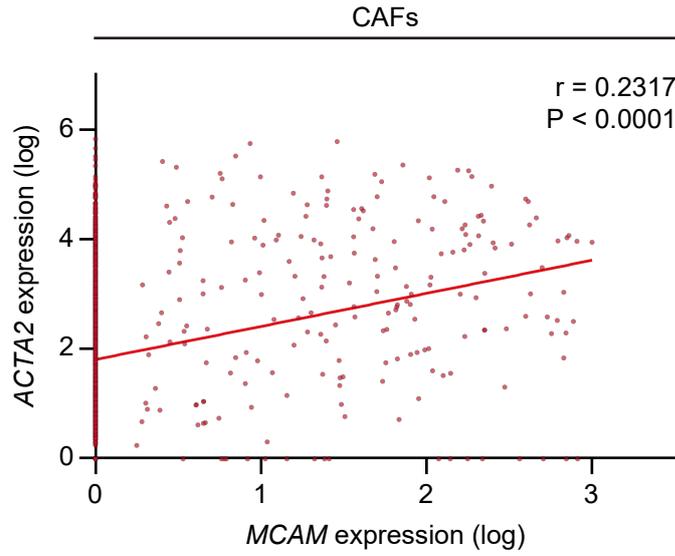
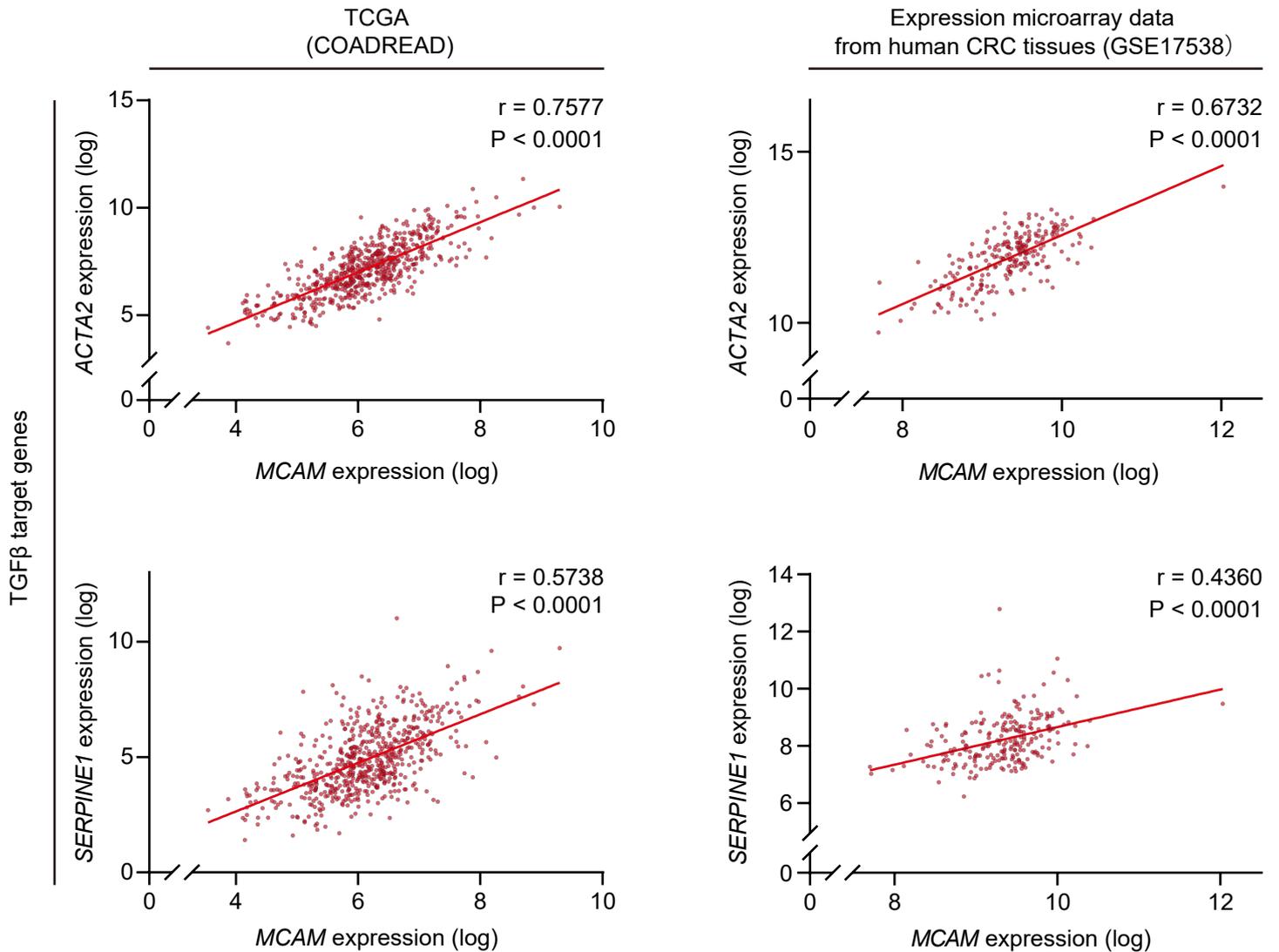
**(C, D)** Violin plots depict *MCAM* expression levels in fibroblasts from human normal colorectal mucosa and CRC tissues. ScRNA-seq data from GSE132465 **(C)** and GSE144735 **(D)** were analyzed<sup>226</sup>. In the scRNA-seq datasets, fibroblasts are defined as pericytes, myofibroblasts, and Stromal 1-3 fibroblasts.

False discovery rate (FDR)-adjusted P-values are shown (A and B). Wilcoxon rank-sum test (C and D).

Solid black lines, median; Dotted black lines, quartiles.

**A**

ScRNA-seq data from human CRC tissues (GSE144735)

**B**

**Supplementary Figure 15: *MCAM* transcript levels are positively correlated with *ACTA2* expression in human CRC expression datasets.**

**(A)** Spearman correlation analyses of *MCAM* and *ACTA2* expression levels in scRNA-seq data from human CRC tissues (GSE144735)<sup>226</sup>. n = 1483 CAFs.

Solid line, linear regression

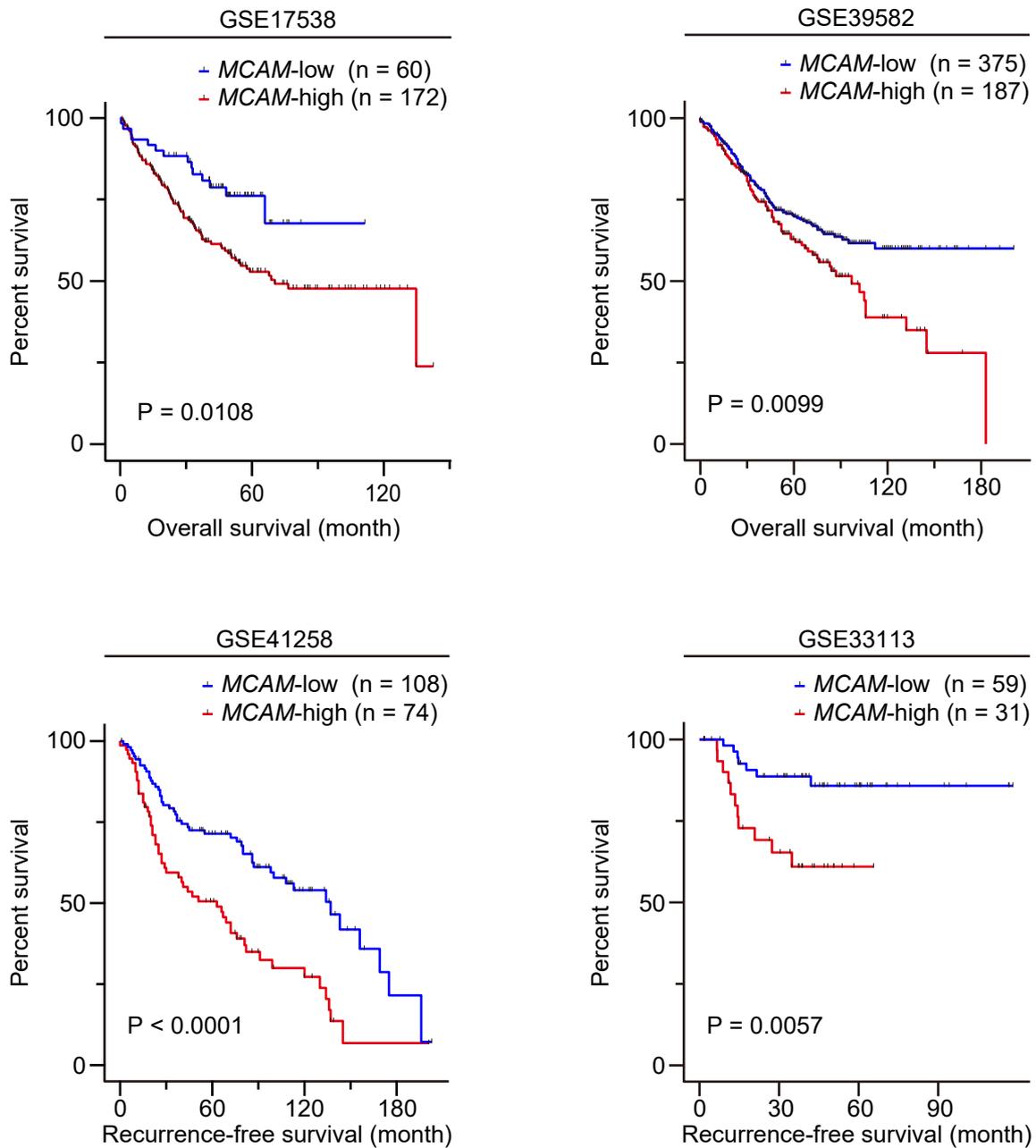
**(B)** *MCAM* expression levels positively correlate with expression levels with TGF- $\beta$  target genes such as *ACTA2* and *SERPINE1*. TCGA (left) and expression microarray data (right)<sup>252</sup> were analyzed.

n = 622 patients (TCGA data) and 232 patients (expression microarray data from GSE17538).

Solid line, linear regression

Spearman correlation (A and B).

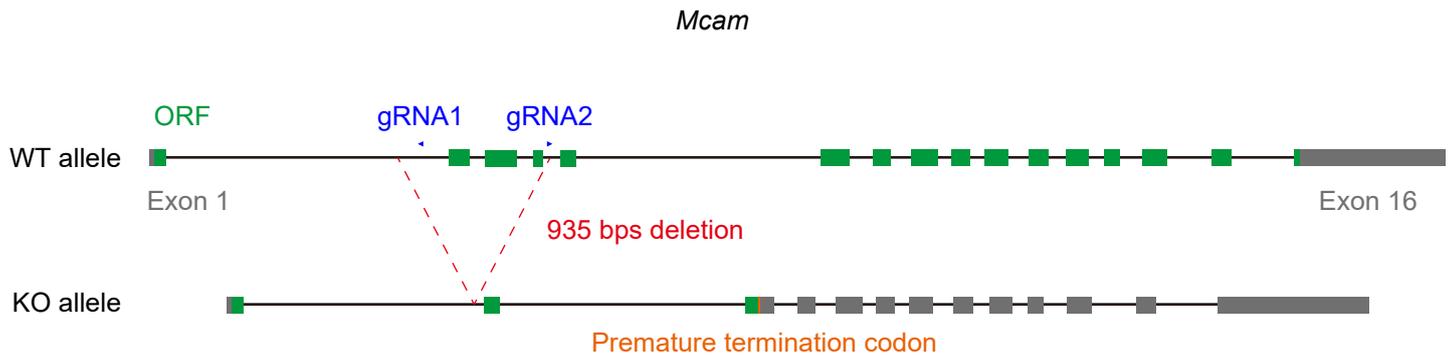
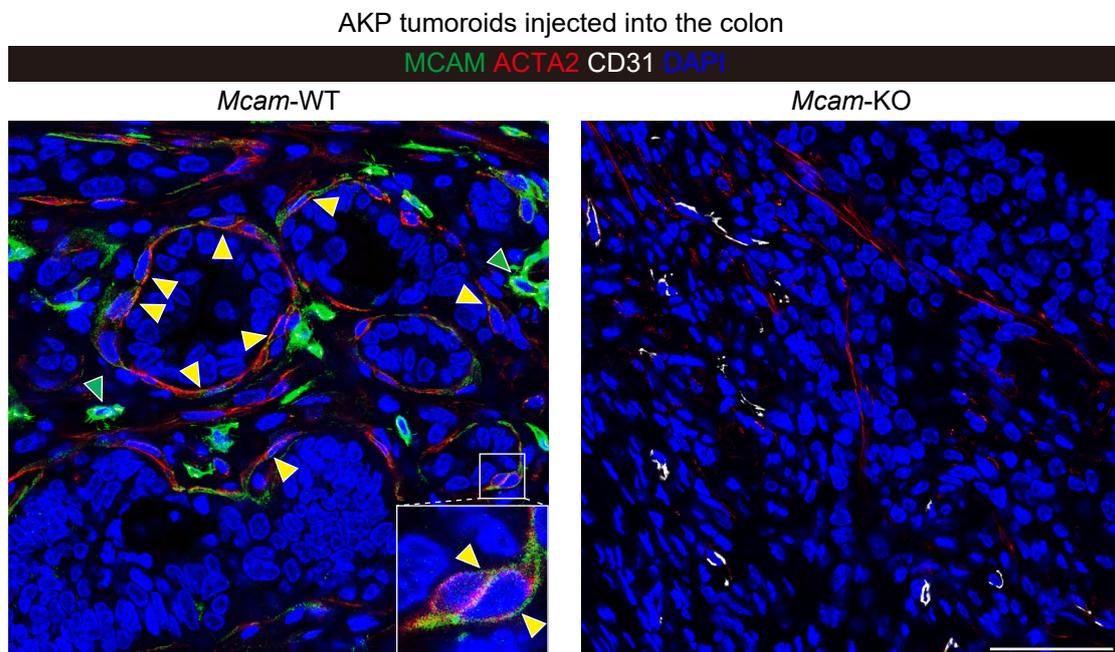
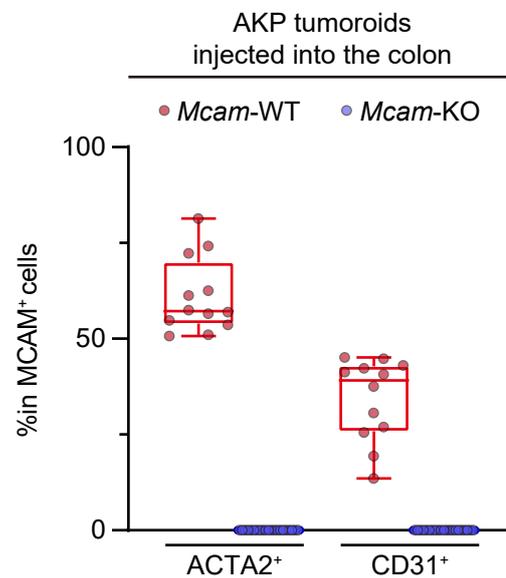
Human colorectal cancer, Expression microarray



(GSE17538, Smith *et al.*, *Gastroenterology*, 2010;  
GSE39582, Marisa *et al.*, *PLoS Med*, 2013;  
GSE41258, Sheffer *et al.*, *Proc Natl Acad Sci USA*, 2009;  
GSE33113, de Sousa *et al.*, *Cell Stem Cell*, 2011)

**Supplementary Figure 16: High *MCAM* expression is associated with poor survival in patients with CRC.**

Kaplan-Meier survival curves showing survival rates of CRC patients classified according to *MCAM* expression levels. Four independent expression microarray datasets from human CRC patients were analyzed<sup>250-252,288</sup>. Log-rank test.

**A****B****C**

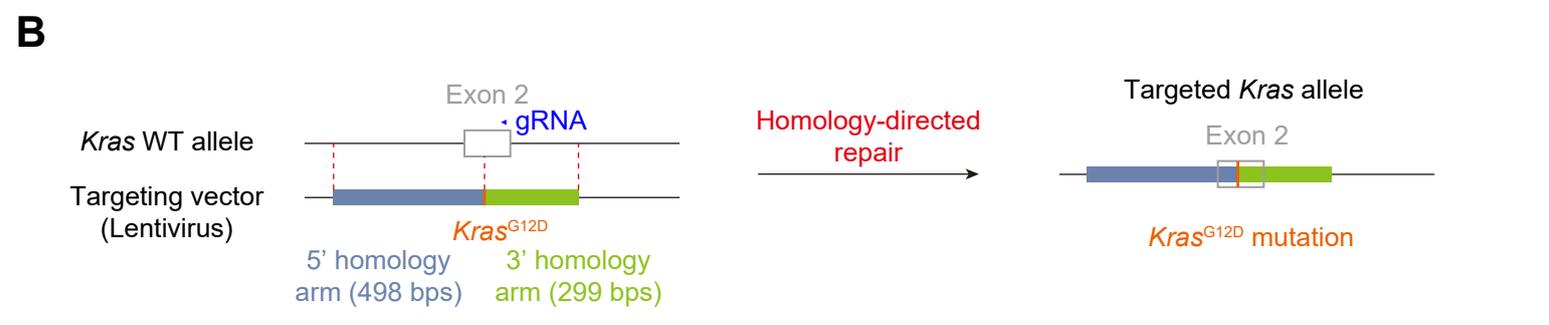
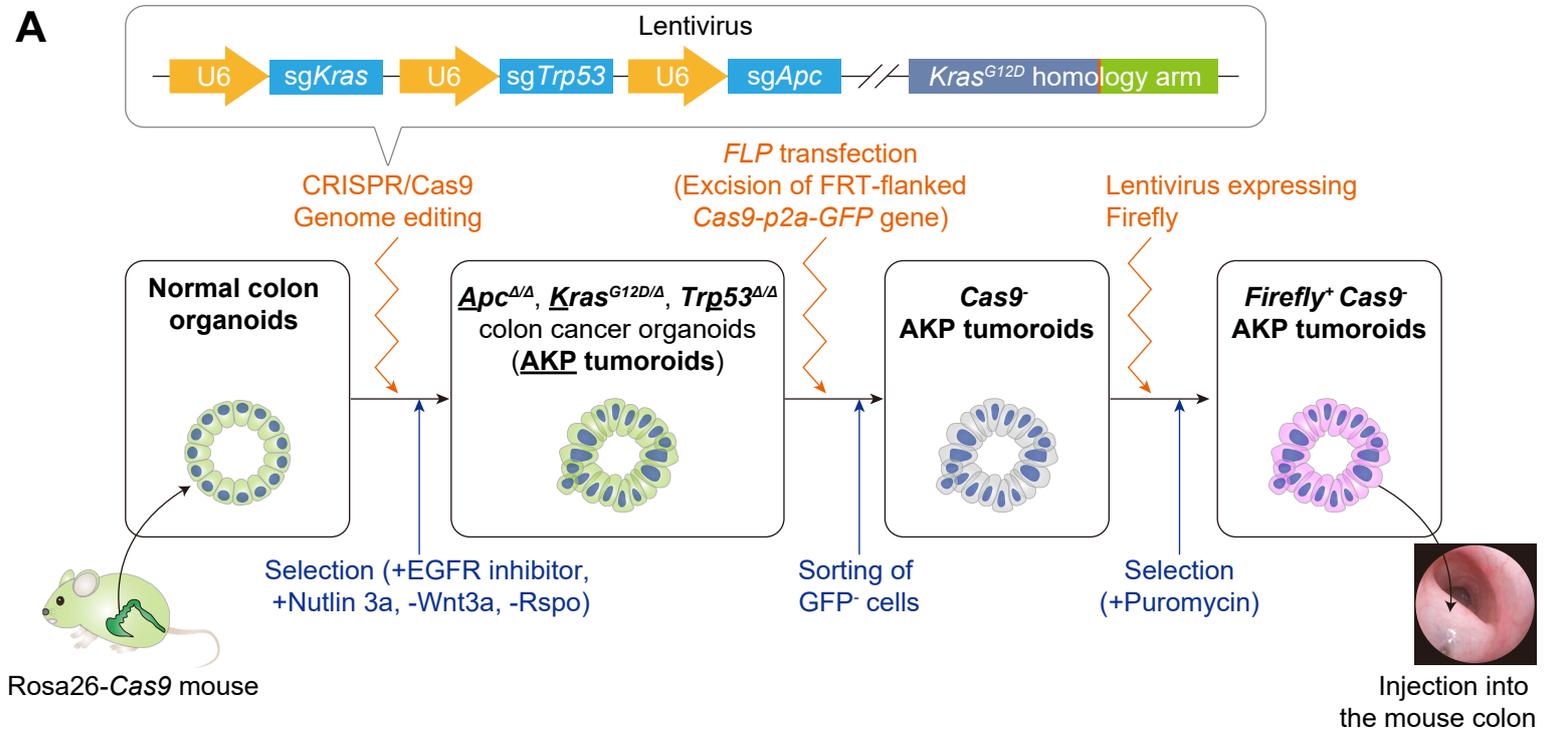
**Supplementary Figure 17: Generation of *Mcam*-knockout mice by CRISPR/Cas9-mediated genome engineering and validation of no MCAM protein expression in colorectal cancer tissues.**

**(A)** A schematic illustration showing CRISPR/Cas9-mediated genome editing of an *Mcam* allele.

A 935 bps large deletion, inclusive of exons 2-4, caused a frameshift mutation, resulting in a premature stop codon in exon 6.

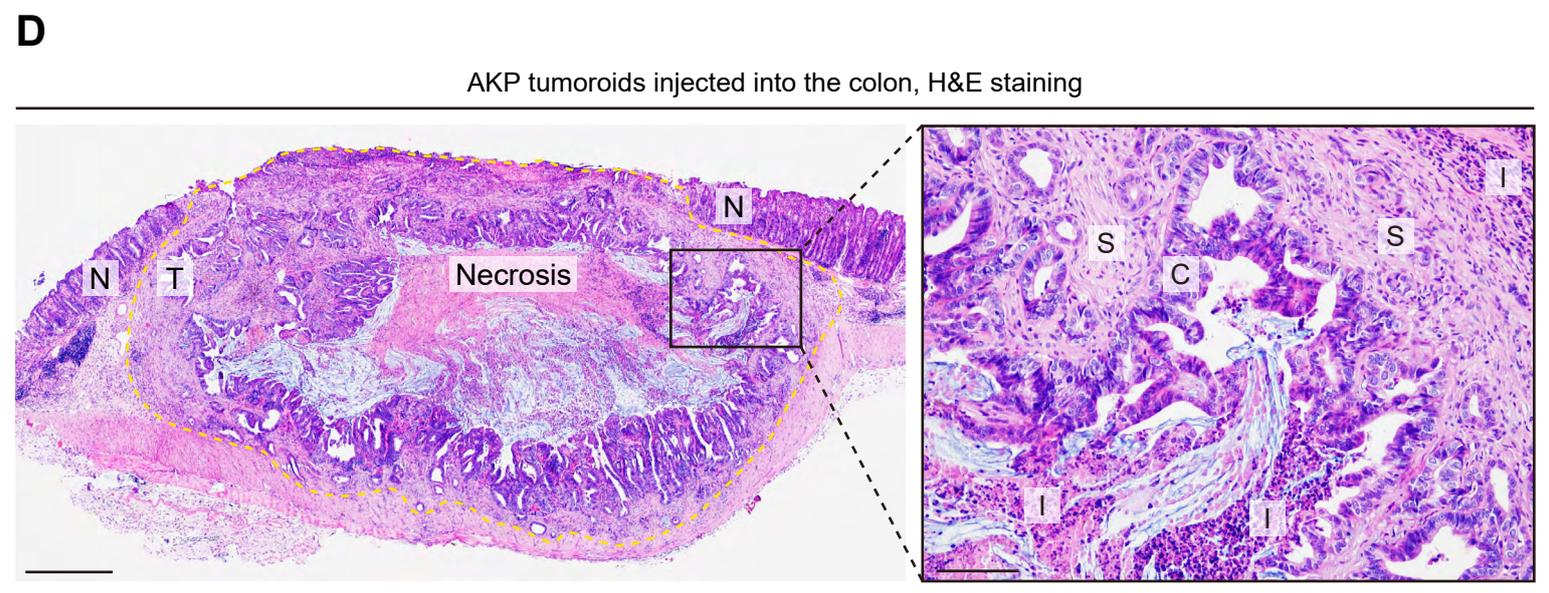
ORF, open reading frame; gRNA, guide RNA; WT, wild type; KO, knockout; CRISPR, clustered regularly interspaced short palindromic repeats.

**(B, C)** Co-immunofluorescence for MCAM, ACTA2, and CD31 using AKP (*Apc*<sup>ΔΔ</sup>, *Kras*<sup>G12D/Δ</sup>, *Trp53*<sup>ΔΔ</sup>) colon tumors from *Mcam*-WT and KO mice. **(B)** Representative pictures. Yellow and green arrowheads denote MCAM<sup>+</sup>ACTA2<sup>+</sup> CAFs and MCAM<sup>+</sup>CD31<sup>+</sup> endothelial cells, respectively. Boxed areas are magnified in the inset. Scale bar, 50 μm. **(C)** Quantification of the ratio of ACTA2<sup>+</sup> CAFs and CD31<sup>+</sup> endothelial cells in MCAM<sup>+</sup> cells. Note that no MCAM protein expression was seen in tumors from *Mcam*-KO mice as assessed by MCAM immunofluorescence. 3 high-power fields (HPFs; 400x)/tumor, 1-2 tumors/mouse, 3-4 mice each group



**C**

|                                      | Allele 1                     |                   | Allele 2       |                    |
|--------------------------------------|------------------------------|-------------------|----------------|--------------------|
|                                      | DNA change                   | Change to Protein | DNA change     | Change to Protein  |
| <i>Apc</i><br>(WT protein, 2860 aa)  | c.2206_2212del               | p.Asp716GlyfsTer3 | c.2209_2216del | p.Met717SerfsTer35 |
| <i>Trp53</i><br>(WT protein, 390 aa) | c.871_876del<br>insTTGGGCTTA | p.Cys239TrpfsTer3 | c.875del       | p.Met240TrpfsTer4  |



**Supplementary Figure 18: Colonic orthotopic injection of *Apc*<sup>Δ/Δ</sup>, *Kras*<sup>G12D/Δ</sup>, *Trp53*<sup>Δ/Δ</sup> CRC organoids (AKP tumoroids) generates desmoplastic stromal reaction accompanied with immune cell infiltration.**

**(A)** Schematic illustration of the CRISPR/Cas9-mediated genome engineering of AKP tumoroids from normal colon organoids isolated from a Rosa26-Cas9 mouse. After *Apc*, *Trp53*, and *Kras* genome editing with lentiviral transduction, correctly targeted clones were enriched by selection medium as indicated. Then, monoclonal lines were handpicked, followed by sequencing *Apc*, *p53*, and *Kras* alleles. To exclude a potential effect of Cas9 expression on tumor growth<sup>291</sup>, an FRT-flanked Cas9-p2a-GFP cassette in the Rosa26 locus was excised by flippase (FLP) overexpression using transient transfection of an FLP plasmid. Finally, lentivirus encoding Firefly was transduced to the Cas9- AKP tumoroids to enable monitoring of tumor cell growth. sg, single guide RNA.

**(B)** Strategy to introduce *Kras*<sup>G12D</sup> mutation into normal mouse colonic organoids. Note that the following mutations were observed in the other *Kras* allele. DNA changes, c.290T>C and c.300\_304del; Protein change, p.Ile36GlyfsTer10. WT, wild type; gRNA, guide RNA.

**(C)** *Apc* and *Trp53* sequence verification. *Apc* and *Trp53* biallelic insertion(ins)/deletion(del) mutations result in prematurely truncated proteins.

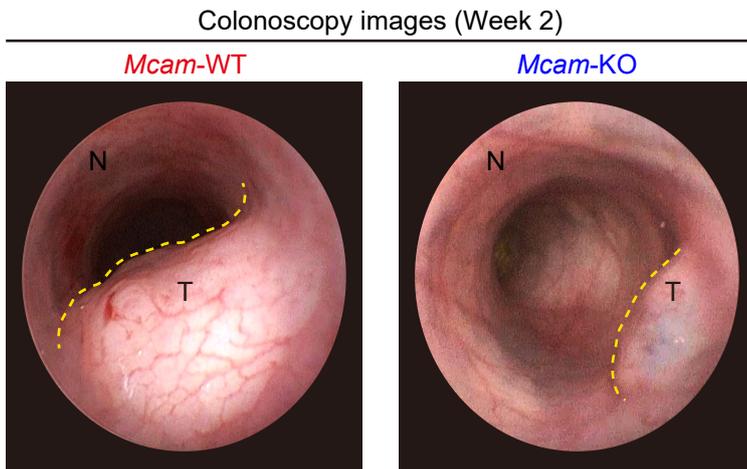
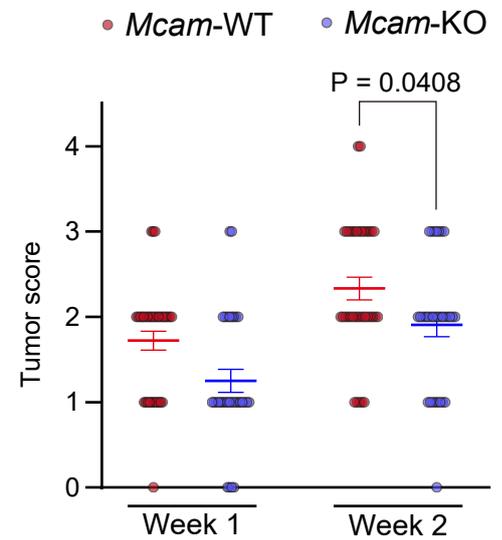
aa, amino acids; fs, frameshift mutation; Ter, translation termination (stop) codon

**(D)** Representative picture showing hematoxylin and eosin (H&E)-stained AKP tumoroids injected into the mouse colon. Histologically, the colon tumors showed prominent stromal reaction and immune cell infiltration, accompanied with necrosis. Yellow dotted line denotes the border between normal mucosa (N) and tumor areas (T). S, Stroma; I, immune cells; C; Cancer cells.

Scale bar, 500 μm (left) and 100 μm (right).

Changes to proteins are reported according to the Human Genome Variation Society nomenclature.

Reference sequences are as follows. *Kras*, NM\_021284.6; *Apc*, NM\_001360980.1; *Trp53*, NM\_011640.3.

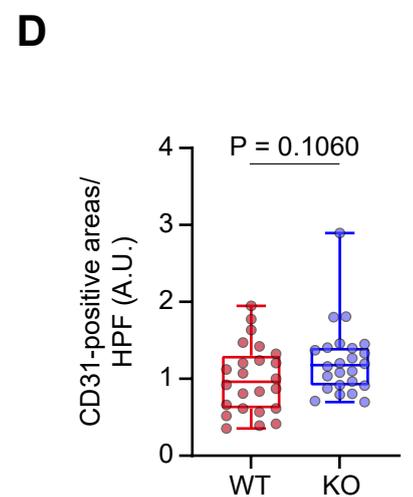
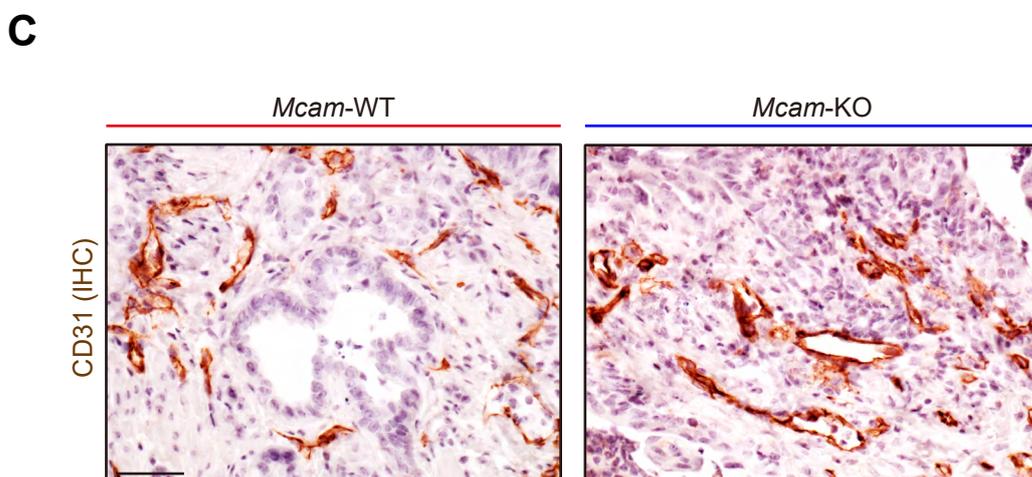
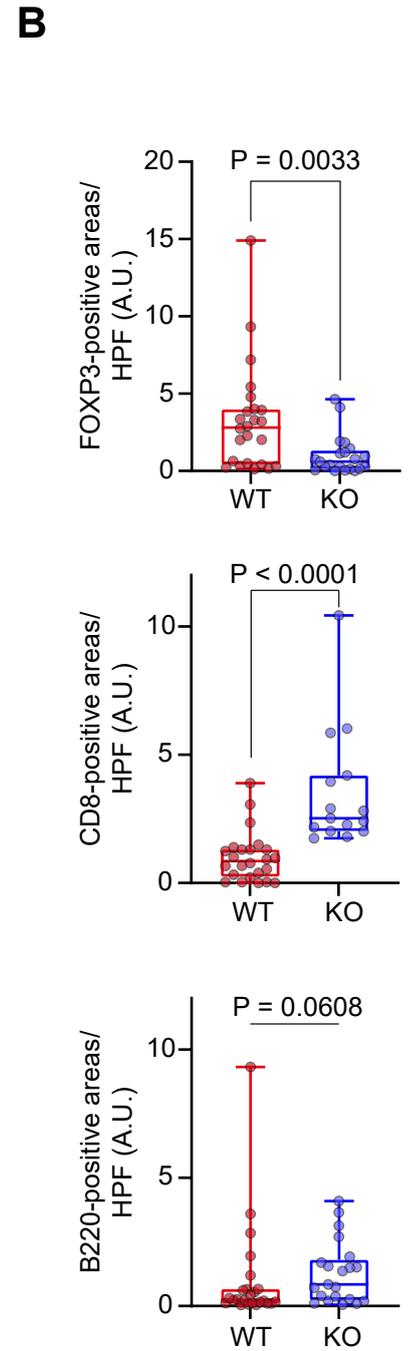
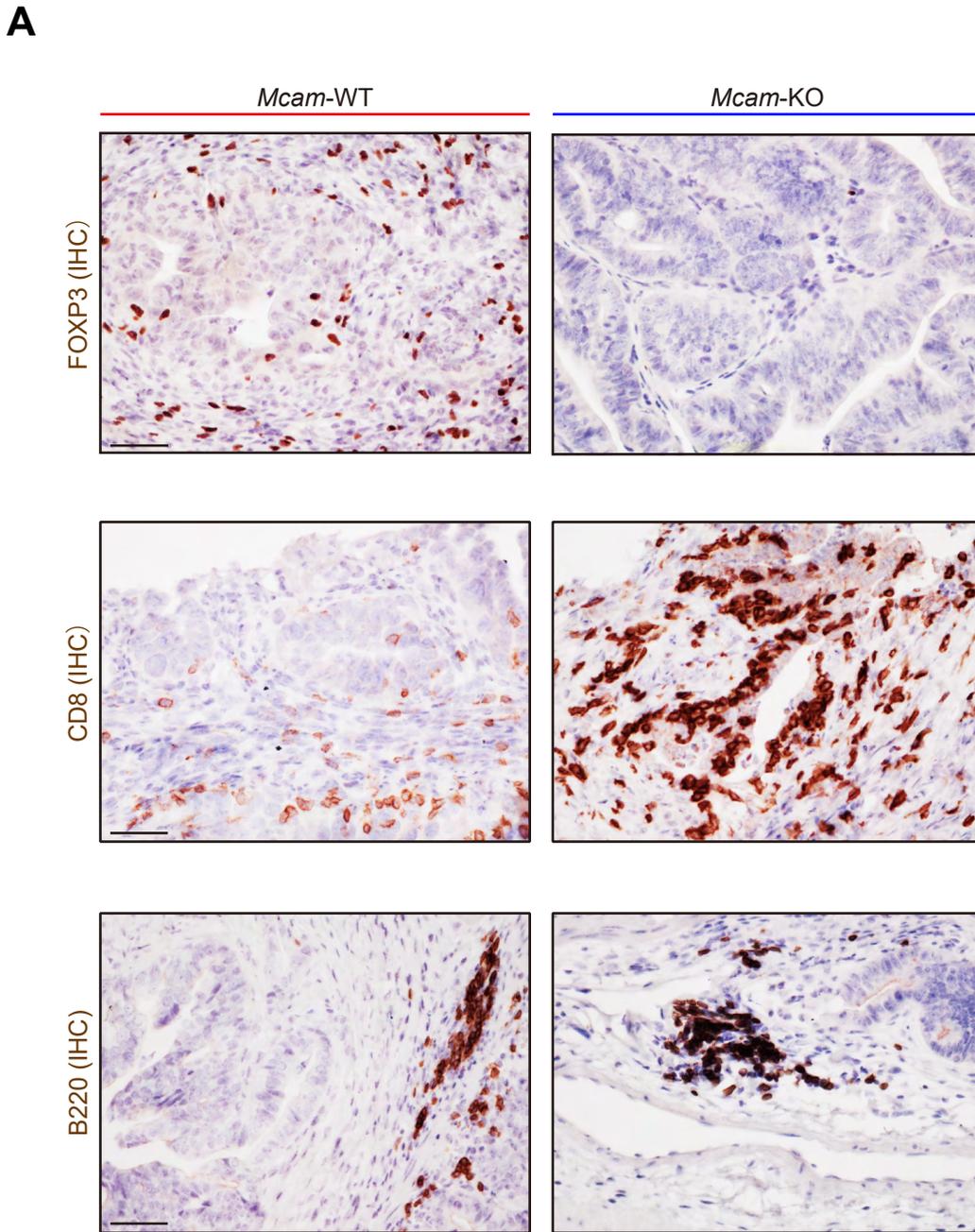
**A****B**

**Supplementary Figure 19: AKP colon tumors from *Mcam*-knockout mice show decreased tumor scores as assessed by colonoscopy.**

**(A, B)** Colonoscopic evaluation of orthotopically injected tumors. **(A)** Representative colonoscopy images of tumors in *Mcam*-WT and KO mice. Yellow dotted lines show boundaries between the normal mucosa (N) and tumors (T). **(B)** Colonoscopic tumor scoring. 36 injections into 18 *Mcam*-WT mice and 32 injections into 16 *Mcam*-KO mice (two injections per mouse).

Tumor scoring was performed as described<sup>292</sup>.

Mean  $\pm$  s.e.m. Mann-Whitney U-test



**Supplementary Figure 20: AKP tumors from *Mcam*-knockout mice exhibit decreased FOXP3<sup>+</sup> regulatory T cells and increased CD8<sup>+</sup> cytotoxic T cells.**

**(A, B)** Immunohistochemistry (IHC) for FOXP3 (regulatory T cell marker), CD8 (cytotoxic T cell marker) and B220 (B cell marker) using AKP tumor from *Mcam*-WT and KO mice. **(A)** Representative pictures. **(B)** Quantification of 3,3'-Diaminobenzidine (DAB)-positive areas using ImageJ. 3 high-power fields (HPFs; 400x)/tumor, 1-2 tumors/mouse, 5 mice each group. A.U., arbitrary unit.

**(C, D)** IHC for CD31 using AKP tumors from *Mcam*-WT and KO mice. **(C)** Representative pictures. **(D)** Quantification of DAB-positive areas using ImageJ. 3 HPFs/tumor, 1-2 tumors/mouse, 5 mice each group.

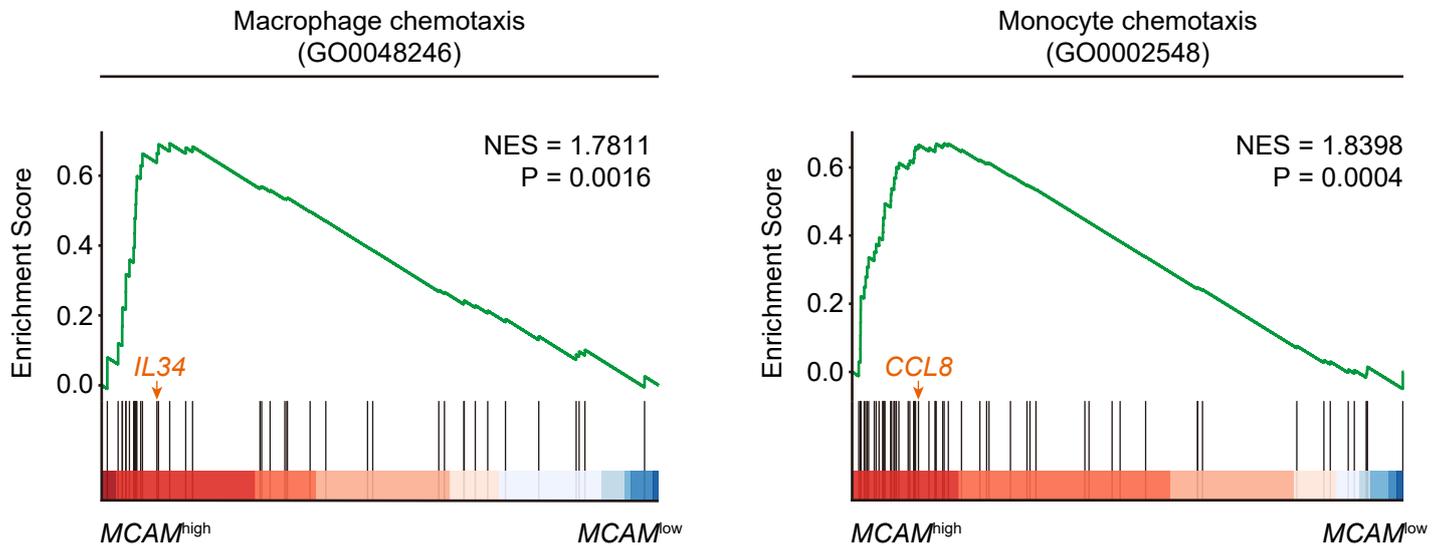
Note that, in Figure 6 and Supplementary Figure 20, all immunohistological analyses were performed using mice harvested three weeks after tumor injection.

Scaler bars, 50  $\mu$ m.

Mann-Whitney U-test (B and D).

Box plots have whiskers of maximum and minimum values; the boxes represent first, second (median), and third quartiles.

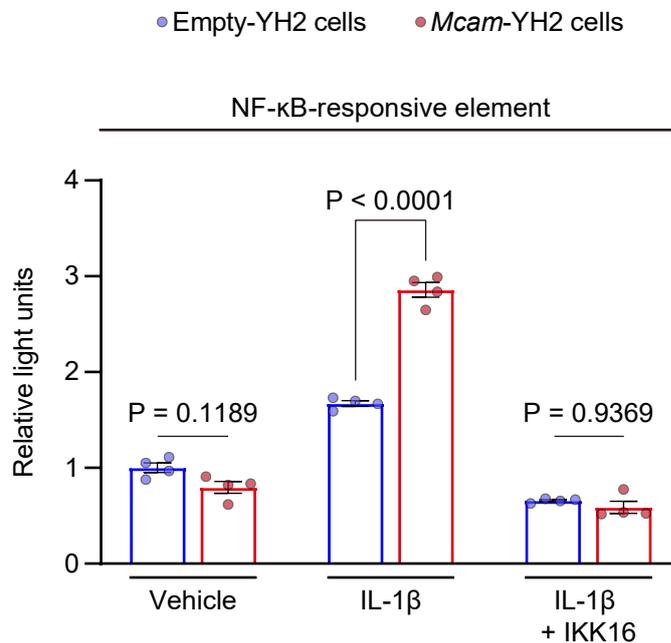
TCGA (COAD and READ)  
Gene set enrichment analysis (GSEA)



**Supplementary Figure 21: Gene set enrichment analyses (GSEA) using TCGA data reveal that genes involved with macrophage and monocyte chemotaxis are enriched in  $MCAM^{high}$  cancers compared with  $MCAM^{low}$  tumors.**

Gene set enrichment analyses for macrophage and monocyte chemotaxis between  $MCAM^{high}$  and  $MCAM^{low}$  colorectal cancers. For differential gene expression analyses, the median  $MCAM$  expression value was used as a cutoff value.

NES, normalized enrichment score. Benjamini and Hochberg adjusted p-values are shown. Orange arrows denote *IL34* (left) and *CCL8* (right) genes.

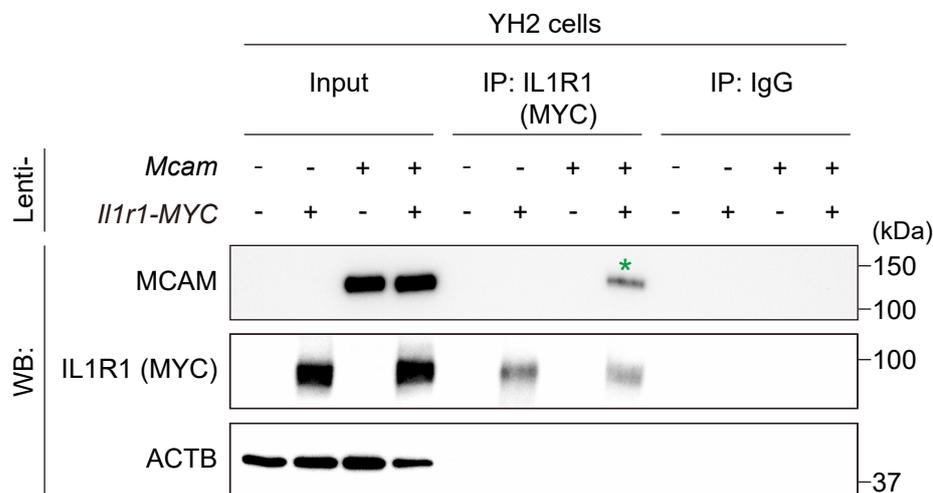


**Supplementary Figure 22: IL-1β-treated MCAM-overexpressing YH2 cells show increased NF-κB signaling as assessed by luciferase assays of NF-κB-responsive elements.**

Luciferase assays of nuclear factor-κB (NF-κB)-responsive elements.

Empty or MCAM-overexpressing YH2 cells were transduced with lentivirus encoding NF-κB-responsive elements-NanoLuc-Ubiquitin promoter-Firefly. These cells were treated with vehicle, IL-1β, or IL-1β + IKK16, followed by dual-luciferase assays. NanoLuc signals were normalized to Firefly luminescence values.

n = 4 each. Mean ± s.e.m. Two-way ANOVA followed by Tukey's post-hoc multiple comparison test

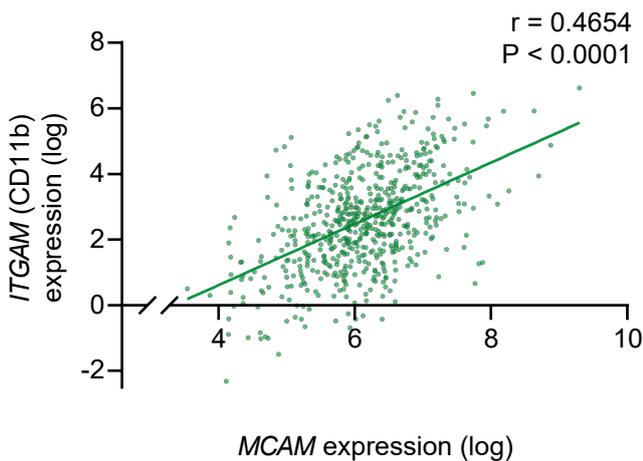
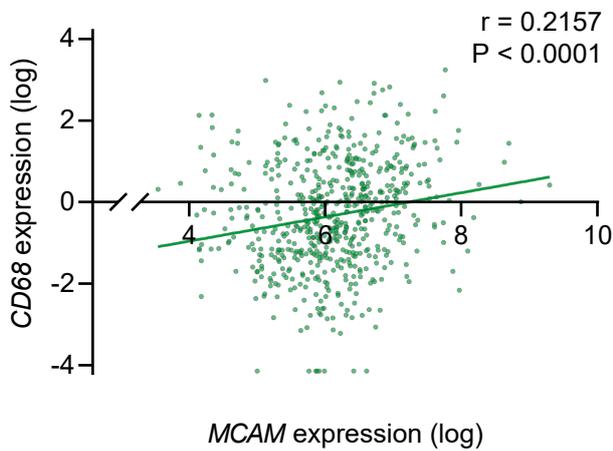
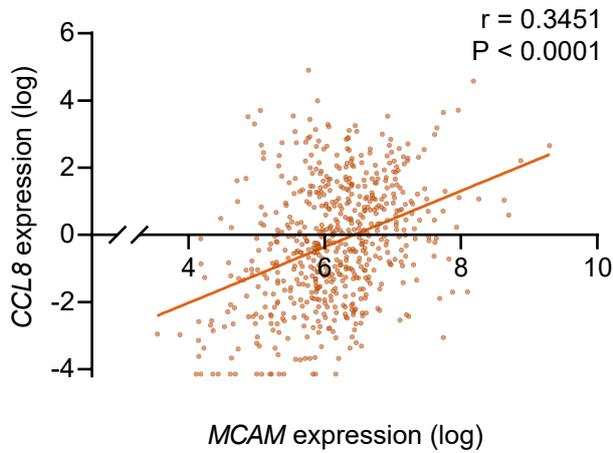
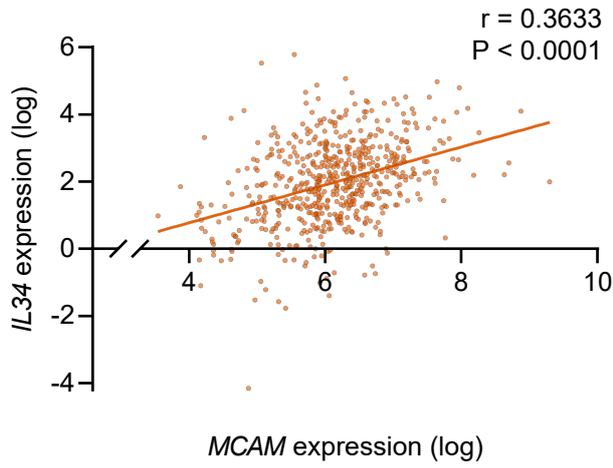


**Supplementary Figure 23: Co-immunoprecipitation of IL1R1 shows that IL1R1 interacts with MCAM in mouse colonic fibroblasts.**

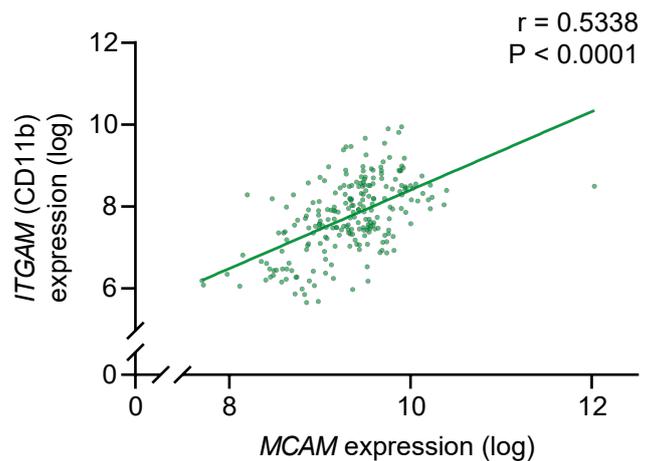
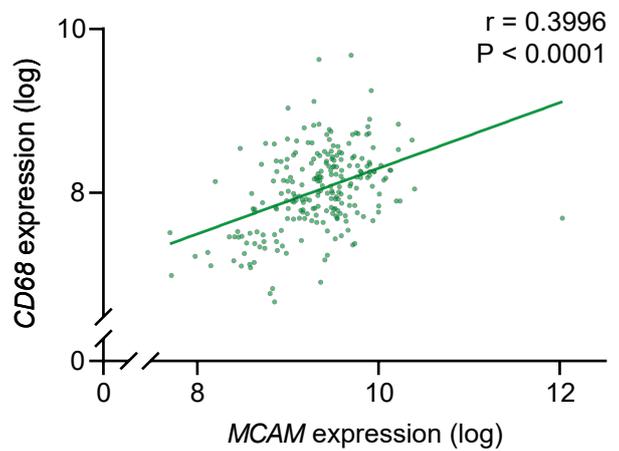
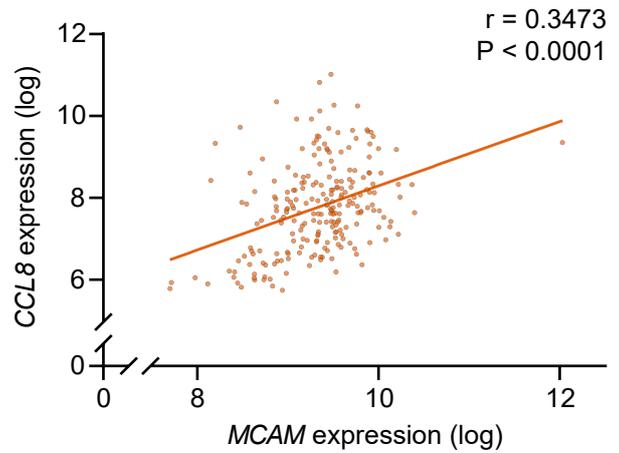
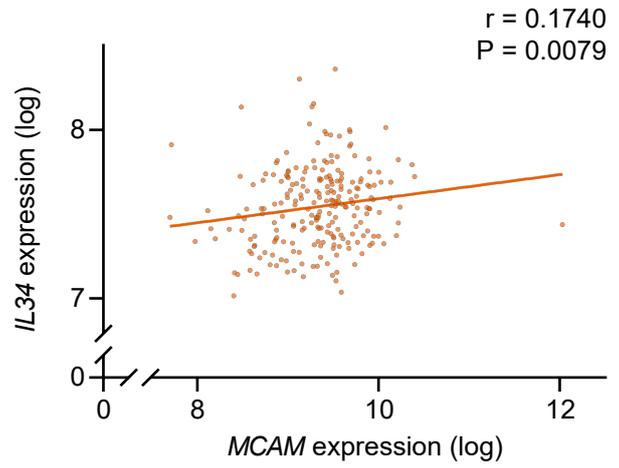
Interleukin-1 receptor 1 (IL1R1) tagged with a MYC epitope was immunoprecipitated using an anti-MYC antibody, followed by Western blotting (WB). *Il1r1-Myc* and/or *Mcam* was overexpressed in YH2 cells using lentiviral transduction. Green asterisk denotes the interaction of IL1R1 with MCAM. Immunoprecipitation with an IgG isotype antibody was performed as a negative control. In the western blotting, an anti-MYC antibody was used to detect IL1R1 protein tagged with MYC. IP, immunoprecipitation.

Macrophage/Monocyte chemoattractants

TCGA  
(COADREAD)



Expression microarray data  
from human CRC tissues (GSE17538)



Macrophage and myeloid cell markers

**Supplementary Figure 24: Analyses of human CRC datasets demonstrate that *MCAM* expression levels are positively correlated with *IL34* and *CCL8* as well as *CD68* and *ITGAM* expression.**

Spearman correlation analyses in TCGA data and expression microarray data from human CRC tissues.

n = 622 patients (TCGA data) and 232 patients (expression microarray data from GSE17538) <sup>252</sup>.

Solid line, linear regression

## Supplementary Table 1

List of 46 genes upregulated in both tumorigenesis and development,  
which are associated with human CRC survival

|               |                  |                 |                |               |
|---------------|------------------|-----------------|----------------|---------------|
| <i>ABCC9</i>  | <i>COX4I2</i>    | <i>MEOX2</i>    | <i>PCDHB7</i>  | <i>TLX2</i>   |
| <i>AGER</i>   | <i>FJX1</i>      | <i>MPP2</i>     | <i>PLCG2</i>   | <i>TMEM40</i> |
| <i>ARL4C</i>  | <i>FOXS1</i>     | <i>NAT8L</i>    | <i>PLCH2</i>   | <i>TNNT2</i>  |
| <i>AVPI1</i>  | <i>FUT1</i>      | <i>NDUFA4L2</i> | <i>RHOV</i>    | <i>TREX2</i>  |
| <i>BNIPL</i>  | <i>GJA4</i>      | <i>NGFR</i>     | <i>SEMA5B</i>  | <i>TUBA8</i>  |
| <i>CCBE1</i>  | <i>GSTO1</i>     | <i>NOTCH3</i>   | <i>SHC4</i>    | <i>UPK3BL</i> |
| <i>CDH4</i>   | <i>HIST1H2AE</i> | <i>NRXN1</i>    | <i>SIX4</i>    |               |
| <i>CDKN2A</i> | <i>HIST1H3D</i>  | <i>NXN</i>      | <i>SLC6A17</i> |               |
| <i>CELF4</i>  | <i>MACC1</i>     | <i>OLFM2</i>    | <i>SOX2</i>    |               |
| <i>CHST3</i>  | <i>MCAM</i>      | <i>PCDHB5</i>   | <i>TGFA</i>    |               |

The list above represents 46 genes selected in **Figure 4B (2)**.

The 18 genes that are upregulated in fibroblasts compared to epithelial cells,  
in both normal adult colon and AOM/DSS tumor, are highlighted in red (**Figure 4B (3)**).

## Supplementary Table 2

Clinicopathological characteristics of colorectal cancer patients analysed by MCAM immunohistochemistry

|   | All<br>(n = 101) | MCAM low<br>(n = 28) | MCAM high<br>(n = 73) | P value<br>(Fisher's exact test) |
|---|------------------|----------------------|-----------------------|----------------------------------|
| Median age at diagnosis<br>(range), years     | 72 (39-82)       | 67.5 (39-79)         | 74 (40-82)            |                                  |
| Median survival months<br>(range)             | 106 (2-202)      | 142 (6-197)          | 91 (2-202)            |                                  |
| Median primary tumour<br>diameter (range), mm | 45 (15-110)      | 45 (15-110)          | 43 (15-100)           |                                  |
| Age, n (%)                                    |                  |                      |                       | 0.091                            |
| ≤ 65  |                  | 12 (42.9%)           | 18 (24.7%)            |                                  |
| > 65  |                  | 16 (57.1%)           | 55 (75.3%)            |                                  |
| Gender, n (%)                                 |                  |                      |                       | 0.508                            |
| Male  |                  | 13 (46.4%)           | 40 (54.8%)            |                                  |
| Female  |                  | 15 (53.6%)           | 33 (45.2%)            |                                  |
| T status, n (%)                               |                  |                      |                       | 0.480                            |
| T3  |                  | 27 (96.4%)           | 72 (98.6%)            |                                  |
| T4  |                  | 1 (3.6%)             | 1 (1.4%)              |                                  |
| N status, n (%)                               |                  |                      |                       | 1.000                            |
| N0  |                  | 15 (53.6%)           | 40 (54.8%)            |                                  |
| N1  |                  | 9 (32.1%)            | 22 (30.1%)            |                                  |
| N2  |                  | 4 (14.3%)            | 11 (15.1%)            |                                  |
| M status, n (%)                               |                  |                      |                       | <b>0.032</b>                     |
| M0  |                  | 28 (100.0%)          | 62 (84.9%)            |                                  |
| M1  |                  | 0 (0.0%)             | 11 (15.1%)            |                                  |
| TNM stage, n (%)                              |                  |                      |                       | 0.051                            |
| Stage II                                      |                  | 15 (53.6%)           | 38 (52.1%)            |                                  |
| Stage III                                     |                  | 13 (46.4%)           | 24 (32.9%)            |                                  |
| Stage IV                                      |                  | 0 (0.0%)             | 11 (15.1%)            |                                  |
| Tumour localization, n (%)                    |                  |                      |                       | 0.647                            |
| Left  |                  | 19 (67.9%)           | 44 (60.3%)            |                                  |
| Right   |                  | 9 (32.1%)            | 29 (39.7%)            |                                  |
| Differentiation, n (%)                        |                  |                      |                       | 0.251                            |
| Moderate                                      |                  | 15 (53.6%)           | 49 (67.1%)            |                                  |
| Poor  |                  | 13 (46.4%)           | 24 (32.9%)            |                                  |
| Treatment, n (%)                              |                  |                      |                       | 0.699                            |
| No  |                  | 21 (75.0%)           | 56 (76.7%)            |                                  |
| Chemotherapy                                  |                  | 6 (21.4%)            | 16 (21.9%)            |                                  |
| Radiotherapy                                  |                  | 1 (3.6%)             | 1 (1.4%)              |                                  |

P value less than 0.05 is shown in bold characters.

### Supplementary Table 3

Univariate and multivariate Cox regression analysis of MCAM protein expression levels and overall survival of colorectal cancer patients

| Variable                           | Overall survival    |              |                       |              |
|------------------------------------|---------------------|--------------|-----------------------|--------------|
|                                    | Univariate analysis |              | Multivariate analysis |              |
|                                    | HR (95% CI)         | P Value      | HR (95% CI)           | P Value      |
| MCAM                               |                     | <b>0.008</b> |                       | <b>0.041</b> |
| Low                                | 1.000               |              | 1.000                 |              |
| High                               | 2.341 (1.252-4.377) |              | 1.974 (1.030-3.786)   |              |
| Age                                |                     | <b>0.007</b> |                       | <b>0.018</b> |
| ≤ 65                               | 1.000               |              | 1.000                 |              |
| > 65                               | 2.294 (1.251-4.206) |              | 2.105 (1.137-3.896)   |              |
| Gender                             |                     | 0.205        |                       |              |
| Female                             | 1.000               |              |                       |              |
| Male                               | 1.363 (0.844-2.201) |              |                       |              |
| N status                           |                     | 0.260        |                       | 0.410        |
| N0                                 | 1.000               |              | 1.000                 |              |
| N1 and N2                          | 1.316 (0.817-2.119) |              | 1.253 (0.733-2.141)   |              |
| M status                           |                     | <b>0.002</b> |                       | <b>0.031</b> |
| M0                                 | 1.000               |              | 1.000                 |              |
| M1                                 | 3.022 (1.488-6.138) |              | 2.393 (1.082-5.289)   |              |
| Tumour localization                |                     | <b>0.046</b> |                       | 0.101        |
| Left                               | 1.000               |              | 1.000                 |              |
| Right                              | 1.631 (1.008-2.641) |              | 1.514 (0.922-2.487)   |              |
| Differentiation                    |                     | 0.825        |                       |              |
| Moderate                           | 1.000               |              |                       |              |
| Poor                               | 1.057 (0.646-1.731) |              |                       |              |
| Treatment                          |                     | 0.716        |                       |              |
| No                                 | 1.000               |              |                       |              |
| Yes (Chemotherapy or Radiotherapy) | 0.901 (0.514-1.579) |              |                       |              |

P values less than 0.05 are shown in bold characters.

## Supplementary Materials and Methods

### Data and materials availability

Data that support the findings of this study are available from the corresponding authors upon reasonable request. Noncommercially available materials described in this study may be obtained with a material transfer agreement (MTA). Requests for materials should be addressed to Susan L. Woods ([susan.woods@adelaide.edu.au](mailto:susan.woods@adelaide.edu.au)).

### Human and animal Ethics

Human colorectal samples were obtained at the time of surgery for routine pathological analysis. This study was approved by the Southern Adelaide Clinical Human Research Ethics Committee (HREC/19/SAC/70, OFR62.19). All animal protocols were approved by the Animal Ethics committees of SAHMRI (SAM205, 236, 303, 387.19), University of Oxford (P0B63BC4D), University of Western Ontario (2018-180), and Nagoya University Graduate School of Medicine (31434).

### Antibodies

The following antibodies were used in this study. Rabbit monoclonal anti-MCAM (1:500 for immunohistochemistry (IHC) and immunofluorescence (IF), ab75769, Abcam), Rabbit polyclonal anti-RFP (1:500 for IF, 600-401-379, Rockland), Goat polyclonal anti-mouse PDGFRA (1:50 for IF, AF1062, R&D systems), Mouse monoclonal anti- $\alpha$ SMA (1:500 for IHC and IF, clone 1A4, Dako), Rabbit polyclonal anti- $\alpha$ SMA (1:500 for IF, ab5694, Abcam), Rat monoclonal anti-BrdU (1:600 for IF, ab6326, Abcam), Rat monoclonal anti-EPCAM (1:100 for IF, 14-5791-81, eBioscience), Rabbit monoclonal anti-Ki67 (1:200 for IF, ab16667, Abcam), Mouse monoclonal ZsGreen (1:1000 for IF, TA180002, OriGene), Rat monoclonal anti-CD31 (1:50 for IF, DIA-310, Dianova), Rat monoclonal anti-CD45-FITC (1:50 for IF, 103108, BioLegend), Rabbit polyclonal anti-FJX1 (1:800 for IHC, PA5-63687, Invitrogen), Rabbit polyclonal anti-NGFR (1:1500 for IHC, ab8874, Abcam), Rabbit polyclonal anti-CD68 (1:1000 for IHC, ab125212, Abcam), Rabbit monoclonal anti-CD11b (1:1000 for IHC, ab133357, Abcam), Rabbit monoclonal anti-FOXP3 (1:50 for IHC, 12653, CST), Rabbit

monoclonal anti-CD8 (1:2000 for IHC, ab209775, Abcam), Rat monoclonal anti-B220 (1:500 for IHC, clone RA3-6B2, eBioscience), Mouse monoclonal anti-IkBa (1:1000 for WB, 4814, CST), Rabbit monoclonal anti-phosphorylated p65 (1:2000 for WB, 3033, CST), Mouse monoclonal anti-ACTB (1:5000 for WB, sc-47778, Santa Cruz), Rat monoclonal anti-HA epitope tag (1:2500 for WB, clone 3F10, Roche), Rabbit polyclonal anti-Myc epitope tag (1:5000 for WB, ab9106, Abcam), Rabbit polyclonal anti- $\alpha$ -Tubulin (1:2000 for WB, 2144, CST), and Rabbit polyclonal anti-phosphorylated p65 (1:50 for IHC, ab194726, Abcam).

### **Immunohistochemistry (IHC) and immunofluorescence (IF)**

Tissues were either fixed with formalin overnight followed by dehydration with 70% ethanol and paraffin-embedding, or fixed with 4% paraformaldehyde overnight followed by dehydration with 30% sucrose and embedding with optimal cutting temperature (O.C.T.) compound. Formalin-fixed and paraffin-embedded tissue sections were deparaffinized with xylene and rehydrated with phosphate-buffered saline (PBS), followed by antigen retrieval by boiling samples in antigen retrieval buffer (pH 6, H-3300; Vector Laboratories) for 30 min.

For IHC, inactivation of endogenous peroxidase was performed with 0.5% H<sub>2</sub>O<sub>2</sub> in methanol for 15 min, followed by washing with PBS. Then, sections were treated with blocking buffer (X0909, Dako) for 30 min, incubated with the indicated primary antibodies overnight at 4 °C, and washed with PBS. Sections were incubated with horseradish peroxidase (HRP)-polymer secondary antibody (ab214879, ab214880, or ab214882, Abcam) for 30 min, followed by signal detection with diaminobenzidine (DAB) solution (K3468, DAKO).

For IF studies, sections were treated with blocking buffer (X0909, Dako) for 30 min, incubated with the indicated primary antibodies overnight at 4 °C, and washed with PBS. Sections were then incubated with Alexa Fluor 488/594/647-conjugated secondary antibodies (All from Thermo Fisher Scientific) for 1 h at room temperature. The sections were then mounted with ProLong Gold antifade reagent containing 4',6-diamidino-2-phenylindole (DAPI; Thermo Fisher Scientific), and fluorescence was examined using a confocal laser-scanning microscope (TCS SP8 MP, Leica) or an inverse immunofluorescence microscope BZ-X710 (Keyence) with optical sectioning.

## **Combined single-molecule fluorescent *in situ* hybridization (smFISH) and immunofluorescence (IF)**

Combined smFISH and IF were implemented by first performing smFISH, followed by IF.

ISH analyses were performed on formalin-fixed and paraffin-embedded mouse tissue samples using RNAscope technology (RNAscope Multiplex Fluorescent Reagent Kit v2, Advanced Cell Diagnostics; ACD) following the manufacturer's instructions. Paraffin blocks were cut into 4-5µm sections. Briefly, tissue sections were baked in a dry oven (HybEZ II Hybridization System, ACD) at 60 °C for 1 h and deparaffinized, followed by incubation with Pretreat 1, 2, and 3 (ACD). Slides were incubated with relevant probes for 1 h at 40 °C, followed by successive incubations with Amp1 to 6 reagents. Staining was visualized with DAB or TSA Plus Cyanine 5 (NEL745001KT, PerkinElmer). For quantification, cells with no ISH staining were defined as "negative," and cells that show 1 dot/cell or more were defined as "positive." *In situ* hybridization (RNAscope) probe used in the study was mouse *Lepr* (Entrez Gene ID: 16847, target region 1361 - 2317, catalog number 402731).

After smFISH, the sections were blocked with blocking buffer (X0909, Dako) and then incubated with a primary antibody overnight at 4 °C. The sections were washed in 1x T-PBS 3 times and then incubated with Alexa Fluor 555-conjugated secondary antibody (Thermo Fisher Scientific) for 60 min at room temperature. The sections were then mounted with ProLong Gold antifade reagent containing 4'6-diamidino-2-phenylindole (DAPI; Thermo Fisher Scientific), and fluorescence was examined using an inverse immunofluorescence microscope BZ-X710 (Keyence, Japan) with optical sectioning.

## **The Human Protein Atlas data**

All immunohistochemistry images in Supplementary Figure 10 were obtained from the Human Protein Atlas<sup>293</sup> (<https://www.proteinatlas.org/>). Genes that were highly expressed in the CRC stroma compared to cancer cells (as visualized by DAB staining and hematoxylin counterstaining) were selected by a clinical pathologist. Immunohistochemistry data for COX4I2, NDUFA4L2, and SHC4 were not available in the Human Protein Atlas (website accessed in April 2020) and thus could not be included in the analysis.

### Quantitative image analysis

In all histopathological quantification using the normal colorectal tissues from human and mice, only lamina propria regions were evaluated. Quantification of fluorescent images was performed using consistent image settings in ImageJ<sup>247</sup>. In Figure 5F, each patient has 3-4 replicate tumor samples obtained from different locations within the tumor. A total score (the sum of intensity score and proportion score)<sup>294</sup> was given to each of the 3-4 replicates/patient by an experienced clinical pathologist in a blinded manner. A patient with an average total score of  $\geq 3.3$  was defined as MCAM-high. The optimal cutoff was determined using the X-tile software<sup>253</sup>. In Figure 3F and G, to evaluate cellular positions of *Lepr*-lineage stromal cells, only well-oriented crypts were evaluated.

In Figure 6H and O and Supplementary Figure 20, immunohistochemistry (IHC) images were processed into separate channels representing nuclei staining (hematoxylin) and IHC staining (DAB) using a color deconvolution function in the Fiji software (ImageJ; National Institutes of Health). Then, binary images were generated by intensity thresholds, and DAB<sup>+</sup> areas were calculated by ImageJ. In Figure 6O, to evaluate phosphor-p65 staining intensity, following an application of the lower threshold, the DAB staining intensity per high power field was calculated by ImageJ.

### Mouse model of genetic fate-mapping

*Lepr*-Cre (JAX Stock No. 008320); Rosa26-LSL (LoxP-stop-LoxP)-tdtomato (JAX Stock No. 007909), *Grem1*-CreERT2 (JAX Stock No. 027039); Rosa26-LSL (LoxP-stop-LoxP)-ZsGreen (JAX Stock No. 007906), *Islr*-CreERT2<sup>213,224</sup>; Rosa26-LSL-tdtomato, *Acta2*-RFP<sup>295</sup>, and *Krt19*-Cre<sup>278</sup>; Rosa26- mT/mG (JAX Stock No. 007676) were used for lineage tracing experiments. *Grem1*-CreERT2 and *Islr*-CreERT2 mice were subcutaneously injected with 2 mg of tamoxifen (T5648, Sigma) dissolved in peanut oil on postnatal day 6.

### An azoxymethane (AOM)/ dextran sulfate sodium (DSS) mouse model of CRC

A colitis-associated AOM/DSS mouse model of CRC was generated as described<sup>256</sup>. Briefly, AOM (Sigma, A5486; 10mg/kg) was administered by intraperitoneal injection to 6-12 week-old mice. One week later, the mice commenced the first of three cycles of DSS to induce inflammation (MP Biomedicals, 160110; Molecular Weight 36,000-50,000 Da). Each cycle was comprised of 5 days 2 - 3% DSS (weight/volume) then 16 days of normal drinking water. To monitor tumor burden, mice were weekly colonoscoped, as previously described<sup>296</sup>. Once mice showed endoscopically visible tumors (for *Lepr*-Cre and *Grem1*-CreERT2 mice) or were finished with the last cycle of DSS administration (for *Islr*-CreERT2 mice), mice were started on BrdU (10280879001, Sigma). Mice received one intraperitoneal injection of BrdU (50 mg/kg) and then were administered continuous BrdU in drinking water (1 mg/ml). 10 weeks after AOM injection, mice were euthanized, and colons were removed, opened longitudinally. Colon tumor tissues were used for RNA isolation, followed by RNA-sequencing. For histopathological analysis, colon tumors and adjacent non-neoplastic tissues were fixed in 10% formalin or 4% paraformaldehyde overnight.

### **Bone marrow transplantation**

Male and female C57BL/6 wild-type (WT) mice and *Acta2*-RFP mice were maintained in pathogen-free conditions in the SAHMRI Bioresources Animal Facility. The C57BL/6 WT mice were used at 8-12 weeks of age, and the *Acta2*-RFP mice were used at 14-21 weeks of age. Mice were irradiated in a filter top cage in a RS 2000 X-Ray Irradiator (Rad Source). The mice were exposed to a single lethal dose of 10 grays. On the day of irradiation, donor mice (C57BL/6 WT or *Acta2*-RFP) were culled, and bone marrow cells were harvested from femurs and tibia and were resuspended in PBS.  $5 - 7.9 \times 10^6$  bone marrow cells were injected into the tail vein of recipient C57BL/6 WT mice using a 30G needle. Bone marrow recipient mice were placed on a prophylactic course of antibiotics commencing three days prior to irradiation. Mice received an initial subcutaneous injection of Baytril (10mg/kg), followed by Baytril Oral Solution administered in drinking water (0.28mg/ml) for 5-6 weeks. Fresh antibiotics were prepared weekly.

To confirm engraftment of bone marrow-derived cells in the bone marrow of the recipient mice, genomic DNA was extracted from harvested bone marrow samples using the QIAamp Mini DNA

Blood kit before qPCR for RFP and *Bcl2*. *Bcl2* was used as a reference gene to normalize the total DNA amount, as described<sup>297</sup>. To determine the percentage of engraftment, a standard curve was prepared from known concentrations of DsRed/wild-type DNA (from 0.26%-100%). The mean values of delta-Ct (CtDsRed-CtBcl2) were plotted against the known concentrations of DNA using linear regression fitting. The percentage of DsRed expression in bone marrow samples from the experimental animals was then calculated against the curve. The following primers were used for qPCR.

*Bcl2* forward, 5'- AAGCTGTCACAGAGGGGCTA -3'; *Bcl2* reverse, 5'- CAGGCTGGAAGGAGAAGATG -3'; RFP forward, 5'- CCCGACTACAAGAAGCTGTCCTTC -3'; RFP reverse, 5'- CGATGAACTTCACCTTG TAGATGAAGCAG -3'.

### **Generation of *Mcam*-knockout mice by CRISPR/Cas9-mediated genome engineering**

C57BL6/J female mice were superovulated with 5 IU Pregnant Mare Serum Gonadotrophin (PMSG; Folligon; Intervet) and 47.5 hours later with 5 IU human chorionic gonadotrophin (hCG; Chorulon; Intervet) before being mated to C57BL6/J males. Presumptive zygotes were collected from oviducts 19 hours post-hCG injection in FHM media (Merk Millipore) and maintained in KSOMAA media (Merk Millipore) under oil at 37 °C in 5% CO<sub>2</sub> 5% O<sub>2</sub> with a Nitrogen balance. Fertilized zygotes were identified in FHM media under oil by the presence of two pronuclei before microinjection of 50 ng/ul SpCas9 protein (PNA Bio), 25 ng/ul SpCas9 single guide RNA (sgRNA) (AGGCTAAGCAGGGACGGGGG), 100 ng/ul SaCas9 KKH mRNA, and 50 ng/ul SaCas9 KKH sgRNA (AGACCTCTCCCAGTGTAAATA) in a buffered solution.

Genomic DNA was isolated from ear notches using a High Pure PCR Template Preparation Kit (Roche) and Knockout (KO) founder screening was performed using a forward primer (GCACATGCCTTACATTCCCC) and a reverse primer (TGACTIONACTCAGCATCACCTGT). KO founder allele was characterized to have 935 bps deletion, inclusive of exons 2-4. This caused a frameshift in the coding sequence, resulting in a premature stop codon in exon 6.

### **Mouse genotyping**

Genomic DNA extracted from mouse ears or tails was used for PCR genotyping. The sequences of the primers were as follows:

*Lepr*-Cre forward, 5'-TTAATCCATATTGGCAGAACGAAAACG-3'; *Lepr*-Cre reverse, 5'-CAGGCTAAGTGCCTTCTCTACA-3'.

*Lepr*-Wild-Type (WT) forward, 5'-CGCACAGTCACAAGATAATGG -3'; *Lepr*-WT reverse, 5'-GCTCTACTGGAATGGAACCTT-3'.

*Krt19*-Cre forward, 5'-TCTCCCTCCTCATCATGTCC-3'; *Krt19*-Cre reverse, 5'-CATGTTTAGCTGGCCCAAAT-3'

*Acta2*-RFP forward, 5'-AGATCCACCAGGCCCTGAA-3'; *Acta2*-RFP reverse, 5'-GTCTTGA ACTCCACCAGGTAG-3'

*Apc*-Min forward , 5'-TGAGAAAGACAGAAGTTA-3'; *Apc*-Min reverse , 5'-TTCCACTTTGGCATAAAGGC -3';

*Apc*-WT forward , 5'- GCCATCCCTTCACGTTAG -3'; *Apc*-WT reverse , 5'-TTCCACTTTGGCATAAAGGC -3'.

*Islr*-CreERT2 forward, 5'-ACACACGACCTTGGCAAGTCCCAGC-3'; *Islr*-CreERT2 reverse, 5'-CGATCCCTGAACATGTCCATCAGG-3';

*Islr*-WT forward, 5'-ACACACGACCTTGGCAAGTCCCAGC-3'; *Islr*-WT reverse,5'-GTCTGCAATCTGGAAGCCATACTTCTCC-3'.

*Grem1*-CreERT2 forward, 5'-TTAATCCATATTGGCAGAACGAAAACG-3'; *Grem1*-CreERT2 reverse, 5'-CAGGCTAAGTGCCTTCTCTACA-3'.

*Rosa26*-LSL-tdtomato forward, 5'-AGATCCACCAGGCCCTGAA-3'; *Rosa26*-LSL-tdtomato reverse, 5'-GTCTTGA ACTCCACCAGGTAG-3';

*Rosa26*-LSL-ZsGreen forward, 5'- GCGCCGTGTGCATCTG-3'; *Rosa26*-LSL-ZsGreen reverse, 5'-ACTCGTGGTACATGCAGTTCTC-3';

*Rosa26*-WT forward, 5'-TTCCCTCGTGATCTGCAACTC-3'; *Rosa26*-WT reverse, 5'-

CTTTAAGCCTGCCAGAAGACT-3' (used for *Rosa26*-LSL-tdtomato and *Rosa26*-LSL-ZsGreen genotyping).

Rosa26-mT/mG forward, 5'-CTCTGCTGCCTCCTGGCTTCT-3'; Rosa26-mT/mG reverse, 5'-TCAATGGGCGGGGGTTCGTT-3'

Rosa26-WT forward, 5'-CTCTGCTGCCTCCTGGCTTCT-3'; Rosa26-WT reverse, 5'-CGAGGCGGATCACAAGCAATA-3' (used for Rosa26-LSL-mT/mG genotyping).

*Mcam*-WT forward, 5'- GCACATGCCTTACATTCCCC -3'; *Mcam*-WT reverse, 5'-TTACCCTCTCCTTACCCTCAACTTC -3'

*Mcam*-KO forward, 5'- GCACATGCCTTACATTCCCC -3'; *Mcam*-KO reverse, 5'-TGACTTACTCAGCATCACCTGT -3'.

### **Generation of *Apc*-, *Trp53*- and *Kras*-mutant mouse colorectal cancer organoids by CRISPR/Cas9-mediated genome editing**

The basal culture medium for mouse colon organoids was Advanced Dulbecco's modified Eagle medium/F12 (Life Technologies) supplemented with 1x gentamicin/antimycotic/antibiotic (Life Technologies), 10 mM HEPES, 2 mM GlutaMAX, 1 x B27 (Life Technologies), 1 x N2 (Life Technologies). The following niche factors were used: 50 ng/ml mouse recombinant EGF (Peprotech), 100 ng/ml mouse recombinant noggin (Peprotech), 20% R-spondin-2 conditioned medium, 50% Wnt-3A conditioned medium. Organoids were plated in 50 ul growth factor-reduced Matrigel (356231, Corning) on a 24-well dish, and 500 ul of the medium was added to organoids

Guide RNAs (gRNAs) specific for each target gene were either previously published or designed de novo using the CRISPR Design Tool<sup>242</sup>. The following gRNA oligos and *Kras*<sup>G12D</sup> homology arms (498-bps 5' homology arm and 299-bps 3' homology arm) were cloned into pLenti-Guide-CMV-dtomato (Modified from; Addgene Plasmid #17452).

gRNA sequences

*Apc*: 5'- GGAAGCCTTGTGGGACATGG -3'

*Trp53*: 5'- GTGTAATAGCTCCTGCATGG -3'<sup>243</sup>

*Kras*: 5'- GCAGCGTTACCTCTATCGTA -3'

Normal colon organoids isolated from a Rosa26-Cas9 mouse (JAX Stock Number 024858)<sup>243</sup> were transduced with the lentivirus that expresses single guide RNAs (*sgApc*, *sgTrp53*, and *sgKras*) and

contains *Kras*<sup>G12D</sup> homology arms. Three days later, media was changed to ADMEM (+EGFR inhibitor, 324840 Calbiochem; + Nutlin-3a, SML0580 Sigma; no Wnt3a; no Rspo) to enrich for correctly targeted clones. The correctly targeted monoclonal line was confirmed by amplicon sequence, as described below. Then, to remove the FRT-flanked Cas9-p2a-GFP gene in the Rosa26 locus, pCAG-FLPe (13787, Addgene) plasmid was transiently transfected to the AKP (*Apc*<sup>Δ/Δ</sup>, *Kras*<sup>G12D/Δ</sup> *Trp53*<sup>Δ/Δ</sup>) tumoroid line using Lipofectamine 2000 (11668019, Thermofisher). Five days after FPLe transfection, recombined GFP-negative AKP tumoroids were collected by the BD FACSAria Fusion Cell Sorter (BD Biosciences) using untransfected AKP tumoroids to set a GFP-negative gate. Finally, the Cas9<sup>-</sup> GFP<sup>-</sup> AKP tumoroids were transduced with lentivirus expressing Firefly and a puromycin resistance gene (pLenti-EF1α-*Firefly*-PGK-Puro). 3 days later, 4 ug/ml of puromycin was added to the medium (ADMEM; +EGFR inhibitor, 324840 Calbiochem; + Nutlin-3a, SML0580 Sigma; no Wnt3a; no Rspo) to enrich for transduced cells.

After CRISPR/Cas9-mediated genome editing, handpicked monoclonal organoid lines were screened for *Apc* and *Trp53* loss of function insertions/deletions using amplicon sequence. The following primers were used, as described<sup>6,246</sup>. Overhang sequences are underlined.

*Apc* forward: 5'- CTGAGACTTGACATCGCAGCTTAATTCAGGCAAATCCTAAGAGAG -3'

*Apc* reverse: 5'- GTGACCTATGAACTCAGGAGTCGGTCTGTTTGCCATGAGATTCC -3'

*Trp53* forward: 5'- GTGACCTATGAACTCAGGAGTCTAGTGAGGTAGGGAGCGACTTC -3'

*Trp53* reverse: 5'- CTGAGACTTGACATCGCAGCCCAAAGAGCGTTGGGCATGTG -3'

To examine *Kras* mutations in the AKP tumoroids, nested PCR was performed using the following primers. Amplicon sequencing from the second PCR products was performed to screen for a *Kras*<sup>G12D</sup> mutation. Overhang sequences are underlined.

1<sup>st</sup> PCR *Kras* forward: 5'-CTTTATTGCAGAACTGCTCTGATGG -3'

1<sup>st</sup> PCR *Kras* reverse: 5'- GGATGGCATCTTGGACCTTACTC -3'

2<sup>nd</sup> PCR *Kras* forward:

5'-GTGACCTATGAACTCAGGAGTCAGTTTTTGATAATCTTGTGTGAGAC -3'

2<sup>nd</sup> PCR *Kras* reverse:

5'- CTGAGACTTGACATCGCAGCAGCCTTGGAATAAAGGACATC -3'

### **Orthotopic injection of AKP tumoroids into the mouse colon**

8-20 week-old C57BL6/J *Mcam*-WT (wild type) or *Mcam*-KO (knockout) mice housed under pathogen-free conditions at the SAHMRI Bioresources facility were used for tumoroid injection.

Given that our AKP tumoroids were derived from a male Rosa26-Cas9 mouse, only male mice were used for injection to exclude potential effects of sex-related differences in tumor rejection and growth<sup>298,299</sup>.

Cell preparation and colonoscopy-guided orthotopic injection into the colon wall were undertaken as previously described<sup>296</sup> with modifications. AKP tumoroids were isolated from matrigel and dissociated to small clusters using TrypLE (12605028, Thermofisher). The cell clusters (equivalent to 600 whole organoids/injection) were washed three times with cold PBS containing 10  $\mu$ M Y-27632 and then resuspended in 20  $\mu$ L of PBS containing 10% Matrigel, 1:1000 India ink, and 10  $\mu$ M Y-27632. The 20  $\mu$ L of cell suspension was injected into the mucosa of the distal colon of anesthetized mice. A customized needle (Hamilton Inc. part number 7803-05, removable needle, 33 gauge, 12 inches long, point 4, 12-degree bevel) was used. Colonoscopy was performed using a Karl Storz Image 1 Camera System comprised of: Image1 HDTV HUB CCU; Cold Light Fountain LED Nova 150 light source; Full HD Image1 3 Chip H3-Z Camera Head; Hopkins Telescope, 1.9mm, 0 degrees. A sealed luer lock was placed on the working channel of the telescope sheath to ensure minimal air leakage (Coherent Scientific, part number 14034-40). Two injections using a total of 40  $\mu$ L volume were performed per mouse.

Tumor scoring by colonoscopy was performed as described<sup>292</sup>. All macroscopic tumor volume measurement and immunohistological analyses were performed using mice harvested three weeks after tumor injection. Tumor volumes of formalin-fixed tissues were calculated using the ellipsoid volume formula, tumor volume ( $\text{mm}^3$ ) =  $\pi/6$  x Length (mm) x Width (mm) x Height (mm).

### ***In vivo* imaging system (IVIS)**

Two weeks after orthotopic injection of AKP tumoroids into the colon, tumoroid-derived luciferase signals were assessed by an *in vivo* imaging system (IVIS) by using a Xenogen IVIS Spectrum

Imaging System (Perkin Elmer Inc). Ten minutes after intraperitoneal injection of 150 mg/kg of D-Luciferin (L-8220, Biosynth Carbosynth). Luciferase activity was quantified using Living Image software (Perkin Elmer Inc).

### **A humane endpoint in animal experiments**

In survival analyses, to determine the humane endpoint in animal experiments, a clinical record score was used. Scores were obtained by one point being given for the presence of each of the following observations: weight loss > 15%; hunched/ruffled coat; dehydrated; absence of movement; or facial grimace. Once a score of 3 was reached, the mice were euthanased.

### **RNA-sequencing from fluorescence-activated cell sorting (FACS)-purified mouse colonic cells**

Four 8-week-old adult mice (2 males and 2 females), four postnatal day 14 mice (2 males and 2 females), and four AOM/DSS-treated mice (2 males and 2 females) were used to isolate colonic cells. Colons were washed three times in Hank's Balanced Salt Solution (H9394, Sigma) and cut into 0.5 mm pieces with a pair of scissors. Colonic tissue pieces were spun at 450 g for 4 minutes. After discarding the supernatant, the colonic tissues were enzymatically digested with 35.5 mg of type IV collagenase (17104019, Life Technologies) and 1 mg of dispase (17105041, Life Technologies) in 10 ml of HBSS, at 37 °C for 45 minutes. The digested colonic cells were then filtered through a 40- $\mu$ m-nylon strainer to remove cell clumps. Cells were stained with the following antibodies: anti-EPCAM-Alexa Fluor 488 (118210, Biolegend), anti-CD31-Alexa Fluor 594 (102520, Biolegend), anti-CD45-APC (103112, Biolegend), and anti-TER119-APC (116212, Biolegend). Dead cells were excluded based on their positive staining for DAPI. FACS-sorting of CD45<sup>+</sup> immune cells, EPCAM<sup>+</sup> epithelial cells, CD31<sup>+</sup> endothelial cells, and EPCAM<sup>-</sup>CD31<sup>-</sup>CD45<sup>-</sup>TER119<sup>-</sup> stromal cells was performed using FACSAria Fusion (BD Bioscience). Data analysis was performed using FlowJo version 10.5.3 (LLC, USA).

Collected cells were spun at 800 g for 5 minutes. Cells were lysed with 1 ml of Trizol, and RNA was isolated using PicoPure™ RNA Isolation Kit (KIT0204, Life Technologies) according to the

manufacture's protocol. DNase treatment was performed with RNase-Free DNase Set (79254, Qiagen) on PicoPure spin columns.

The purity of the isolated RNA was assessed using Bioanalyzer (Agilent, US) to ensure high-quality RNA (RNA integrity number > 7). 200-250 ng of RNA was used for stranded library preparation with RiboZero ribosomal RNA depletion (Kapa). Whole transcriptome 1 x 75 bp read sequencing was performed on a NextSeq 500 (Illumina) by the Australian Cancer Research Facility (ACRF). The RNA-sequencing reads were trimmed for adaptor sequence, low-quality sequences were removed, and remaining reads were aligned to a mouse reference genome (GRCm38), using STAR aligner 2.4.2a, allowing 2 mismatches. Gene expression levels were quantified using STAR with quantmode option. For differential expression (DE) analysis, read counts were normalized by the trimmed mean of M values (TMM) method implemented in the R package edgeR (version 3.28.0). Differentially upregulated genes were defined as genes that showed log<sub>2</sub> fold change greater than or equal to 1 and false discovery rate less than 0.10. The RNA-seq data are available at the Gene Expression Omnibus (GEO) under accession number GSE162508.

### **Analyses of the Cancer Genome Atlas (TCGA) data**

10-year-survival analyses of The Cancer Genome Atlas (TCGA) data (COAD and READ cohorts) were performed on the R program with survival and survminer packages (<https://CRAN.R-project.org/package=survminer>) using the GDC colorectal cancer patient RNA-sequencing data (<https://portal.gdc.cancer.gov/>). Gene expression levels analyzed by RSEM (RNA-Seq by Expectation-Maximization) were downloaded from the website and then normalized using the voom function<sup>300</sup>. Optimal cutoff levels for survival analyses were determined using survminer packages (<https://CRAN.R-project.org/package=survminer>). Log-rank tests were performed using the R survival package (<https://cran.r-project.org/web/packages/survival/index.html>). In Figure 4B(2), median expression levels were used as cutoff values.

To perform differential gene expression analyses and correlation analyses, HTseq raw counts files were downloaded from Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/>). These analyses were performed using the R software (version 4.0.0). TMM normalization and log counts

per million (log CPM) value calculation were performed by edgeR package version 3.30.3<sup>301</sup>. Differential gene expression analyses were performed with glmLRT() function. Log fold change (logFC) was used to generate a rank list for gene set enrichment (GSEA) analyses. The R package clusterProfiler version 3.16.1<sup>302</sup> was used to perform the GSEA analyses. Benjamini & Hochberg adjusted p-values were used for statistical analyses. Gene sets were downloaded using the Molecular signatures database (<http://www.gsea-msigdb.org/gsea/msigdb/index.jsp>; gmt file, version 7.2). In Spearman correlation coefficient analyses, the log CPM values were used as expression values.

The Consensus Molecular Subtypes (CMS)<sup>31</sup> labels were downloaded from the supplementary materials of previous publications (n = 511 patients with CMS1-4 CRC)<sup>31,254</sup>. Then, expression Z-scores for tumor samples were calculated based on the standard deviation and mean value for the normal colorectal tissues. Violin plots were then generated using the ggplot2 package (version 3.2.1).

### **Analyses of publicly available cDNA gene expression microarray datasets**

Expression microarray data (GSE17538<sup>252</sup>, 33113<sup>288</sup>, 39396<sup>103</sup>, 39582<sup>251</sup>, and 41258<sup>250</sup>) and patient information were obtained from Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>). Normalized gene expression levels for the following probes were used. P\_200974\_at, *ACTA2*; P\_209955\_s\_at, *FAP*; P\_215305\_at, *PDGFRA*; P\_216572\_at, *FOXL1*; P\_209087\_x\_at, *MCAM*; 202628\_s\_at, *SERPINE1*; 237046\_x\_at, *IL34*; 214038\_at, *CCL8*; 203507\_at, *CD68*; 205786\_s\_at, *ITGAM* (GSE17538, 33113, 39582, and 41258); 209087\_PM\_x\_at, *MCAM* (GSE39396). In GSE41258, primary colorectal cancer samples included in a previous analysis<sup>250</sup> were analyzed by choosing samples with a label of “Included in analysis; Yes” on the GEO website. Optimal cutoff levels for survival analyses were determined using survminer packages (<https://CRAN.R-project.org/package=survminer>). Survival analyses (Log-rank test) were performed using an R program with survival (<https://cran.r-project.org/web/packages/survival/index.html>). For differential expression analyses, false discovery rate (FDR)-adjusted P-values were calculated by GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>).

## Analyses of publicly available single-cell RNA-sequencing data

Single-cell RNA-sequencing (ScRNA-seq) data from GSE132465, GSE144735<sup>226</sup>, and GSE114374<sup>290</sup> were analyzed using BBrowser2 (Version 2.4.10; BioTuring, US; <https://bioturing.com/>). Unique molecular identifier (UMI) counts available on the GEO website (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) were used for normalization by the software. In GSE132465 and 144735, fibroblasts are defined as pericytes, myofibroblasts, and Stromal 1-3 fibroblasts<sup>226</sup>. Endothelial cells include tip-like endothelial cells, stalk-like endothelial cells, proliferative endothelial cells, and lymphatic endothelial cells<sup>226</sup>. Immune cells comprise T cells, B cells, myeloid cells, and mast cells<sup>226</sup>. In GSE114374, the default cell type annotations in the BBrowser2 software were used. In Supplementary Figure 4, t-distributed stochastic neighbor embedding (t-SNE) plots were generated on the User-friendly InteRface tool to Explore Cell Atlas (URECA) website (<http://ureca-singlecell.kr>)<sup>226</sup>.

Differential gene expression analysis was performed between 411 *MCAM*<sup>high</sup> CAFs and 1443 *MCAM*<sup>low</sup> CAFs using the mean *MCAM* expression value in CAFs as a cutoff. Wilcoxon rank-sum test was performed in the BBrowser2 software to identify differentially upregulated genes in *MCAM*<sup>high</sup> CAFs. Differentially upregulated genes were defined as genes that show p-values less than 0.05.

Cytokines/chemokines involved with macrophage/monocyte chemotaxis were defined by combining the following Gene Ontologies: Macrophage chemotaxis (GO0048246), Monocyte chemotaxis (GO0002548), and Cytokine activity (GO0005125). To examine for cytokines/chemokines (i.e. extracellular soluble factors) with roles in macrophage/monocyte chemotaxis, macrophage/monocyte chemotaxis genes (GO0048246 or GO0002548) that are found in GO0005125 (Cytokine activity; gene sets composed of extracellular soluble factors) were selected. This resulted in the selection of the 41 cytokines/chemokines involved in macrophage/monocyte recruitment: *C5*, *CCL1*, *CCL2*, *CCL3*, *CCL4*, *CCL5*, *CCL7*, *CCL8*, *CCL11*, *CCL13*, *CCL14*, *CCL15*, *CCL16*, *CCL17*, *CCL18*, *CCL19*, *CCL20*, *CCL21*, *CCL22*, *CCL23*, *CCL24*, *CCL25*, *CCL26*, *CCL3L1*, *CCL4L1*, *CKLF*, *CSF1*, *CX3CL1*, *CXCL10*, *CXCL12*, *GREM1*, *HMGB1*,

*IL6*, *IL34*, *MIF*, *MSMP*, *MSTN*, *TNFSF11*, *TNFSF18*, *XCL1*, and *XCL2*. Human gene lists in the Gene Ontologies were downloaded from AmiGO 2 (<http://amigo.geneontology.org/amigo/search/ontology>).

## Cell Culture

A mouse colonic fibroblast cell line, YH2 cells<sup>240</sup>, was generously provided by Professor Tony Burgess (The Walter and Eliza Hall Institute of Medical Research). YH2 cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Sigma) supplemented with 10% fetal bovine serum (FBS; Gibco). In Figure 5D, YH2 cells were stimulated with 10 ng/ml of recombinant human TGF- $\beta$ 1 (240-B-002, R&D systems) and/or 10  $\mu$ M of Galunisertib (ADV465749242; AChemBlock) in DMEM containing 1%FBS for 24 hours, as described<sup>6</sup>. In Figure 6J and K, YH2 cells were serum-starved in DMEM containing 0.5% FBS for 24 hours before stimulation with (a) 0.2 ng/ml of recombinant human IL-1 $\beta$  (PHC0814, Thermo Fisher) or (b) 0.2 ng/ml of IL-1 $\beta$  + 1.0  $\mu$ M of IKK-16 (SML1138, Sigma) in DMEM without FBS for 15 minutes. In Figure 6L, YH2 cells were serum-starved in DMEM containing 0.5% FBS for 24 hours before stimulation with (a) 0.2 ng/ml of recombinant human IL-1 $\beta$  or (b) 0.2 ng/ml of IL-1 $\beta$  + 0.1  $\mu$ M of IKK-16 in DMEM containing 0.5% FBS for 18 hours.

All cell lines used were routinely screened for Mycoplasma contamination by MycoAlert Mycoplasma Detection Kit (LT07-118, Lonza).

## Plasmids

To overexpress *mScarlet-1*- Hemagglutinin (HA) epitope tags or *Mcam*-HA in YH2 cells, DNA fragments for mouse codon-optimized *mScarlet-1* or mouse *Mcam* gene with 3 x HA tags at the amino-terminus were generated using custom gene synthesis service by Gene Universal (New Jersey, USA). These genes were subcloned into unidirectional pLenti-EF1 $\alpha$  (Human elongation factor-1 alpha promoter)-GFP-p2A-Puro (puromycin resistance gene)-EF1 $\alpha$ -MCS (multiple cloning site) vectors<sup>6</sup>, resulting in the generation of pLenti-EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -*mScarlet*-HA and pLenti-EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -*Mcam*-HA.

To overexpress *I1r1* in YH2 cells, DNA fragments for codon-optimized mouse *I1r1* with 3 x *Myc* epitope tags at the amino-terminus was also generated using custom gene synthesis service by Gene Universal (New Jersey, USA). The *I1r1-Myc* tag gene was subcloned into a pLenti-EF1 $\alpha$ -Empty-p2a-BSD (blasticidin resistance gene) vector, resulting in the generation of pLenti-EF1 $\alpha$ -*I1r1-Myc*-p2a-BSD vector.

To overexpress *Firefly* in AKP tumoroids, a Firefly gene in the pGL4.20 [luc2/Puro] Vector (E6751, Promega) was subcloned to bidirectional pLenti-EF1 $\alpha$ -MCS-PGK-Puro vector. This resulted in the generation of pLenti-EF1 $\alpha$ -*Firefly*-PGK-Puro.

### **Lentivirus production and transduction**

293T cells were co-transfected with psPAX2 (Addgene; plasmid #12260), pMD2.G (Addgene; plasmid #12259), and a lentivirus vector plasmid. At 48 and 72 h after transfection, viral supernatants were harvested, filtered through a 0.45- $\mu$ m filter, and concentrated using Amicon Ultra Centrifugal Filters (Merck Millipore; UFC910024). Concentrated lentivirus particles were used for transduction. 48 h after transduction, positively transduced cells were selected with 4  $\mu$ g/mL puromycin and/or blasticidin if the lentivirus vectors contain an antibiotics resistance gene.

### **Quantitative PCR**

Total RNA was extracted using TRIzol reagent (Invitrogen) and an RNeasy Mini Kit (Qiagen). Purified RNA samples were reverse-transcribed using Transcriptor Universal cDNA Master (Roche) according to the manufacturer's instruction, followed by the dilution of cDNA at 1:5. Quantitative reverse-transcription PCR (qRT-PCR) of the complementary DNAs (cDNAs) was performed with KAPA PROBE or SYBER FAST qPCR Master Mix (KAPABiosystems) and was run on a QuantStudio 7 Flex Real-Time PCR System (Thermo Scientific). The data were analyzed using the  $2^{-\Delta\Delta C_t}$  method and normalized to *Gapdh* expression levels.

Probes or primers (generated by IDT) used in this study are as follows. Primers and *Cc/8* probes were designed using the Primer-BLAST online program<sup>248</sup>.

Primers:

Mouse *Mcam* (Forward: 5'-GCCCTTGCAAGAGGAGGAGAA-3', Reverse: 5'-ACTAGGCGTGCACTCAGAACA-3'), mouse *Acta2* (Forward: 5'-CTCTTCCAGCCATCTTTCATTG-3', Reverse: 5'-AATGCCTGGGTACATGGTG-3'), and mouse *Gapdh* (Forward: 5'-CCTCGTCCCGTAGACAAAATG-3', Reverse: 5'-TGTAGTTGAGGTCAATGAAGGG-3').

Probes:

Mouse *Cd2* (Mm.PT.58.42401764), mouse *Ptprc* (Mm.PT.58.7583849), mouse *Cdh5* (Mm.PT.58.8747496), mouse *Vwf* (Mm.PT.58.32874548), mouse *Grem1* (Mm.PT.58.11631114), mouse *Acta2* (Mm.PT.58.16320644), mouse *Fap* (Mm.PT.58.31960536), mouse *Epcam* (Mm.PT.58.11851150), mouse *Cdh1* (Mm.PT.58.41847659), mouse *Il34* (Mm.PT.58.32379406), mouse *Gapdh* (Mm.PT.39a.1), and mouse *Ccl8* (Forward: 5'-CTCCAGTCACCTGCTGCTTTC-3', Reverse: 5'-AGAGAGACATACCCTGCTTGGTC-3', Probe: 5'-AGCTGAAGATCCCCCTTCGGGTGCTGAAAAGCT-3').

### Western blot analysis

Cells were lysed in a lysis buffer (78501, Thermo Fisher) supplemented with cOmplete Protease Inhibitor (Roche) and PhosSTOP Phosphatase Inhibitor cocktails (4906845001, Roche). Lysates were clarified by centrifugation at 12,000 × g for 10 min at 4 °C. Then, sodium dodecyl sulfate (SDS) sample buffer (10 mM Tris-HCl, 2% SDS, 2 mM EDTA, 0.02% bromophenol blue, 6% glycerol; pH 6.8) was added.

The separation was performed by SDS-polyacrylamide gel electrophoresis using a precast gel (4568094 or 4568095, Bio-rad). Proteins were then transferred to Polyvinylidene Difluoride (PVDF) membranes (1620177, Bio-rad) using a semi-dry transfer system (1703940, Bio-rad). The membranes were blocked in 5% milk in PBS containing 0.05% Tween 20, and then incubated with primary antibodies. Proteins were detected by horseradish peroxidase (HRP)-conjugated secondary antibodies (NA9310V, GE Healthcare; N934VS, GE Healthcare; ab97057, Abcam), followed by signal development using an HRP substrate (WBLUR0500, Millipore). The blots were imaged using ChemiDoc MP (Bio-Rad) and quantified using the Image Lab software (Version 6.0.1, Bio-rad).

## Immunoprecipitation

YH2 cells were lysed in IP lysis/Wash Buffer (26149, Pierce Co-Immunoprecipitation Kit, Thermofisher), supplemented with cOmplete Protease Inhibitor (Roche) and PhosSTOP Phosphatase Inhibitor cocktails (4906845001, Roche). In Figure 6M, lysates were immunoprecipitated with an HA antibody using Pierce HA Tag IP/Co-IP Kit (26180, Thermofisher), according to the manufacturer's instruction. In Supplementary Figure 23, immunoprecipitation was performed using the Pierce Co-IP Kit (26149, Thermofisher). An anti-MYC tag antibody (ab9106, Abcam) or IgG isotype (ab171870, Abcam) was coupled to agarose resin, following the manufacturer's protocol.

In Figure 6M, YH2 cells used in the immunoprecipitation experiment are as follows (left to right); (1) Non-transduced YH2 cells (as a negative control for HA immunoprecipitation), (2) EF1 $\alpha$ -Empty-p2a-BSD and EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -mScarlet-HA YH2 cells, (3) EF1 $\alpha$ -Empty-p2a-BSD and EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -Mcam-HA YH2 cells (4) EF1 $\alpha$ -Il1r1-Myc-p2a-BSD and EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -mScarlet-HA YH2 cells, (5) EF1 $\alpha$ -Il1r1-Myc-p2a-BSD and EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -Mcam-HA YH2 cells. In Supplementary Figure 23, the following YH2 cells were used (left to right); (1) EF1 $\alpha$ -Empty-p2a-BSD and EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -mScarlet-HA YH2 cells (as a negative control), (2) EF1 $\alpha$ -Il1r1-Myc-p2a-BSD and EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -mScarlet-HA YH2 cells, (3) EF1 $\alpha$ -Empty-p2a-BSD and EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -Mcam-HA YH2 cells, (4) EF1 $\alpha$ -Il1r1-Myc-p2a-BSD and EF1 $\alpha$ -GFP-p2A-Puro-EF1 $\alpha$ -Mcam-HA YH2 cells.

## Luciferase reporter assay

A lentivirus vector containing 10 sequential copies of a nuclear factor- $\kappa$ B (NF- $\kappa$ B) responsive element (GGGAATTTCC or GGGACTTTCC)-NanoLuc-Ubiquitin promoter-Firefly-p2a-BSD (blasticidin resistance gene) was generated by using custom gene synthesis service by Gene Universal (New Jersey, USA). Empty or MCAM-overexpressing YH2 cells were transduced with the lentivirus, followed by selection with 4  $\mu$ g/ml of blasticidin. These YH2 cells were serum-starved in DMEM containing 0.5% FBS for 24 hours before stimulation with (a) 0.005 ng/ml of recombinant human IL-1 $\beta$  (PHC0814, Thermo Fisher), or (b) 0.005 ng/ml of IL-1 $\beta$  + 0.5  $\mu$ M of IKK-16 (SML1138,

Sigma) in DMEM containing 0.5% FBS for 18 hours. Cell lysates were collected using a passive lysis buffer (E1941, Promega). Luminescence values of cell lysates were measured using a Nano-Glo Dual-Luciferase Reporter Assay System (N1610, Promega) and a GloMax microplate reader (GM3000, Promega) following the manufacturer's instruction. The Nanoluc luminescence levels were normalized to the Firefly luminescence levels.

## Chapter 4: Discussion

In this thesis, I found that BMP signalling regulated by CAF-specific genes, *GREM1* and *ISLR*, modulates CRC progression and predicts patient survival. Furthermore, this work identified *Lepr*-lineage cells as a major source of colorectal CAFs in a mouse model of CRC. These *Lepr*-lineage CAFs express MCAM, which promotes TAM-recruitment and tumour progression. My data suggest that targeting these stromal genes could be an attractive therapeutic strategy to inhibit CRC progression.

Recent reports have shown that CAF subpopulations show their distinct gene expression profiles and have characterized signaling pathways responsible for inducing the specific CAF phenotypes<sup>54,265,268</sup>. These studies have proposed several CAF classifications, including myofibroblastic and inflammatory CAFs. However, functional CAF heterogeneity with regards to tumour progression or inhibition (i.e., tumour-promoting v.s. tumour-inhibiting CAFs) remains to be fully elucidated. Here, my study revealed that BMP signalling mediated by CAF-derived *GREM1* and *ISLR* drives CRC progression and determines patient outcomes. I also found that MCAM<sup>+</sup> pericyte-like CAFs, which are distinct from conventional CAFs such as myofibroblastic and inflammatory CAFs<sup>265,266,276,277</sup>, alter immune landscapes to accelerate CRC growth. These studies pave the way for better understanding functional CAF heterogeneity and provide a rationale for selectively targeting these cancer-promoting or -restraining CAFs to treat CRC.

Stromal TGF- $\beta$  is a key regulator of CRC initiation, progression and metastasis<sup>103,303</sup>. In keeping with this, my study showed that TGF- $\beta$  induces expression of tumour-promoting *Grem1* and *Mcam* at the expense of tumour-restraining *Islr*. This novel CAF polarization/differentiation mechanism mediated by TGF- $\beta$  might, in part, explain the tumour-promoting actions of TGF $\beta$ . Given that stromal TGF- $\beta$  confers resistance to immunotherapy<sup>303</sup>, it would be intriguing to investigate whether TGF $\beta$ -responsive CAFs that express *GREM1* or *MCAM* contribute to the immune evasion and resistance to the immune checkpoint blockades. Conversely, it is plausible that *ISLR*<sup>+</sup> CAFs in the TGF- $\beta$ <sup>low</sup>

tumour milieu could exert an immunomodulatory effect in enhancing efficacies of immune checkpoint inhibition, which requires further studies.

One limitation of this study is that I have not comprehensively elucidated an overlapping or lineage relationship between *Grem1*/*Islr*<sup>+</sup> CAFs and *Mcam*<sup>+</sup> CAFs, despite revealing that TGF- $\beta$  is a critical upstream regulator of these genes. Recent advances in single-cell transcriptome analyses have enabled CAF phenotyping based on gene expression profiles and trajectory analyses<sup>266,287</sup>. To gain more insights into stromal evolution during colorectal carcinogenesis, future studies should combine scRNA-seq pseudo-time trajectory analysis and fate-mapping mouse models including *Lepr*-Cre, *Grem1*-CreERT2 and *Islr*-CreERT2 lines.

In conclusion, this study uncovered the hitherto unknown roles of CAF subsets identified by *GREM1*, *ISLR* or *MCAM* in colorectal carcinogenesis. Future approaches to selectively target these CAF subsets or phenotypically remodel these CAFs through TGF- $\beta$  inhibition might have clinical significance to limit CRC progression and improve CRC patient outcomes.

## Bibliography

- 1 Ganesh, K. *et al.* Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol* **16**, 361-375, doi:10.1038/s41575-019-0126-x (2019).
- 2 van Pelt, G. W. *et al.* The tumour-stroma ratio in colon cancer: the biological role and its prognostic impact. *Histopathology* **73**, 197-206, doi:10.1111/his.13489 (2018).
- 3 Kobayashi, H. *et al.* Cancer-associated fibroblasts in gastrointestinal cancer. *Nat Rev Gastroenterol Hepatol* **16**, 282-295, doi:10.1038/s41575-019-0115-0 (2019).
- 4 Ozdemir, B. C. *et al.* Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell* **25**, 719-734, doi:10.1016/j.ccr.2014.04.005 (2014).
- 5 Rhim, A. D. *et al.* Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell* **25**, 735-747, doi:10.1016/j.ccr.2014.04.021 (2014).
- 6 Kobayashi, H. *et al.* The Balance of Stromal BMP Signaling Mediated by GREM1 and ISLR Drives Colorectal Carcinogenesis. *Gastroenterology* **160**, 1224-1239 e1230, doi:10.1053/j.gastro.2020.11.011 (2021).
- 7 Dvorak, H. F. Tumors: wounds that do not heal-redux. *Cancer Immunol Res* **3**, 1-11, doi:10.1158/2326-6066.CIR-14-0209 (2015).
- 8 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
- 9 Junttila, M. R. & de Sauvage, F. J. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* **501**, 346-354, doi:10.1038/nature12626 (2013).
- 10 Kalluri, R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer* **16**, 582-598, doi:10.1038/nrc.2016.73 (2016).
- 11 Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nat Rev Cancer* **6**, 392-401, doi:10.1038/nrc1877 (2006).
- 12 Neesse, A., Algul, H., Tuveson, D. A. & Gress, T. M. Stromal biology and therapy in pancreatic cancer: a changing paradigm. *Gut* **64**, 1476-1484, doi:10.1136/gutjnl-2015-309304 (2015).

- 13 Vennin, C. *et al.* Reshaping the Tumor Stroma for Treatment of Pancreatic Cancer. *Gastroenterology* **154**, 820-838, doi:10.1053/j.gastro.2017.11.280 (2018).
- 14 Gerling, M. *et al.* Stromal Hedgehog signalling is downregulated in colon cancer and its restoration restrains tumour growth. *Nat Commun* **7**, 12321, doi:10.1038/ncomms12321 (2016).
- 15 Lee, J. J. *et al.* Control of inflammation by stromal Hedgehog pathway activation restrains colitis. *Proc Natl Acad Sci U S A* **113**, E7545-E7553, doi:10.1073/pnas.1616447113 (2016).
- 16 Orimo, A. *et al.* Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**, 335-348, doi:10.1016/j.cell.2005.02.034 (2005).
- 17 Olumi, A. F. *et al.* Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res* **59**, 5002-5011, doi:10.1186/bcr138 (1999).
- 18 Quante, M. *et al.* Bone marrow-derived myofibroblasts contribute to the mesenchymal stem cell niche and promote tumor growth. *Cancer Cell* **19**, 257-272, doi:10.1016/j.ccr.2011.01.020 (2011).
- 19 Tauriello, D. V. F. *et al.* TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature*, doi:10.1038/nature25492 (2018).
- 20 Lannagan, T. R. M. *et al.* Genetic editing of colonic organoids provides a molecularly distinct and orthotopic preclinical model of serrated carcinogenesis. *Gut*, doi:10.1136/gutjnl-2017-315920 (2018).
- 21 Bhowmick, N. A., Neilson, E. G. & Moses, H. L. Stromal fibroblasts in cancer initiation and progression. *Nature* **432**, 332-337, doi:10.1038/nature03096 (2004).
- 22 Olive, K. P. *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* **324**, 1457-1461, doi:10.1126/science.1171362 (2009).
- 23 Provenzano, P. P. *et al.* Enzymatic targeting of the stroma ablates physical barriers to treatment of pancreatic ductal adenocarcinoma. *Cancer Cell* **21**, 418-429, doi:10.1016/j.ccr.2012.01.007 (2012).

- 24 Jacobetz, M. A. *et al.* Hyaluronan impairs vascular function and drug delivery in a mouse model of pancreatic cancer. *Gut* **62**, 112-120, doi:10.1136/gutjnl-2012-302529 (2013).
- 25 Feig, C. *et al.* Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. *Proc Natl Acad Sci U S A* **110**, 20212-20217, doi:10.1073/pnas.1320318110 (2013).
- 26 Aiello, N. M. *et al.* Metastatic progression is associated with dynamic changes in the local microenvironment. *Nat Commun* **7**, 12819, doi:10.1038/ncomms12819 (2016).
- 27 Cox, T. R. *et al.* LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis. *Cancer Res* **73**, 1721-1732, doi:10.1158/0008-5472.CAN-12-2233 (2013).
- 28 Peinado, H. *et al.* Pre-metastatic niches: organ-specific homes for metastases. *Nat Rev Cancer* **17**, 302-317, doi:10.1038/nrc.2017.6 (2017).
- 29 Bijlsma, M. F., Sadanandam, A., Tan, P. & Vermeulen, L. Molecular subtypes in cancers of the gastrointestinal tract. *Nat Rev Gastroenterol Hepatol* **14**, 333-342, doi:10.1038/nrgastro.2017.33 (2017).
- 30 Kumar, V. *et al.* Cancer-Associated Fibroblasts Neutralize the Anti-tumor Effect of CSF1 Receptor Blockade by Inducing PMN-MDSC Infiltration of Tumors. *Cancer Cell* **32**, 654-668 e655, doi:10.1016/j.ccell.2017.10.005 (2017).
- 31 Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350-1356, doi:10.1038/nm.3967 (2015).
- 32 Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet* **47**, 320-329, doi:10.1038/ng.3225 (2015).
- 33 Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* **47**, 1168-1178, doi:10.1038/ng.3398 (2015).
- 34 Ji, J. *et al.* Hepatic stellate cell and monocyte interaction contributes to poor prognosis in hepatocellular carcinoma. *Hepatology* **62**, 481-495, doi:10.1002/hep.27822 (2015).
- 35 Isella, C. *et al.* Stromal contribution to the colorectal cancer transcriptome. *Nat Genet* **47**, 312-319, doi:10.1038/ng.3224 (2015).

- 36 Infinity Pharmaceuticals. Infinity Reports Update from Phase 2 Study of Saridegib Plus Gemcitabine in Patients with Metastatic Pancreatic Cancer. <http://www.businesswire.com/news/home/20120127005146/en/Infinity-Reports-Update-Phase-2-Study-Saridegib#U-DoOICSy6w> (2012).
- 37 Bailey, J. M. *et al.* Sonic hedgehog promotes desmoplasia in pancreatic cancer. *Clin Cancer Res* **14**, 5995-6004, doi:10.1158/1078-0432.CCR-08-0291 (2008).
- 38 Neesse, A. *et al.* Stromal biology and therapy in pancreatic cancer: ready for clinical translation? *Gut*, doi:10.1136/gutjnl-2018-316451 (2018).
- 39 Pallangyo, C. K., Ziegler, P. K. & Greten, F. R. IKKbeta acts as a tumor suppressor in cancer-associated fibroblasts during intestinal tumorigenesis. *J Exp Med* **212**, 2253-2266, doi:10.1084/jem.20150576 (2015).
- 40 Maris, P. *et al.* Asporin Is a Fibroblast-Derived TGF-beta1 Inhibitor and a Tumor Suppressor Associated with Good Prognosis in Breast Cancer. *PLoS Med* **12**, e1001871, doi:10.1371/journal.pmed.1001871 (2015).
- 41 Zhang, J. *et al.* Fibroblast-specific protein 1/S100A4-positive cells prevent carcinoma through collagen production and encapsulation of carcinogens. *Cancer Res* **73**, 2770-2781, doi:10.1158/0008-5472.CAN-12-3022 (2013).
- 42 Gore, J. & Korc, M. Pancreatic cancer stroma: friend or foe? *Cancer Cell* **25**, 711-712, doi:10.1016/j.ccr.2014.05.026 (2014).
- 43 Tarin, D. & Croft, C. B. Ultrastructural features of wound healing in mouse skin. *J Anat* **105**, 189-190 (1969).
- 44 Furuya, S. & Furuya, K. Subepithelial fibroblasts in intestinal villi: roles in intercellular communication. *Int Rev Cytol* **264**, 165-223, doi:10.1016/S0074-7696(07)64004-2 (2007).
- 45 Shoshkes-Carmel, M. *et al.* Subepithelial telocytes are an important source of Wnts that supports intestinal crypts. *Nature*, doi:10.1038/s41586-018-0084-4 (2018).
- 46 Worthley, D. L. *et al.* Gremlin 1 identifies a skeletal stem cell with bone, cartilage, and reticular stromal potential. *Cell* **160**, 269-284, doi:10.1016/j.cell.2014.11.042 (2015).
- 47 Degirmenci, B., Valenta, T., Dimitrieva, S., Hausmann, G. & Basler, K. GLI1-expressing

- mesenchymal cells form the essential Wnt-secreting niche for colon stem cells. *Nature*, doi:10.1038/s41586-018-0190-3 (2018).
- 48 Powell, D. W., Pinchuk, I. V., Saada, J. I., Chen, X. & Mifflin, R. C. Mesenchymal cells of the intestinal lamina propria. *Annu Rev Physiol* **73**, 213-237, doi:10.1146/annurev.physiol.70.113006.100646 (2011).
- 49 Koliaraki, V., Pallangyo, C. K., Greten, F. R. & Kollias, G. Mesenchymal Cells in Colon Cancer. *Gastroenterology* **152**, 964-979, doi:10.1053/j.gastro.2016.11.049 (2017).
- 50 Biswas, S. *et al.* Microenvironmental control of stem cell fate in intestinal homeostasis and disease. *J Pathol* **237**, 135-145, doi:10.1002/path.4563 (2015).
- 51 Ohlund, D., Elyada, E. & Tuveson, D. Fibroblast heterogeneity in the cancer wound. *J Exp Med* **211**, 1503-1523, doi:10.1084/jem.20140692 (2014).
- 52 Sugimoto, H., Mundel, T. M., Kieran, M. W. & Kalluri, R. Identification of fibroblast heterogeneity in the tumor microenvironment. *Cancer Biology & Therapy* **5**, 1640-1646, doi:10.4161/cbt.5.12.3354 (2006).
- 53 Madar, S., Goldstein, I. & Rotter, V. 'Cancer associated fibroblasts'-more than meets the eye. *Trends Mol Med* **19**, 447-453, doi:10.1016/j.molmed.2013.05.004 (2013).
- 54 Ohlund, D. *et al.* Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med* **214**, 579-596, doi:10.1084/jem.20162024 (2017).
- 55 Arnold, J. N., Magiera, L., Kraman, M. & Fearon, D. T. Tumoral immune suppression by macrophages expressing fibroblast activation protein- $\alpha$  and heme oxygenase-1. *Cancer Immunol Res* **2**, 121-126, doi:10.1158/2326-6066.CIR-13-0150 (2014).
- 56 Tchou, J. *et al.* Fibroblast activation protein expression by stromal cells and tumor-associated macrophages in human breast cancer. *Hum Pathol* **44**, 2549-2557, doi:10.1016/j.humpath.2013.06.016 (2013).
- 57 Okada, H., Danoff, T. M., Kalluri, R. & Neilson, E. G. Early role of Fsp1 in epithelial-mesenchymal transformation. *Am J Physiol* **273**, F563-574, doi:10.1152/ajprenal.1997.273.4.F563 (1997).
- 58 Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis

- but contributes to chemoresistance. *Nature* **527**, 472-476, doi:10.1038/nature15748 (2015).
- 59 Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349-361, doi:10.1016/j.cell.2011.11.025 (2012).
- 60 Osterreicher, C. H. *et al.* Fibroblast-specific protein 1 identifies an inflammatory subpopulation of macrophages in the liver. *Proc Natl Acad Sci U S A* **108**, 308-313, doi:10.1073/pnas.1017547108 (2011).
- 61 Costa, A. *et al.* Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *Cancer Cell*, doi:10.1016/j.ccell.2018.01.011 (2018).
- 62 Yamamura, Y. *et al.* Akt-Girdin signaling in cancer-associated fibroblasts contributes to tumor progression. *Cancer Res* **75**, 813-823, doi:10.1158/0008-5472.CAN-14-1317 (2015).
- 63 Worthley, D. L. *et al.* Bone marrow cells as precursors of the tumor stroma. *Exp Cell Res* **319**, 1650-1656, doi:10.1016/j.yexcr.2013.03.006 (2013).
- 64 Koliaraki, V., Pasparakis, M. & Kollias, G. IKKbeta in intestinal mesenchymal cells promotes initiation of colitis-associated cancer. *J Exp Med* **212**, 2235-2251, doi:10.1084/jem.20150542 (2015).
- 65 Worthley, D. L., Giraud, A. S. & Wang, T. C. Stromal fibroblasts in digestive cancer. *Cancer Microenviron* **3**, 117-125, doi:10.1007/s12307-009-0033-8 (2010).
- 66 Gaggioli, C. *et al.* Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. *Nat Cell Biol* **9**, 1392-1400, doi:10.1038/ncb1658 (2007).
- 67 Catenacci, D. V. *et al.* Randomized Phase Ib/II Study of Gemcitabine Plus Placebo or Vismodegib, a Hedgehog Pathway Inhibitor, in Patients With Metastatic Pancreatic Cancer. *J Clin Oncol* **33**, 4284-4292, doi:10.1200/JCO.2015.62.8719 (2015).
- 68 Lee, J. J. *et al.* Stromal response to Hedgehog signaling restrains pancreatic cancer progression. *Proc Natl Acad Sci U S A* **111**, E3091-3100, doi:10.1073/pnas.1411679111 (2014).
- 69 Roberts, K. J., Kershner, A. M. & Beachy, P. A. The Stromal Niche for Epithelial Stem Cells: A Template for Regeneration and a Brake on Malignancy. *Cancer Cell* **32**, 404-410,

doi:10.1016/j.ccell.2017.08.007 (2017).

- 70 Erez, N., Truitt, M., Olson, P., Arron, S. T. & Hanahan, D. Cancer-Associated Fibroblasts Are Activated in Incipient Neoplasia to Orchestrate Tumor-Promoting Inflammation in an NF-kappaB-Dependent Manner. *Cancer Cell* **17**, 135-147, doi:10.1016/j.ccr.2009.12.041 (2010).
- 71 Satoyoshi, R., Kuriyama, S., Aiba, N., Yashiro, M. & Tanaka, M. Asporin activates coordinated invasion of scirrhous gastric cancer and cancer-associated fibroblasts. *Oncogene* **34**, 650-660, doi:10.1038/onc.2013.584 (2015).
- 72 Lo, A. *et al.* Tumor-Promoting Desmoplasia Is Disrupted by Depleting FAP-Expressing Stromal Cells. *Cancer Res* **75**, 2800-2810, doi:10.1158/0008-5472.CAN-14-3041 (2015).
- 73 Kraman, M. *et al.* Suppression of antitumor immunity by stromal cells expressing fibroblast activation protein-alpha. *Science* **330**, 827-830, doi:10.1126/science.1195300 (2010).
- 74 Lo, A. *et al.* Fibroblast activation protein augments progression and metastasis of pancreatic ductal adenocarcinoma. *JCI Insight* **2**, doi:10.1172/jci.insight.92232 (2017).
- 75 Santos, A. M., Jung, J., Aziz, N., Kissil, J. L. & Pure, E. Targeting fibroblast activation protein inhibits tumor stromagenesis and growth in mice. *J Clin Invest* **119**, 3613-3625, doi:10.1172/JCI38988 (2009).
- 76 Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328-337, doi:10.1038/nature12624 (2013).
- 77 Alkasalias, T., Moyano-Galceran, L., Arsenian-Henriksson, M. & Lehti, K. Fibroblasts in the Tumor Microenvironment: Shield or Spear? *Int J Mol Sci* **19**, doi:10.3390/ijms19051532 (2018).
- 78 Biffi, G. *et al.* IL-1-induced JAK/STAT signaling is antagonized by TGF-beta to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. *Cancer Discov*, doi:10.1158/2159-8290.CD-18-0710 (2018).
- 79 Potenta, S., Zeisberg, E. & Kalluri, R. The role of endothelial-to-mesenchymal transition in cancer progression. *Br J Cancer* **99**, 1375-1379, doi:10.1038/sj.bjc.6604662 (2008).
- 80 Li, Y., Wang, J. & Asahina, K. Mesothelial cells give rise to hepatic stellate cells and myofibroblasts via mesothelial-mesenchymal transition in liver injury. *Proc Natl Acad Sci U S A* **110**, 2324-2329, doi:10.1073/pnas.1214136110 (2013).

- 81 Rinkevich, Y. *et al.* Identification and prospective isolation of a mesothelial precursor lineage giving rise to smooth muscle cells and fibroblasts for mammalian internal organs, and their vasculature. *Nat Cell Biol* **14**, 1251-1260, doi:10.1038/ncb2610 (2012).
- 82 Wilm, B., Ipenberg, A., Hastie, N. D., Burch, J. B. & Bader, D. M. The serosal mesothelium is a major source of smooth muscle cells of the gut vasculature. *Development* **132**, 5317-5328, doi:10.1242/dev.02141 (2005).
- 83 Hosaka, K. *et al.* Pericyte-fibroblast transition promotes tumor growth and metastasis. *Proc Natl Acad Sci U S A* **113**, E5618-5627, doi:10.1073/pnas.1608384113 (2016).
- 84 Arina, A. *et al.* Tumor-associated fibroblasts predominantly come from local and not circulating precursors. *Proc Natl Acad Sci U S A* **113**, 7551-7556, doi:10.1073/pnas.1600363113 (2016).
- 85 Ren, G. *et al.* CCR2-dependent recruitment of macrophages by tumor-educated mesenchymal stromal cells promotes tumor development and is mimicked by TNFalpha. *Cell Stem Cell* **11**, 812-824, doi:10.1016/j.stem.2012.08.013 (2012).
- 86 Worthley, D. L. *et al.* Human gastrointestinal neoplasia-associated myofibroblasts can develop from bone marrow-derived cells following allogeneic stem cell transplantation. *Stem Cells* **27**, 1463-1468, doi:10.1002/stem.63 (2009).
- 87 Shi, Y., Du, L., Lin, L. & Wang, Y. Tumour-associated mesenchymal stem/stromal cells: emerging therapeutic targets. *Nat Rev Drug Discov* **16**, 35-52, doi:10.1038/nrd.2016.193 (2017).
- 88 Jung, Y. *et al.* Recruitment of mesenchymal stem cells into prostate tumours promotes metastasis. *Nat Commun* **4**, 1795, doi:10.1038/ncomms2766 (2013).
- 89 Karnoub, A. E. *et al.* Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. *Nature* **449**, 557-563, doi:10.1038/nature06188 (2007).
- 90 Fujiwara, K. *et al.* CD271(+) subpopulation of pancreatic stellate cells correlates with prognosis of pancreatic cancer and is regulated by interaction with cancer cells. *PLoS One* **7**, e52682, doi:10.1371/journal.pone.0052682 (2012).
- 91 Labernadie, A. *et al.* A mechanically active heterotypic E-cadherin/N-cadherin adhesion enables fibroblasts to drive cancer cell invasion. *Nat Cell Biol* **19**, 224-237,

doi:10.1038/ncb3478 (2017).

- 92 Lau, E. Y. *et al.* Cancer-Associated Fibroblasts Regulate Tumor-Initiating Cell Plasticity in Hepatocellular Carcinoma through c-Met/FRA1/HEY1 Signaling. *Cell Rep* **15**, 1175-1189, doi:10.1016/j.celrep.2016.04.019 (2016).
- 93 Vermeulen, L. *et al.* Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat Cell Biol* **12**, 468-476, doi:10.1038/ncb2048 (2010).
- 94 Luraghi, P. *et al.* MET signaling in colon cancer stem-like cells blunts the therapeutic response to EGFR inhibitors. *Cancer Res* **74**, 1857-1869, doi:10.1158/0008-5472.CAN-13-2340-T (2014).
- 95 Rhee, H. *et al.* Keratin 19 expression in hepatocellular carcinoma is regulated by fibroblast-derived HGF via a MET-ERK1/2-AP1 and SP1 axis. *Cancer Res*, doi:10.1158/0008-5472.CAN-17-0988 (2018).
- 96 Claperon, A. *et al.* Hepatic myofibroblasts promote the progression of human cholangiocarcinoma through activation of epidermal growth factor receptor. *Hepatology* **58**, 2001-2011, doi:10.1002/hep.26585 (2013).
- 97 Neufert, C. *et al.* Tumor fibroblast-derived epiregulin promotes growth of colitis-associated neoplasms through ERK. *J Clin Invest* **123**, 1428-1443, doi:10.1172/JCI63748 (2013).
- 98 Chen, W. J. *et al.* Cancer-associated fibroblasts regulate the plasticity of lung cancer stemness via paracrine signalling. *Nat Commun* **5**, 3472, doi:10.1038/ncomms4472 (2014).
- 99 Vaquero, J. *et al.* IGF2/IR/IGF1R pathway in tumor cells and myofibroblasts mediates resistance to EGFR inhibition in cholangiocarcinoma. *Clin Cancer Res*, doi:10.1158/1078-0432.CCR-17-3725 (2018).
- 100 Grivennikov, S. *et al.* IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer. *Cancer Cell* **15**, 103-113, doi:10.1016/j.ccr.2009.01.001 (2009).
- 101 Bollrath, J. *et al.* gp130-mediated Stat3 activation in enterocytes regulates cell survival and cell-cycle progression during colitis-associated tumorigenesis. *Cancer Cell* **15**, 91-102, doi:10.1016/j.ccr.2009.01.002 (2009).

- 102 Putoczki, T. L. *et al.* Interleukin-11 is the dominant IL-6 family cytokine during gastrointestinal tumorigenesis and can be targeted therapeutically. *Cancer Cell* **24**, 257-271, doi:10.1016/j.ccr.2013.06.017 (2013).
- 103 Calon, A. *et al.* Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. *Cancer Cell* **22**, 571-584, doi:10.1016/j.ccr.2012.08.013 (2012).
- 104 Su, S. *et al.* CD10(+)GPR77(+) Cancer-Associated Fibroblasts Promote Cancer Formation and Chemoresistance by Sustaining Cancer Stemness. *Cell* **172**, 841-856 e816, doi:10.1016/j.cell.2018.01.009 (2018).
- 105 Sneddon, J. B. *et al.* Bone morphogenetic protein antagonist gremlin 1 is widely expressed by cancer-associated stromal cells and can promote tumor cell proliferation. *Proc Natl Acad Sci U S A* **103**, 14842-14847, doi:10.1073/pnas.0606857103 (2006).
- 106 Davis, H. *et al.* Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche. *Nat Med* **21**, 62-70, doi:10.1038/nm.3750 (2015).
- 107 Ruivo, C. F., Adem, B., Silva, M. & Melo, S. A. The Biology of Cancer Exosomes: Insights and New Perspectives. *Cancer Res* **77**, 6480-6488, doi:10.1158/0008-5472.CAN-17-0994 (2017).
- 108 Gu, J. *et al.* Gastric cancer exosomes trigger differentiation of umbilical cord derived mesenchymal stem cells to carcinoma-associated fibroblasts through TGF-beta/Smad pathway. *PLoS One* **7**, e52465, doi:10.1371/journal.pone.0052465 (2012).
- 109 Fang, T. *et al.* Tumor-derived exosomal miR-1247-3p induces cancer-associated fibroblast activation to foster lung metastasis of liver cancer. *Nat Commun* **9**, 191, doi:10.1038/s41467-017-02583-0 (2018).
- 110 Luga, V. *et al.* Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. *Cell* **151**, 1542-1556, doi:10.1016/j.cell.2012.11.024 (2012).
- 111 Shimoda, M. *et al.* Loss of the Timp gene family is sufficient for the acquisition of the CAF-like cell state. *Nat Cell Biol* **16**, 889-901, doi:10.1038/ncb3021 (2014).
- 112 Richards, K. E. *et al.* Cancer-associated fibroblast exosomes regulate survival and proliferation of pancreatic cancer cells. *Oncogene* **36**, 1770-1778, doi:10.1038/onc.2016.353 (2017).

- 113 Wang, X., Enomoto, A., Asai, N., Kato, T. & Takahashi, M. Collective invasion of cancer: Perspectives from pathology and development. *Pathol Int* **66**, 183-192, doi:10.1111/pin.12391 (2016).
- 114 Turley, S. J., Cremasco, V. & Astarita, J. L. Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat Rev Immunol* **15**, 669-682, doi:10.1038/nri3902 (2015).
- 115 Salmon, H. *et al.* Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. *J Clin Invest* **122**, 899-910, doi:10.1172/JCI45817 (2012).
- 116 Yang, X. *et al.* FAP Promotes Immunosuppression by Cancer-Associated Fibroblasts in the Tumor Microenvironment via STAT3-CCL2 Signaling. *Cancer Res* **76**, 4124-4135, doi:10.1158/0008-5472.CAN-15-2973 (2016).
- 117 Tan, W. *et al.* Tumour-infiltrating regulatory T cells stimulate mammary cancer metastasis through RANKL-RANK signalling. *Nature* **470**, 548-553, doi:10.1038/nature09707 (2011).
- 118 Cannarile, M. A. *et al.* Colony-stimulating factor 1 receptor (CSF1R) inhibitors in cancer therapy. *J Immunother Cancer* **5**, 53, doi:10.1186/s40425-017-0257-y (2017).
- 119 Ene-Obong, A. *et al.* Activated pancreatic stellate cells sequester CD8<sup>+</sup> T cells to reduce their infiltration of the juxtatumoral compartment of pancreatic ductal adenocarcinoma. *Gastroenterology* **145**, 1121-1132, doi:10.1053/j.gastro.2013.07.025 (2013).
- 120 Carstens, J. L. *et al.* Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. *Nat Commun* **8**, 15095, doi:10.1038/ncomms15095 (2017).
- 121 Nazareth, M. R. *et al.* Characterization of Human Lung Tumor-Associated Fibroblasts and Their Ability to Modulate the Activation of Tumor-Associated T Cells. *The Journal of Immunology* **178**, 5552-5562, doi:10.4049/jimmunol.178.9.5552 (2007).
- 122 Pinchuk, I. V. *et al.* PD-1 ligand expression by human colonic myofibroblasts/fibroblasts regulates CD4<sup>+</sup> T-cell activity. *Gastroenterology* **135**, 1228-1237, 1237 e1221-1222, doi:10.1053/j.gastro.2008.07.016 (2008).
- 123 Comito, G. *et al.* Cancer-associated fibroblasts and M2-polarized macrophages synergize during prostate carcinoma progression. *Oncogene* **33**, 2423-2431, doi:10.1038/onc.2013.191

- (2014).
- 124 Kim, J. H. *et al.* The role of myofibroblasts in upregulation of S100A8 and S100A9 and the differentiation of myeloid cells in the colorectal cancer microenvironment. *Biochem Biophys Res Commun* **423**, 60-66, doi:10.1016/j.bbrc.2012.05.081 (2012).
- 125 Nielsen, S. R. *et al.* Macrophage-secreted granulins supports pancreatic cancer metastasis by inducing liver fibrosis. *Nat Cell Biol* **18**, 549-560, doi:10.1038/ncb3340 (2016).
- 126 Mariathasan, S. *et al.* TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*, doi:10.1038/nature25501 (2018).
- 127 Thomas, D. A. & Massague, J. TGF-beta directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. *Cancer Cell* **8**, 369-380, doi:10.1016/j.ccr.2005.10.012 (2005).
- 128 US National Library of Medicine. Clinicaltrials.gov. <https://clinicaltrials.gov/ct2/show/NCT02734160> (2016).
- 129 De Palma, M., Biziato, D. & Petrova, T. V. Microenvironmental regulation of tumour angiogenesis. *Nat Rev Cancer* **17**, 457-474, doi:10.1038/nrc.2017.51 (2017).
- 130 Fukumura, D. *et al.* Tumor induction of VEGF promoter activity in stromal cells. *Cell* **94**, 715-725, doi:10.1016/s0092-8674(00)81731-6 (1998).
- 131 Pietras, K., Pahler, J., Bergers, G. & Hanahan, D. Functions of paracrine PDGF signaling in the proangiogenic tumor stroma revealed by pharmacological targeting. *PLoS Med* **5**, e19, doi:10.1371/journal.pmed.0050019 (2008).
- 132 Anderberg, C. *et al.* Paracrine signaling by platelet-derived growth factor-CC promotes tumor growth by recruitment of cancer-associated fibroblasts. *Cancer Res* **69**, 369-378, doi:10.1158/0008-5472.CAN-08-2724 (2009).
- 133 Crawford, Y. *et al.* PDGF-C mediates the angiogenic and tumorigenic properties of fibroblasts associated with tumors refractory to anti-VEGF treatment. *Cancer Cell* **15**, 21-34, doi:10.1016/j.ccr.2008.12.004 (2009).
- 134 Yan, M. & Jurasz, P. The role of platelets in the tumor microenvironment: From solid tumors to leukemia. *Biochim Biophys Acta* **1863**, 392-400, doi:10.1016/j.bbamcr.2015.07.008 (2016).
- 135 Rupp, T. *et al.* Tenascin-C Orchestrates Glioblastoma Angiogenesis by Modulation of Pro- and

- Anti-angiogenic Signaling. *Cell Rep* **17**, 2607-2619, doi:10.1016/j.celrep.2016.11.012 (2016).
- 136 Mitsuhashi, A. *et al.* Fibrocyte-like cells mediate acquired resistance to anti-angiogenic therapy with bevacizumab. *Nat Commun* **6**, 8792, doi:10.1038/ncomms9792 (2015).
- 137 Leung, C. S. *et al.* Cancer-associated fibroblasts regulate endothelial adhesion protein LPP to promote ovarian cancer chemoresistance. *J Clin Invest* **128**, 589-606, doi:10.1172/JCI95200 (2018).
- 138 Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* **15**, 786-801, doi:10.1038/nrm3904 (2014).
- 139 Calvo, F. *et al.* Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nat Cell Biol* **15**, 637-646, doi:10.1038/ncb2756 (2013).
- 140 Zanconato, F., Cordenonsi, M. & Piccolo, S. YAP/TAZ at the Roots of Cancer. *Cancer Cell* **29**, 783-803, doi:10.1016/j.ccell.2016.05.005 (2016).
- 141 Pickup, M. W., Mouw, J. K. & Weaver, V. M. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep* **15**, 1243-1253, doi:10.15252/embr.201439246 (2014).
- 142 Levental, K. R. *et al.* Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139**, 891-906, doi:10.1016/j.cell.2009.10.027 (2009).
- 143 Rice, A. J. *et al.* Matrix stiffness induces epithelial-mesenchymal transition and promotes chemoresistance in pancreatic cancer cells. *Oncogenesis* **6**, e352, doi:10.1038/oncsis.2017.54 (2017).
- 144 Schrader, J. *et al.* Matrix stiffness modulates proliferation, chemotherapeutic response, and dormancy in hepatocellular carcinoma cells. *Hepatology* **53**, 1192-1205, doi:10.1002/hep.24108 (2011).
- 145 Bordeleau, F. *et al.* Matrix stiffening promotes a tumor vasculature phenotype. *Proc Natl Acad Sci U S A* **114**, 492-497, doi:10.1073/pnas.1613855114 (2017).
- 146 Veenstra, V. L. *et al.* Stromal SPOCK1 supports invasive pancreatic cancer growth. *Mol Oncol* **11**, 1050-1064, doi:10.1002/1878-0261.12073 (2017).
- 147 Panciera, T., Azzolin, L., Cordenonsi, M. & Piccolo, S. Mechanobiology of YAP and TAZ in

- physiology and disease. *Nat Rev Mol Cell Biol* **18**, 758-770, doi:10.1038/nrm.2017.87 (2017).
- 148 Barker, H. E., Cox, T. R. & Epler, J. T. The rationale for targeting the LOX family in cancer. *Nat Rev Cancer* **12**, 540-552, doi:10.1038/nrc3319 (2012).
- 149 Ishihara, S., Inman, D. R., Li, W. J., Ponik, S. M. & Keely, P. J. Mechano-Signal Transduction in Mesenchymal Stem Cells Induces Prosaposin Secretion to Drive the Proliferation of Breast Cancer Cells. *Cancer Res* **77**, 6179-6189, doi:10.1158/0008-5472.CAN-17-0569 (2017).
- 150 Miller, B. W. *et al.* Targeting the LOX/hypoxia axis reverses many of the features that make pancreatic cancer deadly: inhibition of LOX abrogates metastasis and enhances drug efficacy. *EMBO Mol Med* **7**, 1063-1076, doi:10.15252/emmm.201404827 (2015).
- 151 Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer* **13**, 714-726, doi:10.1038/nrc3599 (2013).
- 152 Hessmann, E. *et al.* Fibroblast drug scavenging increases intratumoural gemcitabine accumulation in murine pancreas cancer. *Gut* **67**, 497-507, doi:10.1136/gutjnl-2016-311954 (2018).
- 153 Roodhart, J. M. *et al.* Mesenchymal stem cells induce resistance to chemotherapy through the release of platinum-induced fatty acids. *Cancer Cell* **20**, 370-383, doi:10.1016/j.ccr.2011.08.010 (2011).
- 154 Duluc, C. *et al.* Pharmacological targeting of the protein synthesis mTOR/4E-BP1 pathway in cancer-associated fibroblasts abrogates pancreatic tumour chemoresistance. *EMBO Mol Med* **7**, 735-753, doi:10.15252/emmm.201404346 (2015).
- 155 Lotti, F. *et al.* Chemotherapy activates cancer-associated fibroblasts to maintain colorectal cancer-initiating cells by IL-17A. *J Exp Med* **210**, 2851-2872, doi:10.1084/jem.20131195 (2013).
- 156 Ireland, L. *et al.* Chemoresistance in Pancreatic Cancer Is Driven by Stroma-Derived Insulin-Like Growth Factors. *Cancer Res* **76**, 6851-6863, doi:10.1158/0008-5472.CAN-16-1201 (2016).
- 157 Mürköster, S. *et al.* Tumor Stroma Interactions Induce Chemoresistance in Pancreatic Ductal Carcinoma Cells Involving Increased Secretion and Paracrine Effects of Nitric Oxide and

- Interleukin-1 $\beta$ . *Cancer Research* **64**, 1331-1337, doi:10.1158/0008-5472.Can-03-1860 (2004).
- 158 Wang, W. *et al.* Effector T Cells Abrogate Stroma-Mediated Chemoresistance in Ovarian Cancer. *Cell* **165**, 1092-1105, doi:10.1016/j.cell.2016.04.009 (2016).
- 159 Zhang, H. *et al.* Cancer associated fibroblasts-promoted LncRNA DN3OS confers radioresistance by regulating DNA damage response in esophageal squamous cell carcinoma. *Clin Cancer Res*, doi:10.1158/1078-0432.CCR-18-0773 (2018).
- 160 Barker, H. E., Paget, J. T., Khan, A. A. & Harrington, K. J. The tumour microenvironment after radiotherapy: mechanisms of resistance and recurrence. *Nat Rev Cancer* **15**, 409-425, doi:10.1038/nrc3958 (2015).
- 161 Tommelein, J. *et al.* Radiotherapy-Activated Cancer-Associated Fibroblasts Promote Tumor Progression through Paracrine IGF1R Activation. *Cancer Res* **78**, 659-670, doi:10.1158/0008-5472.CAN-17-0524 (2018).
- 162 Paulsson, J. & Micke, P. Prognostic relevance of cancer-associated fibroblasts in human cancer. *Semin Cancer Biol* **25**, 61-68, doi:10.1016/j.semcancer.2014.02.006 (2014).
- 163 Ao, Z. *et al.* Identification of Cancer-Associated Fibroblasts in Circulating Blood from Patients with Metastatic Breast Cancer. *Cancer Res* **75**, 4681-4687, doi:10.1158/0008-5472.CAN-15-1633 (2015).
- 164 Resovi, A. *et al.* Soluble stroma-related biomarkers of pancreatic cancer. *EMBO Mol Med*, doi:10.15252/emmm.201708741 (2018).
- 165 Torres, S. *et al.* Proteome profiling of cancer-associated fibroblasts identifies novel proinflammatory signatures and prognostic markers for colorectal cancer. *Clin Cancer Res* **19**, 6006-6019, doi:10.1158/1078-0432.CCR-13-1130 (2013).
- 166 Zhang, D. Y. *et al.* A hepatic stellate cell gene expression signature associated with outcomes in hepatitis C cirrhosis and hepatocellular carcinoma after curative resection. *Gut* **65**, 1754-1764, doi:10.1136/gutjnl-2015-309655 (2016).
- 167 Huijbers, A. *et al.* The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial. *Ann Oncol* **24**, 179-185, doi:10.1093/annonc/mds246 (2013).

- 168 Tsujino, T. *et al.* Stromal myofibroblasts predict disease recurrence for colorectal cancer. *Clin Cancer Res* **13**, 2082-2090, doi:10.1158/1078-0432.CCR-06-2191 (2007).
- 169 Erkan, M. *et al.* The activated stroma index is a novel and independent prognostic marker in pancreatic ductal adenocarcinoma. *Clin Gastroenterol Hepatol* **6**, 1155-1161, doi:10.1016/j.cgh.2008.05.006 (2008).
- 170 Liao, R. *et al.* Clinical significance and gene expression study of human hepatic stellate cells in HBV related-hepatocellular carcinoma. *J Exp Clin Cancer Res* **32**, 22, doi:10.1186/1756-9966-32-22 (2013).
- 171 Loeffler, M., Kruger, J. A., Niethammer, A. G. & Reisfeld, R. A. Targeting tumor-associated fibroblasts improves cancer chemotherapy by increasing intratumoral drug uptake. *J Clin Invest* **116**, 1955-1962, doi:10.1172/JCI26532 (2006).
- 172 Sherman, M. H. *et al.* Vitamin D receptor-mediated stromal reprogramming suppresses pancreatitis and enhances pancreatic cancer therapy. *Cell* **159**, 80-93, doi:10.1016/j.cell.2014.08.007 (2014).
- 173 Ferrer-Mayorga, G. *et al.* Vitamin D receptor expression and associated gene signature in tumour stromal fibroblasts predict clinical outcome in colorectal cancer. *Gut* **66**, 1449-1462, doi:10.1136/gutjnl-2015-310977 (2017).
- 174 Froeling, F. E. *et al.* Retinoic acid-induced pancreatic stellate cell quiescence reduces paracrine Wnt-beta-catenin signaling to slow tumor progression. *Gastroenterology* **141**, 1486-1497, doi:10.1053/j.gastro.2011.06.047 (2011).
- 175 Carapuca, E. F. *et al.* Anti-stromal treatment together with chemotherapy targets multiple signalling pathways in pancreatic adenocarcinoma. *J Pathol* **239**, 286-296, doi:10.1002/path.4727 (2016).
- 176 US National Library of Medicine. Clinicaltrials.gov. <https://clinicaltrials.gov/ct2/show/NCT03331562> (2017).
- 177 Ostermann, E. *et al.* Effective immunoconjugate therapy in cancer models targeting a serine protease of tumor fibroblasts. *Clin Cancer Res* **14**, 4584-4592, doi:10.1158/1078-0432.CCR-07-5211 (2008).

- 178 Waldhauer, I. *et al.* Novel Tumor-Targeted, Engineered IL-2 Variant (IL2v)-Based Immunocytokines For Immunotherapy Of Cancer. *Blood* **122**, 2278-2278 (2013).
- 179 US National Library of Medicine. Clinicaltrials.gov. <https://clinicaltrials.gov/ct2/show/NCT02627274> (2015).
- 180 Xu, C. *et al.* Interferon-alpha-secreting mesenchymal stem cells exert potent antitumor effect in vivo. *Oncogene* **33**, 5047-5052, doi:10.1038/onc.2013.458 (2014).
- 181 Loebinger, M. R., Eddaoudi, A., Davies, D. & Janes, S. M. Mesenchymal stem cell delivery of TRAIL can eliminate metastatic cancer. *Cancer Res* **69**, 4134-4142, doi:10.1158/0008-5472.CAN-08-4698 (2009).
- 182 Grisendi, G. *et al.* Adipose-derived mesenchymal stem cells as stable source of tumor necrosis factor-related apoptosis-inducing ligand delivery for cancer therapy. *Cancer Res* **70**, 3718-3729, doi:10.1158/0008-5472.CAN-09-1865 (2010).
- 183 Lee, R. H., Yoon, N., Reneau, J. C. & Prockop, D. J. Preactivation of human MSCs with TNF-alpha enhances tumor-suppressive activity. *Cell Stem Cell* **11**, 825-835, doi:10.1016/j.stem.2012.10.001 (2012).
- 184 von Einem, J. C. *et al.* Treatment of advanced gastrointestinal cancer with genetically modified autologous mesenchymal stem cells - TREAT-ME-1 - a phase I, first in human, first in class trial. *Oncotarget* **8**, 80156-80166, doi:10.18632/oncotarget.20964 (2017).
- 185 Theiss, A. L., Simmons, J. G., Jobin, C. & Lund, P. K. Tumor necrosis factor (TNF) alpha increases collagen accumulation and proliferation in intestinal myofibroblasts via TNF receptor 2. *J Biol Chem* **280**, 36099-36109, doi:10.1074/jbc.M505291200 (2005).
- 186 Tape, C. J. *et al.* Oncogenic KRAS Regulates Tumor Cell Signaling via Stromal Reciprocation. *Cell* **165**, 910-920, doi:10.1016/j.cell.2016.03.029 (2016).
- 187 McLin, V. A., Henning, S. J. & Jamrich, M. The role of the visceral mesoderm in the development of the gastrointestinal tract. *Gastroenterology* **136**, 2074-2091, doi:10.1053/j.gastro.2009.03.001 (2009).
- 188 Albregues, J. *et al.* Epigenetic switch drives the conversion of fibroblasts into proinvasive cancer-associated fibroblasts. *Nat Commun* **6**, 10204, doi:10.1038/ncomms10204 (2015).

- 189 Zhang, D. *et al.* Metabolic reprogramming of cancer-associated fibroblasts by IDH3alpha downregulation. *Cell Rep* **10**, 1335-1348, doi:10.1016/j.celrep.2015.02.006 (2015).
- 190 El Agha, E. *et al.* Mesenchymal Stem Cells in Fibrotic Disease. *Cell Stem Cell* **21**, 166-177, doi:10.1016/j.stem.2017.07.011 (2017).
- 191 Nombela-Arrieta, C., Ritz, J. & Silberstein, L. E. The elusive nature and function of mesenchymal stem cells. *Nat Rev Mol Cell Biol* **12**, 126-131, doi:10.1038/nrm3049 (2011).
- 192 Crisan, M. *et al.* A perivascular origin for mesenchymal stem cells in multiple human organs. *Cell Stem Cell* **3**, 301-313, doi:10.1016/j.stem.2008.07.003 (2008).
- 193 Kfoury, Y. & Scadden, D. T. Mesenchymal cell contributions to the stem cell niche. *Cell Stem Cell* **16**, 239-253, doi:10.1016/j.stem.2015.02.019 (2015).
- 194 Dominici, M. *et al.* Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* **8**, 315-317, doi:10.1080/14653240600855905 (2006).
- 195 Kramann, R. *et al.* Perivascular Gli1+ progenitors are key contributors to injury-induced organ fibrosis. *Cell Stem Cell* **16**, 51-66, doi:10.1016/j.stem.2014.11.004 (2015).
- 196 LeBleu, V. S. *et al.* Origin and function of myofibroblasts in kidney fibrosis. *Nat Med* **19**, 1047-1053, doi:10.1038/nm.3218 (2013).
- 197 Conklin, L. S., Hanley, P. J., Galipeau, J., Barrett, J. & Bollard, C. M. Intravenous mesenchymal stromal cell therapy for inflammatory bowel disease: Lessons from the acute graft versus host disease experience. *Cytotherapy* **19**, 655-667, doi:10.1016/j.jcyt.2017.03.006 (2017).
- 198 Manieri, N. A. *et al.* Mucosally transplanted mesenchymal stem cells stimulate intestinal healing by promoting angiogenesis. *J Clin Invest* **125**, 3606-3618, doi:10.1172/JCI81423 (2015).
- 199 Torres, S. *et al.* LOXL2 Is Highly Expressed in Cancer-Associated Fibroblasts and Associates to Poor Colon Cancer Survival. *Clin Cancer Res* **21**, 4892-4902, doi:10.1158/1078-0432.CCR-14-3096 (2015).
- 200 Weissmueller, S. *et al.* Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor beta signaling. *Cell* **157**, 382-394, doi:10.1016/j.cell.2014.01.066

- (2014).
- 201 Krishnan, H. *et al.* Podoplanin - an emerging cancer biomarker and therapeutic target. *Cancer Sci*, doi:10.1111/cas.13580 (2018).
- 202 Neri, S. *et al.* Podoplanin-expressing cancer-associated fibroblasts lead and enhance the local invasion of cancer cells in lung adenocarcinoma. *Int J Cancer* **137**, 784-796, doi:10.1002/ijc.29464 (2015).
- 203 Kurahashi, M. *et al.* A novel population of subepithelial platelet-derived growth factor receptor alpha-positive cells in the mouse and human colon. *Am J Physiol Gastrointest Liver Physiol* **304**, G823-834, doi:10.1152/ajpgi.00001.2013 (2013).
- 204 Zepp, J. A. *et al.* Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung. *Cell* **170**, 1134-1148 e1110, doi:10.1016/j.cell.2017.07.034 (2017).
- 205 Morikawa, S. *et al.* Prospective identification, isolation, and systemic transplantation of multipotent mesenchymal stem cells in murine bone marrow. *J Exp Med* **206**, 2483-2496, doi:10.1084/jem.20091046 (2009).
- 206 Satelli, A. & Li, S. Vimentin in cancer and its potential as a molecular target for cancer therapy. *Cell Mol Life Sci* **68**, 3033-3046, doi:10.1007/s00018-011-0735-1 (2011).
- 207 Varnat, F., Siegl-Cachedenier, I., Malerba, M., Gervaz, P. & Ruiz i Altaba, A. Loss of WNT-TCF addiction and enhancement of HH-GLI1 signalling define the metastatic transition of human colon carcinomas. *EMBO Mol Med* **2**, 440-457, doi:10.1002/emmm.201000098 (2010).
- 208 Ding, L. & Morrison, S. J. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* **495**, 231-235, doi:10.1038/nature11885 (2013).
- 209 Mitola, S. *et al.* Gremlin is a novel agonist of the major proangiogenic receptor VEGFR2. *Blood* **116**, 3677-3680, doi:10.1182/blood-2010-06-291930 (2010).
- 210 Kalajzic, I. *et al.* Use of type I collagen green fluorescent protein transgenes to identify subpopulations of cells at different stages of the osteoblast lineage. *J Bone Miner Res* **17**, 15-25, doi:10.1359/jbmr.2002.17.1.15 (2002).
- 211 Malanchi, I. *et al.* Interactions between cancer stem cells and their niche govern metastatic

- colonization. *Nature* **481**, 85-89, doi:10.1038/nature10694 (2011).
- 212 Liu, A. Y., Zheng, H. & Ouyang, G. Periostin, a multifunctional matricellular protein in inflammatory and tumor microenvironments. *Matrix Biol* **37**, 150-156, doi:10.1016/j.matbio.2014.04.007 (2014).
- 213 Hara, A. *et al.* Roles of the Mesenchymal Stromal/Stem Cell Marker Mefflin in Cardiac Tissue Repair and the Development of Diastolic Dysfunction. *Circ Res* **125**, 414-430, doi:10.1161/CIRCRESAHA.119.314806 (2019).
- 214 Bach, D. H., Park, H. J. & Lee, S. K. The Dual Role of Bone Morphogenetic Proteins in Cancer. *Mol Ther Oncolytics* **8**, 1-13, doi:10.1016/j.omto.2017.10.002 (2018).
- 215 Jung, B., Staudacher, J. J. & Beauchamp, D. Transforming Growth Factor beta Superfamily Signaling in Development of Colorectal Cancer. *Gastroenterology* **152**, 36-52, doi:10.1053/j.gastro.2016.10.015 (2017).
- 216 McCarthy, N. *et al.* Distinct Mesenchymal Cell Populations Generate the Essential Intestinal BMP Signaling Gradient. *Cell Stem Cell* **26**, 391-402 e395, doi:10.1016/j.stem.2020.01.008 (2020).
- 217 He, X. C. *et al.* BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nat Genet* **36**, 1117-1121, doi:10.1038/ng1430 (2004).
- 218 Qi, Z. *et al.* BMP restricts stemness of intestinal Lgr5(+) stem cells by directly suppressing their signature genes. *Nat Commun* **8**, 13824, doi:10.1038/ncomms13824 (2017).
- 219 Dutton, L. R. *et al.* Fibroblast-derived Gremlin1 localises to epithelial cells at the base of the intestinal crypt. *Oncotarget* **10**, 4630-4639, doi:10.18632/oncotarget.27050 (2019).
- 220 Ren, J. *et al.* Cancer-associated fibroblast-derived Gremlin 1 promotes breast cancer progression. *Breast Cancer Res* **21**, 109, doi:10.1186/s13058-019-1194-0 (2019).
- 221 Neckmann, U. *et al.* GREM1 is associated with metastasis and predicts poor prognosis in ER-negative breast cancer patients. *Cell Commun Signal* **17**, 140, doi:10.1186/s12964-019-0467-7 (2019).
- 222 Yan, K. *et al.* Glioma cancer stem cells secrete Gremlin1 to promote their maintenance within the tumor hierarchy. *Genes Dev* **28**, 1085-1100, doi:10.1101/gad.235515.113 (2014).

- 223 Kawasaki, K. *et al.* Chromosome Engineering of Human Colon-Derived Organoids to Develop a Model of Traditional Serrated Adenoma. *Gastroenterology* **158**, 638-651 e638, doi:10.1053/j.gastro.2019.10.009 (2020).
- 224 Mizutani, Y. *et al.* Meflin-Positive Cancer-Associated Fibroblasts Inhibit Pancreatic Carcinogenesis. *Cancer Res* **79**, 5367-5381, doi:10.1158/0008-5472.CAN-19-0454 (2019).
- 225 Xu, J. *et al.* Secreted stromal protein ISLR promotes intestinal regeneration by suppressing epithelial Hippo signaling. *EMBO J* **39**, e103255, doi:10.15252/embj.2019103255 (2020).
- 226 Lee, H. O. *et al.* Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet* **52**, 594-603, doi:10.1038/s41588-020-0636-z (2020).
- 227 Shoshkes-Carmel, M. *et al.* Subepithelial telocytes are an important source of Wnts that supports intestinal crypts. *Nature* **557**, 242-246, doi:10.1038/s41586-018-0084-4 (2018).
- 228 Shin, K. *et al.* Hedgehog signaling restrains bladder cancer progression by eliciting stromal production of urothelial differentiation factors. *Cancer Cell* **26**, 521-533, doi:10.1016/j.ccell.2014.09.001 (2014).
- 229 Dunbar, C. E. *et al.* Gene therapy comes of age. *Science* **359**, doi:10.1126/science.aan4672 (2018).
- 230 George, L. A. *et al.* Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant. *N Engl J Med* **377**, 2215-2227, doi:10.1056/NEJMoa1708538 (2017).
- 231 Lee, J. W. *et al.* Hepatocytes direct the formation of a pro-metastatic niche in the liver. *Nature* **567**, 249-252, doi:10.1038/s41586-019-1004-y (2019).
- 232 Lisowski, L. *et al.* Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* **506**, 382-386, doi:10.1038/nature12875 (2014).
- 233 Urist, M. R. Bone: formation by autoinduction. *Science* **150**, 893-899, doi:10.1126/science.150.3698.893 (1965).
- 234 Pickup, M. W. *et al.* BMPR2 loss in fibroblasts promotes mammary carcinoma metastasis via increased inflammation. *Mol Oncol* **9**, 179-191, doi:10.1016/j.molonc.2014.08.004 (2015).
- 235 Beppu, H. *et al.* Stromal inactivation of BMPRII leads to colorectal epithelial overgrowth and

- polyp formation. *Oncogene* **27**, 1063-1070, doi:10.1038/sj.onc.1210720 (2008).
- 236 Xu, R. *et al.* Long-term expression of angiostatin suppresses metastatic liver cancer in mice. *Hepatology* **37**, 1451-1460, doi:10.1053/jhep.2003.50244 (2003).
- 237 Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92, doi:10.1093/nar/gkz1001 (2020).
- 238 Maeda, K. *et al.* Identification of Meflin as a Potential Marker for Mesenchymal Stromal Cells. *Sci Rep* **6**, 22288, doi:10.1038/srep22288 (2016).
- 239 Persons, D. A., Mehaffey, M. G., Kaleko, M., Nienhuis, A. W. & Vanin, E. F. An improved method for generating retroviral producer clones for vectors lacking a selectable marker gene. *Blood Cells Mol Dis* **24**, 167-182, doi:10.1006/bcmd.1998.0184 (1998).
- 240 Hirokawa, Y., Yip, K. H., Tan, C. W. & Burgess, A. W. Colonic myofibroblast cell line stimulates colonoid formation. *Am J Physiol Gastrointest Liver Physiol* **306**, G547-556, doi:10.1152/ajpgi.00267.2013 (2014).
- 241 Khalil, H., Nie, W., Edwards, R. A. & Yoo, J. Isolation of primary myofibroblasts from mouse and human colon tissue. *J Vis Exp*, doi:10.3791/50611 (2013).
- 242 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
- 243 Platt, R. J. *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440-455, doi:10.1016/j.cell.2014.09.014 (2014).
- 244 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016).
- 245 Ponomarev, V. *et al.* A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Eur J Nucl Med Mol Imaging* **31**, 740-751, doi:10.1007/s00259-003-1441-5 (2004).
- 246 Aubrey, B. J. *et al.* An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo. *Cell Rep* **10**, 1422-1432, doi:10.1016/j.celrep.2015.02.002 (2015).
- 247 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image

- analysis. *Nat Methods* **9**, 671-675, doi:10.1038/nmeth.2089 (2012).
- 248 Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134, doi:10.1186/1471-2105-13-134 (2012).
- 249 Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect Differential Expression. *Ann Appl Stat* **10**, 946-963, doi:10.1214/16-AOAS920 (2016).
- 250 Sheffer, M. *et al.* Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proc Natl Acad Sci U S A* **106**, 7131-7136, doi:10.1073/pnas.0902232106 (2009).
- 251 Marisa, L. *et al.* Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* **10**, e1001453, doi:10.1371/journal.pmed.1001453 (2013).
- 252 Smith, J. J. *et al.* Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* **138**, 958-968, doi:10.1053/j.gastro.2009.11.005 (2010).
- 253 Camp, R. L., Dolled-Filhart, M. & Rimm, D. L. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* **10**, 7252-7259, doi:10.1158/1078-0432.CCR-04-0713 (2004).
- 254 Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337, doi:10.1038/nature11252 (2012).
- 255 Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* **49**, 708-718, doi:10.1038/ng.3818 (2017).
- 256 Neufert, C., Becker, C. & Neurath, M. F. An inducible mouse model of colon carcinogenesis for the analysis of sporadic and inflammation-driven tumor progression. *Nat Protoc* **2**, 1998-2004, doi:10.1038/nprot.2007.279 (2007).
- 257 Xiao, X., Li, J. & Samulski, R. J. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J Virol* **72**, 2224-2232,

doi:10.1128/JVI.72.3.2224-2232.1998 (1998).

- 258 Parmiani, G. Immunological approach to gene therapy of human cancer: improvements through the understanding of mechanism(s). *Gene Ther* **5**, 863-864, doi:10.1038/sj.gt.3300692 (1998).
- 259 Strobel, B., Miller, F. D., Rist, W. & Lamla, T. Comparative Analysis of Cesium Chloride- and Iodixanol-Based Purification of Recombinant Adeno-Associated Viral Vectors for Preclinical Applications. *Hum Gene Ther Methods* **26**, 147-157, doi:10.1089/hgtb.2015.051 (2015).
- 260 Wang, Q. *et al.* Efficient production of dual recombinant adeno-associated viral vectors for factor VIII delivery. *Hum Gene Ther Methods* **25**, 261-268, doi:10.1089/hgtb.2014.093 (2014).
- 261 Gilles, J. F., Dos Santos, M., Boudier, T., Bolte, S. & Heck, N. DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis. *Methods* **115**, 55-64, doi:10.1016/j.ymeth.2016.11.016 (2017).
- 262 Jackstadt, R. *et al.* Epithelial NOTCH Signaling Rewires the Tumor Microenvironment of Colorectal Cancer to Drive Poor-Prognosis Subtypes and Metastasis. *Cancer Cell* **36**, 319-336 e317, doi:10.1016/j.ccell.2019.08.003 (2019).
- 263 Lugli, A. *et al.* Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. *Mod Pathol* **30**, 1299-1311, doi:10.1038/modpathol.2017.46 (2017).
- 264 Fleming, N. I. *et al.* SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer. *Cancer Res* **73**, 725-735, doi:10.1158/0008-5472.CAN-12-2706 (2013).
- 265 Elyada, E. *et al.* Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov* **9**, 1102-1123, doi:10.1158/2159-8290.CD-19-0094 (2019).
- 266 Dominguez, C. X. *et al.* Single-Cell RNA Sequencing Reveals Stromal Evolution into LRRC15(+) Myofibroblasts as a Determinant of Patient Response to Cancer Immunotherapy. *Cancer Discov* **10**, 232-253, doi:10.1158/2159-8290.CD-19-0644 (2020).
- 267 Bartoschek, M. *et al.* Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat Commun* **9**, 5150,

doi:10.1038/s41467-018-07582-3 (2018).

- 268 Biffi, G. *et al.* IL1-Induced JAK/STAT Signaling Is Antagonized by TGFbeta to Shape CAF Heterogeneity in Pancreatic Ductal Adenocarcinoma. *Cancer Discov* **9**, 282-301, doi:10.1158/2159-8290.CD-18-0710 (2019).
- 269 Zhou, B. O., Yue, R., Murphy, M. M., Peyer, J. G. & Morrison, S. J. Leptin-receptor-expressing mesenchymal stromal cells represent the main source of bone formed by adult bone marrow. *Cell Stem Cell* **15**, 154-168, doi:10.1016/j.stem.2014.06.008 (2014).
- 270 Decker, M. *et al.* Leptin-receptor-expressing bone marrow stromal cells are myofibroblasts in primary myelofibrosis. *Nat Cell Biol* **19**, 677-688, doi:10.1038/ncb3530 (2017).
- 271 Corselli, M. *et al.* Perivascular support of human hematopoietic stem/progenitor cells. *Blood* **121**, 2891-2901, doi:10.1182/blood-2012-08-451864 (2013).
- 272 Brechbuhl, H. M. *et al.* Fibroblast Subtypes Regulate Responsiveness of Luminal Breast Cancer to Estrogen. *Clin Cancer Res* **23**, 1710-1721, doi:10.1158/1078-0432.CCR-15-2851 (2017).
- 273 Wang, Z. *et al.* CD146, from a melanoma cell adhesion molecule to a signaling receptor. *Signal Transduct Target Ther* **5**, 148, doi:10.1038/s41392-020-00259-8 (2020).
- 274 Xing, S. *et al.* Targeting endothelial CD146 attenuates colitis and prevents colitis-associated carcinogenesis. *Am J Pathol* **184**, 1604-1616, doi:10.1016/j.ajpath.2014.01.031 (2014).
- 275 Jiang, T. *et al.* CD146 is a coreceptor for VEGFR-2 in tumor angiogenesis. *Blood* **120**, 2330-2339, doi:10.1182/blood-2012-01-406108 (2012).
- 276 Zhang, M. *et al.* Single-cell transcriptomic architecture and intercellular crosstalk of human intrahepatic cholangiocarcinoma. *J Hepatol* **73**, 1118-1130, doi:10.1016/j.jhep.2020.05.039 (2020).
- 277 Wu, S. Z. *et al.* Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. *EMBO J* **39**, e104063, doi:10.15252/embj.2019104063 (2020).
- 278 Asfaha, S. *et al.* Krt19(+)/Lgr5(-) Cells Are Radioresistant Cancer-Initiating Stem Cells in the Colon and Intestine. *Cell Stem Cell* **16**, 627-638, doi:10.1016/j.stem.2015.04.013 (2015).
- 279 Reichert, M. *et al.* The Prrx1 homeodomain transcription factor plays a central role in

- pancreatic regeneration and carcinogenesis. *Genes Dev* **27**, 288-300, doi:10.1101/gad.204453.112 (2013).
- 280 Sancho, E., Batlle, E. & Clevers, H. Signaling pathways in intestinal development and cancer. *Annu Rev Cell Dev Biol* **20**, 695-723, doi:10.1146/annurev.cellbio.20.010403.092805 (2004).
- 281 Tian, B., Zhang, Y. & Li, N. CD146 protein as a marker to predict postoperative liver metastasis in colorectal cancer. *Cancer Biother Radiopharm* **28**, 466-470, doi:10.1089/cbr.2012.1426 (2013).
- 282 Struyf, S. *et al.* Synergistic induction of MCP-1 and -2 by IL-1beta and interferons in fibroblasts and epithelial cells. *J Leukoc Biol* **63**, 364-372, doi:10.1002/jlb.63.3.364 (1998).
- 283 Baghdadi, M. *et al.* Interleukin-34, a comprehensive review. *J Leukoc Biol* **104**, 931-951, doi:10.1002/JLB.MR1117-457R (2018).
- 284 Raz, Y. *et al.* Bone marrow-derived fibroblasts are a functionally distinct stromal cell population in breast cancer. *J Exp Med* **215**, 3075-3093, doi:10.1084/jem.20180818 (2018).
- 285 Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat Med* **23**, 1124-1134, doi:10.1038/nm.4409 (2017).
- 286 Garcia, P. E. *et al.* Differential Contribution of Pancreatic Fibroblast Subsets to the Pancreatic Cancer Stroma. *Cell Mol Gastroenterol Hepatol* **10**, 581-599, doi:10.1016/j.jcmgh.2020.05.004 (2020).
- 287 Friedman, G. *et al.* Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4+ and PDPN+ CAFs to clinical outcome. *Nature Cancer* **1**, 692-708, doi:10.1038/s43018-020-0082-y (2020).
- 288 de Sousa, E. M. F. *et al.* Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. *Cell Stem Cell* **9**, 476-485, doi:10.1016/j.stem.2011.10.008 (2011).
- 289 Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, doi:10.1126/science.aan2507 (2017).
- 290 Kinchen, J. *et al.* Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. *Cell* **175**, 372-386 e317, doi:10.1016/j.cell.2018.08.067 (2018).

- 291 Ajina, R. *et al.* SpCas9-expression by tumor cells can cause T cell-dependent tumor rejection in immunocompetent mice. *Oncoimmunology* **8**, e1577127, doi:10.1080/2162402X.2019.1577127 (2019).
- 292 Becker, C. *et al.* In vivo imaging of colitis and colon cancer development in mice using high resolution chromoendoscopy. *Gut* **54**, 950-954, doi:10.1136/gut.2004.061283 (2005).
- 293 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419, doi:10.1126/science.1260419 (2015).
- 294 Taki, T. *et al.* CD109 regulates in vivo tumor invasion in lung adenocarcinoma through TGF-beta signaling. *Cancer Sci* **111**, 4616-4628, doi:10.1111/cas.14673 (2020).
- 295 Magness, S. T., Bataller, R., Yang, L. & Brenner, D. A. A dual reporter gene transgenic mouse demonstrates heterogeneity in hepatic fibrogenic cell populations. *Hepatology* **40**, 1151-1159, doi:10.1002/hep.20427 (2004).
- 296 Lannagan, T. R. M. *et al.* Genetic editing of colonic organoids provides a molecularly distinct and orthotopic preclinical model of serrated carcinogenesis. *Gut* **68**, 684-692, doi:10.1136/gutjnl-2017-315920 (2019).
- 297 An, N. & Kang, Y. Using quantitative real-time PCR to determine donor cell engraftment in a competitive murine bone marrow transplantation model. *J Vis Exp*, e50193, doi:10.3791/50193 (2013).
- 298 Eichwald, E. J., Silmsler, C. R. & Weissman, I. Sex-linked rejection of normal and neoplastic tissue. I. Distribution and specificity. *J Natl Cancer Inst* **20**, 563-575 (1958).
- 299 Dakup, P. P., Porter, K. I., Little, A. A., Zhang, H. & Gaddameedhi, S. Sex differences in the association between tumor growth and T cell response in a melanoma mouse model. *Cancer Immunol Immunother* **69**, 2157-2162, doi:10.1007/s00262-020-02643-3 (2020).
- 300 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
- 301 McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297, doi:10.1093/nar/gks042 (2012).

- 302 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
- 303 Tauriello, D. V. F. *et al.* TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature* **554**, 538-543, doi:10.1038/nature25492 (2018).