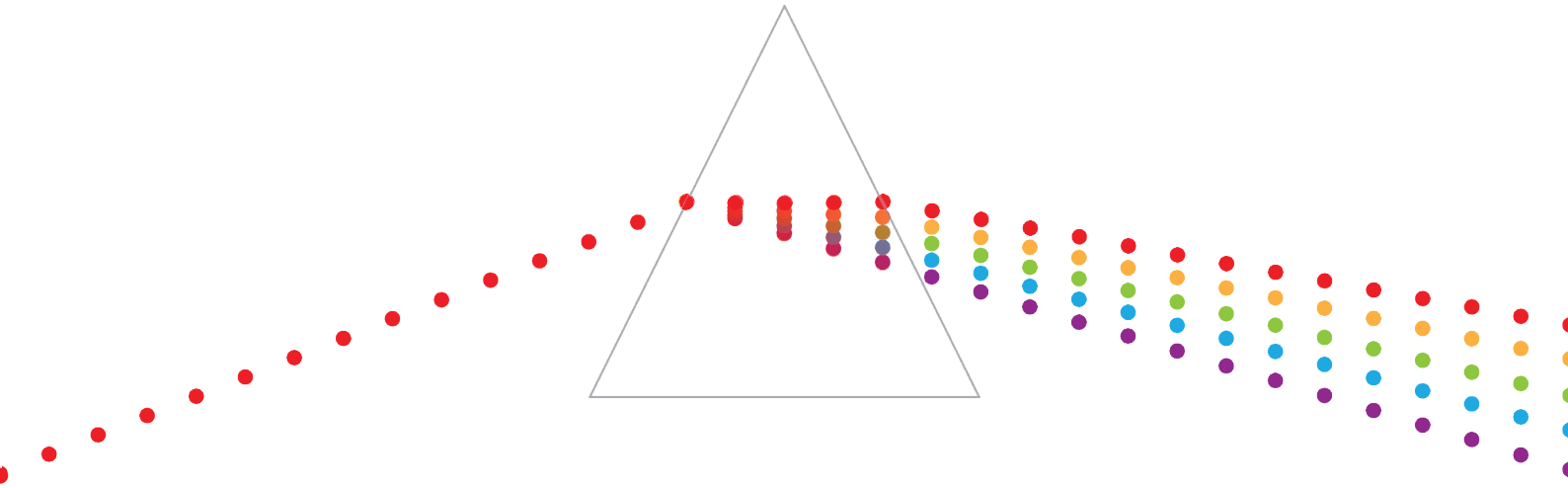
# Modeling and targeting metastatic relapse in colorectal cancer

Adrià Cañellas Socias

# Modeling and targeting metastatic relapse in colorectal cancer



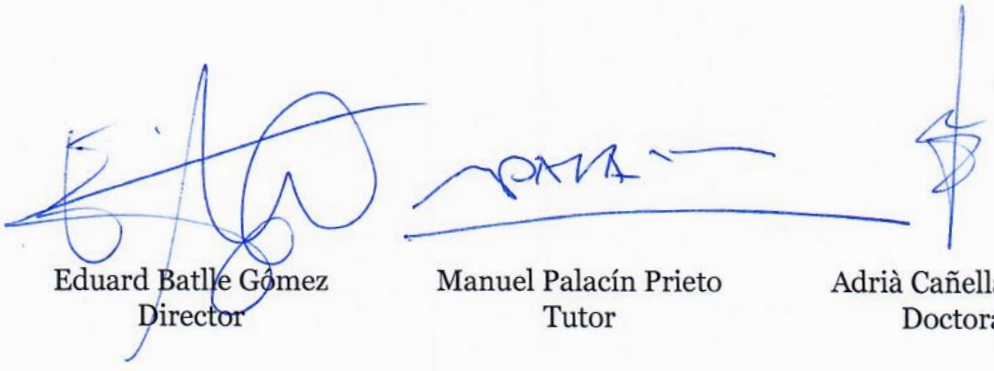
Adrià Cañellas Socias

Universitat de Barcelona
Facultat de Biologia
IRB Barcelona
Programa de Doctorat en Biomedicina

# Modeling and targeting metastatic relapse in colorectal cancer

Memòria presentada per Adrià Cañellas Socias per optar al grau de doctor per la Universitat de Barcelona

Aquesta tesi ha estat realitzada sota la direcció del Dr. Eduard Batlle Gómez a l'Institut de Recerca Biomèdica de Barcelona



| Eduard Batlle Gómez | Manuel Palacín Prieto | Adrià Cañellas Socias |
|---|---|---|
| Director | Tutor | Doctorand |

Barcelona, 2022

*A sa senyora Sato, en Dobbie i es senyoret Nilson*

# Table of Contents

# Abbreviations

| | |
|---|---|
| **AJCC** | American Joint Committee on Cancer |
| **αSMA** | Alpha Smooth Muscle Actin |
| **ANXA** | Annexin |
| **APC** | Adenomatous Polyposis Syndrome |
| **ASCL2** | Achaete-Scute Family BHLH Transcription Factor 2 |
| **BMPR** | Bone Morphogenic Protein Receptor |
| **bp** | base pairs |
| **CAFs** | Cancer-associated fibroblasts |
| **Cas9** | CRISPR associated protein 9 |
| **CBC** | Crypt Base Columnar Cell |
| **CD31** | Platelet endothelial cell adhesion molecule |
| **CD45** | Lymphocyte common antigen |
| **CDX2** | Caudal Type Homeobox 2 |
| **CID** | Chronic Inducer of Dimerization |
| **CIMP** | CpG Island Methilator Phenotype |
| **CIN** | Chromosomal Instability |
| **CMS** | Consensus Molecular Subtypes |
| **CRC** | Colorectal Cancer |
| **CRISPR** | Clustered Regularly Interspaced Short Palindromic Repeats |
| **CSC** | Cancer Stem Cell |
| **CTL** | Cytotoxic T lymphocytes |
| **CTLA-4** | Cytotoxic T-lymphocyte protein 4 |
| **DAPI** | 4'6'-Diamidino-2-phenylindole |
| **DSB** | Double Strand Break |
| **DSS** | Dextran Sodium Sulphate |
| **DTCs** | Disseminated Tumor Cells |
| **DVL** | Disheveled |
| **ECM** | Extracellular Matrix |
| **EdU** | Ethynyl Deoxyuridine |
| **EGF** | Epidermal Growth Factor |
| **EGFP** | Enhanced Green Fluorescent Protein |
| **EGFR** | EGF Receptor |
| **EMP1** | Epithelial Membrane Protein 1 |
| **EMP1-iCT** | EMP1-inducibleCaspase9-tdTOMATO knock-in |
| **EMT** | Epithelial to Mesenchymal Transition |
| **EPCAM** | Epithelial Cell Adhesion Molecule |
| **EPHB2** | Ephrin Receptor B2 |
| **EpiHR** | Epithelial-specific High Hazard-Ratio |
| **ESC** | Embryonic Stem Cell |
| **FABP1** | Fatty Acid Binding Protein 1 |
| **FACS** | Fluorescent Activated Cell Sorting |
| **FAP** | Familial Adenomatous Polyposis |
| **Gal** | TGF-β inhibitor LY2157299 |
| **GEMM** | Genetic Engineered Mouse Model |
| **GFP** | Green Fluorescent Protein |
| **GO** | Gene Ontology |

| | |
|---|---|
| **gRNA** | guide RNA for Cas9 (tracrRNA+crRNA) |
| **GSEA** | Gene-Set Enrichment Analysis |
| **GZMB** | Granzyme B |
| **HE** | Hematoxilin-Eosin |
| **HNPCC** | Hereditary Non Polyposis Colorectal Cancer |
| **HR** | Homologous Recombination |
| **HRCs** | High Relapse Cells |
| **HSC** | Hematopoietic Stem Cell |
| **iCASP9** | Inducible Caspase 9 |
| **IFNα** | Interferon alpha |
| **IFNγ** | Interferon gamma |
| **IRES** | Internal ribosome entry site |
| **ISC** | Intestinal Stem Cell |
| **kb** | kilobases |
| **KI67** | Marker of Proliferation Ki67 |
| **KO** | Knock-Out |
| **KRAS** | Kirsten Rat Sarcoma Virus |
| **KRT20** | Keratin 20 |
| **L1CAM** | L1 Cell Adhesion Molecule |
| **LAMC2** | Laminin Subunit Gamma 2 |
| **LGR5** | Leucine Rich Repeat Containing G Protein-Coupled Receptor 5 |
| **LOH** | Loss of Heterozigosis |
| **LRC** | Label Retaining Cell |
| **Luc** | Luciferin |
| **LY6C** | Lymphocyte antigen 6 complex locus G6C |
| **LY6G** | Lymphocyte antigen 6 complex locus G6D |
| **MAPK** | Mitogen-Activated Protein Kinase Patwhay |
| **MDSCs** | Myeloid Derived Supressor Cells |
| **MEK** | Mitogen-activated Protein Kinase Kinase |
| **MHC** | Major histocompatibility complex |
| **Min** | Multiple Intestinal Neoplasia |
| **MLH1** | MutL Homolog 1 |
| **MMP7** | Matrix Metallopeptidase 7 |
| **MMR** | Mismatch Repair |
| **mRNA** | messenger RNA |
| **MSH2** | MutS Homolog 2 |
| **MSI** | microsatellite instability |
| **MTO** | Mouse tumor organoid |
| **mTORC1** | mammalian/mechanistic target of rapamycin complex 1 |
| **MUC2** | Mucin 2 |
| **MYC** | MYC Proto-Oncogene, BHLH Transcription Factor |
| **NHEJ** | Non Homologous End Joining |
| **NK** | Natural killer |
| **NT** | Non treated |
| **OLFM4** | Olfactomedin 4 |
| **PCR** | polymerase chain reaction |
| **PD1** | Programmed Cell Death 1 |
| **PD-1** | Programmed cell death protein 1 |
| **PD-L1** | Programmed death-ligand 1 |

| | |
|---|---|
| **PDO** | Patient Derived Organoid |
| **PDX** | Patient Derived Xenograft |
| **PIK3CA** | Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha |
| **PPBP** | Pro-Platelet Basic Protein |
| **RNF43** | Ring Finger Protein 43 |
| **Rspo** | Rspondin |
| **RT-qPCR** | Real time quantitative polymerase chain reaction |
| **SC** | Stem Cell |
| **scRNAseq** | Single cell RNA sequencing |
| **SDCBP2** | Syndecan Binding Protein 2 |
| **SMOC2** | SPARC Related Modular Calcium Binding 2 |
| **SPIM** | Selective Plane Illumination Microscopy |
| **SSA** | Sessile Serrated Adenomas |
| **StrHR** | Stromal-specific High Hazard-Ratio |
| **TA** | Transient Amplifying |
| **Tam** | Tamoxifen |
| **TCGA** | The Cancer Genome Atlas |
| **TCR** | T Cell receptor |
| **Teff** | Effector T cell |
| **TGF-β** | Transforming Growth Factor Beta |
| **TIL** | Tumor infiltrating lymphocyte |
| **TME** | Tumor microenvironment |
| **TNFα** | Tumor necrosis factor alpha |
| **TNM** | Tumor invasiveness, lymphatic node dissemination and metastases grade |
| **TOM** | TdTOMATO Fluorescent Protein |
| **TOP2A** | DNA Topoisomerase II Alpha |
| **TP53** | Tumor Protein p53 |
| **Treg** | Regulatory T cell |
| **VEGF** | Vascular Endothelial Growth Factor |
| **WT** | Wild Type |

# Abstract

Colorectal cancer (CRC) kills around 700,000 people worldwide every year. The majority of these deaths are the result of dissemination of the disease to foreign organs. Despite undergoing curative resection of the primary tumor, 30-40% CRC patients will relapse in the following years. In these patients, residual disseminated tumor cells (DTCs) are undetectable until they regenerate metastatic disease. The identity and features of residual tumor cells and their niches have remained elusive due to the impossibility of analyzing this clinically occult population in patients.

In the first chapter of this thesis, we discovered that genes associated with elevated risk of relapse in human patients are expressed by a defined subset of primary tumor cells that we named High Relapse Cells (HRCs). HRCs are abundant at invasion fronts, retain an epithelial program and express genes involved in cell adhesion, locomotion and extracellular matrix remodeling. To investigate HRCs, we established a human-like CRC mouse model that undergoes metastatic relapse following surgical resection of the primary tumor. We also developed methodology to isolate residual disseminated tumor cells before metastases are detectable. Single cell profiling demonstrated that residual tumor cells occult in mouse livers after primary CRC surgery resembled the HRCs present in patients. Using Emp1 (epithelial morphogenic protein 1) as a marker gene for HRCs, we tracked and selectively eliminated this cell population. Genetic ablation of HRCs prior to extirpation of the primary CRC prevented metastatic recurrence and mice remained disease-free after surgery.

In the second chapter, we tackled how the tumor microenvironment (TME) changes over time during the formation of metastases. We discovered that at the onset of metastasis, pre-existing T cell immunity against the primary CRC can eliminate DTCs as they reach the liver. During this phase, neoadjuvant checkpoint immunotherapy given before surgical removal of the primary tumor is sufficient to eradicate micrometastases. Yet, this curative effect is restricted to a narrow temporal window due to a rapidly evolving TME. By profiling at the single-cell level metastatic lesions at different stages, we found that the TME of metastases becomes complex over time. Thus, to be efficacious, immunotherapeutic treatment must be tailored to these different stages of progression.

The last chapter describes how kinetics of metastatic relapse are influenced by the genetic makeup of CRCs. Tumors with an incomplete set of driver mutations retain niche dependencies. As a result, DTCs transiently enter a latent state resistant to chemotherapy. Eventually, after cessation of therapy, DTCs resume growth and regenerate metastases.

Our findings reveal the features of the tumor cell population responsible for CRC recurrence and anticipate that therapies targeting HRCs may prevent disease relapse. Moreover, we demonstrated that residual disease is a particular state of unique vulnerability that can be targeted using immunotherapies, when applied in a timely manner. The timing of such therapeutic window will be likely determined by the genetics of each CRC.

# Introduction

## Intestinal biology

### Tissue homeostasis and adult stem cells.

The transition from unicellular to multicellular organisms was one of the most notable events in the evolution of life. The shift to multicellularity is estimated to have occurred 1.5 billion years ago and it gave rise to increasingly complex life forms. While unicellular organisms depend on just one cell for all, multicellular organisms have a diverse set of cell types with specialized functions. In metazoans, groups of cells with common objectives assemble in tissues, which in turn are structured into organs and systems to work collectively to perform specific functions.

Inherent to biological function is cell damage, which poses risks to the organism, such as organ failure or cancer development. To preserve homeostasis, tissues are organized hierarchically to separate tissue function and maintenance. Specialized cells are typically short-lived to minimize the workload endured, resulting in the loss of over 50 billion cells every day in our bodies (Fuchs and Blau, 2020). Thus, our tissues require constant replenishment to stay alive which is achieved by the existence of long-lived undifferentiated cells, named adult stem cells (SCs).

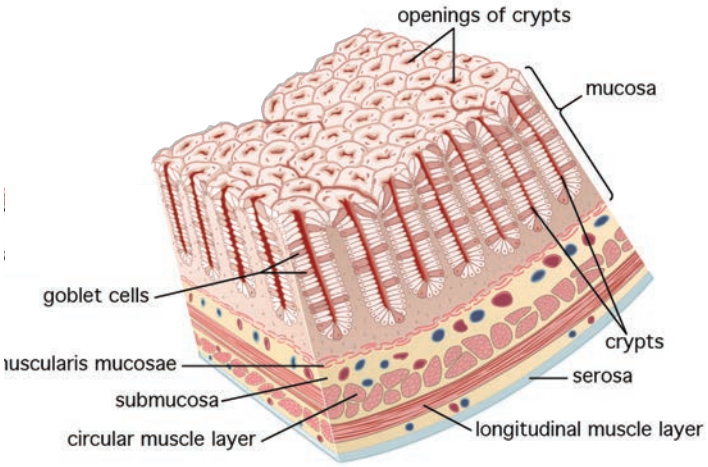
Adult SCs reside in most tissues of our body. Although more restricted than embryonic stem cells in the types of tissues they form, adult SCs show long-term survival and regenerate both themselves and cells in their resident tissues during homeostasis and after injury. Tissue SCs typically reside in specialized microenvironments (niches) and their biology differs depending on the tissues. Those under constant environmental insults such as the skin, the hematopoietic system and the intestine have very high rates of self-renewal. In contrast, in other tissues, such as muscle, pancreas and liver, SCs are used sparingly, but are mobilized quickly and expand to repair wounds when the tissue is damaged (Fuchs, 2009).

### The mammalian intestine

The intestine is a tubular structure within the gastrointestinal tract that connects the end of the stomach to the anus. The mammalian intestine is thus the last component of the digestive system and it is divided into three main regions, namely the small intestine, the large intestine or colon and the rectum. The small intestine comprises three consecutive regions: the duodenum, jejunum and ileum, which are responsible for the final steps of food digestion and nutrient uptake. The colon performs functions related mainly to water absorption and stool compaction.

The intestinal tube consists of various concentric layers creating a central lumen through which digested food and water pass. From the outside to the inside, the following cell layers are found: the serosa, two thick muscle layers, the submucosa, the muscularis mucosae, the lamina propria and the intestinal epithelium (I-Figure 1).

The outer layer contains innervated smooth muscle involved in peristaltic movements and it is surrounded by a thin protective sheet of squamous epithelial cells named serosa. The submucosa contains blood and lymphatic vessels as well as nerves and myofibroblasts. After the muscularis mucosae - a fine line of smooth muscle fibers that separates both regions - we find the lamina propria, which consists of connective tissue and stromal cells. The first layer in contact with the lumen is the intestinal epithelium or mucosa, a monostratified epithelium with protrusions and invaginations called crypts of Lieberkühn (usually referred to as crypts). The intestinal epithelium needs to coordinate the uptake of metabolites and protection against multiple environmental insults (e.g. trillions of bacteria, an acidic pH and mechanical stress). This complex problem is solved by an intricate tissue architecture that defines tissue function and self-renewal in two physically separate regions: the crypt and the villus (I-Figure 2).

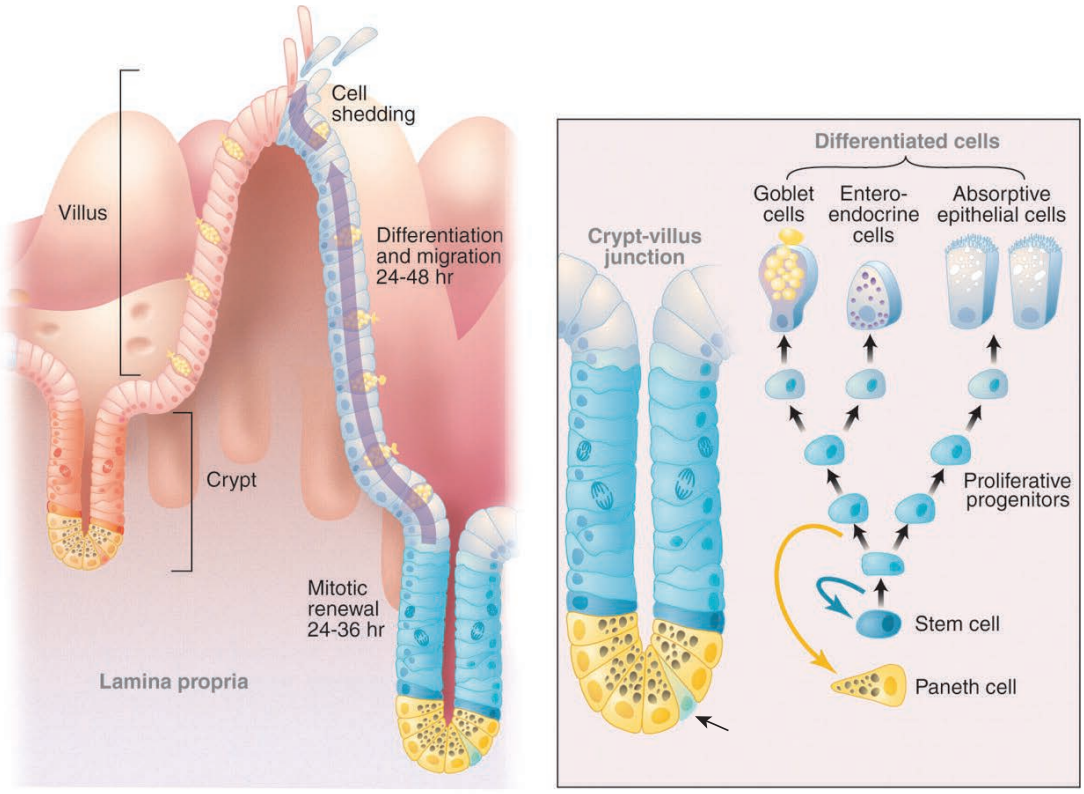



**I-Figure 1 | Anatomy of the intestinal tract.** Scheme of the intestinal layers that compose the colon. Adapted from the Encyclopedia Britannica.

The intestinal epithelium contains millions of crypt-villus units (Gehart and Clevers, 2019). Each one consists of a villus- a protrusion of the intestinal wall- surrounded by multiple invaginations named crypts (I-Figure 2). These protuberances massively increase the surface area of the intestinal wall thus enhancing nutrient absorption. Villus length decreases along the intestinal track; villi in the duodenum are >1 mm but they are absent in the colon, which has a flat epithelial surface. Contouring the epithelium, a network of capillaries and lymph vessels make absorbed nutrients instantly available to the

rest of the body. Nonetheless, increasing the absorptive surface area also raises exposure to various sources of stress. To minimize such exposure, epithelial cells in the villi are short-lived (3-5 days). This renewal rate is the highest of all mammalian tissues and is driven by multipotent SCs that reside in the bottom of the intestinal crypts.

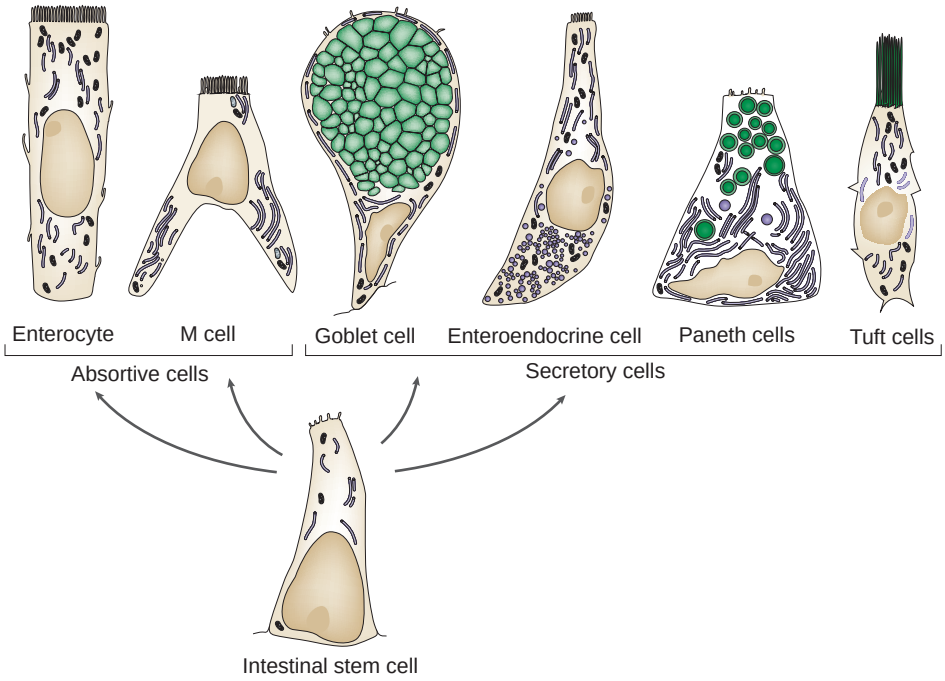





**I-Figure 2 | Crypt-villus structures.** The small intestine is shaped into crypts and villi (left). Cells in the villi are short-lived and fueled by ISCs that reside at the bottom of the crypts (arrow). ISCs give rise to transient amplifying, which differentiate into specialized cell lineages (right). Adapted from (Radtke & Clevers, 2005).

Intestinal crypts are invaginations of the intestinal wall that create an environment well protected from the hazards of the digestive process. Secluded in the bottom of the crypts, continuously dividing SCs give rise to progenitor cells (or transient amplifying cells) that rapidly proliferate and differentiate into mature intestinal cells. In a matter of days, these cells then migrate from the crypt to the tip of the villus, where they serve their function and eventually undergo apoptosis. This design enables the protection of actively proliferating cells and restricts exposure to hazards to short-lived post mitotic cells. The single-layered epithelium contains six distinct mature cell types, which are divided into absorptive (enterocytes and M cells) and secretory (Paneth, goblet, enteroendocrine and tuft cells) lineages (I-Figure 3). These cell populations perform specialized functions such as nutrient uptake (enterocytes) and secretion of protective mucus (goblet cells), bactericidal peptides (Paneth cells), hormones (enteroendocrine cells) and immunomodulatory

agents (M cells). The intestinal SC (ISC) is the engine that continuously supplies all cell types in appropriate ratios and, in recent years, many studies have been devoted to understanding the nature of this prototype SC system (I-Figure 2).




**I-Figure 3 | Cell types in the intestinal epithelium.** ISCs give rise to six classes of terminally differentiated cells. The most abundant cells are enterocytes, followed by goblet and Paneth cells. Enteroendocrine, Tuft and M cells are present at lower frequencies. Modified from (Crosnier et al., 2006)

## The intestinal stem cell

ISCs have been object of study in adult stem cell research for many decades. Since 1974, cumulative evidence gathered using radiation and chemical mutagenesis to label cells in the intestinal epithelium suggested that fast-proliferating crypt base columnar (CBC) cells were precursors of the differentiated cell types in the epithelium of mice (Bjerknes and Cheng, 1999; Cheng and Leblond, 1974; Winton et al., 1988). At that time, knowledge emanating from other systems, such as the hematopoietic system and the hair follicle, made this view highly controversial, since adult stem cells were commonly regarded as slow-proliferating or quiescent (Clevers, 2013). In fact, other researchers proposed that slow-growing cells capable of label retention (label retaining cells or LRCs) that locate immediately above Paneth cells (in position “+4”) were the true ISCs (Gehart & Clevers, 2019).

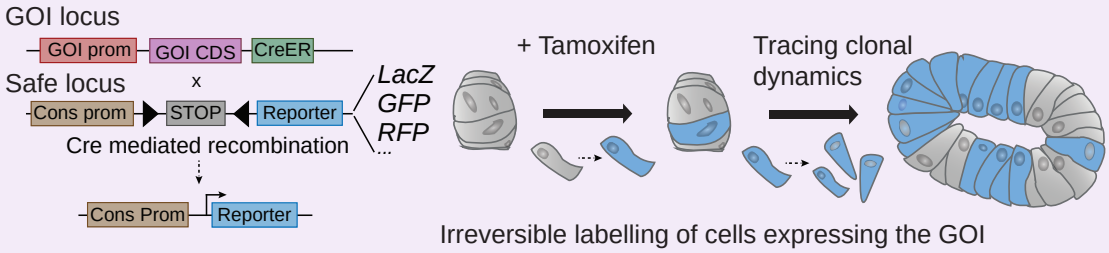
## Box 1 | Approaches to study stem cell potential in vivo

### Transplantation assays

In limiting dilution assays, a given subpopulation of cells is first isolated based on marker protein expression and then transplanted in vivo to assess its ability to regenerate a given tissue or tumor. The expected result is that repopulating capacity will be limited to subpopulations of self-renewing cells. An important limitation of this study is the disruption of the original niche. Transplantation-based approaches may reveal the potential of stem cells, but may not necessarily unveil the fate of these cells under steady-state conditions.

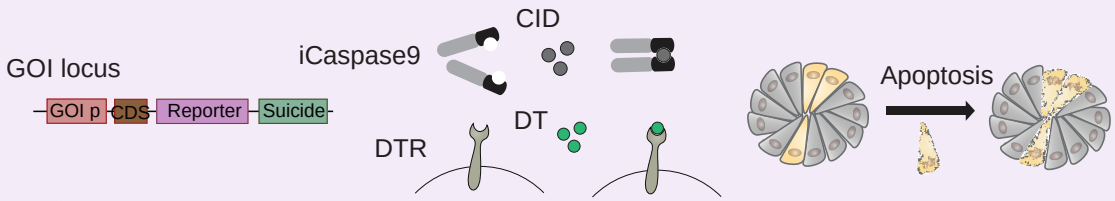
### Lineage tracing

Lineage tracing experiments follow the progeny of a specific subpopulation in its native environment. Modern lineage tracing tools consist of two components: (i) an inducible Cre recombinase under the transcriptional control of a marker gene of interest (GOI), and (ii) a reporter gene that will be permanently expressed upon recombination of loxP sites. In order to make the system inducible, the Cre recombinase is fused to the estrogen receptor (ER; CreERT2) : upon tamoxifen administration the recombinase translocates to the nucleus where it becomes active. The reporter cassette typically contains a strong, ubiquitous promoter and a reporter gene, both separated by a stop signal that is in turn flanked by loxP sites (><). In this experimental setting, upon tamoxifen administration only cells expressing the marker gene will lose the stop signal and thus the reporter gene will start being expressed. Importantly, the expression of the reporter is constitutive and will be irreversibly expressed by all the progeny of labeled cells, thereby allowing the study of the clonal behavior of particular cell populations.

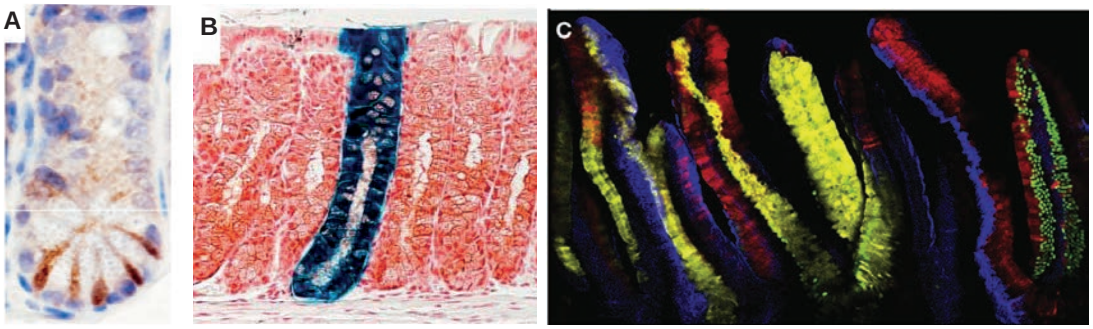



### Genetic cell ablation

Cell ablation is an approach to study the relevance of specific cell lineages based on selective induction of cell death. The marker GOI drives the expression of pro-caspases or diphtheria-toxin (DT) receptors, which will exert their suicide function only upon ligand binding. In caspase-mediated cell ablation, chemical-induced dimerization (CID) compounds induce the interaction and consequent activation of pro-caspases, resulting in apoptosis of the cells of interest. In DT mediated ablations, the toxin causes the death of cells expressing its receptor. These approaches can be used to ascertain whether the population of interest is required for tissue or tumor growth.



With the advent of lineage tracing tools (Box 1) and the use of $Lgr5^{EGFP\_IRES\_CreERT2}$ mice, the Clevers lab provided definitive proof that adult ISCs were CBC cells in the base of the crypt (Barker et al., 2007). Lgr5 was first identified as a direct target of the WNT signaling pathway (Van de Wetering et al., 2002) and was later found to be a specific marker of CBC cells (Barker et al., 2007) (I-Figure 4a). Lineage tracing experiments using $Lgr5^{EGFP\_IRES\_CreERT2}$ R26R-lacZ mice, demonstrated that Lgr5+ cells generated offspring that spread from the bottom of the crypt to the tip of the villus within 5 days. The traced clones generated ribbons that contained cells of all lineages and many persisted throughout life, thereby demonstrating the stemness nature of Lgr5+ cells (I-Figure 4b). At the time, this discovery challenged the prevailing dogma in the stem cell field by showing that quiescence is not an essential attribute of adult SCs.




**I-Figure 4 | Lgr5 expression and lineage tracing in the intestine. a**, Immunohistochemical analysis of EGFP expression in Lgr5-EGFP-IRES-creERT2 knock-in mouse. From (Barker et al., 2007). Observe the restricted expression of Lgr5 gene to CBC cells, which are intermingled with Paneth cells in the intestine. **b**, Histological analysis of LacZ activity in Lgr5-EGFP-IRES-creERT2 mice crossed with Rosa26-lacZ reporter mice 60 days after induction with tamoxifen Tracing from Lgr5+ CBC cells generated fully labelled ribbons, suggesting that it represented the stem cell of the intestine. From (Barker et al., 2007). **c**, Whole-mount imaging of the intestine showing multi-color ribbons generated from the "confetti" Cre-reporter mice. From (Snippert et al., 2010).

## Intestinal regeneration and plasticity

To study what would happen upon the death of ISCs, Frederic de Sauvage and colleagues designed $Lgr5^{DTR\text{-}EGFP}$ mice. In these animals, the Lgr5 promoter drives diphtheria toxin receptor (DTR) expression, so that Lgr5+ cells can be selectively killed upon treatment with the toxin (Tian et al., 2011) (Box 1). Surprisingly, the intestinal epithelium successfully regenerated after discrete ablation of Lgr5+ SCs. However, restoration failed when Lgr5+ cells were continuously ablated, thus suggesting that a different subpopulation of intestinal cells replenishes the SC pool (Metcalfe et al., 2014). Repopulation of the SC niche has been attributed to +4 cells, as well as to other cells typically considered

lineage-committed, such as enteroendocrine cells and enterocytes (Barriga et al., 2017; Buczacki et al., 2013; Murata et al., 2020; Tetteh et al., 2016; Tian et al., 2011).

The ability of mature cells to dedifferentiate is an important safety mechanism to ensure epithelial integrity. Analysis of the epigenetic landscape of intestinal stem cells and differentiated cells showed that permissive open chromatin is maintained throughout differentiation (Kim et al., 2014). Thus, the absence of epigenetic fate locking enables a highly plastic behavior instructed by the niche environment. Fate mapping of individual stem cells by a multicolor Cre-reporter demonstrated that, in homeostasis, crypts drift towards clonality (I-Figure 4c) (Snippert et al., 2010). These cellular dynamics are consistent with a model in which ISCs divide symmetrically and stochastically adopt either stem or TA fates. The stem cell zone model takes as its central premise that the crypt bottom is a SC-inducing microenvironment, in which virtually any epithelial cell entering will acquire stem properties. Stem cell daughters that exit the stem cell zone and pass through position 5 will start their way to differentiation toward various individual lineages and commitment will occur during the upward migration to the villus. Upon intestinal damage and stem cell loss, remodeling of the extracellular matrix induces a regenerative fetal-like state, through YAP/TAZ activation, which mediates the reconstruction of the intestinal epithelium (Ayyaz et al., 2019; Yui et al., 2018)

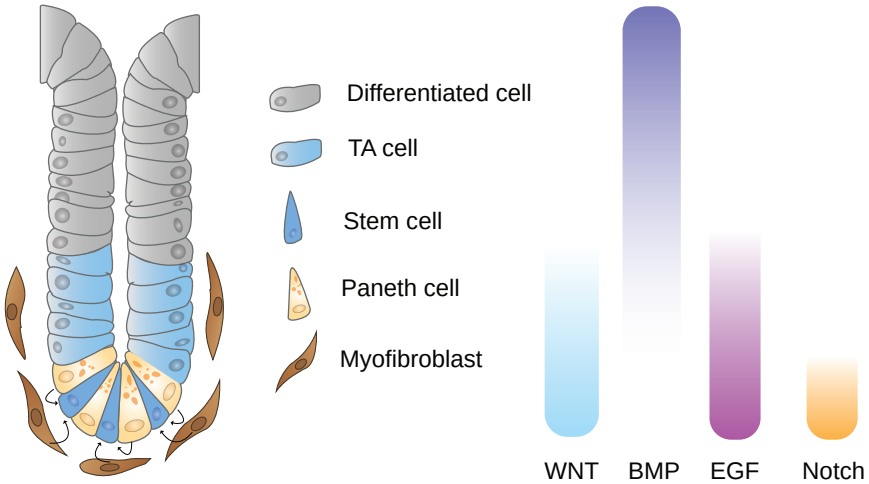


## The stem cell niche

The fact that ISCs divide every day is at odds with the risk of DNA damage and resulting malignancies. This might explain why the crypt developed into a highly protective environment that supports proliferation and minimizes damage to stem cells. The spatial properties of crypts ensure that stem cells do not come into direct contact with the digestive process. To avoid such exposure, multiple tightly regulated gradients of signaling factors are built that restrict the hyperproliferative state to the crypt base. We now know that deregulation of such microenvironmental signals through oncogenic mutations pose a great risk for colorectal cancer.

Intestinal stemness and differentiation is regulated mostly by four well-characterized signaling pathways, namely WNT, EGF, Notch and BMP (Clevers, 2013; Medema and Vermeulen, 2011). In contrast to other lineage decisions, cell hierarchy in the intestine is not hard-routed, but rather microenvironmentally regulated (Sato et al., 2013). Signaling molecules emanating from the surrounding crypt niche tightly regulate the activation or inactivation of these pathways (Crosnier et al., 2006; McCarthy et al., 2020a). Paneth cells and the sub-epithelial mesenchyme are the key players in the establishment of the crypt niche (McCarthy et al., 2020b; Sato et al., 2011). These cells generate gradients

of WNT, epidermal growth factor (EGF), Notch ligands and bone morphogenic protein (BMP) inhibitors that shape epithelial cell fates across the intestinal epithelium (I-Figure 5).



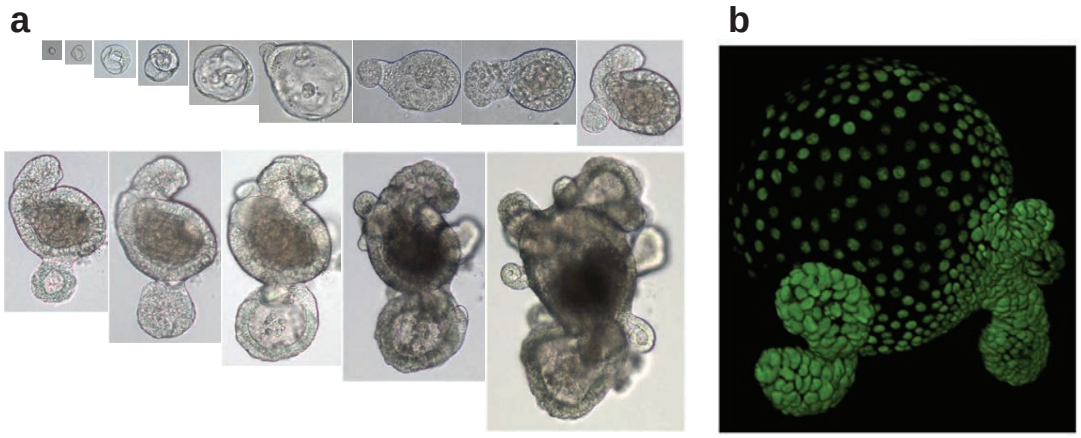

**I-Figure 5 | The stem cell niche.** Cartoon of the intestinal epithelial structure (left) and the spatial gradients of Wnt, BMP, EGF and Notch signals formed along the crypt axis (right). Paneth cells and sub-epithelial myofibroblasts produce the signals for stem cell maintenance (arrows).

Paneth cells (PC) are intermingled cells at the bottom of the crypt, with each stem cell touching at least one PC. In fact, the position of an ISC relative to a PC directly influences its fate: stem cells that reside centrally in the crypt, surrounded by PCs, have higher chances of becoming the dominant clone than those at the border of the crypt (Ritsma et al., 2014). PCs express WNT3, EGF and Notch ligands DLL1 and DLL4, all of which are essential components of the stem cell niche. However, experiments performing ablation of PCs in vivo have shown that Notch ligands are the only Paneth-born stem cell signal required in vivo, thus demonstrating the redundancy of WNT and EGF with other sources (Durand et al., 2012; Sasaki et al., 2016). Indeed, the intestinal epithelium receives structural support and soluble signals from the underlying mesenchyme, formed by fibroblasts, pericytes, endothelial cells, neural cells and smooth muscle cells. R-spondins and BMP inhibitors such as noggin are produced solely by the surrounding mesenchyme. BMP-2 and -4 ligands are expressed in the mesenchyme of villi, whereas BMP inhibitors are expressed in the mesenchyme around crypts (Medema and Vermeulen, 2011).

## Organoids

For decades, it was considered impossible to establish long-term cultures from primary adult tissues without inducing malignant transformation. However, in 2009 Toshiro Sato, Hans Clevers and colleagues developed a three-dimensional (3D) culture system

that indefinitely maintains intestinal epithelial structures that spontaneously rebuild the crypt structure (I-Figure 6) (Sato et al., 2009). Single Lgr5+ cells *in vitro* generate miniguts that are formed by all the intestinal cell types, thus recapitulating the hierarchical organization of the tissue of origin. Importantly, these *in vitro* grown miniguts, commonly referred to as organoids, regenerated epithelial patches upon transplantation into mice with a damaged intestinal mucosa (Yui et al., 2012). The resulting epithelial patches were indiscernible from the surrounding recipient epithelium, implying an unprecedented similarity of organoid cultures to *in vivo* organs (Yui et al., 2012).




**I-Figure 6 | Miniguts from a single ISC. a,** Sequential imaging of an intestinal organoid grown from a single sorted Lgr5-GFP cell (left). Organoid are 3D epithelial in vitro cultures that reflect the structural and hierarchical properties of the intestinal epithelium. From (Sato et al., 2009). **b,** Expression of histone 2B-GFP in a mini-gut organoid. Crypts emanate from a large central body that contains the lumen. Adapted from (Clevers, 2013).

The ability to expand intestinal stem cells *in vitro* provided the definitive intestinal stem cell niche requirements. Supplementation of only three factors, R-spondin, EGF and the BMP inhibitor noggin, is sufficient to maintain the organoid culture (Sato et al., 2009). Notably, in addition to the aforementioned factors, only the presence of Paneth cells is necessary for efficient culture (Sato et al., 2011). Organoid cultures can be passaged indefinitely and they remain genetically stable. These features have prompted the expansion of large living biobanks of normal and cancer organoids that are used for drug development and personalized medicine (Gehart and Clevers, 2019).

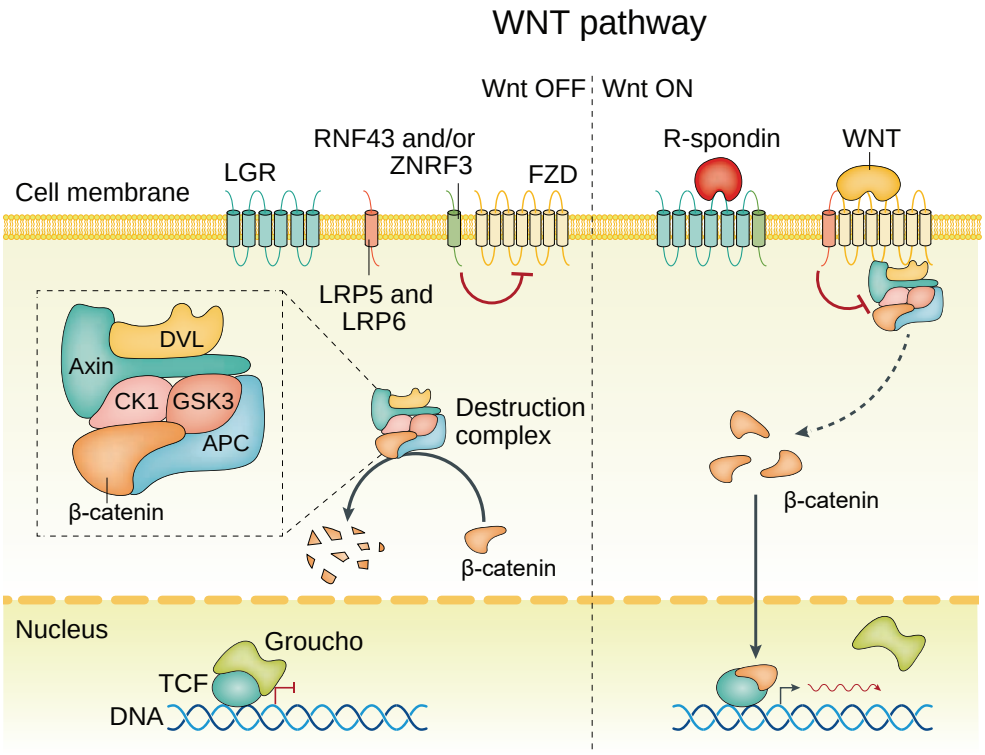
Nowadays, organoid cultures are no longer limited to the intestine. The Clevers lab has pioneered the development of similar three-dimensional (3D) epithelial structures that recapitulate fundamental aspects of various organs of origin, including pancreas, liver, prostate, kidney and others (Huch et al., 2013a, 2013b; Karthaus et al., 2014; Schutgens et al., 2019). In the last decade, organoids have revolutionized science and have been extensively applied from stem cell research (Fujii et al., 2016a; Pérez-González et al.,

2020; Qin et al., 2020; Sato et al., 2011; Serra et al., 2019; Sprangers et al., 2021), to the modeling of multiple diseases including colorectal cancer (Drost et al., 2015; Fujii et al., 2016b; Fumagalli et al., 2017; Matano et al., 2015; O'Rourke et al., 2017), to personalized medicine (Qu et al., 2021; Yao et al., 2020) and to regenerative medicine (Jackstadt and Sansom, 2016; Jung et al., 2011; Lukonin et al., 2020; Yui et al., 2012).

## Signaling pathways in the stem cell niche

### WNT

The WNT signaling pathway is evolutionarily conserved and controls countless biological processes during development and adult life of all animals (Clevers and Nusse, 2012). Most mammalian genomes harbor 19 Wnt genes that typically signal over a short distance. Canonical WNT signals regulate the abundance of β-catenin protein, which acts as a transcriptional co-activator of the TCF/LEF transcription factor family (Clevers and Nusse, 2012).




**I-Figure 7 | The WNT pathway.** When no WNT ligands bind FZD, the destruction complex breaks down beta-catenin. In the absence of R-spondin, the WNT receptor FZD is continuously degraded through ubiquitination by RNF43 or ZNRF3. Once WNT binds to FZD and LRP5–LRP6, the destruction complex is inhibited and beta-catenin accumulates, enters the nucleus and drives transcription by binding to T cell factor (TCF) and displacing Groucho. Sustained WNT pathway activation is only possible in the presence of R-spondin, as its binding to receptors of the LGR family sequesters RNF43–ZNRF3 and thus stabilizes FZD expression. Adapted from (Gehart and Clevers, 2019).

WNT proteins bind a heterodimeric receptor complex consisting of a Frizzled (Fz) and an LRP5/6 protein (I-Figure 7). In the absence of active Wnt signaling, β-catenin is phosphorylated, ubiquitinated and targeted for degradation by a destruction complex formed by adenomatous polyposis coli (APC), Axin, casein kinase 1 (CK1), glycogen synthase kinase-3 (GSK3) and Dishevelled. In this scenario, TCF/LEF proteins are inactive due to interaction with the transcriptional repressor Groucho (Nusse and Clevers, 2017). Engagement by Wnt causes a ligand-induced conformational change of the receptor that enables its association with Axin. Without axin, the destruction complex falls apart, β-catenin is stabilized and translocates to the nucleus, where it binds to TCF to up-regulate the expression of target genes.

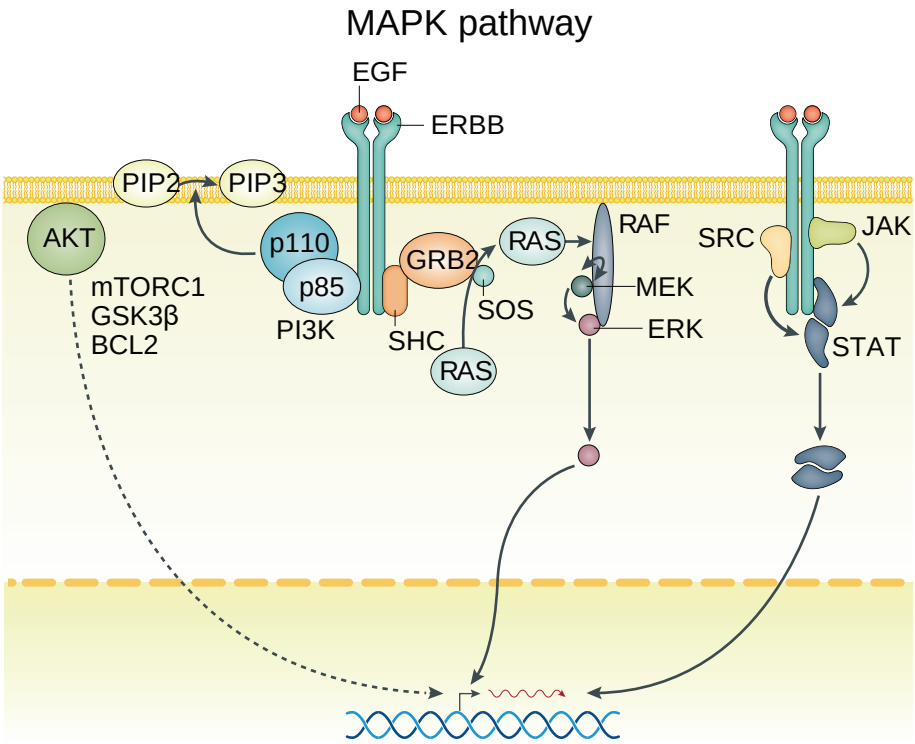
WNT is the key pathway for the maintenance of stem cell fate and for the proliferation of stem and transient-amplifying cells. Prior to the discovery of ISC markers, the Clevers lab showed that the Wnt signaling pathway was essential for the development and maintenance of the mouse intestine (Korinek et al., 1998). Tcf4 knock-out mice lack proliferative crypts, implying that Wnt signaling is required for the establishment of the SC compartment (Korinek et al., 1998). Furthermore, constitutive activation of Wnt leads to the formation of adenomas in mice through the acquisition of a constitutive progenitor phenotype (Van de Wetering et al., 2002).

Despite the abundance of WNT ligands at the bottom of the crypt, their ability to activate the WNT pathway is dependent on R-spondins (RSPO), another essential component of the ISC niche. R-Spondins are Wnt agonists that interact on the cell surface with members of the LGR5 family to enhance Wnt signaling. ZNRF3 and RNF43 are two transmembrane molecules with E3 ubiquitin ligase activity that downregulate Wnt signaling by causing the turn-over of Frizzled receptors through endocytosis. Once R-spondin binds to LGRs, it sequesters ZNRF3 and RNF43, and the amount of Frizzled receptors in the cell membrane increases (Gehart and Clevers, 2019). ISCs express both LGR4 and LGR5 and knock-out of both phenocopies the loss of WNT signaling (De Lau et al., 2011). Thus, the R-spondin/LGR axis amplifies the WNT pathway and helps to fine-tune its activation (Huels and Sansom, 2017; De Lau et al., 2011).

**EGF**

EGF is a core component of the ISC niche that controls the division rate of ISCs but does not appear to be necessary for the maintenance of SC identity (Basak et al., 2017). Increased activity of the EGF pathway, observed in KRas mutant clones, increases SC proliferation and gives mutated stem cells a selective advantage for dominance in their crypt. Paneth cells and mesenchymal pericryptal cells produce EGF and transforming growth factor-α (TGF-α) ligands that bind the ERBB1/EGFR, which is highly expressed

in CBC cells (Gehart and Clevers, 2019). EGF ligands are pro-proteins and they must be cleaved in order to become activated and secreted. When bound to their ligand, EGFRs dimerize and transactivate their constitutively inactive kinase domains (I-Figure 8).

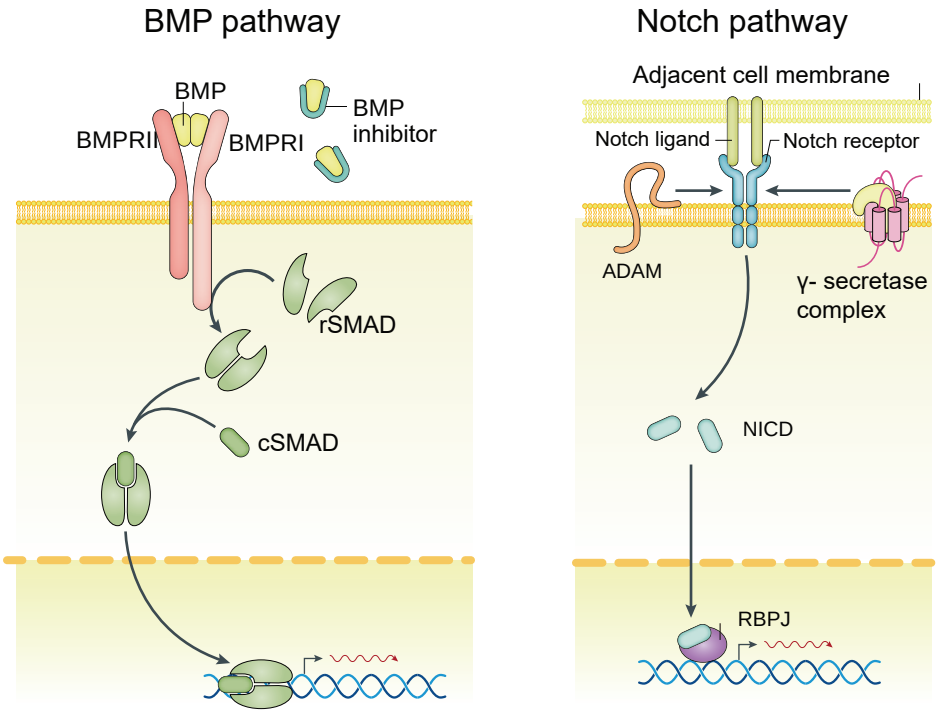



**I-Figure 8 | EGF signaling cascade.** Activation of epidermal growth factor (EGF) receptors (ERBB) by EGF activates the mitogen-activated protein kinase (MAPK) cascade to activate ERK. ERK then phosphorylates a wide array of nuclear targets to promote proliferation and counteract apoptosis. Activated EGF receptor complexes also recruit Janus kinase (JAK) and SRC. These tyrosine kinases phosphorylate signal transducer and activator of transcription (STAT) proteins, which dimerize and enter the nucleus to regulate transcription. In parallel, EGF signaling also enables binding of phosphoinositide 3-kinase (PI3K), which converts PIP2 into PIP3. PIP3 serves as binding site and activator of AKT. AKT has several downstream targets that regulate proliferation and survival, such as mTORC1, BCL2 or GSK3. Adapted from (Gehart and Clevers, 2019).

### BMP

The TGF-β family comprises more than 30 factors, which include TGF-β itself, activins, BMPs, myostatins and other molecules (Massagué, 2012). These proteins engage heterotypic dimerization of their receptors, a process that activates their kinase domains (I-Figure 9). The receptors for BMPs are BMPR1 and BMPR2, which phosphorylate SMAD1, SMAD5 and SMAD8. In the case of TGFβ signaling, TGFBR1 and TGFBR2 phosphorylate SMAD2 and SMAD3. Once activated, the receptor substrates Smads (RSmads) shuttle to the nucleus and form a complex with SMAD4, a binding partner common to all RSmads,

that regulates gene expression (Massagué, 2008).




**I-Figure 9 | BMP and Notch signaling pathways.** Binding of bone morphogenetic protein (BMP) couples its receptors (BMPR1 and BMPRII). This dimerization event leads to phosphorylation and subsequent activation of receptor-regulated mothers against decapentaplegic homologues (rSMADs). These then bind common SMAD (cSMAD or SMAD4) in a heterotrimeric complex and translocate to the nucleus, where the complex regulates target gene expression. BMP signaling can be inhibited by sequestration of BMPs by secreted BMP inhibitors such as Noggin, or Gremlin. Notch signaling is initiated by binding of Notch receptors to Notch ligands. Two subsequent cleavage events release Notch intracellular domain (NICD) into the cytoplasm. NICD translocates to the nucleus, where it binds RBPJ to activate downstream genes. Adapted from (Gehart and Clevers, 2019).

BMP signaling suppresses the growth of ISCs and induces differentiation (Qi et al., 2017). Distinct sub-epithelial fibroblast populations generate the essential BMP intestinal gradient to limit the hyperproliferation of ISCs. Telocytes in the crypt-villus junction produce abundant BMP2 and BMP4, whereas trophocytes lie near ISCs at the crypt base and express the BMP inhibitor Gremlin1 (McCarthy et al., 2020b). Full intestinal blockade of BMP results in the formation of ectopic crypt units in the villus and spontaneous benign overgrowth lesions called hamartomas (Haramis et al., 2004). Indeed, patients with germline mutations that inactivate SMAD4 of BMPR1 develop juvenile polyposis, a cancer predisposition syndrome that includes frequent occurrence of intraepithelial neoplasia (Haramis et al., 2004).

### Notch signaling

Notch signaling requires direct membrane contact between two cells, one expressing Notch ligands (such as DLL1 or DLL4) and the other a Notch receptor (for example NOTCH1) (I-Figure 9). Binding of Notch receptors to Notch ligands induces the cleavage of Notch intracellular domain (NICD) into the cytoplasm. NICD then translocates to the nucleus, where it binds RBPJ to activate downstream genes, the most important being HES1. HES1 in turn represses the transcription factor of ATOH1, a key element in secretory differentiation that drives the expression of Notch ligands (Durand et al., 2012; Yang et al., 2001). This mechanism, known as lateral inhibition, stabilizes stochastic differences in pathway activation by translating them into a binary 'on' or 'off" state.

Given that both receptor and ligands are transmembrane proteins, Notch signaling has the shortest range of all microenvironmental cues in the intestine. Notch blocks the differentiation of SCs to the secretory lineage, thus regulating the ratio of secretory to absorptive progeny. Moreover, lateral inhibition of Notch signaling shapes the fascinating arrangement of Lgr5+ cells with Paneth cells by ensuring contact between both cell lineages. Paneth cells are the only known source of Notch ligands in the intestinal crypt (Sato et al., 2011).

### Other stem cell regulators

Hippo signaling, best known for regulating tissue size during development, has received increasing attention in the past few years due to its role in epithelial regeneration and cancer (Gehart and Clevers, 2019). This pathway integrates a multitude of different inputs including mechanical forces, growth factors and inflammatory signals through G protein-coupled receptor signaling. These signals converge to control the activity of LATS1/2 kinases, which in turn target transcriptional regulators YAP and TAZ for degradation. Thus, YAP and TAZ are transcriptionally active when LATS are turned off. The role of Hippo signaling in homeostasis is still under debate, but its importance during intestinal regeneration is well established (Cheung et al., 2020; Gregorieff et al., 2015; Yui et al., 2018). Interestingly, YAP and TAZ promote regeneration by inhibiting WNT activation through direct repression of classical WNT target genes like LGR5 and AXIN2 (Barry et al., 2013). Although several studies have addressed the role of Hippo in colorectal cancer, its role is still controversial (Barry et al., 2013; Cheung et al., 2020).

Other important regulators of stem cell biology are inflammatory cues produced by immune cells or nutrient availability. For example, low caloric intake slows the proliferation of ISCs and TA cells, while nutrient overabundance through high-fat diets increases the number of stem cells (Beyaz et al., 2016).

# Colorectal cancer

## Colorectal cancer epidemiology

Cancer is a leading cause of premature death and a major impediment to increasing life expectancy worldwide (Sung et al., 2021).The global burden of cancer incidence is rising, reflecting population growth and aging, as well as changes in risk factors associated with socioeconomic development. Even when corrected for an aged population, the proportion of people with cancer has been slowly increasing over recent decades (GBD, 2017). This observation reflects changes in population habits such as diet, lack of exercise, alcohol consumption, and others. On the other hand, cancer-related death rates corrected for changes in age-profiles are consistently decreasing, thereby revealing the impact of efforts devoted to earlier detection of cancer and novel therapeutic agents.

Colorectal cancer (CRC) is one of the most common human cancers, with an estimated 1.2 million new cases diagnosed worldwide per year (Dekker et al., 2019). Given that almost all epithelial cells in the intestinal lining are replaced on a weekly basis, this cell population is at serious risk of malignant conversion. Indeed, around 50% of the Western population develops an adenomatous polyp by the age of 70, a small proportion of which will progress to cancer, with a lifetime cancer risk estimated to be 5% (Radtke and Clevers, 2005). The prognosis of patients with CRC has slowly improved during recent decades in many countries. Five-year relative survival has reached almost 65% in high-income countries, such as Australia, USA and several European countries, but has remained less than 50% in low-income countries (Rahbari et al., 2019). At present, surgical removal of the primary tumor remains the mainstay of treatment for solid tumors (Dekker et al., 2019).

## Environmental and genetic risk factors

Cancer is a genetic disease, in which a medley of mutations accumulates to the point that cells reach a state of unchecked growth. Such mutations may be inherited, induced by environmental factors or result from DNA replication errors. Although still controversial, work from Tomasetti and Vogelstein suggested that around two thirds of cancer mutations arise due to “bad luck”. In their studies, they found that the more stem cells and the more rapidly a particular organ’s stem cells replicate, the higher the risk of cancer in that tissue (Tomasetti and Vogelstein, 2015; Tomasetti et al., 2017). These numbers are consistent with epidemiological estimates of the fraction of cancers that can be prevented by changes in the environment (around 40%) (Tomasetti and Vogelstein, 2015).

Indeed, 60% of CRC cases are sporadic, meaning that there is no family history or apparent genetic predisposition. Familial CRC (30% of the total) refers to patients who have at least one blood relative with CRC or an adenoma, but with no specific germline mutation. Hereditary CRC syndromes (10%) result from germline inheritance of mutations in highly penetrant cancer susceptibility genes. The two most common forms of hereditary cancers are hereditary non-polyposis colon cancer (HNPCC or Lynch syndrome) and familial adenomatous polyposis (FAP).

About 2-5% of all CRC cases are attributed to Lynch syndrome, which is caused by a germline mutation in one of several mismatch repair (MMR) genes. The combination of the germline mutation with the inactivation of the remaining normal allele results in loss of MMR function. This leads to an accumulation of mutations in microsatellites (microsatellite instability, MSI), a hallmark of HNPCC. On the other hand, FAP is characterized by high numbers of adenomatous colorectal polyps and it accounts for approximately 1% of CRC cases. FAP is inherited by a germline mutation in the adenomatous polyposis coli (APC) gene.

In sporadic CRC, the preeminent risk factor is age, as 90% of diagnoses occur after 50 years of age (Asociación Española Contra el Cáncer, 2018). Nonetheless, additional risk factors have been identified, such as male sex, inflammatory bowel disease, smoking, excessive alcohol consumption, high consumption of red and processed meat, obesity and diabetes (Dekker et al., 2019).

## Staging

The American Joint Committee on Cancer (AJCC) has established a general cancer staging protocol called the TNM Staging System (I-Figure 10). T stands for tumor and evaluates the original primary tumor in grades of aggressiveness; N stands for lymph node and qualifies the presence of cancer cells spreading to nearby lymph nodes; and M addresses the presence of distant metastasis (spreading of cancer to distant organs). Based on these three categories, patients are classified into four stages (I, II, III, and IV). While statistically representing significantly different risk groups for recurrence and cancer-related death, the staging system does not accurately predict recurrence at the level of individual patients.

- Stage I CRC involves small non-invasive lesions and has a survival rate of 88%. Surgical resection is the only treatment required.

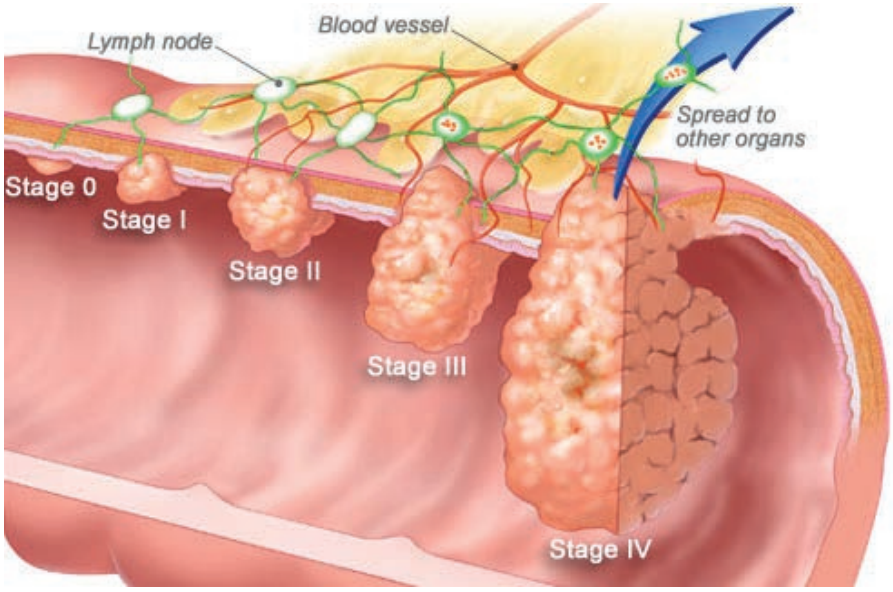
- In Stage II CRC, tumors are larger and invade the muscular layer of the intestine, yet no distant growth is evident. Patients diagnosed at this stage have a survival rate of 50

– 75%. Treatment with adjuvant chemotherapy is optional and decided by the clinician.

- In the case of Stage III patients, the tumor is invasive and local metastasis is also present in adjacent lymph nodes. This group has a survival rate of 25 – 55 % and treatment consists of surgical resection of the tumors as well as systemic adjuvant chemotherapy.

The major cause of death in patients with stage II and III tumors is disease relapse. Relapses occur mostly in the form of metastasis and are due to the presence of disseminated cells that remained occult and resisted therapy. Around 40% of all stage II and III cases will relapse up to 5 years after the end of therapy.

- Stage IV CRCs are large, invasive tumors that already present colonization of distant organs, mainly liver and lungs, at the time of diagnosis. Patients diagnosed at this stage have a survival rate lower than 10% and there is no effective treatment for this stage of the disease.




**I-Figure 10 | Colorectal cancer staging.** Clinical classification based on the TNM system groups CRCs based on their degree of muscular invasion and spreading to other organs.

## Current treatments for CRC

Most CRC start as a polyp, a small abnormal growth in the lining of the colon. There are several types of colonic polyps, and most of them will turn out not to be cancerous. Only the adenomatous polyps or adenomas, which comprise 10% of all polyps, are considered pre-malignant lesions likely to develop into colon cancer. The fact that most CRCs originate from a polyp explains why the screening for colonic polyps has become such an important part of disease management. In fact, screening programs are effective in reducing the number of deaths caused by CRC (Dekker et al., 2019).

The current standard of care for CRC patients involves radical surgical resection of the primary tumor, which cures a large proportion of early-stage patients. However over the following months around 20% stage II and 40% stage III patients will develop metastatic disease, as a result of tumor cells that disseminated before resection (Dekker et al., 2019). Unlike the primary tumor, metastases are less frequently removed by surgery and they are the main cause of death. Patients that present with metastases at the time of diagnosis (stage IV) and those at perceived risk of relapse (stage III and a fraction of stage II) receive cytotoxic chemotherapy: in most cases a combination of folinic acid, 5-fluorouracil, and oxaliplatin or irinotecan (FOLFOX and FOLFIRI, respectively). This strategy aims to kill highly proliferative cancer cells and has been a staple in the treatment of solid tumors for decades (Dekker et al., 2019; Moertel et al., 1990). Adjuvant chemotherapy (given after surgery) reduces the risk of relapse by 30% to stage II and III patients (Moertel et al., 1990), and it performs poorly in the metastatic setting, almost invariably giving rise to drug resistance and disease progression (André et al., 2009; Varghese, 2015).

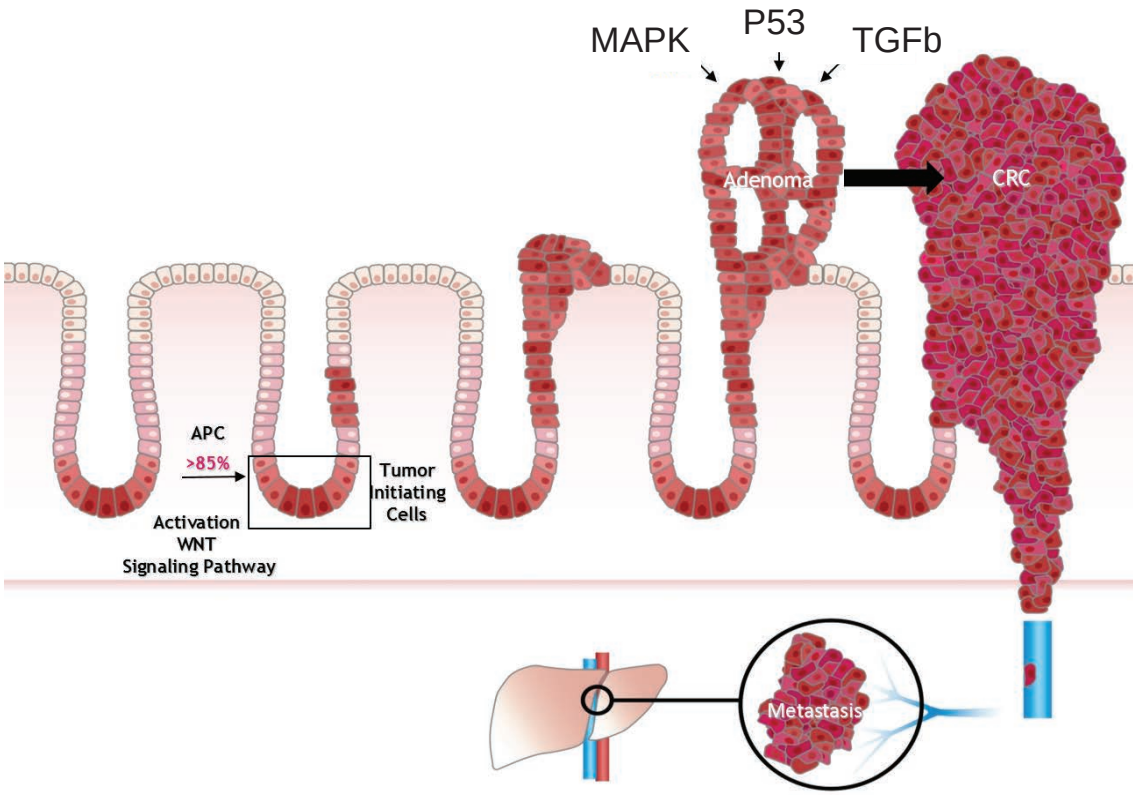
Standard systemic chemotherapy is increasingly combined with targeted treatments that eliminate specific dysregulated pathways that are crucial for cancer growth and survival. For example, inhibitors of EGFR signaling such as cetuximab and panitumumab improve survival in patients with CRC. Unfortunately, these therapies are effective only in patients without mutations in the MAPK signaling pathways such as KRAS (Douillard et al., 2013). Moreover, often these therapies meet with acquired resistance through mutations in antibody-binding sites of EGFR or downstream signaling (Bardelli and Siena, 2010; Siravegna et al., 2015). Recently, Russo et al. showed that in response to EGFR targeted therapy, CRC tumor cells employed a mechanism of adaptive mutability — well-known in bacteria — to increase the chance of resistant subclones emerging (Russo et al., 2019). Other examples of targeted therapies are BRAF inhibitors (Vemurafenib) or anti-vascular endothelial growth factor (VEGF) inhibitors (Bevacizumab) (Elamin et al., 2015; Hurwitz et al., 2004).

A better understanding of advanced cancer may lead to more effective therapies for metastatic CRC. However, improvements in the management of earlier stages of the disease are also highly relevant. Arguably the most important question for stage I–III patients is whether or not to treat the patient and which therapeutic strategy will be beneficial to prevent recurrence in each case.

## The adenoma to carcinoma model: the genetics of CRC

Back in 1990, Vogelstein described the adenoma to carcinoma model, which states that the stepwise accumulation of multiple genetic alterations drives the transformation of

a healthy colonic epithelium to colon adenocarcinoma (Fearon and Vogelstein, 1990). Three decades of work and the sequencing of thousands of tumor samples by The Cancer Genome Atlas Network (TCGA) have provided further information but the main message remains unaltered. The first hit that induces the transition from normal to polyploid tissue is seen in the WNT pathway, whereas progression to adenomas and carcinomas depends on activating mutations in the RAS pathway and inactivation of p53 and/or TGF-β/BMP signaling pathways (I-Figure 11).




**I-Figure 11 | Genetic model of CRC progression.** Schematic representation of the multistep acquisition of genetic alterations that underlie the progression from a healthy intestinal epithelium to a metastatic CRC.

APC mutations are the first event in the multistep process of CRC formation and they occur in more than 80% of colorectal carcinomas (Muzny et al., 2012). Inactivating mutations in APC cause hyperactivation of WNT signaling leading to premalignant lesions called adenomas. Other less common mutated WNT signaling pathway components are CTNNB1 (β-catenin), RNF43, ZNRF3, RSPO2 and RSPO3 (Yaeger et al., 2018). The following steps are activating mutations in the Mitogen-Activated Protein Kinase (MAPK) pathway and inactivating mutations of the TP53 pathway. Point mutations of KRAS are the most common oncogenic alteration to hyperactive MAPK signaling, which may also be achieved through mutations in NRAS, HRAS, EGFR, PIK3CA, PTEN and others. Finally, mutations in SMAD4, and less frequently in SMAD2, SMAD3 and TGF-β receptors confer insensitivity to cytostatic TGF-β and BMP (Bailey et al., 2018; Campbell et al.,

2020; Yaeger et al., 2018).

The above mentioned sequence of carcinogenesis represent the canonical route for CRC development. However, CRCs can also arise through alternative routes and mutational processes. One of them are serrated precursor lesions, which follow a distinct molecular pathway and differing in histological architecture from the conventional adenoma–carcinoma sequence (Janssen et al., 2020). Serrated tumors often accumulate mutations in BRAF and become deficient in DNA mismatch repair, which can lead to hypermutated CRCs that also acquire atypical numbers of tandem repeats (Jones et al., 2008). These cancers are also called microsatellite instable (MSI) tumors, a portion of which arise from hereditary mutations in DNA mismatch repair genes (Lynch syndrome). In contrast to hypermutated/MSI tumors, CRCs that are microsatellite stable (MSS) typically accumulate moderately low numbers of mutations and are characterized by chromosomal instability.

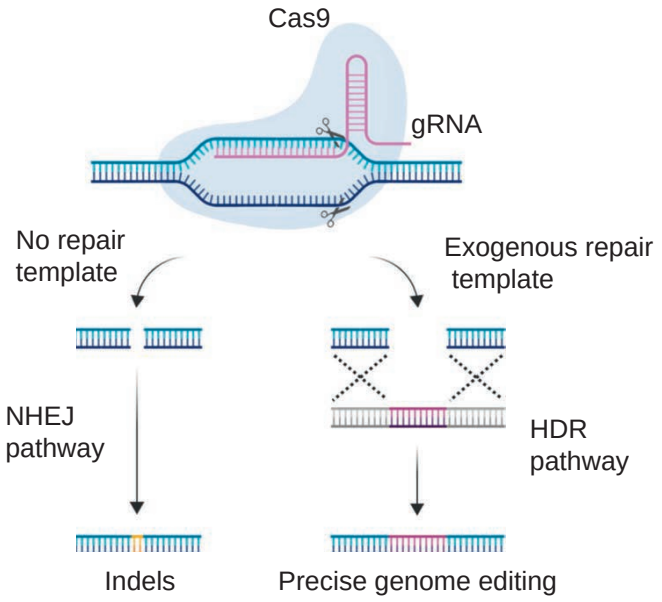
Interestingly, most alterations involved in CRC progression in patients fall into those signaling pathways regulating the self-renewal of normal ISCs. As cells progress along the normal adenoma–carcinoma sequence, they become gradually less dependent on their niches, a process that enables their growth in foreign environments. In fact, colorectal tumor organoid libraries demonstrate the progressive loss of niche factor requirements during tumorigenesis (Fujii et al., 2016b). Thus, in CRCs, aberrant driver pathway signals delimit the niche-restricted growth of the cancer cells and permit their dominant overgrowth in the hostile environments of the remote tissues they invade or metastasize.

Although 30 years later the main message prevails, today it is less clear whether the acquisition of mutations is indeed sequential. Bioinformatic studies suggest that some CRCs may be 'born to be bad', wherein the driver mutations that confer invasive and even metastatic potential are specified early on (Sottoriva et al., 2015). In addition, it is currently disputed that the metastatic capacity is only acquired late in time, with the complete accumulation of somatic alterations (Hu et al., 2019, 2020). In fact, large sequencing efforts comparing paired metastatic lesions and primary tumors have failed to find metastases-specific oncogenic mutations for most CRCs (Brannon et al., 2014; Zehir et al., 2017).

## Colorectal cancer models

Faithful disease modeling is an essential milestone in the discovery of novel therapeutic agents. In the last seven years, our field has witnessed a revolution in the way we model CRC, driven by the development of two technologies: CRISPR-Cas9 genome editing (Box 2) (Jinek et al., 2012) and 3D organoid cultures (Sato et al., 2009).

**Box 2 | Genome engineering with CRISPR-Cas9**



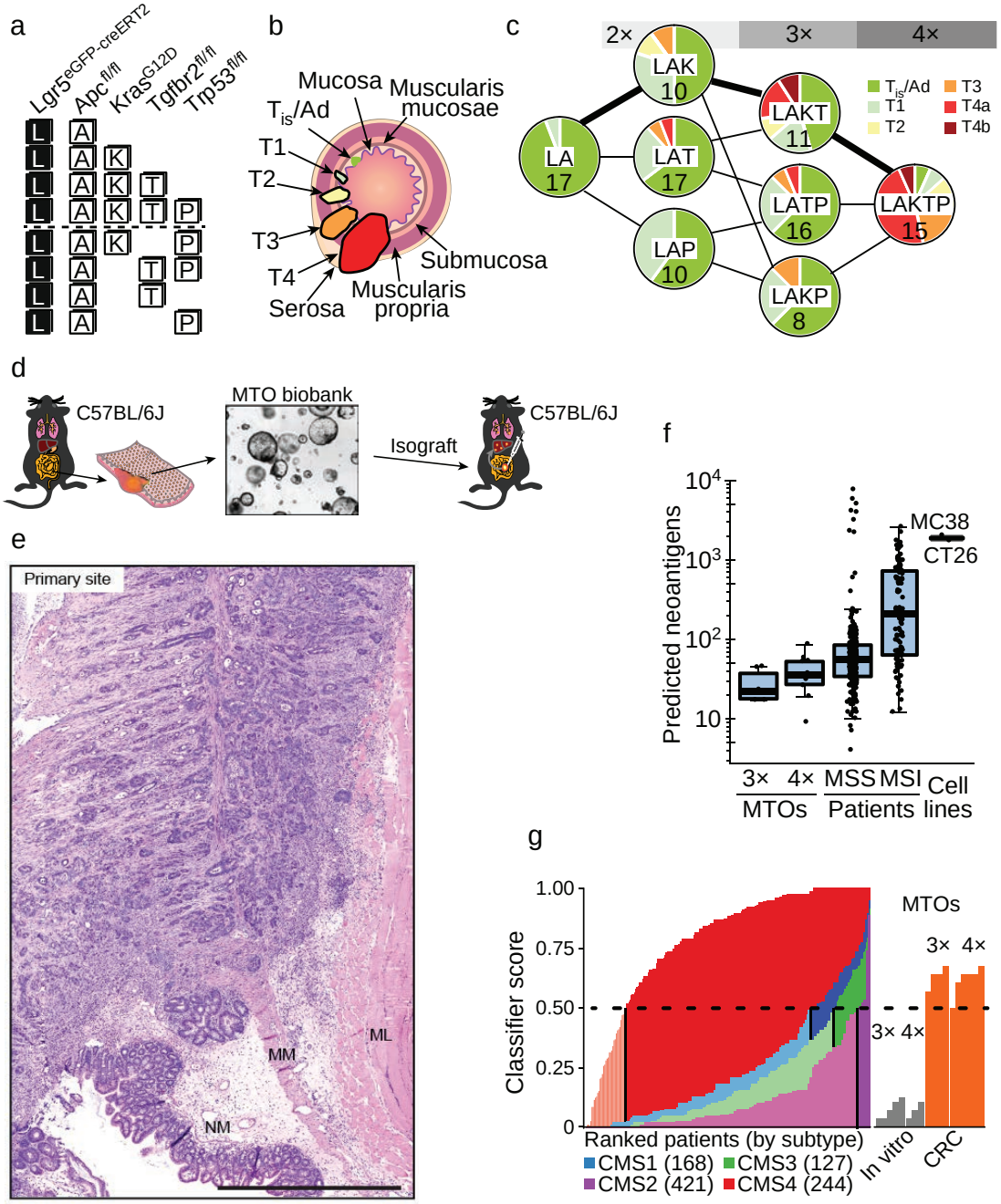

The sgRNA binds to the Cas nuclease, directing it towards specific genomic regions. Upon localization, the Cas protein creates a DSB that is repaired via the NHEJ or HDR pathway. NHEJ-mediated repair occurs throughout the cell cycle and is error prone in the absence of a repair template, ultimately resulting in small insertions or deletions at the cut site, thereby affecting gene function. The HDR pathway, in contrast, is a high-fidelity repair mechanism that utilizes an endogenous (e.g., sister chromatid) or exogenous DNA template to accurately repair the DSB. This property can be used to incorporate DNA elements of foreign origin, such as nucleotide variants, tags and loxP sites. Abbreviations: DSB, double-strand break; gDNA, genomic DNA; HDR, homology-directed repair; Indels, insertions/deletions; NHEJ, nonhomologous end-joining; sgRNA, single guide RNA.

For decades, the discovery of anti-cancer drugs relied on human cancer cell lines cultured in vitro or xenotransplanted into immunodeficient mice. However, cell lines represent selected subpopulations of the original tumor grown flat in plastic, and often they do not reflect the histopathological features of human cancer. On the other hand, mice engineered to carry up to two genetic alterations in oncogenes or tumor suppressor genes have been instrumental to model the early steps of intestinal carcinogenesis. APCmin, APC+/- Smad4-/-, BrafV637E, APCfl/+ TP53fl/fl, APCfl/+ KRasG12D/+, and many other genetically engineered mice (Heijstek et al., 2005; Jackstadt and Sansom, 2016) have been widely used to study the formation of adenomas and further progression to invasive carcinomas. Yet, these compound mutant mice were not useful to model late stage human CRC as they develop metastasis with low penetrance and long latency. Thus, for many years, late-stage CRC modelling in a fully immunocompetent background relied on two mouse cell lines. In this regard, CT26 (or Colon26, BALB/C) and MC38 (BL/6) are 2D cultured cell lines from the 1970s derived from chemical carcinogenesis. These cell lines give rise to solid tumors with little stroma upon transplantation and, puzzlingly, they display a sarcoma-like histology that differs substantially from the glandular organization characteristic of human CRCs (Tauriello et al., 2018). In addition, due to chemical-exposure these cell lines harbor thousands of genetic alterations that do not

reproduce the genetic makeup of CRCs (Tauriello et al., 2018, I-Figure 12f).

The recent development of 3D epithelial organoid cultures provides a nearly unlimited in vitro source of genetically stable tissue (Sato et al., 2009). Organoids can be easily maintained and manipulated in vitro and they faithfully recapitulate the characteristics of tissues of origin (Sato et al., 2013). The establishment of living tumor organoid biobanks containing collections of patient-derived tumor organoid (PDO) cultures offers a platform for high-throughput drug screens (Yao et al., 2020). This potentially allows correlations between tumor behavior and patient prognosis, as well as allowing the development of patient-specific treatment regimens (personalized medicine) (Fujii et al., 2016b; Roerink et al., 2018). Moreover, with the onset of CRISPR technology, it has recently become feasible to grow healthy human or mouse-colon epithelium in vitro and to sequentially introduce four of the most common mutations in CRC: in APC, KRAS, TP53, and SMAD4 genes (Drost et al., 2015; Matano et al., 2015; Melo et al., 2017). Orthotopic transplantation of murine and human tumor organoids allows the study of both primary tumor formation and spontaneous development of metastases. Importantly, it circumvents the drawbacks of GEMMs which — due to the high tumor burden through the whole intestinal tract — die before tumors have progressed to a metastatic stage. In fact, transplantation of tumor organoids was instrumental to demonstrate that the acquisition of alterations in these four driver pathways is necessary for tumors to metastasize efficiently (Drost et al., 2015; Fumagalli et al., 2018; Matano et al., 2015). In addition, the introduction of mutations in MMR genes by means of CRISPR/Cas9 has enabled the definition of a mutational signature specific of MSI tumors (Drost et al., 2017).

Although PDO technology represents a powerful resource for finding effective therapeutic strategies, xenografting into immunodeficient mice precludes the study of an intact TME and cancer immunity in particular. To overcome this limitation, we recently built a preclinical model of advanced CRC in immunocompetent mice (Tauriello et al., 2018). Our lab crossed mice bearing the four main genetic alterations present in human CRC — Apcfl/fl, KrasLSL-G12D, Trp53fl/fl, Tgfbr2fl/fl (A, K, P, T) — and conditionally triggered recombination in ISC (Lgr5-CreERT2) (I-Figure 12). Quadruple compound mice developed aggressive, human-like MSS CRC tumors that formed metastasis in 40% of the cases (I-Figure 12). Given the ethical and practical constraints of working with mice bearing multiple mutated alleles, we generated a biobank of mouse tumor organoids (MTOs) with different genetic alterations isolated from CRC primary tumors and liver metastases (LiMs). MTOs can be injected into the caecum wall of syngeneic immunocompetent mice to generate primary tumors that faithfully reproduce the traits of the tumor of origin. In addition, MTOs can be injected into the portal circulation to assess liver metastatic potential.




**I-Figure 12 | Immunocompetent mouse model for advanced CRCs. a**, Combination of mutant alleles used to generate mouse models. **b**, Classification of tumors according to the AJCC–TNM system. **c,** T- diagnosis per mouse, number of mice is indicated in the circles. The number of mutant pathways is indicated on top. d, Schematic of MTO biobank and isograft technology. **e**, Hematoxylin and eosin staining of an AKTP-T4 carcinoma developed from an orthotopically isografted 4x MTO. **f**, Predicted Major Histocompatibility Complex-I (MHC-I) neoantigens in MTOs, human MSS and MSI CRCs and mouse CRC cell lines. **g**, Left, cross-validation of consensus molecular subtype 4 (CMS4) classifier on human CRCs. Patients (shown as vertical bars) are ranked by subtype on the basis of classifier score. Right, classifier applied to 3× and 4× mutant MTOs in vitro and to orthotopic isografts of those MTOs. Adapted from (Tauriello et al., 2018).

# Tumor heterogeneity

## Cancer stem cells

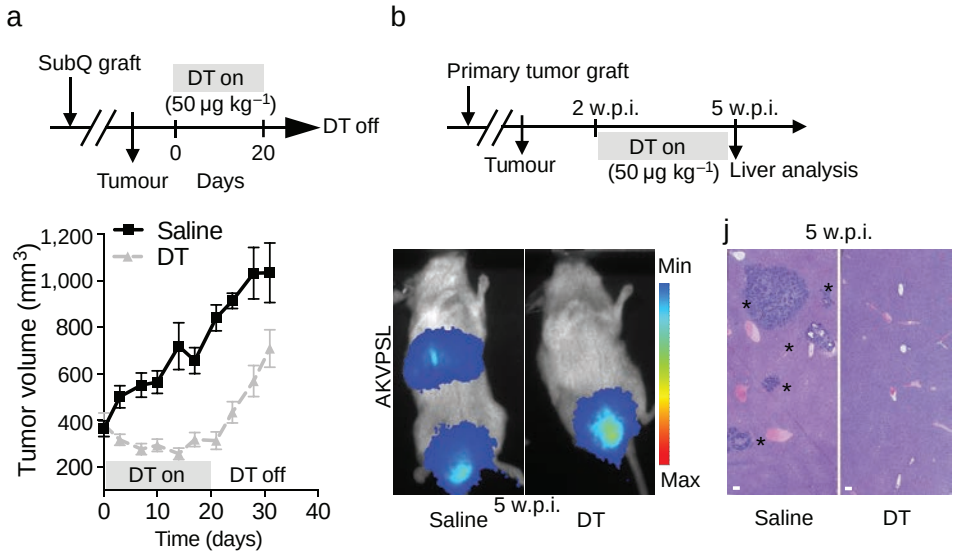
Cellular heterogeneity is a well-established hallmark of advanced cancers that fuels resistance to therapies (Hanahan and Weinberg, 2011). The advent of single cell expression profiling technology has uncovered strikingly high levels of intra-tumor heterogeneity across multiple cancers. Interestingly, phenotypic and functional heterogeneity arise among cancer cells within the same tumor not necessarily as a consequence of genetic diversity, but often due to environmental cues and footprints of homeostatic differentiation programs (Dalerba et al., 2011; Tirosh et al., 2016). In this framework, the Cancer Stem Cell (CSC) concept states that tumor growth, analogous to the renewal of healthy tissues, is fueled by small numbers of dedicated stem cells. In the last decade, CSCs have been identified in many cancer types, including leukemia, breast, brain, pancreatic, prostate, lung, ovarian and colorectal cancer (Batlle and Clevers, 2017).

Despite the acquisition of genetic alterations in key driver pathways that deregulate microenvironmental control of stemness and proliferation, colorectal tumors maintain a CSC hierarchy (Melo et al., 2017; Merlos-Suárez et al., 2011; Morral et al., 2020; Schepers et al., 2012; Shimokawa et al., 2017). The presence of cancer stem cells within these tumors was first suggested in three parallel studies published in 2007 (Dalerba et al., 2007c; O'Brien et al., 2007; Ricci-Vitiani et al., 2007a). These studies centered on the ability of the cell surface markers CD133 or CD44 to separate tumor initiating cells (TICs) from non-TICs. Studies from our lab and others later corroborated that human CRCs comprise heterogeneous cell populations and showed that cell hierarchies were reminiscent of the normal colonic epithelium (Dalerba et al., 2011; Merlos-Suárez et al., 2011).

More recently, two studies formally demonstrated the dependence of colorectal tumors on CSCs for metastatic outgrowth. These studies used CRISPR–Cas9 gene-editing technology to insert cassettes into the LGR5 locus of CRC patient–derived organoids for lineage tracing and ablation experiments (Box 1) (Melo et al., 2017; Shimokawa et al., 2017). These experiments revealed that the output progeny of Lgr5+ colorectal cancer cells produce is proportional to the size of xenografts, whereas tumor cells expressing the terminal differentiation marker gene keratin 20 (KRT20) produced progeny that mostly persisted as single cells or disappeared over time (Shimokawa et al., 2017). In addition, experiments using diphtheria-toxin (DT) to ablate $Lgr5^{+}$ cells showed that tumor progression was halted during cell ablation but rapidly resumed growth upon treatment discontinuation (I-Figure 13a) (Melo et al., 2017). These observations suggest that $Lgr5^{+}$

cells drive tumor growth in advanced CRC but in their absence, tumors are maintained by Lgr5$^-$ cancer cells, which constantly attempt to replenish the CSC pool. Lgr5$^+$ CSCs re-appear upon treatment discontinuation, leading to rapid tumor regrowth. This plasticity is reminiscent of the normal intestine, where homeostasis is maintained by alternative cell compartments that compensate for Lgr5$^+$ cell loss (Tian et al., 2011).






**I-Figure 13 | Lgr5+ CSCs in primary and metastatic CRC. a**, Experimental setup scheme showing Diphtheria toxin (DT) dosing schedule. Growth rate of subcutaneous AKP tumors in response to treatment. **b**, DT dosing schedule for Lgr5-lineage ablation in orthotopic tumors. Representative bioluminescent images of mice with orthotopic AKPS tumors after the indicated time and treatment. Images of liver sections stained with hematoxylin and eosin from saline- and DT-treated animals with orthotopic AKPS tumors. Adapted from (Melo et al., 2017).

Taken altogether, these studies demonstrate that colorectal tumors present a heterogeneous composition and a hierarchical but plastic cell organization. Why and how these hierarchies are maintained in WNT aberrant backgrounds remains unresolved. The balance between stemness and the differentiation phenotype in CRCs may depend on additional signaling pathways or the surrounding stroma. Intriguingly, De Sousa e Melo et al. showed that ablation of Lgr5+ CSCs halted the growth of AKTP liver metastases but was ineffective in AKTP intestinal primary tumors (I-Figure 13a,b) (Melo et al., 2017). This data underscores the importance of the environment in the dynamics of cellular heterogeneity and tumor growth.

CSC targeting is now a focus of drug discovery efforts that are in the initial stages of clinical evaluation. Nonetheless, the high plasticity of stem cell hierarchies is complicating the success of CSC-based therapies (Pastushenko et al., 2018; de Sousa e Melo and de

Sauvage, 2019). It has gradually become clear that, in many tumors, CSCs are conceived in dedicated niches (Lenos et al., 2018). Thus, given that CSCs will always be re-created as long as the tumor stem cell niche remains intact, targeting the niche emerges as a more attractive therapeutic approach than the pursuit of the continuous elimination of the CSC population. Still, this strategy might fail in aggressive tumors that become completely independent of niche signals (Drost et al., 2015; Fujii et al., 2016b; Matano et al., 2015).

## Cell of origin of metastases

Extensive genome sequencing efforts comparing metastatic lesions with primary tumors have failed to find any metastasis-specific mutations (Brannon et al., 2014; Zehir et al., 2017). This suggests that the ability to colonize foreign organs does not depend on the acquisition of further genetic alterations. Given that tumors retain cell-specialization (up to a certain degree), many researchers have started to wonder whether the ability to migrate and disseminate throughout the body is restricted to a specific cell lineage.

Cancer stem cells have been put forward as the most probable origin of metastatic relapse in CRC due to their self-renewal capacity (Batlle and Clevers, 2017; Merlos-Suárez et al., 2011). To study this notion, Fumagalli et al. used intravital multiphoton microscopy to visualize an Lgr5-GFP reporter in orthotopically transplanted AKP CRC tumors (Fumagalli et al., 2020). Interestingly, they observed that most migratory cells escaping from the primary tumor, in blood circulation and seeding the liver were Lgr5-GFP negative. During liver metastatic growth they observed the reappearance of Lgr5+ cells, suggesting a plastic conversion of Lgr5- into Lgr5+ cells. These observations are consistent with those made by De Sousa E Melo et al. who showed a requirement for Lgr5+ cells during metastatic outgrowth (I-Figure 13b) (Melo et al., 2017). However, a major caveat of this study is that the vast majority of cells in primary tumors are already Lgr5- (93.9%), in similar levels to invasion fronts, circulation or micrometastases (Fumagalli et al., 2020). Thus, authors cannot conclude that there is an enrichment of Lgr5- cells in metastatic settings compared to an advanced primary tumor.

Recently, Joan Massagué's laboratory proposed that L1 cell adhesion molecule (L1CAM) is essential for CRC organoid growth, epithelial regeneration after colitis, orthotopic carcinoma growth, liver metastases colonization and chemoresistance (Ganesh et al., 2020). Ganesh et al. found that L1CAM knock-down (KD) cells were less efficient to form organoids in part due to deficient adhesion to laminins. L1CAM expression was induced upon organoid disruption and L1CAM$^{high}$ cells had higher organoid-initiating capacity, suggesting increased self-renewal capacity. However, the relationship with canonical

stem cell markers such as LGR5 was unclear, since subpopulations of L1CAM$^{high}$ LGR5$^{high}$ and L1CAM$^{high}$ LGR5$^{low}$ cells were present in the 4 patient samples analysed. Moreover, the detrimental effect of L1CAM deficiency in steady-state organoids and primary tumors raises some doubts regarding the conclusions drawn in the metastatic setting. In liver metastases colonization experiments, organoids were pretreated with doxycycline for 48 hours to downregulate L1CAM before dissociation and injection, thereby already generating differences in cellular fitness unrelated to the metastatic context. With regards to orthotopic experiments, L1CAM KD primary tumors were consistently smaller, which might explain the reduced metastatic burden observed in liver and lungs. In my opinion, the breadth of phenotypes observed suggests that downregulating L1CAM compromises cell viability in virtually every context and thus is not particularly required in metastatic processes.

Other studies aiming to find the cell of origin of metastases have compared transcriptomic profiles of recurrent versus non-recurrent stage I-III colorectal cancer patients (Merlos-Suárez et al., 2011, De Sousa E Melo et al., 2013; Marisa et al., 2013; Sadanandam et al., 2013). However, as I will discuss in the next section, our lab discovered that most genes predicting poor prognosis are expressed by the tumor microenvironment. Next, I will summarize the importance of non-mutated stromal cells in the progression of cancer.



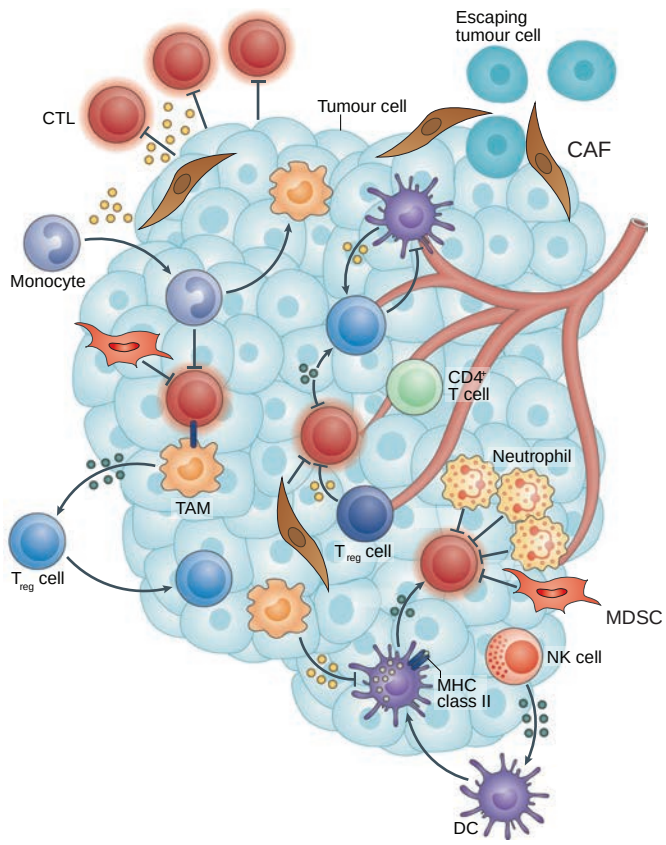
## Tumor microenvironment

### The tumor microenvironment (TME)

The idea that cancer is regulated by interactions of cancer cells ("seed") with their microenvironment ("soil") was first postulated by Stephen Paget over a century ago (reviewed by Maman & Witz, 2018). Nonetheless, cancer was mostly considered a mass of growing epithelial cells until the 1980s (Hanahan and Weinberg, 2000). Nowadays, it is well established that — despite losing growth constraints and abandoning their native niches — tumor cells engage new interactions with surrounding non-malignant cells and these are essential to understand cancer biology. Thus, as in tissue homeostasis — where a plethora of cells coordinate to effect normal physiology —, the cancer ecosystem comprises a rich variety of non-epithelial cells orchestrated by tumor cells.

The tumor microenvironment (TME) consists of various cellular types surrounding mutated cancer cells, as well as a structural network of extracellular matrix (ECM), signaling molecules, metabolites, physical conditions (e.g., pH, oxygen, stiffness), and other host

## Box 3 | Inhabitants of the tumor microenvironment




**Tumor-associated macrophages (TAMs):** 'Classically activated' M1 macrophages contribute to tumor rejection through type 1 cytokine production and antigen presentation, whereas M2 macrophages enhance angiogenesis and matrix remodeling through type 2 cytokine production. TAMs share M2 characteristics; their presence is often tumor-promoting and they are associated with poor prognosis.

**Myeloid-derived suppressor cells (MDSCs):** MDSCs accumulate in almost all patients and animal models with cancer. MDSCs suppress anti-tumor immune responses and promote the invasive potential of tumor cells. Factors such as IL-6 or GM-CSF induce the recruitment of these cells.

**Tumor-associated neutrophils (TANs):** Neutrophils are the most abundant cell type in the blood and the first to reach a site of infection to protect us (Tüting and De Visser, 2016). They also frequently accumulate in solid tumors, and studies in lung or brain metastases suggest that they play a major role during metastatic dissemination (Coffelt et al., 2015; Wellenstein et al., 2019).

**Dendritic cells (DCs):** Mature DCs are the most proficient antigen-presenting cells. Hence, unsurprisingly their infiltration is associated with a favorable prognosis in colorectal, head and neck and other types of cancer (Gardner and Ruffell, 2016).

**Endothelial cells:** Endothelial cells comprise the blood vasculature of the tumor, and increased density is frequently associated with poor prognosis (Joyce and Pollard, 2009).

**Cancer-associated fibroblasts (CAFs):** CAFs represent a large component of the stroma and they are regarded as tumor promoting in most types of cancer (Calon et al., 2012; Kalluri, 2016).

**CD4+ T cells:** T helper type 1 cells support CD8+ T cells in tumor rejection, whereas T helper type 2 cells polarize immunity away from an anti-tumor response. Regulatory T cells dampen CD8+ cell activation and Natural Killer (NK)-mediated cell killing.

**CD8+ T cells:** Also known as cytotoxic T lymphocytes (CTLs), they are the effector cells of the adaptive immune system and they specifically recognize and destroy cancer cells through perforin- and granzyme-mediated apoptosis.

**B cells:** B lymphocytes are another main player in adaptive immunity. They recognize circulating antigens in its native form and respond by secreting protective antibodies. Nonetheless, their role in cancer immunity is less prominent.

**Natural Killer (NK) cells:** NK cells are effector lymphocytes of the innate immune system that are cytotoxic to cancer cells through the perforin– granzyme pathway. NK cells contribute to cancer immune surveillance by sensing MHC-I loss or deregulation.

Drawing adapted from (Garner and de Visser, 2020)

factors such as inflammation and the microbiota. Non-mutant (i.e. *wild-type*) neighboring cells typically comprise distinct cell types, such as tissue parenchymal cells, vascular cells, stromal fibroblasts and a variety of bone marrow-derived cells (BMDCs), including lymphocytes, monocytes, macrophages, myeloid-derived suppressor cells (MDSCs), neutrophils, mast cells, dendritic cells (DCs) and others (Box 3). In this section I will discuss the evidence that the TME can exert inhibitory effects on malignant cells and, how tumors circumvent these inhibitory signals during cancer progression and instead exploit the TME to their own ends (Hanahan and Weinberg, 2011).

Interactions with the TME are essential to promote an innumerable variety of processes that are vital for cancer cells, such as the maintenance of cell proliferation, induction of stemness, promotion of invasion and metastasis, and evasion of anti-tumor immunity (Hanahan and Coussens, 2012). In this thesis, I will specially focus on mechanisms that mediate cancer immune surveillance and how cancer cells cooperate with the TME to avoid it.



## Anti-tumor immunity, immune escape and immunotherapies

Even the simplest forms of life have mechanisms to protect themselves against infections. The immune system is a network of biological processes that work in coordination to protect an organism from pathogens, cancer cells and foreign objects. In vertebrates, there are two major immune subsystems, namely the innate and the adaptive. The innate immune system provides a preconfigured fast response that detains a broad group of stimuli. In the meantime, the adaptive immune system builds a tailored response to each individual stimulus by recognizing specific molecules. These two components must be coordinated to produce an effective immune response.

The immune system is capable of recognizing and killing tumor cells. However, its anti-tumoral role remained unappreciated for years because tumors not only effectively suppress immune responses but also corrupt them to their advantage (Chen and Mellman, 2017; Mellman et al., 2011).

Early observations in cancer patients indicated that tumors were more prone to arise in chronic inflammation sites, and subsequent studies in mouse models demonstrated a link between inflammation and cancer initiation and progression (Coussens et al., 2013; Hanahan and Weinberg, 2011). In an inflammatory process, immune cells are first recruited to destroy pathogens and clear out damaged cells. A second phase then induces tissue regeneration through the secretion of growth factors and ECM. In 1986, thinking ahead of his time, Harold Dvorak described tumors as “wounds that do not heal” (Dvorak, 1986). Dvorak suggested that tumors were able to subvert and benefit from biolog-

ical processes similar to those occurring during the second wave of a wound healing response. Indeed, collective evidence now supports that inflammation is malleable, and that tumors reprogram immune cells towards pro-tumorigenic and immunosuppressive responses (Braun et al., 2021; Joyce and Fearon, 2015; Thorsson et al., 2018).

The first experimental evidence that the endogenous T cell compartment could help control tumor growth (a process named cancer immune surveillance) was obtained by Robert Schreiber's group, who showed that mice lacking an intact immune system were more susceptible to carcinogen-induced and spontaneous cancers (Shankaran et al., 2001). Multiple studies in mice have since demonstrated the existence of endogenous tumor-specific responses, which can be boosted with drugs named immunotherapies (Koebel et al., 2007; Laouar et al., 2002; Mcgranahan et al., 2016). In recent years, cancer immunotherapies have revolutionized the field of oncology by showing unprecedented efficacy in several human malignancies (Brahmer et al., 2012; Topalian et al., 2012, 2020).

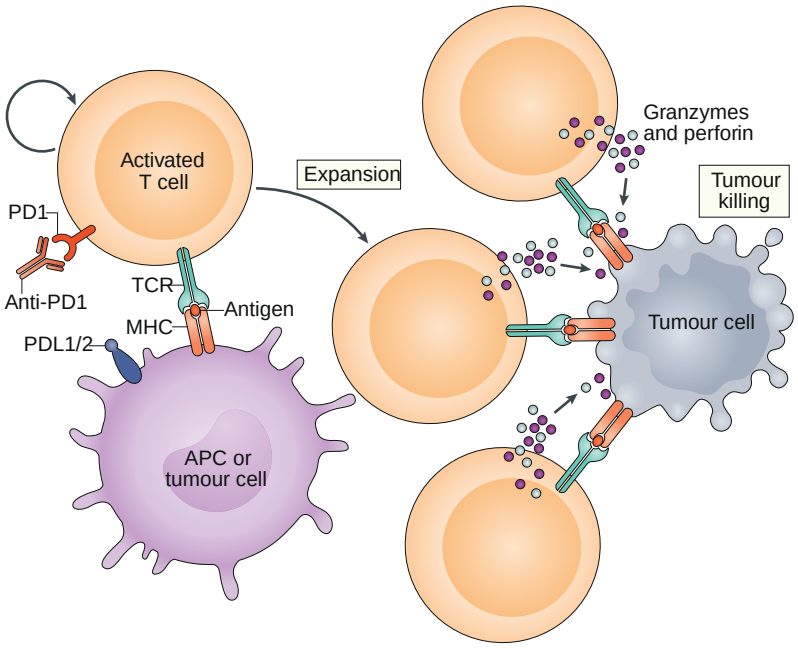
The paradoxical properties of the immune system in cancer have puzzled the research community for many years and in part owing to the functional plasticity of leukocytes. In this regard, considerable effort is being devoted to gaining a better understanding of the immune landscape of tumors, and single-cell RNA sequencing technologies are resulting pivotal to obtain comprehensive, unbiased analyses of cellular diversity. The more we learn, the more we discover that, depending on the tissue context and stimuli present, individual cell types have opposing functions. For example, type 1 CD4+ T cells (TH1) support CD8+ T cells in tumor rejection, whereas type 2 CD4+ T cells (TH2) and CD4+ T regulatory cells block the activation of CD8+ T cells. This huge complexity results in the lack of clear associations between the presence of major immune cell types and defined outcomes across different TMEs and cancer patients (Gentles et al., 2015).

In the following three sections, I will describe the normal functioning of cancer immunity and the mechanisms co-opted by tumors. Finally, I will review some of the most successful immunotherapies that are already changing the lives of some cancer patients.

## Adaptive anti-tumor immunity

The existence of anti-cancer immunity was debated for decades, yet it is now well established that the immune system can recognize and kill cancer cells (Dunn et al., 2004; Schumacher and Schreiber, 2015). In previous sections, I have reviewed how tumor cells accumulate genetic alterations that provide a selective advantage. Nonetheless, this increasing diversity comes at a cost: the further a cancer cell diverges from a normal cell, the more likely it is to be recognized by the immune system. It has been demonstrated

that the mutational burden of tumors contributes to their immune recognition and that it partly determines the response to cancer immunotherapy (Le et al., 2015; Mcgranahan et al., 2016; Rizvi et al., 2015). T cells are the central mediators of adaptive immune responses in tumors and they are distinguished by having antigen receptors (TCRs). TCRs are generated by random rearrangement of gene segments, which, followed by selective processes, result in a vast repertoire of T cell clones that recognize uncountable non-autologous antigens. Thus, through TCRs, T cells are able to recognize new epitopes translated by malignant cells that are displayed on major histocompatibility complexes (MHCs), on their own surface or on professional antigen-presenting cells (APCs) (I-Figure 14).

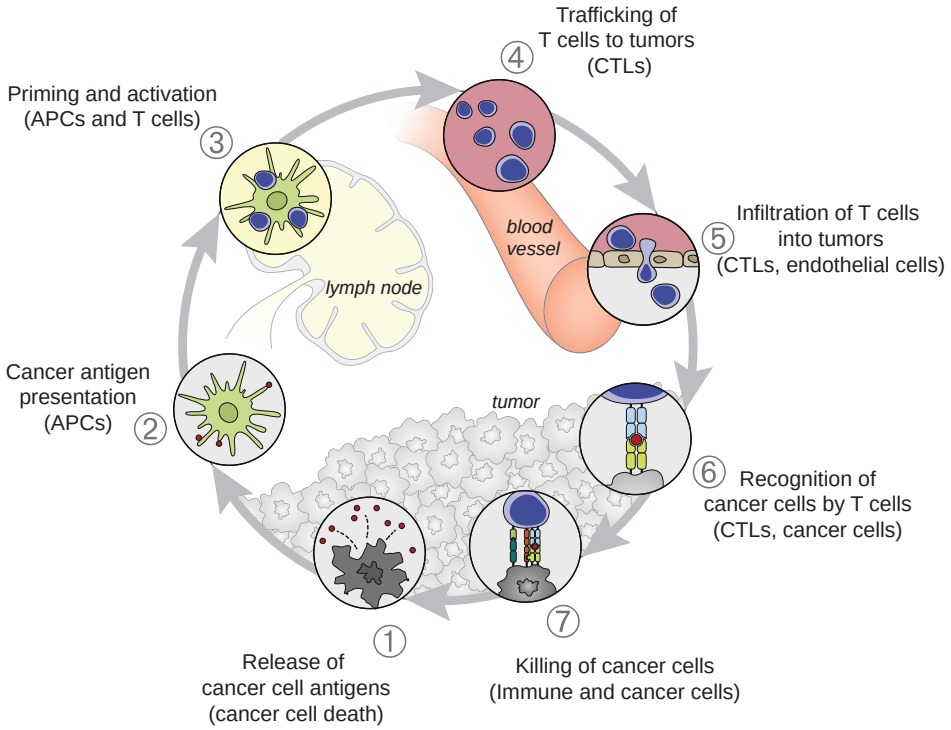



**I-Figure 14 | T cell priming, effector activity and the PD1 axis.** Antigen-presenting cells (APCs) load antigens onto MHC molecules to prepare for contact with CD4 T cells that display cognate T cell receptors (TCRs). For proper activation, antigen presentation must be in conjunction with CD28 costimulatory binding to B7-1/B7-2. CD8 T cells bind to tumor antigens presented on cancer cells by MHC-I, which ultimately leads to the release of cytolytic mediators, such as perforin and granzyme, causing enhanced tumor killing. Late after activation in peripheral tissues, programmed cell death 1 (PD1) expression is induced, thereby promoting T cell exhaustion upon binding to its ligands PDL1 and PDL2. Blocking the PD1 axis through the administration of anti-PD1 (or anti-PDL1 or anti-PDL2) antibodies prevents this inhibitory interaction and unleashes antitumoral T cell activity. Adapted from (Waldman et al., 2020).

Tumor antigens can be grouped into two classes: tumor-specific antigens (TSAs) and tumor-associated antigens (TAAs). TSAs arise mainly due to oncogenic (driver or passenger) non-silent mutations that lead to the production of new protein sequences, known

as neoantigens. In tumors with viral etiology, such as cervical cancer, neoepitopes can also be derived from viral open reading frames. The second group is formed by non-mutated proteins to which T cell tolerance is incomplete, for example because of low or tissue-restricted expression (Schumacher and Schreiber, 2015). Antigens are loaded into proteins belonging to MHC class I (MHC-I) and MHC class II (MHC-II). MHC-I is expressed in the surface of all nucleated cells in our body and displays endogenous peptide fragments to cytotoxic CD8+ T cells (CTLs). On the other hand, MHC-II is expressed mostly by APCs - DCs, macrophages and B cells -, which uptake extracellular peptides and present them to helper CD4+ T cells. Importantly, TCR binding to MHCs is insufficient for full T cell activation and requires the interaction of costimulatory molecules CD28 and B7. While CD8+ T cells carry out direct cytotoxic reactions that kill infected or neoplastic cells, CD4+ T cells orchestrate adaptive immunity by producing cytokines with chemotactic and pro-inflammatory properties (Waldman et al., 2020).

Thus, an effective anticancer immune response consists of a series of events that occur iteratively, commonly known as the Cancer-Immunity Cycle (I-Figure 15) (Chen and Mellman, 2013). In the first step, cancer antigens are captured by DCs in the lymph nodes. These cells then present the captured antigens on MHC-II molecules to T cells, resulting in priming and activation of effector T cell responses against cancer-specific antigens. Environmental signals will specify the nature of the immune response, balanced either towards attack or tolerance of the insult, which will be reflected by the ratio of T effector cells versus T regulatory cells. Finally, activated T cells traffic to the tumor site and




**I-Figure 15 | The Cancer-Immunity Cycle.** Immune surveillance is a cyclic process that

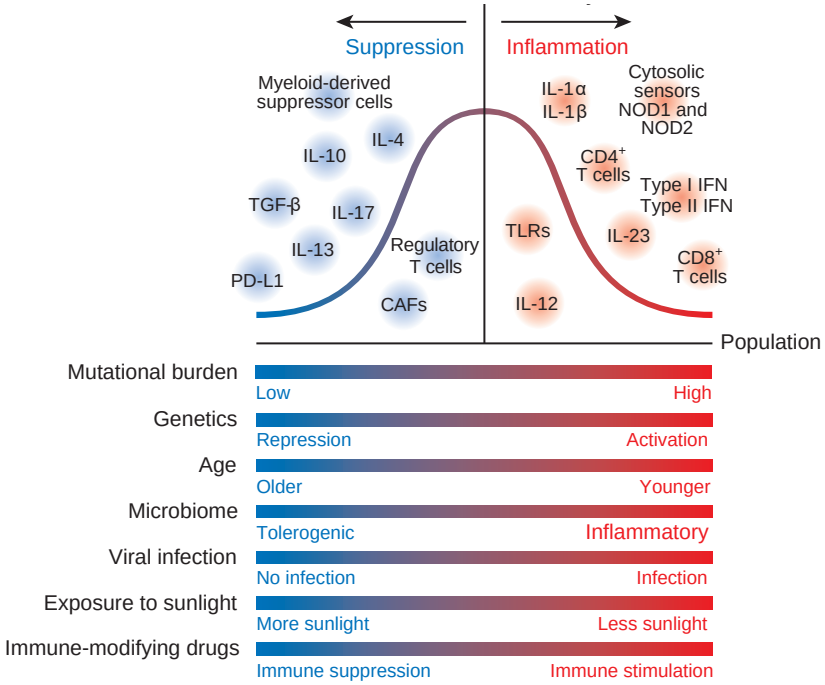
should be self-propagating, leading to an accumulation of immune-stimulatory factors that amplify and broaden T cell responses. Yet, the cycle is also characterized by inhibitory factors that halt or limit the development of immunity. This cycle can be divided into seven major steps, starting with the release of antigens from the cancer cell and ending with the killing of cancer cells. Each step is described above, with the primary cell types involved. Adapted from (Chen and Mellman, 2013).



infiltrate the tumor. They will then recognize tumor cells through TCR-MHC-I binding and secrete cytotoxic proteins to kill their target cancer cells. Effective killing of cancer cells releases additional tumor-associated antigens, which will increase the depth of the response in subsequent revolutions of the cycle (I-Figure 15).

## Elements of cancer immune evasion

In cancer patients, virtually every step of the Cancer-Immunity Cycle can be hijacked by tumors (tumor cells + the TME), thus impeding optimal performance. Tumor antigens may not exist or their presentation might be suppressed. In addition, environmental signals can influence T cell priming to treat antigens as self rather than foreign, thereby creating regulatory rather than effector T cell responses. T cells may not migrate properly to tumors or they may be inhibited from infiltrating them. Finally, factors in the TME might suppress effector cells and inhibit their cytotoxic capacity. Thus, the immune profile of a tumor reflects the contribution of an array of factors, including intrinsic properties of




**I-Figure 16 | Multiple factors balance tolerance and immunity.** Anticancer immunity is subject to a multitude of factors that contribute to the generation of activated T-cell

immunity or tolerance. These factors are distributed in a continuum (red–blue bars) and act to determine the set point by directly or indirectly controlling the expression of tolerogenic (blue circles) or immunogenic (red circles) cytokines and cell types. Adapted from (Chen and Mellman, 2017).

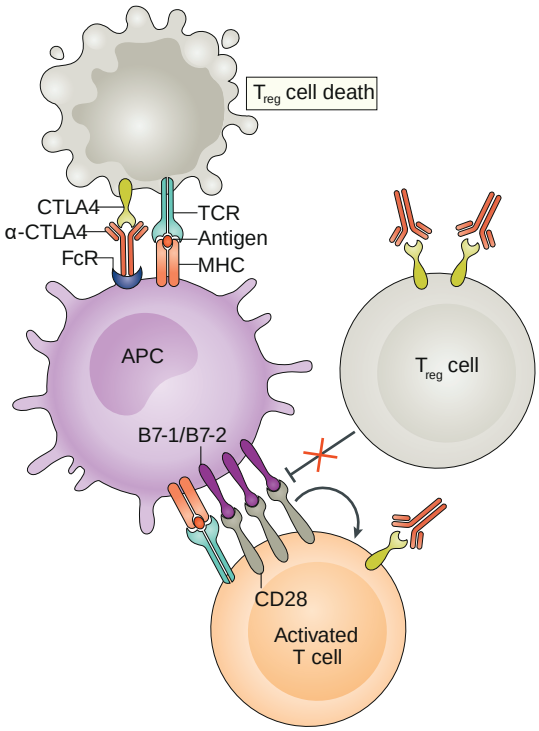
tumor cells (such as mutational landscape and cytokine secretion), the composition and features of the TME, and extrinsic factors of the host, such as the gut microbiota, the presence of infection or exposure to sunlight. Small variations in these factors — rather than dramatic alterations — may be sufficient to tip the balance between tolerance and immunity (Chen and Mellman, 2017) (I-Figure 16).

Large exome-sequencing datasets across distinct cancer types have revealed that the number of observed neoantigens is lower than expected based on mutation rates, a process of subclonal selection known as immunoediting (Dupage et al., 2012; Matsushita et al., 2012). In addition, cancer cells may acquire resistance to adaptive immunity by, for instance, reducing or abrogating antigen presentation (Kalbasi and Ribas, 2020; Yamamoto et al., 2020). Interestingly, induced quiescence of metastatic tumor cells upon arrival to a foreign organ has been associated with downregulation of antigen presentation (Agudo et al., 2018; Malladi et al., 2016). Finally, some tumors with a high mutational burden (TMB) might not have sufficient neoantigens to be recognized by the immune system (Schumacher and Schreiber, 2015). In fact, several studies argue that many T cells in tumors are unable to recognize cancer cell antigens and instead recognize a wide range of epitopes unrelated to cancer. These T cells may be alternatively associated with tissue disruption and inflammation, and it is currently unknown whether they are indeed innocent “bystanders” (Simoni et al., 2018). Identifying which tumors are candidates to be recognized by the patient immune system is an unresolved medical issue and an urgent unmet clinical need.

Multiple environmental signals — produced by tumor cells or the TME — can suppress the priming or activation of the immune system. Preclinical studies in mouse models of cancer have attributed a major role to stromal cell types, such as cancer-associated fibroblasts (CAFs), macrophages, neutrophils and MDSCs (reviewed in Box 3), in favoring such immune evasion (Kaneda et al., 2016; Moynihan et al., 2016; Saha et al., 2017) .

One of the most potent mechanisms to suppress anti-tumor immune responses is activating negative regulatory pathways (called checkpoints) associated with immune homeostasis. Upon resolution of an inflammatory process, the immune system uses checkpoints to cease T cell priming and activity and thus prevent the onset of autoimmunity. Seminal work by Nobel-awardees James Allison and Tasuku Honjo led to the discovery of Cytotoxic T lymphocyte antigen 4 (CTLA-4) and programmed cell death 1 (PD-1), the

most prominent examples of T cell immune checkpoint molecules. CTLA-4 receptor is expressed in activated T cells and it competes with CD28 for costimulatory B7 ligands,




**I-Figure 17 | CTLA-4 checkpoint inhibition.** CTLA4 on Treg cells leads to trans-endocytosis of B7 ligands and interferes with the CD28 co-stimulatory ability of APCs. Cytotoxic T lymphocyte antigen 4 (CTLA4)-blocking antibodies (α-CTLA4), especially when bound to an Fc receptor (FcR) on an antigen-presenting cell (APC), can promote antibody-dependent cellular cytotoxicity (ADCC). In addition, α-CTLA4 can bind to CTLA4 on the surface of the Treg cell and prevent it from counter-regulating the CD28-mediated co-stimulatory pathways that are playing a role in T cell activation. Adapted from (Waldman et al., 2020).

thereby inhibiting proliferation and IL-2 secretion by T cells (I-Figure 17). PD-1 is also expressed by activated T-cells and, upon binding to its ligands PD-L1 and PDL-2, it attenuates the activity of T cells, leading to an exhausted phenotype (I-Figure 14). PD-L1 is induced in many tumors to evade immune attack and it can be expressed by tumor or stromal cells (Cerezo-Wallis et al., 2020; Noguchi et al., 2017; Saha et al., 2017; Tauriello et al., 2018).

Other important mechanisms of immune evasion include the secretion of proteins that may act directly on T cells and/or polarize the TME into immunosuppression. Known examples include TGF-β, IL-4, IL-10, IL-17, IL-33, G-CSF, GM-CSF, MDK or prostaglandins (Cerezo-Wallis et al., 2020; Coffelt et al., 2015; Massagué, 2008; Ruffell et al., 2014; Taniguchi et al., 2020). In addition, nutrient and oxygen availability, as well as metabolites such as tryptophan, arginine and adenosine, have been shown to affect the quality of the immune response (I-Figure 16) (Chen and Mellman, 2017).

In addition to the cytokines and hormones secreted by tumors, the nature of the host also influences the immune composition of a tumor. Some important components that have emerged over recent years are diet, microbiota and aging, which in turn are heavily interconnected. The gastrointestinal tract contains trillions of microbes that have diverse

roles in the maintenance of immunity and metabolism. Groundbreaking studies by Laurence Zitvogel showed that fecal microbiota transplantation (FMT) from immunotherapy-responsive patients increases responses to PD-1 therapy in mice, whereas FMT from non-responsive patients does not (Routy et al., 2018). Obese individuals are at higher risk of developing cancer and high-fat diets are related to the malfunctioning of our immune system (Beyaz et al., 2016; Pascual et al., 2017; Ringel et al., 2020). In addition, as we age, the activity of our immune system decreases due to processes like immune senescence (Chen and Mellman, 2017).

## Cancer immunotherapies

Genetic instability combined with phenotypic diversity and adaptability of cancer cells have hampered the anticipated success of targeted therapies in cancer treatment due to acquired resistances. Moreover, personalized therapies are highly time-consuming, hence making it difficult to keep pace with the tempo of cancer progression. T cells are a unique system endowed with sufficient diversity and adaptability to match tumors.

Cancer immunotherapies are drugs intentioned to leverage the cytotoxic potential of the human immune system. Importantly, the roadblocks in the cancer-immunity cycle can be multiple and diverse across tumor types and patients, thus bringing about a great variety of immune landscapes in cancer. There are various types of immunotherapies that aim to release the brakes that hold inactive anti-tumor immunity. Here I will focus on immune checkpoint blockade, immune modulators, adoptive cellular therapies and cancer vaccines.

Among the different types of cancer immunotherapy, immune checkpoint blockade (ICB) has had the broadest impact, with several antibodies targeting CTLA4 or the PD1/PD-L1 axis approved for use in different types of cancer (I-Figure 14 and 17). In 2010, the field was astounded by a landmark clinical trial that demonstrated that treatment with ipilimumab, an antibody against CTLA-4, improved the survival of patients with metastatic melanoma (Hodi et al., 2010). Since then, monoclonal antibodies that block PD-1 and PD-L1 have been approved for use by the U.S. Food and Drug Administration (FDA) across 17 types of advanced cancers, showing curative and durable responses (Topalian et al., 2020). A major advantage of ICB is that drug personalization towards tumor types or patients is not required, thus greatly reducing the costs associated with other immunotherapies (Waldman et al., 2020).

Unfortunately, most patients do not derive clinical benefit from ICB yet. A number of biological factors affect ICB efficacy, and one of the most recognized features is the TMB (Chowell et al., 2021; Rizvi et al., 2015). In fact, those cancer types with better response

rates have a higher mutational load per megabase of coding DNA (Schumacher and Schreiber, 2015). Some examples are tumors induced by ultraviolet light (melanoma) or other carcinogens such as tobacco (lung cancer smoker patients). However, none of these factors act in isolation and, thus, alone they cannot optimally identify patients who could benefit from ICB across different cancer types (Chowell et al., 2021).

Moreover, ICB might only work in combination with radiation or immunomodulatory drugs that impinge in other steps of the immune cycle. As I previously described, a milieu of factors present in the TME act locally or systemically to inhibit the immune response (I-Figure 17). Many preclinical studies have tested the role of several immunomodulatory agents in combination with ICB. These drugs either target specific immunosuppressive proteins or ablate subpopulations responsible for immunosuppression. Some examples of these drugs include: TGF-β inhibitors (Tauriello et al., 2018), IL-10 inhibitors (Ruffell et al., 2014), anti-CSF1 antibodies that block the recruitment of TAMs (Ries et al., 2014), PI3k inhibitors that reprogram myeloid cells (De Henau et al., 2016; Kaneda et al., 2016) and anti-GM-CSF antibodies that block the recruitment of MDSCs (Wu et al., 2014). Other articles have shown that antigen-release induced by radiation of tumor cells synergizes with ICB by enhancing the priming and activation of naïve T cells thus diversifying the anti-tumor TCR repertoire (Hwang et al., 2018). More recently, engineered bacteria have been used to reprogram the metabolism of immunosuppressed tumors and make them responsive to ICB (Canale et al., 2021). Overcoming these restrictions is key to revealing the potential anti-tumor effects of checkpoint antagonists, which might be regarded as ineffective when administered as monotherapy. This field of research is now enjoying unprecedented bench-to-bedside clinical success, with currently more than 3500 active clinical trials evaluating diverse combinations of T cell modulators (Waldman et al., 2020).

In the area of adoptive cell transfer, a successful immunotherapy is the transfer of T cells with engineered chimeric antigen receptors (CAR T cells) that aim to foster recognition of tumor antigens. CAR T cells are generated by isolating circulating T cells from cancer patients and introducing ex vivo the expression of recombinant TCRs that recognize tumor-specific antigens. On the basis of sensational results, autologous T cells engineered to express a CAR specific for the CD19 B lymphocyte molecule were recently approved by the FDA for treatment of B cell malignancies (refractory pre-B cell acute lymphoblastic leukemia and diffuse large B cell lymphoma) (June et al., 2018). Importantly, this therapy is not directed from a tumor-specific antigen but against a B-cell specific antigen, and it benefits from the fact that B cell ablation is tolerable in humans. On the other hand, the development of CAR T cell therapies in other cancers is limited by the requirement of a distinct tumor-restricted target antigen on the cell surface, as well as the penetration

of CAR-Ts in solid tumors (June et al., 2018). In fact, attempts to target tumor-associated antigens in solid tumors have achieved little success so far (Rosenberg and Restifo, 2015).

Cancer vaccines, which intend to promote antigen presentation by delivering antigenic peptides or nucleic acid into the host, are another immunotherapy that requires TSA. Cancer vaccines gained much attention in the past but they have not been successfully translated into the clinics yet (Palucka and Banchereau, 2012). The first vaccines aimed to elicit responses were designed for TAA, but the immune system tends to recognize such antigens as self, thus leading to weak responses and autoimmune-related toxicities (Waldman et al., 2020). Thus, the focus of current research has switched towards vaccines directed against TSAs (Castle et al., 2012).

Candidate neoantigens can be identified by combining the identification of mutant tumor-expressed proteins with algorithms that evaluate their potential to be neoantigens (Luksza et al., 2017; Schumacher and Schreiber, 2015). The process starts with exon sequencing of a cancer biopsy and normal tissue to identify the missense mutations present in tumor cells. RNA expression data is often collected to verify whether the mutant protein is expressed. Then, in silico methods are used to predict which somatic mutations might be MHC binders. However, most of these predicted neoepitopes will fail to turn-up when tested and only a handful are found to elicit a T-cell response. As a result, neoantigen discovery and application remains an obstacle towards personalized immunotherapies such as CAR-T cells or cancer vaccines (Schumacher and Schreiber, 2015).

Another major drawback of cancer vaccines and CAR-T cells is that — as it occurs with targeted therapies — (i) they rely on one or few tumor antigens, thereby becoming vulnerable to tumor adaptability and mutability, and (ii) they probably need to be tumor-specific (or even patient-specific), which increases their costs and time of production. While another type of adoptive cell transfer — tumor infiltration lymphocyte (TIL) transfer — leverages the complete tumor-reactive TCR repertoire, it still requires extensive experimental work for each patient (Rosenberg and Restifo, 2015).

Autoimmune disorders are one of the main complications derived from immunotherapies. Ironically, cytokine-storm syndrome (CRS) is considered an on-target effect of immunotherapies since its presence demonstrates that active T cells are at work in the body (Hwang et al., 2018).

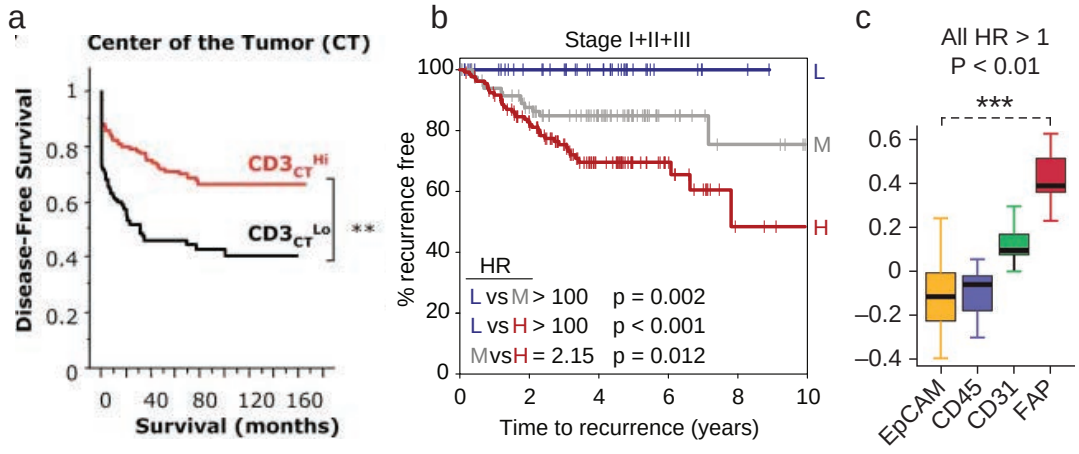
## The tumor microenvironment of advanced colorectal cancers

Colorectal cancer (CRC) is an excellent example of the paradoxical role of the immune

system in cancer. The pathogenesis of CRC is associated with inflammation. Individuals with chronic inflammatory conditions, such as Crohn's disease and ulcerative colitis, are at higher risk of developing CRC (Palucka and Coussens, 2016). In fact, inflammation is induced in mouse models with dextran sodium sulfate (DSS) to exacerbate intestinal tumorigenesis, whereas blockade of inflammatory pathways is linked to decreased incidence in patients and mice (Jackstadt and Sansom, 2016).

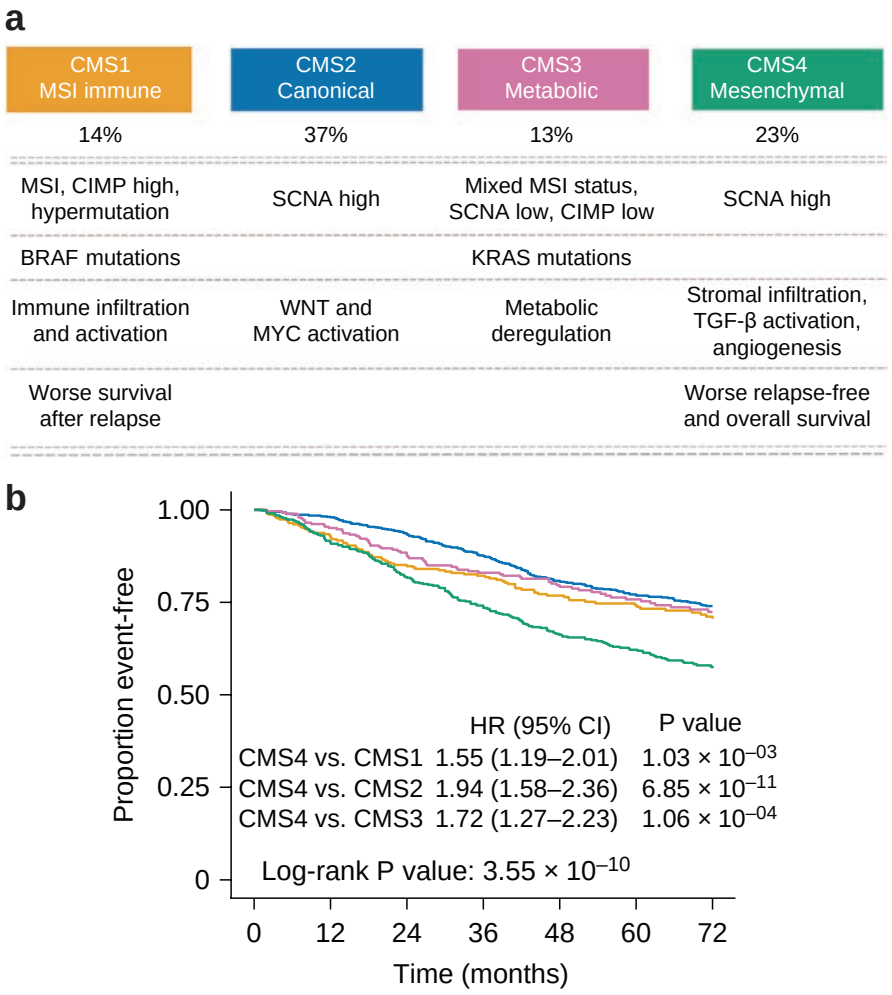
In contrast, Jerome Galon and colleagues showed that the type, density, and location of immune cells within tumors were a better predictor of patient survival than the standard methods used to stage CRC (Galon et al., 2006). More specifically, they found that high infiltration with CD3+ T cells predicted a lower risk of disease-recurrence, which helped to identify patients with a high risk of metastases who would benefit most from adjuvant therapy. This seminal article suggested that the adaptive immune response played a major role in preventing CRC recurrence (I-Figure 18a).

A few years later, our group showed that the amount of TGF-β (TGFβ1+2+3) was another independent biomarker of poor-prognosis beyond staging (Calon et al., 2012) (I-Figure 18b). Calon, Espinet et al. showed that TGF-β activates stromal fibroblasts, which in turn secrete a cocktail of additional pro-metastatic factors such as IL11 (Calon et al., 2012). This study solved a long-standing paradox: why are TGF-β levels increased in CRCs given its cytostatic effects in tumor cells? Our group demonstrated that tumor cells mutate SMAD4 or TGF-β receptors, but the stromal TGF-β response induces a gene program that includes a plethora of cytokines, growth factors, and ECM changes that play key roles during disease progression and metastasis (Calon et al., 2012).




**I-Figure 18 | Disease-relapse in CRC is defined by the TME.** a, Kaplan-Meier curves illustrate the duration of disease-free survival according to CD3 cell density in the tumor center. From (Galon et al., 2006b). b, Kaplan-Meier plots display recurrence-free survival of stage I-III CRC patients according to TGF-β levels. From (Calon et al., 2012). c, Genes defining poor-prognosis in CRC (HR>1) are predominantly expressed in CAFs. From (Calon et al., 2015).

The subsequent generation of large collections of transcriptomic datasets from tumor samples enabled the identification of CRC subtypes on the basis of distinctive global gene expression profiles (Marisa et al., 2013; Sadanandam et al., 2013; De Sousa E Melo et al., 2013). The most widely accepted meta-analysis divided CRCs into four defined consensus molecular subtypes (CMS). These four classes represented MSI-like (CMS1), canonical WNT/MYC (CMS2), metabolically dysregulated (CMS3), and mesenchymal (CMS4) tumors (I-Figure 19) (Guinney et al., 2015). The CMS4 subtype gained most of the attention because CMS4-patients (about 25%) have a poorer prognosis (I-Figure 19). Interestingly, this subtype was characterized by elevated expression of TGF-β signatures and mesenchymal genes, which was hypothesized to be linked with the presence of epithelial-to-mesenchymal transition (EMT) in tumor cells (Guinney et al., 2015). However, tumor transcriptomes contain not only epithelial-expressed genes but also the expression profile of cells present in the TME. Our group and others discovered that the vast majority of genes predicting poor prognosis in patients were actually expressed by CAFs,




**I-Figure 19 | The consensus molecular subtypes (CMS) in CRC. a,** Proposed taxonomy of CRC by Guinney et al., reflecting significant biological differences in the gene expression-based molecular subtypes. **b,** Kaplan-Meier of disease-free survival in the four CMS. Note that CMS4-patients have the poorest prognosis. Adapted from (Guinney et al., 2015).

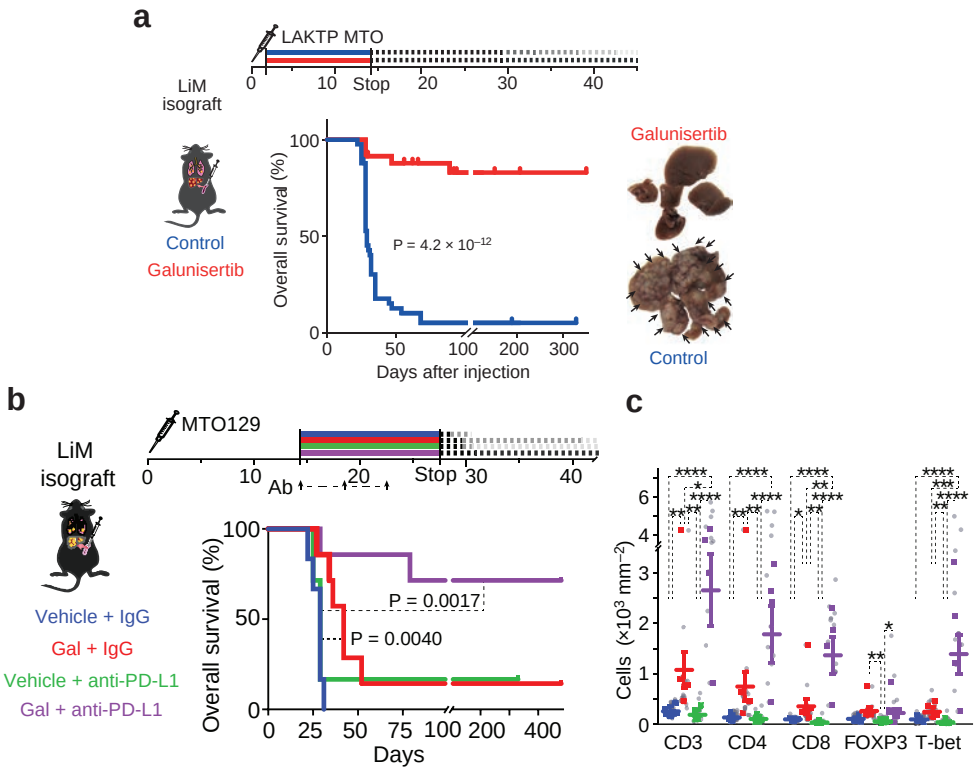
rather than by cancer cells (I-Figure 18c) (Calon et al., 2015; Isella et al., 2015). Indeed, CMS4 tumors show the highest degree of stromal infiltration, while CMS1 tumors are highly infiltrated by lymphocytes. Thus, the poor-prognosis subtype CMS4 is characterized by a TGF-β-activated TME rather than EMT in tumor cells (Calon et al., 2015).



## Immunotherapy in colorectal cancer

In sharp contrast to the remarkable responses to ICB observed in patients with melanomas, renal-cell cancer and lung tumors, the first clinical trial of PD-1 blockade in CRC led to disappointing results: only 1 out of 33 patients responded (Brahmer et al., 2012; Topalian et al., 2012). A follow-up study showed that only CRC patients with deficient mismatch-repair status and high TMB benefited from ICB with pembrolizumab (Le et al., 2015). After these results, it was considered that all MSS CRCs were immunologically "cold", meaning scarcely T cell-infiltrated and non-immunogenic, due to the low number of mutations present in comparison with MSI CRCs. Therefore, MSS CRCs were not considered candidates to benefit from ICB.

In 2018, we revealed that mouse models of MSS CRCs are, however, recognized by the adaptive immune system, yet effector T cell activity is suppressed through high TGF-β levels in the TME. Mice bearing 4 oncogenic mutations (AKTP) — or Mouse Tumor Organoids (MTOs) derived from those mice and transplanted back into immunocompetent mice — generated metastatic intestinal tumors that display key hallmarks of human MSS



**I-Figure 20 | TGF-β drives immunotherapy resistance in MSS mCRCs. a,** Therapeutic effect of TGF-β inhibition with Galunisertib in AKTP tumors. MTOs were injected in the liver and treated two days after with Galunisertib or vehicle (initiation setting). b, TGF-β- and PDL1-blocking is ineffective in overt-metastases unless used in combination. c, Dual therapy induces infiltration of T lymphocytes in CRC liver metastases.

CRC, including a low TMB, T-cell exclusion and TGF-β-activated stroma. Blocking TGF-β signaling with small molecule inhibitors enabled T cell infiltration and was sufficient to confer susceptibility to anti-PD-1–PD-L1 checkpoint- based therapies in advanced CRC mouse models (I-Figure 20).

Consistent with the well-established role of TGF-β signaling in suppressing the differentiation and activity of T cells (Gorelik and Flavell, 2000; Laouar et al., 2002), we observed, both in mice and humans, that a TGF-β-activated TME antagonized a Th1-effector cell phenotype. Thus, we demonstrated that inhibition of TGF-β signaling could render patients with MSS, stroma-rich CRCs and poor prognosis susceptible to benefit from immunotherapy.

# Objectives

The main objective of my PhD was to develop a new CRC relapse model that would allow us to characterize the biology of residual disease. Our ultimate goal is to design new therapies that prevent disease relapse.

Our specific goals were:

1. Develop a mouse model mimicking primary tumor resection with posterior recurrence to clinically relevant metastatic sites, as it occurs in human patients.

2. Identify tumor cells in primary tumors responsible for metastatic dissemination and relapse.

3. Model metastatic latency and explore the phenotypic states adopted by residual CRC cells.

4. Dissect the anatomy and cellular composition of residual disease niches.

5. Characterize tumor-immune crosstalk during metastatic dissemination.

6. Identification of DTCs vulnerabilities and design of therapies to prevent disease relapse.

# Chapter I - High risk of metastatic recurrence in colorectal cancer by residual EMP1+ tumor cells

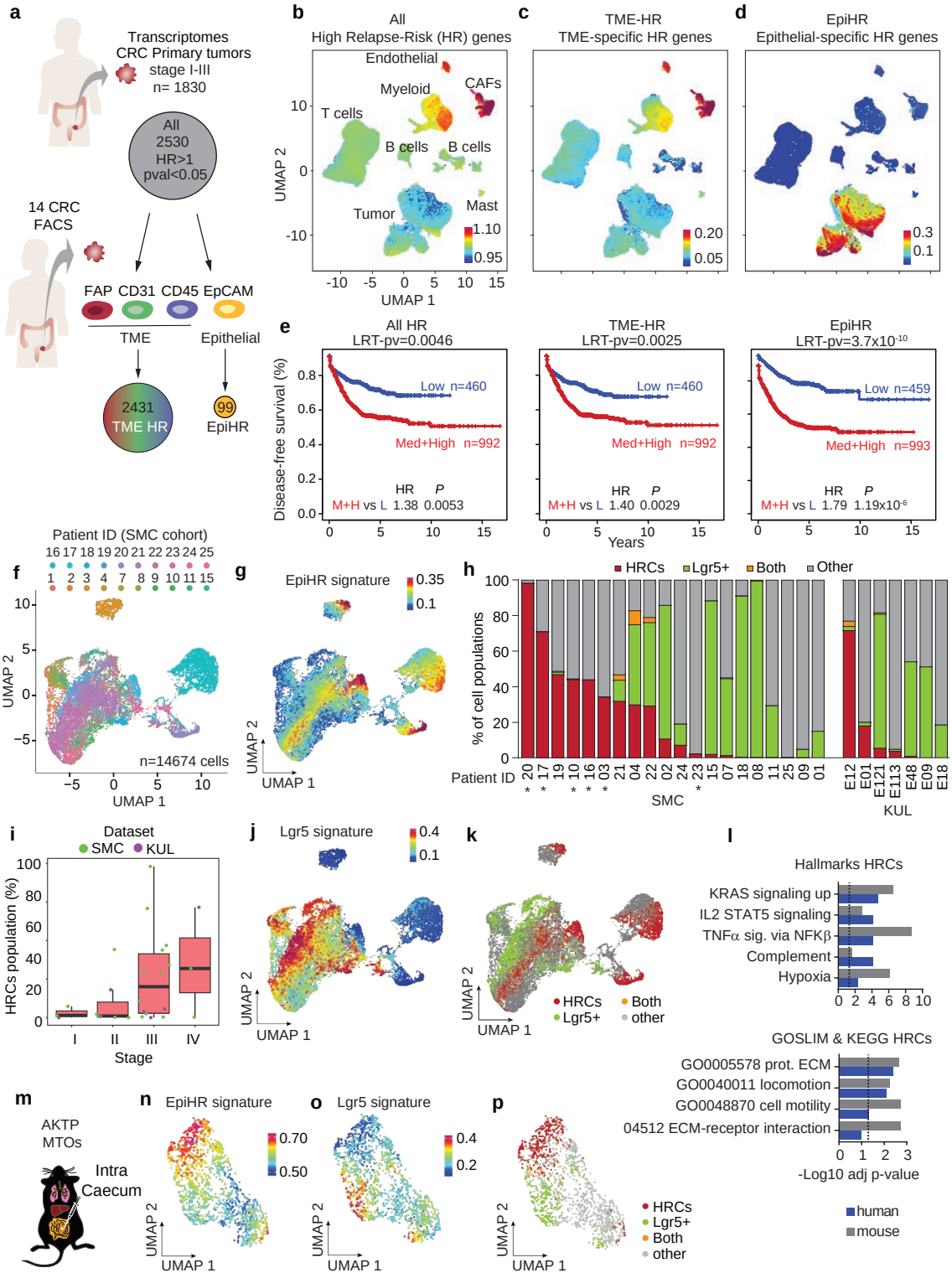
## Results

## Characterization of the epithelial poor prognosis transcriptome in CRC

Surgical resection of the primary CRC effectively cures most patients diagnosed with locoregional disease. However, about 5% AJCC Stage I, 20% Stage II and 40% Stage III patients will develop metastases over the following years. It is possible to identify which CRC patients are at risk of relapse by analyzing the transcriptome of the primary tumor. We and others have previously shown that the vast majority of genes that predict high risk of disease relapse in CRC are expressed by cells of the tumor microenvironment (TME), particularly by cancer-associated fibroblast (CAFs) (Calon et al., 2015; Isella et al., 2015).

To further investigate this finding, we sought to map at single cell level the expression of the poor prognosis CRC transcriptome. Using a large pooled transcriptomic cohort of primary CRC samples (Laibe et al., 2012; Marisa et al., 2013; Muzny et al., 2012; Sadanandam et al., 2013; Solé et al., 2014; De Sousa E Melo et al., 2013; Tripathi et al., 2014) (n=1830 stage I-III CRC, descriptive statistics in C1-Supplementary Table 1), we identified 2530 genes that predicted disease relapse (HR>1, p-val<0.05) (C1-Figure 1a). Subsequently, the expression of this poor prognosis geneset was analyzed in two independent scRNAseq CRC datasets that included both tumor epithelial and microenvironment cells; 20 patients corresponding to the Samsung Medical Center (SMC) cohort and 7 patients from the Katholieke Universiteit Leuven (KUL) cohort (Lee et al., 2020). Supporting our previous findings, CAFs, endothelial cells and to a lower extent myeloid cells expressed highest levels of the poor prognosis genes (SMC cohort in C1-Figure 1b and KUL cohort in C1-Extended Data Figure 1a). However, a detailed analysis of the recurrence geneset in different cell populations purified from primary CRC patient samples (tumor cells/EPCAM+, leukocytes/CD45+, endothelial/CD31+ cells or CAFs/FAP+) (Calon et al., 2015) revealed that 99 out of the 2530 genes were upregulated in epithelial tumor cells compared to TME cells (C1-Figure 1a).

Indeed, these 99 recurrence-associated genes showed epithelial tumor cell-restricted expression patterns both in the SMC (C1-Figure 1d) and KUL (C1-Extended Data Figure 1c) scRNAseq cohorts. Despite representing a minority of the global poor prognosis transcriptome, the epithelial-specific high risk geneset (EpiHR) predicted recurrence with an accuracy equivalent to the subset of poor prognosis genes expressed in the tumor microenvironment (C1-Figure 1e). In multivariate analysis including the two signatures and clinical variables (AJCC stage, age, gender and MSS/MSI status), the TME and epithelial genesets were independent prognostic factors (EpiHR: HR (+1 SD) =2.26, p-val=$1.2x10^{-7}$; TME-HR: HR (+1 SD) =1.74, p-val= $8x10^{-4}$). Thus, as opposed to the stromal expression pattern of most recurrence-associated genes in CRC, the EpiHR geneset




**C1-Figure 1 | Identification of poor prognosis epithelial CRC cells. a,** A metacohort of 7 pooled human stage I-III CRC datasets (n= 1830 patients, Supplementary Table 1) with disease-free survival follow-up data after primary tumor resection was used to find genes that predicted metastatic recurrence. Out of 2530 genes predicting disease-relapse (HR>1), a subset of 99 genes is expressed by Epcam+ cells in the GSE39397 dataset, whereas the rest are expressed mostly by cells in the tumor microenvironment (TME). **b-d**, Uniform Manifold Approximation and Projection (UMAP) layout of whole tumors (stroma + epithelium cells) belonging to 20 CRC patients from the SMC dataset, colored by gene expression of b)

all high hazard ratio genes (All HR, n=2530), **c)** tumor microenvironment-specific HR genes (TME-HR, n=2431) and **d)** epithelial-specific HR genes (EpiHR, n=99). **e,** Kaplan-Meier survival curves indicating relapse-free survival for patients according to All HR, TME-HR and EpiHR gene signatures. Two-sided Wald test. **f,** UMAP layout of 14674 CRC tumor cells colored by patient ID. **g,** UMAP of tumor cells colored according to the expression of the EpiHR signature. **h,** Barplot quantifying the sub-population composition of each patient in the SMC (left) and KUL (right) datasets. Patient ID is detailed. Patients with low WNT signature scores are marked with an "*". **i,** Boxplot representing the proportion of HRCs in each clinical stage. Box plots have whiskers of maximum 1.5 times the interquartile range; Boxes represent first, second (median) and third quartiles. n= 3, 7, 14, 3, patients from left to right. Patients from the SMC dataset and the KUL dataset are distinguished as indicated. **j,** UMAP of tumor cells colored by expression of the Lgr5 signature. **k,** UMAP of tumor cells labelled according to their classification as HRCs, Lgr5+, double positive or other cells. **l,** Relevant Hallmarks, GOSLIM and KEGG gene signatures enriched in HRCs (hypergeometric test) compared to the rest of tumor cells in human and mouse CRC samples. **m,** Primary tumors were generated in the caecum of C57BL/6J mice by injecting syngeneic mouse tumor organoids with mutated alleles for Apc (A), Kras (K), TGF-beta receptor type 2 (T) and p53 (P), (AKTP-MTOs) constitutively expressing GFP. **n- p,** UMAPs depicting GFP+ tumor cells dissociated from primary tumors colored by expression levels of (**n**) EpiHR signature, (**o**) Lgr5 signature and (**p**) their classification as HRCs, Lgr5+, double positive or other cells.

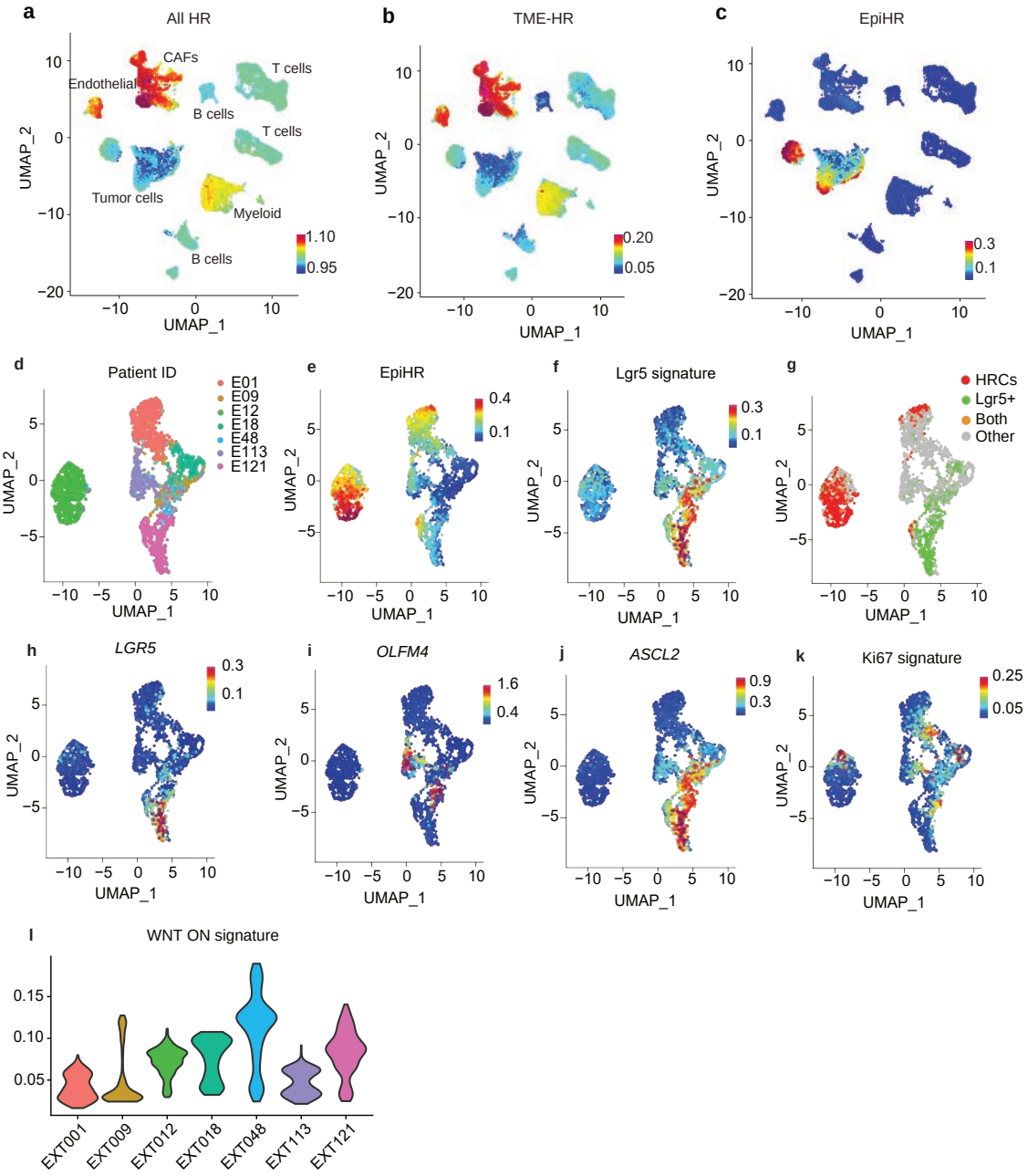
encodes determinants of disease relapse with epithelial tumor cell-specific expression.

## EpiHR marks a discrete tumor cell population in CRC

Representation of epithelial CRC cells using Uniform Manifold Approximation and Projections (UMAPs) showed that 18 out 27 CRCs contained cells labeled with the EpiHR geneset in proportions ranging from 1.4% to 98.1% (C1-Figure 1f-h and C1-Extended Data Figure 1d,e). Reinforcing their association with malignant progression, most samples in the scRNAseq dataset containing abundant EpiHR+ cells (from here onwards named HRCs for High Relapse Cells) were stage III and IV CRCs (C1-Figure 1i – p=0.051 Stage I+II versus III+IV).

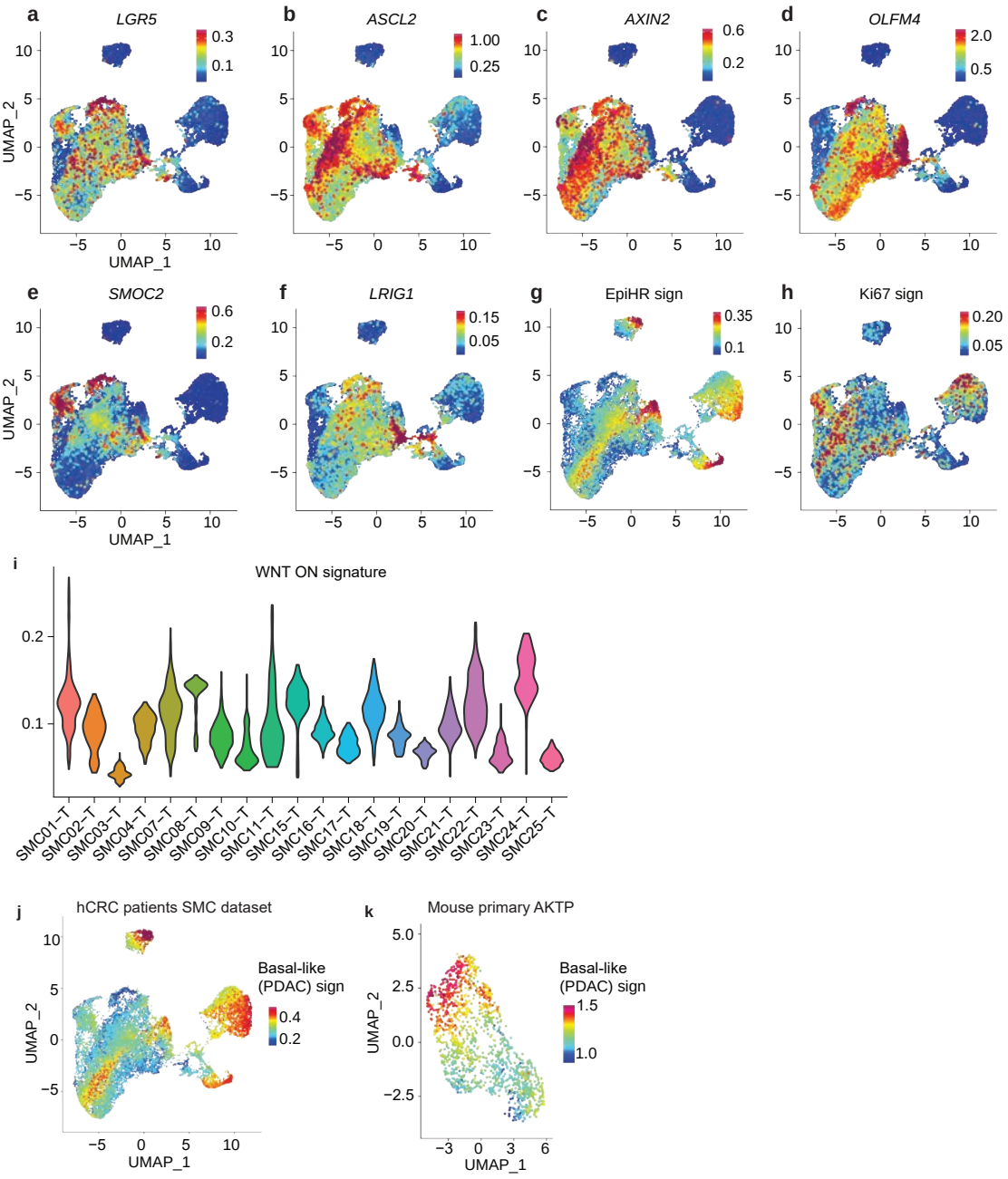
Widespread evidence has demonstrated that CRC growth is driven by a subset of LGR5+ stem cell-like tumor cells (Melo et al., 2017; Shimokawa et al., 2017). Our analyses, however, revealed that HRCs represent a distinct population, shown by the mutually exclusive distribution of the two populations in UMAPs (C1-Figure 1j-k, Extended Data Figure 1e-g). Quantification showed that only one tumor sample exhibited a significant number of HRCs co-expressing the LGR5 signature (7.79% sample SMC04, C1-Figure 1h) whereas five others included a minimal fraction (<3%) of cells marked by both Lgr5 and HRC gene programs. The expression patterns of individual stem cell marker genes such as OLFM4 and ASLC2 and other surrogates of WNT activation confirmed that HRCs were

not LGR5+ stem cell-like cells (C1-Extended Data Figure 1h-k and C1-Extended Data Figure 2a-h). Indeed, a subset of CRCs exhibited marginal WNT target gene expression




**C1-Extended Data Figure 1 | Identification of poor prognosis tumor cells in the KUL dataset. a-c,** UMAP layout of whole tumors (stroma + epithelium cells) from 7 CRC patients in the KUL dataset. Colored by (**a**) gene expression of all high hazard ratio genes (All HR), (**b**) tumor microenvironment-specific HR genes (TME-HR), and (**c**) epithelial-specific HR genes (EpiHR). **d,** UMAP layout of 2718 CRC tumor cells from the KUL cohort colored by patient ID. **e-f,** UMAP of same tumor cells showing the expression of (**e**) EpiHR and (**f**) Lgr5 signatures. **g,** UMAP of same tumor cells labelled according to their classification into HRCs, Lgr5+, double positive or other cells. **h-k,** UMAP of tumor cells showing gene expression levels of canonical intestinal stem cell genes LGR5, OLFM4, ASCL2 and the mKi67 proliferation signature. **i,** Violin plot showing WNT-ON signature expression levels

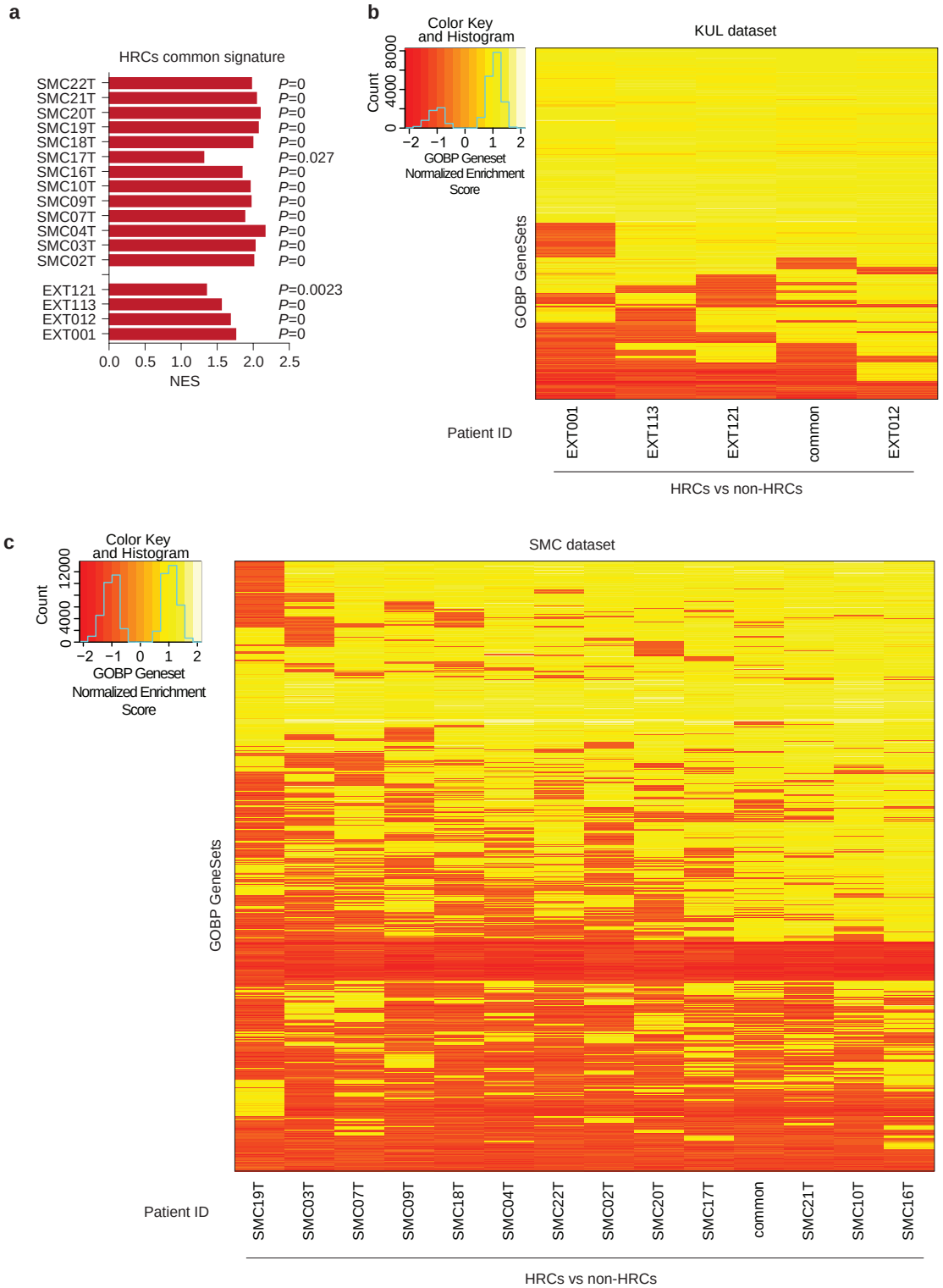
in epithelial tumor cells from patients in the KUL cohort. The WNT-ON signature are the top100 upregulated genes in control vs tamoxifen-treated LS174T and SW403 CRC cells expressing a tamoxifen-inducible dominant negative TCF4 transcription factor to switch-off WNT signaling (Morral et al., 2020).




**C1-Extended Data Figure 2 | Additional description of the single cell SMC dataset. a-h,** UMAPs of tumor cells in the SMC dataset showing gene expression levels of canonical intestinal stem cell genes LGR5, ASCL2, AXIN2, OLFM4, SMOC2, LRIG1 and the EpiHR and mKi67 signatures as indicated. i, Violin plot showing WNT-ON signature expression levels in epithelial tumor cells from patients in the SMC cohort. **j-k,** UMAPs of human CRC tumor cells in the SMC dataset (**j**) and mouse CRC AKTP tumor cells (**k**) painted with the Basal cell state signature in Pancreatic cancer by (Raghavan et al., 2021).

levels (C1-Extended Data Figure 2i), showed no LGR5+ cells, yet contained HRCs (identified with an * in Figure 1h).

Tumor cells labeled by the EpiHR signature in different tumor samples expressed a common gene program (C1-Extended Data Figure 3a) and displayed a shared enrichment pattern of annotated genesets (C1-Extended Data Figure 3b-c) implying that this population exhibits a similar phenotype and plays similar functions across different CRCs.



**C1-Extended Data Figure 3 | Analysis of conserved biological functions in HRCs across patients. a,** Enrichment coefficients (GSEA) of the HRCs common signature (defined as the top 100 HRC-marker genes ordered by fold change) across fold changes of differential expression comparing HRC with all other cells in each patient. **b-c,** Heatmap showing the normalized enrichment scores (NES) for Genesets in Gene Ontology Biological Processes (GOBP) in HRCs from different patients in the KUL (**b**) and SMC (**c**) cohorts. Only GOBP genesets with NES scores above [0.5] are shown. Genesets and patients are ordered by hierarchical clustering.

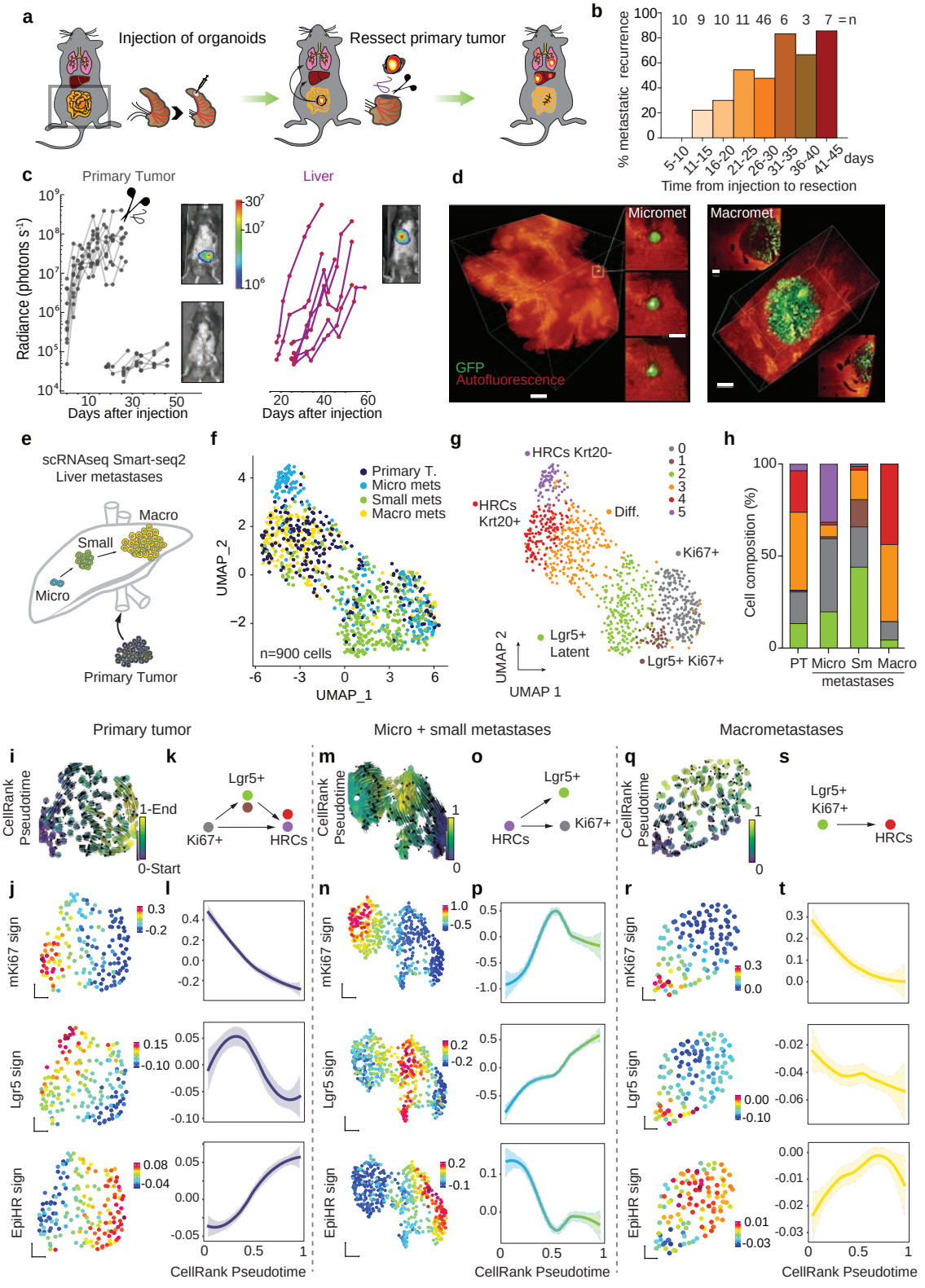
It was evident a significant upregulation of genes related to NF-Kβ signaling, hypoxia, cell locomotion and ECM remodeling in HRCs compared to the rest of the tumor cells (C1-Figure 1l).

In a previous study, we described that mice bearing mutations in Apc, Kras, Tgfbr2 and p53 (AKTP) in Lgr5+ ISCs develop metastatic CRCs. These quadruple mutant CRCs reproduce key features of advanced human disease including recruitment of a TGF-β-activated tumor microenvironment, T cell exclusion and metastatic dissemination (Tauriello et al., 2018). We analyzed by scRNAseq CRCs generated by implantation of AKTP mouse tumor organoids (MTOs) in the caecum of c57BL/6 mice (C1-Figure 1m-p). Mirroring the observations in human tumor samples, we found that mouse CRCs contained abundant HRCs and that this population did not express the Lgr5+ ISC-like expression program (Figure 1n-p). HRCs were enriched in similar gene categories in both species implying functional equivalence (C1-Figure 1l). We also found that the signature of basal-like pancreatic cancer cells (Raghavan et al., 2021) marked both human and mouse HRCs suggesting that they adopt a state akin to the most aggressive subtype of pancreatic cancer (C1-Extended Data Figure 2j-k).

## Dynamic evolution of metastatic cells during relapse

Residual disseminated tumor cells hidden in the liver or lungs of CRC patients cannot be directedly isolated. Moreover, animal models that recapitulate CRC relapse are not well established. Prompted by these drawbacks, we developed a new mouse model of metastatic relapse that allowed us to investigate the contribution of HRCs to metastatic recurrence. In brief, we innovated classical needle-based orthotopic injections by relocating them to the tip of the caecum, which allowed complete surgical excision of singular invasive CRCs (Figure 2a, Extended Data Figure 4c and Supplementary Video). Dual GFP/Luciferase-labelled AKTP MTOs grew rapidly in the caecum of c57BL/6 mice, colonized the adjacent mucosa and generated T3 and T4 invasive cancers (C1-Extended Data Figure 4a,b) (Tauriello et al., 2018). Bioluminescence monitoring revealed that mice remained free of primary disease after surgical resection yet, over the following days, they relapsed in the form of liver metastasis (C1-Figure 2a-c). Occasionally, we also observed metas-

tases in mesenteric lymph nodes, lungs, peritoneum and diaphragm (C1-Extended Data Figure 4f-h). Primary tumor resection shortly after implantation cured all mice, whereas surgery at later points resulted in increased proportions of mice developing metastatic recurrences (C1-Figure 2b). In experiments of early primary tumor resection (day 11-15),

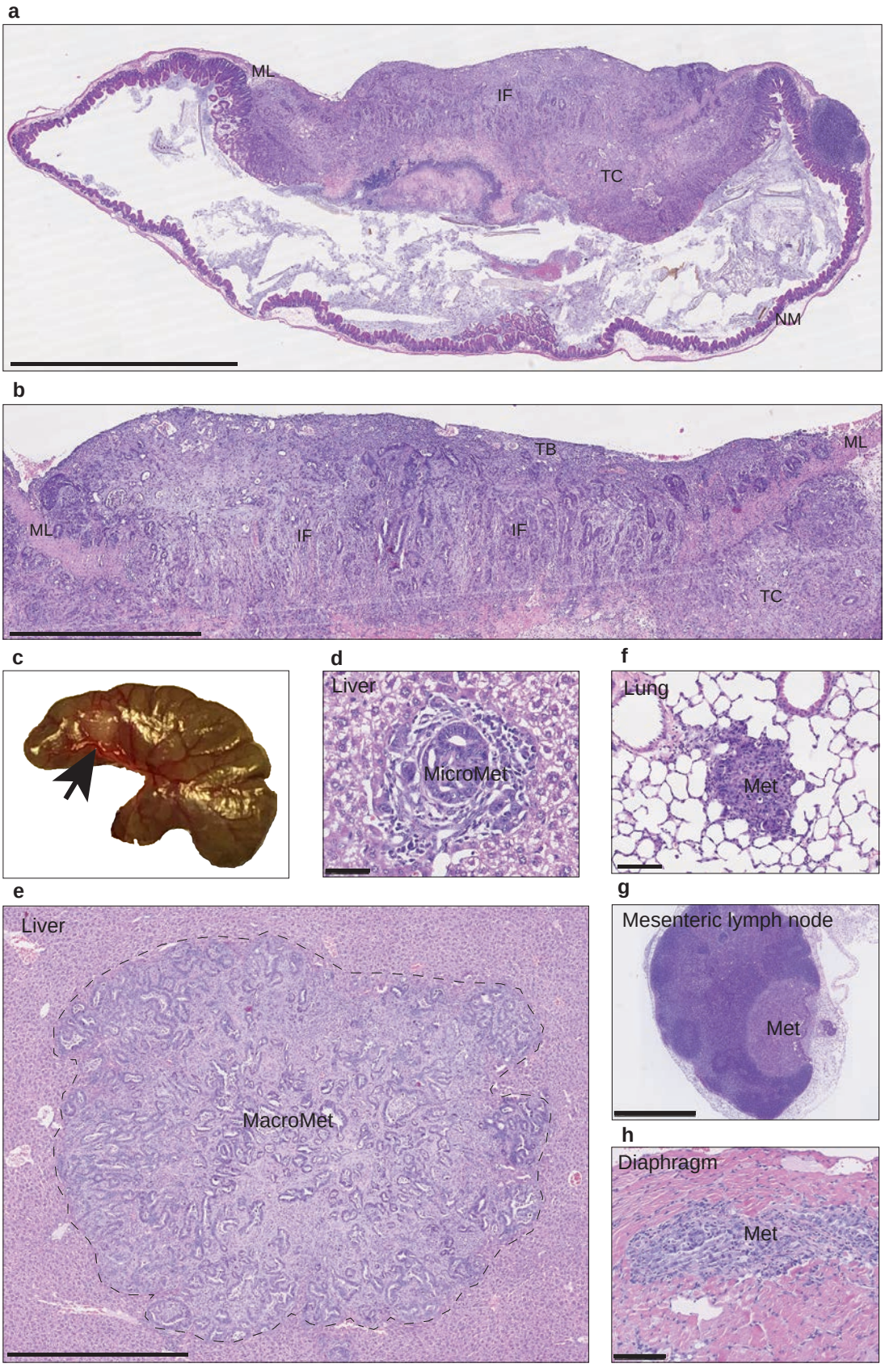



**C1-Figure 2 | Spatiotemporal dynamics of CRC metastases resolved by single cell RNA sequencing. a,** Schematic of the human-like mouse model of CRC metastatic

relapse developed herein. The model is based on the orthotopic implantation and surgical removal of primary tumors, which is then followed by metastatic recurrence to clinically relevant sites. Approximately 50,000 cells are implanted in the distal caecum, where they engraft, proliferate and disseminate mainly into the liver. Primary tumors are fully extirpated by surgery but some mice relapse in the form of metastases. **b,** Percentage of metastatic recurrence depended on time elapsed from injection to primary tumor resection. Number of mice per each time point is detailed above the barplot. **c,** Longitudinal intravital bioluminescence (BLI) imaging quantification (photons s−1) of a representative experiment where AKTP MTOs were implanted in the caecum of c57BL/6J mice. 3-4 weeks post-implantation, primary tumors are excised resulting in reduction of bioluminescence to background levels. Grey points and lines represent bioluminescence in the lower thorax of individual mice. After surgery, bioluminescence reappears in the upper abdomen of mice, indicating the presence of liver metastases (right graph, in purple). Representative images of bioluminescence in the same mouse before, after surgery and upon liver metastases formation are shown. **d,** Whole livers containing GFP-expressing tumor cells were optically cleared for 3D imaging. Using a custom lightsheet microscope, we visualized spontaneous micrometastasis (left) and fully-grown metastases (right) in the mouse CRC relapse model. GFP-fluorescence labels tumor cells, auto fluorescence is used to visualize the liver. Scale bars, left image (300 µm on Maximum Intensity Projections, MIP, and selected single plane insets 50 µm) right (100 µm on MIP and single plane insets 50 µm). **e,** Illustration of the longitudinal single cell RNA-expression analysis of tumor cells along the metastatic cascade. Single cells were collected at four stages and each Smart-seq2 plate contained cells from all conditions to avoid batch effects. Primary tumor samples were matched with micro, small or macro metastases samples. For additional information, see Extended Data Figure 6. **f-g,** UMAP layout of 900 tumor cells isolated from 7 different mice colored by (**f)** metastatic stage and (**g**) Seurat clusters. **h,** Barplot showing Seurat cluster composition by sample stage. **i, m, q,** Vector fields representing RNA velocity (Bergen et al., 2020; La Manno et al., 2018) projected on UMAPs of primary tumors (i), micro + small metastases (m) and macrometastases (q). Colored by the pseudotime estimated for each cell with scVelo (Bergen et al., 2020; Lange et al., 2020). **j, n, r,** UMAPs with cells separated in primary tumors (**j**), micro+small metastases (**n**) and macrometastases (**r**) and colored by gene expression of mKi67, Lgr5 and EpiHR gene signatures. **k, o, s,** Schematics showing distinct hierarchical behavior during the different stages of metastases formation. **l, p, t,** Smoothed mKi67, Lgr5 and EpiHR gene signature expression trends fitted with Generalized Additive Models as a function of pseudotime in primary tumors (**l**), micro+small (**p**) and big metastases (**t**).

faint bioluminescent signal could be detected ex vivo in some livers immediately after surgery implying the presence of residual disseminated tumor cells. Lightsheet 3D fluorescence imaging of cleared livers revealed 3 to 10-cell micrometastases at the time of resection (C1-Figure 2d).

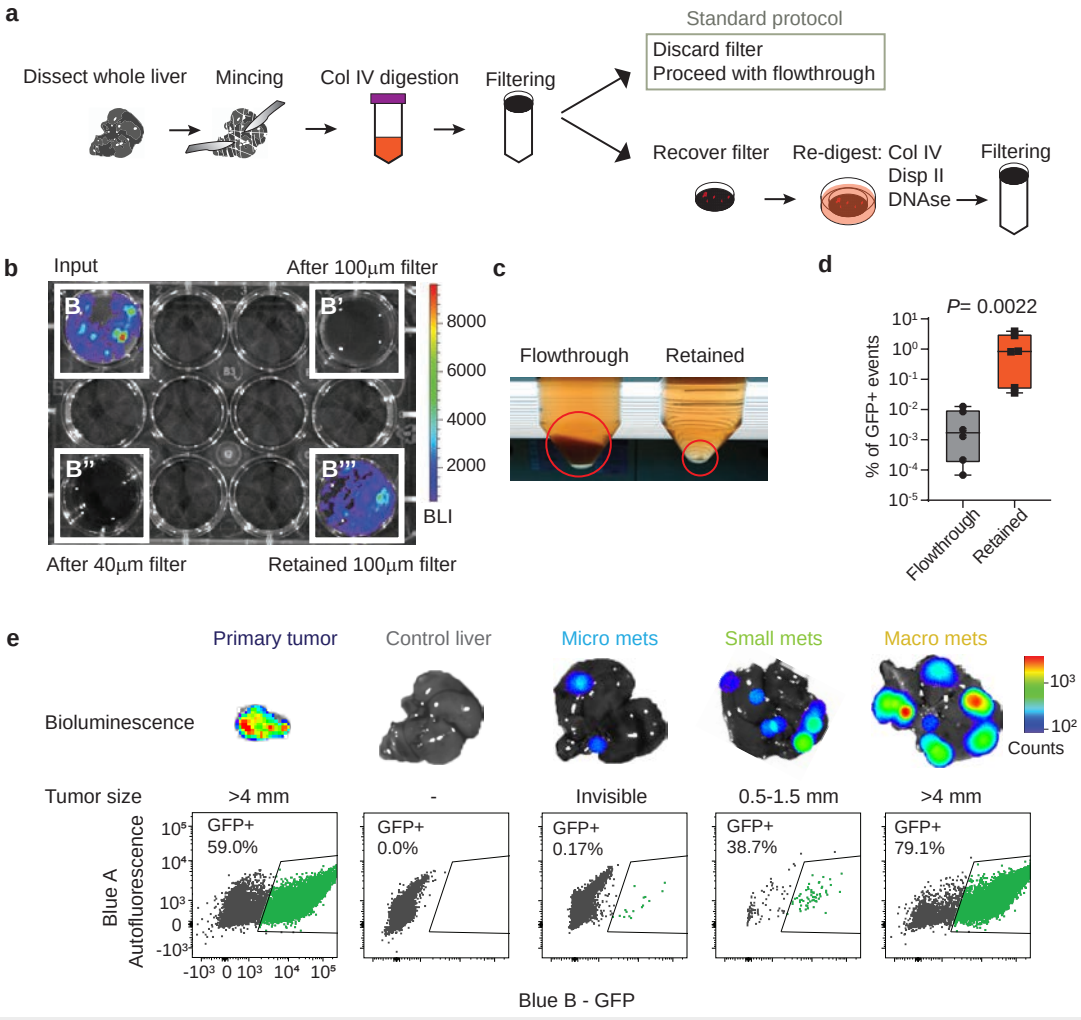
We next sought to profile tumor cells along the process of relapse. Isolation of residual tumor cells from large organs has historically been a major hurdle in cancer research (Massagué and Obenauf, 2016). The paucity of disseminated tumor cells (DTCs) compared to healthy tissue has limited attempts to isolate and directly profile residual cancer




**C1-Extended Data Figure 4 | Histologic analysis of the CRC relapse mouse model. a,** Representative micrograph of an hematoxylin- and eosin-stained adenocarcinoma with subserosal invasion (T4) generated by injection of an AKTP MTO in the mouse caecum. Tumor center (TC), invasive fronts (IF), muscle layer (ML) and normal mucosa (NM) are

indicated. Scale bar, 2.5 mm. **b,** Representative image of a different T4 tumor penetrating the muscle layer (ML) and reaching the serosa layer. Scale bar, 1mm. **c,** Picture of a caecum 21 days after injection and imaged at the time of surgery showing a primary tumor (arrow) in the distal part. **d-e,** Hematoxylin and eosin (HE) staining of micrometastases and large metastases observed in the liver of orthotopic isografted mouse. Scale bars, 50 µm and 1 mm, respectively. In e, tumoral tissue is surrounded by dashed lines. **f-h,** HE stainings of lung, lymph node and diaphragm metastases from orthotopic isografted mice. Scale bars, 100 µm (f and h) 1 mm (g).

cells from large organs. The state of the art approach involves an in vitro selection step to isolate DTCs (Laughney et al., 2020) that may however distort their transcriptional state. To overcome these limitations, we devised a tissue dissociation strategy that enriches for residual tumor cells from whole liver samples (C1-Extended Data Figure 5a). Shortly, we discovered that during tissue preparation for FACS, the vast majority of luciferase+ tumor cells were retained in 100mm filters after mild enzymatic digestions, whereas most parenchymal liver cells flowed through in these conditions (C1-Extended Data Figure 5b,




**C1-Extended Data Figure 5 | Strategy to purify residual disseminated tumor cells from whole livers. a,** Schematic of a novel tissue-dissociation strategy that en-

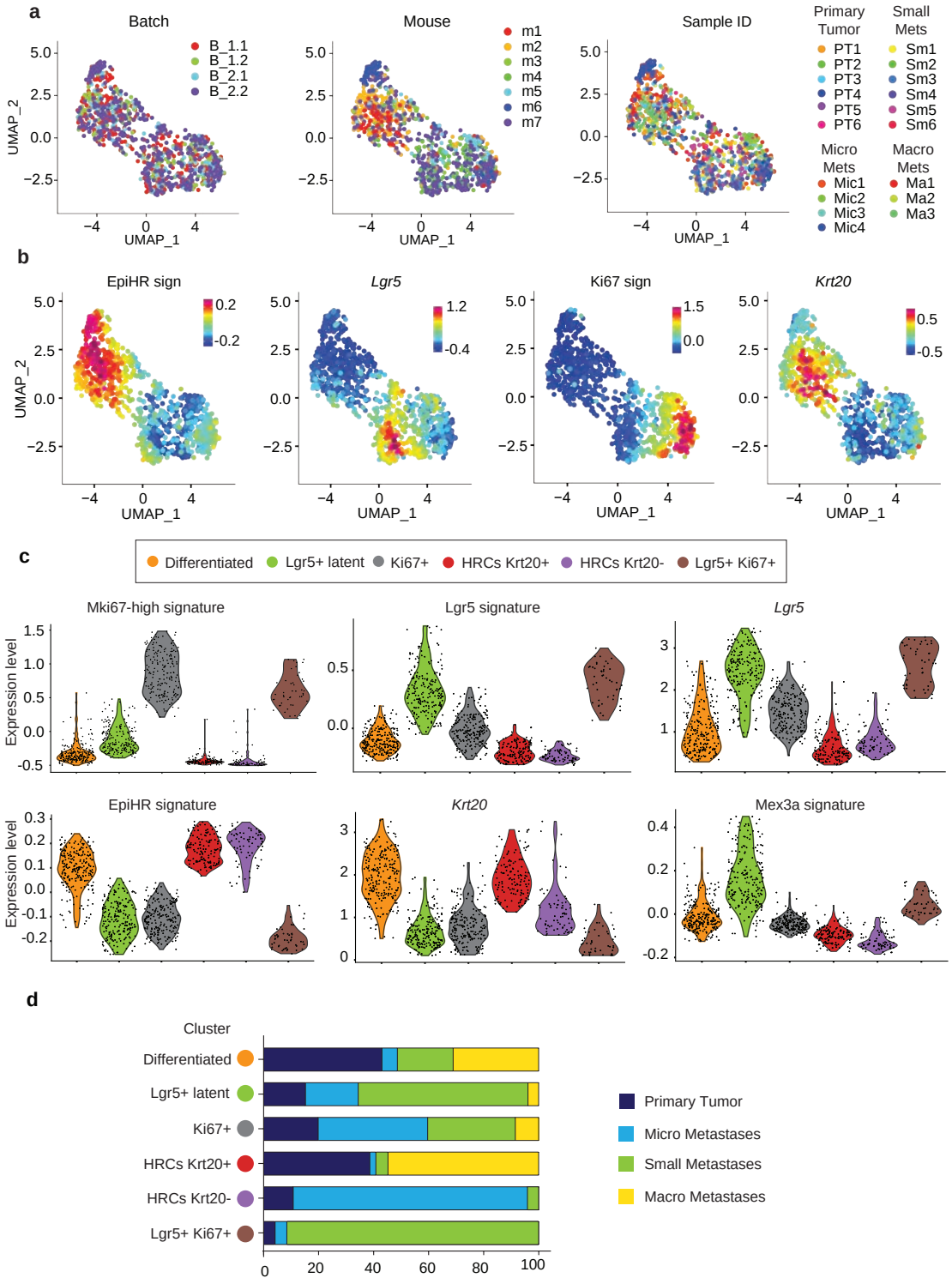
ables recovery of residual disseminated tumor cells (DTCs) from livers. Whole livers are dissected and minced thoroughly. After a mild collagenase IV digestion, samples are filtered through 100 µm meshes and the flow through is discarded. Most liver parenchymal cells flow-through whereas the filter retained sample is highly enriched in tumor cells. Remaining tissue in the filter is re-digested with a stronger enzymatic cocktail to fully digest it, and then re-filtered. **b,** Representative bioluminescent image of a whole liver sample containing luciferase+ tumor cells before enzymatic digestion (B, input), after filtering through 100 µm (B') and 40 µm (B'') meshes (old protocol), and after recovering and re-digesting tissue retained in the 100 µm filter (B'''). **c,** Image showing the large cell pellet containing liver cells after 1 mild digestion and the small pellet in the retained and re-digested sample enriched in DTCs. **d,** Percentage of GFP+ cells measured by flow cytometry in samples with 1 round of digestion compared to re-digested samples. n=6 independent paired samples. Paired two-sided Wilcoxon test on percentages. **e,** Representative bioluminescent images, tumor burden and flow-cytometry plots of the 4 different stages analyzed by single-cell Smart-sequencing described in Figure 2. Micrometastases samples were DTCs collected from livers with absent or low bioluminescence in which metastases were not visible. For small metastases samples, metastatic nodules were visible but small in size (<1.5mm). Macrometastases samples were metastatic nodules larger than 4mm.

c). By re-digesting cells retained in the filter, we obtained a 400-fold GFP-Luciferase+ enriched metastatic tumor cell preparation (C1-Extended Data Figure 5c). This step allowed purification of residual tumor cells from individual livers exhibiting absent or very low ex vivo bioluminescence (C1-Extended Data Figure 5e).

By means of this approach, we profiled by Smart-seq2 single-cell RNA sequencing 900 GFP+ tumor cells with an average coverage of 6,449 genes per cell, derived from livers collected at different time points after implantation of MTOs as well as from their corresponding primary CRCs. We confirmed the presence of micrometastases, small metastases or macrometastases by bioluminescence measurements in the resected livers (C1-Extended Data Figure 5e).

UMAP representation showed that tumor cells from primary tumors and metastases overlapped to a large extent (C1-Figure 2e,f and Extended Data Figure 6). Hierarchical clustering analysis identified 6 cell clusters (C1-Figure 2g and Extended Data Figure 6b, c). Cluster 0 included cells that expressed elevated levels of proliferation and biosynthesis-encoding genes (C1-Figure 2g and Extended Data Figure 6b, c). Lgr5+ ISC-like tumor cells were located in clusters 1 and 2, with those in cluster 1 proliferating and those in cluster 2 expressing a signature of latent Mex3a+ stem cells (Barriga et al., 2017) (C1-Figure 2g and Extended Data Figure 6c). Tumor cells of cluster 3 upregulated the differentiation marker Krt20. Clusters 4 and 5 were largely enriched in HRCs. Some cells in cluster 4 expressed Krt20 suggesting that HRCs can also undergo differentiation (C1-Figure 2g and Extended Data Figure 6b, c). Quantification of cell types revealed a dynamic distri-

bution of cell populations across the metastatic relapse process (C1-Figure 2h). Primary CRCs and macrometastases exhibited similar distribution of cell populations, including




**C1-Extended Data Figure 6 | Additional description of profiled residual CRC cells. a,** UMAPs of colorectal primary tumors and liver metastases at different stages (micro, small and large) colored according to sequencing batch, mouse ID, and sample ID. **b,** UMAPs showing the expression levels of the EpiHR, and mKi67 gene signatures and Lgr5 and Krt20 genes. **c,** Violin plots showing expression of relevant genes used to define the 6 different Seurat clusters. **d,** Fraction of cells (x axis) from each Seurat cluster (y axis) present

in the different sample types: Primary Tumor, micro-, small- and macro- metastases according to the indicated color code. Note the "HRCs Krt20-" are mostly exclusive from micro metastases samples, whereas Lgr5+ cells are highly enriched in small metastases samples.

proliferative cells, Lgr5+ ISC-like cells, Krt20+ differentiated tumor cells and HRCs most of which were also Krt20+ (C1-Figure 2h). In contrast, micrometastases were largely enriched in undifferentiated (Krt20-) HRCs and also contained abundant proliferative cells (Figure 2h). Small metastases were mainly formed by Lgr5+ ISC-like cells in quiescent and proliferative states and contained fewer HRCs than micrometastases (Figure 2h).

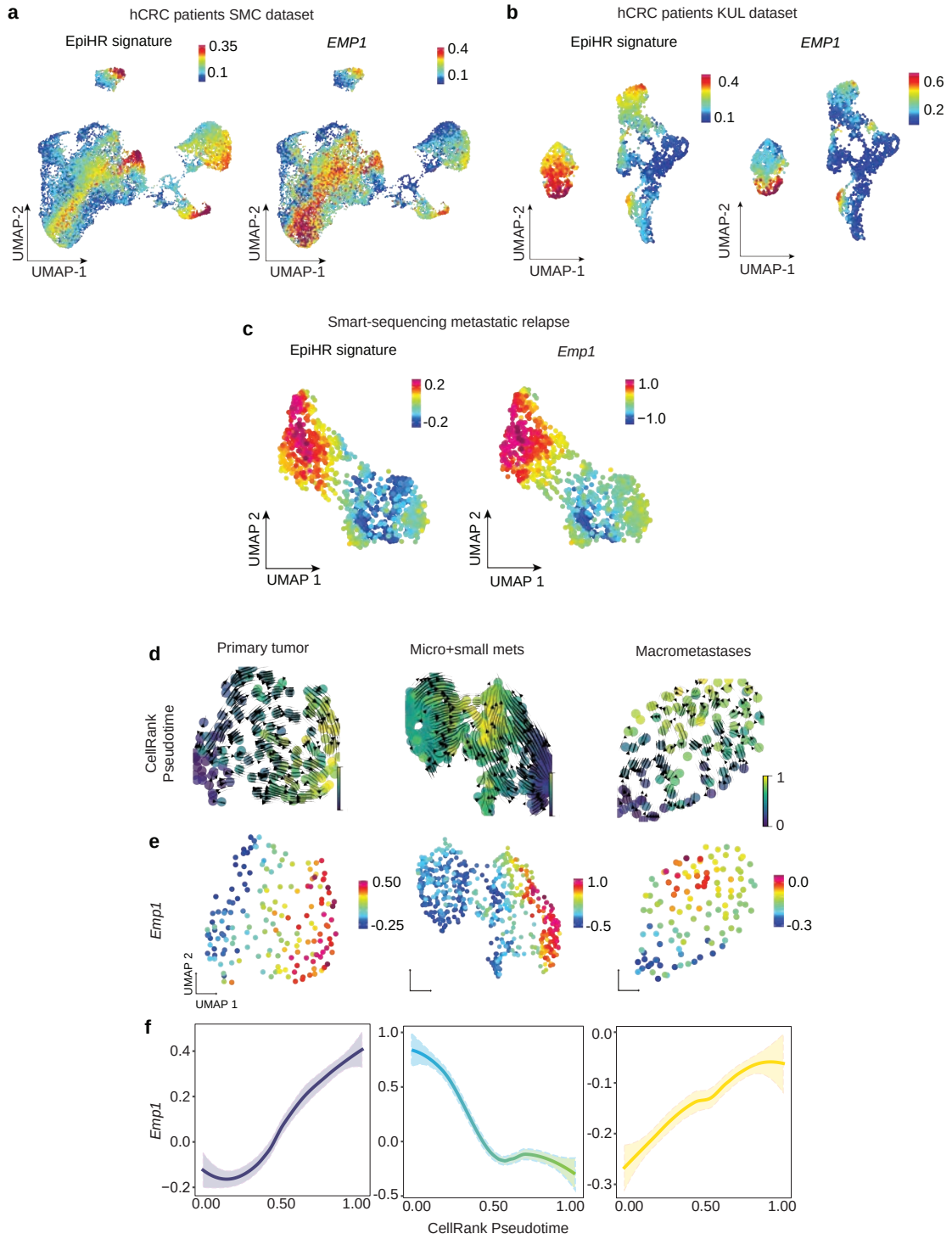
Using CellRank (Lange et al., 2020) - a novel computational method that combines trajectory inference with directional information based on RNA splicing (La Manno et al., 2018) - we calculated cell fate potentials over the timeline of disease relapse (Figure 2i-t). CellRank predicted different hierarchical organizations during metastatic progression. In primary CRCs, proliferative Lgr5-neg tumor cells gave rise to Lgr5+ cells and HRCs (C1-Figure 2i-l). In macrometastases, the apex of the hierarchy was occupied by proliferative Lgr5+ ISC-like cells which generated HRCs overtime (C1-Figure 2q-t). In contrast, the algorithm prognosticated that the cells that initiate the cellular hierarchy of micrometastases correspond to undifferentiated HRCs in cluster 6 (C1-Figure 2m-p). This cell population gave rise to Lgr5+ ISC-like and proliferative tumor cell progeny (C1-Figure 2m-p), which are abundant in small metastases (C1-Figure 2h and Extended Data Figure 6d).

## HRCs are enriched in invasion fronts and micrometastases

We next sought to identify individual markers that enable genetic manipulation of the HRC population in our murine CRC relapse model. As detailed above, comparison of human and mouse scRNAseq datasets revealed a subset of genes expressed consistently by HRCs of the two species (C1-Figure 3a). Among them, we focused on epithelial membrane protein 1 (EMP1) because it was expressed at high levels in HRCs and exhibited a large degree of overlap with the expression of the EpiHR geneset in both human (Extended Data Figure 7a-b) and AKTP mouse CRCs (C1-Figure 3b and Extended Data Figure 7c). Echoing our results with the EpiHR signature, CellRank predicted that the cell origin of metastatic relapse expressed elevated *Emp1* levels (C1-Extended Data Figure 7d-f).

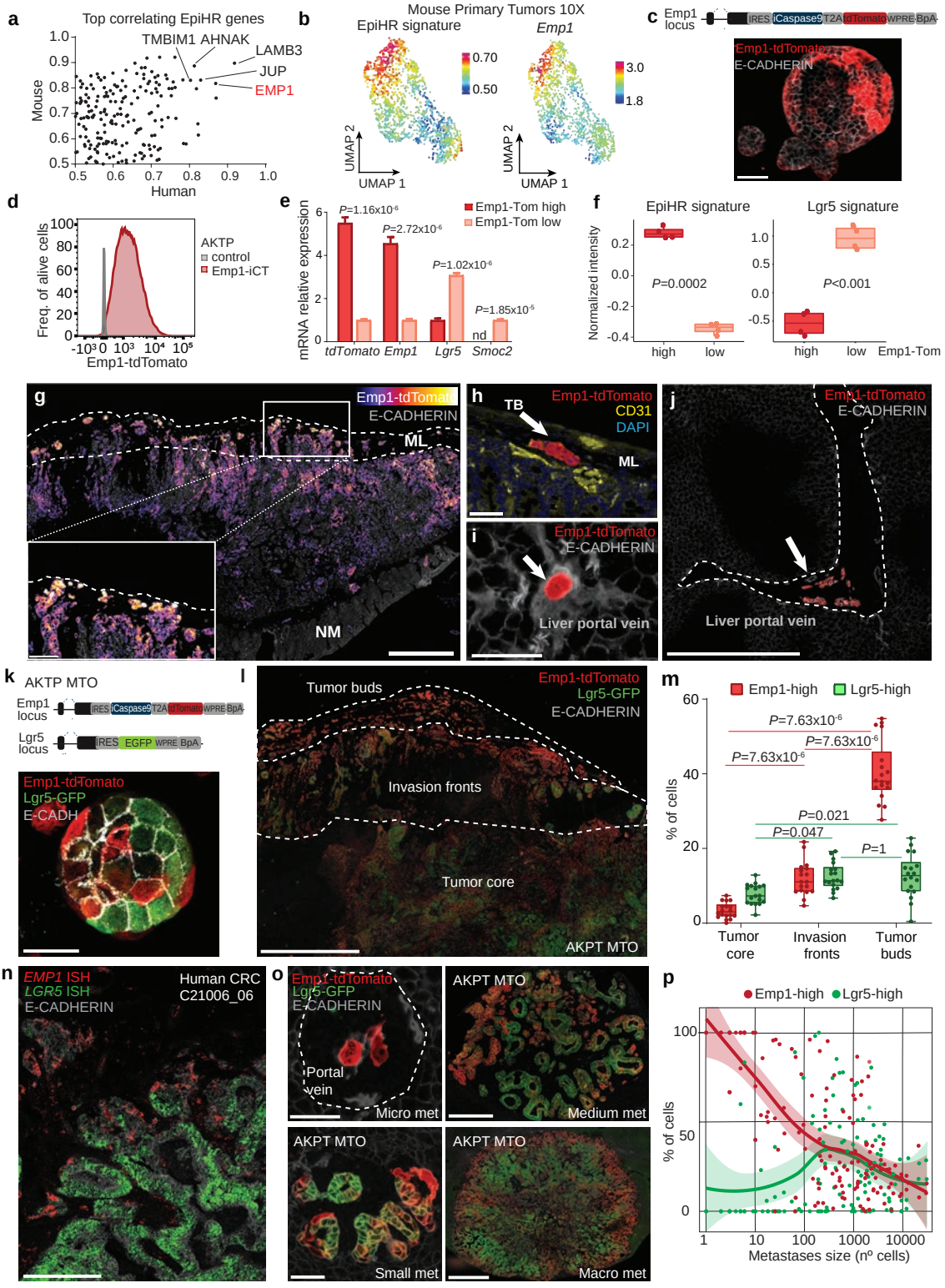
We thus leveraged Emp1 to track HRCs during disease relapse. To this end, we knocked-in an inducible-Caspase9-tdTomato (iCT) cassette (Shimokawa et al., 2017) into the Emp1 locus of AKTP MTOs using CRISPR-Cas9 (C1-Figure 3c). Inspection of knock-in MTOs revealed high tomato expression in a subset of tumor cells (C1-Figure 3c). We inoculated Emp1-iCT AKTP MTOs into the caecum of c57BL/6 mice. Tomato expression

in dissociated epithelial tumor cells measured by flow cytometry revealed heterogeneity in *Emp1* expression (C1-Figure 3d). Tomato-high cells purified by FACS showed large




**C1-Extended Data Figure 7 | Epithelial morphogenic protein 1 (EMP1) is an HRC marker gene. a,** UMAP of tumor cells from CRC patients in the SMC dataset colored according to the expression of EpiHR signature (left) and of EMP1 gene (right). **b,** As in **a,** for CRC tumor cells from the KUL datasets. **c,** UMAP representation of Smart-sequencing single cell data of mouse tumor cells along metastatic relapse sequence colored by the EpiHR signature (left) and Emp1 gene (right). **d,** Vector fields representing RNA velocity(Bergen et al., 2020; La Manno et al., 2018) projected on primary tumor, micro+small and macro me-

tastases UMAPs, colored by the pseudotime estimated for each cell with scVelo (Bergen et al., 2020; Lange et al., 2020). **e,** UMAPs colored by Emp1 gene expression. **f,** Smoothed Emp1 gene expression trends fitted with Generalized Additive Models as a function of pseudotime in primary tumor, micro+small and big metastases samples.

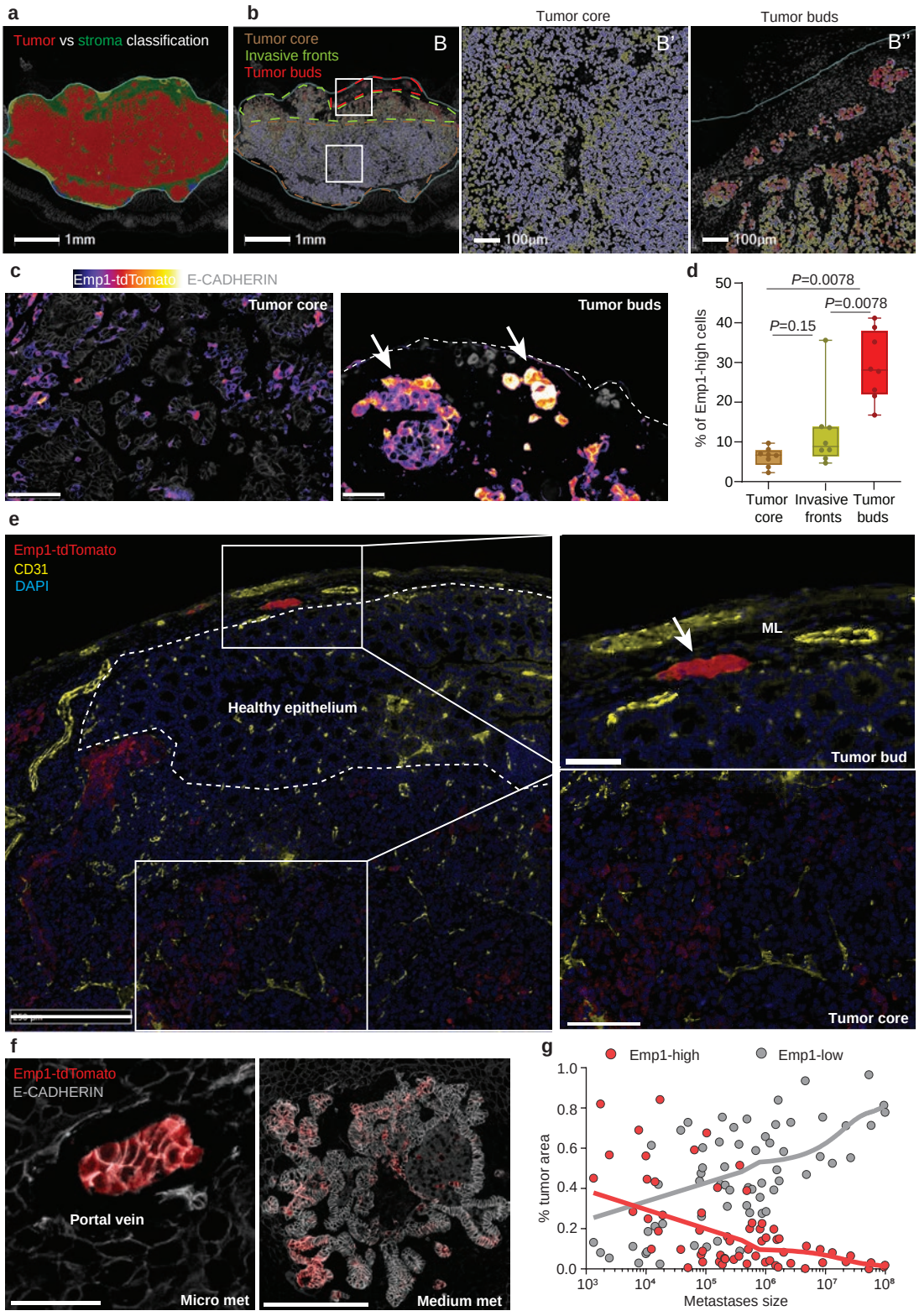



**C1-Figure 3 | Emp1 marks cells enriched in invasion fronts and micrometastases. a,** Scatter plot showing the correlation value between individual genes in the human

SMC cohort (x axis) and in mouse primary tumors (y axis) with the EpiHR signature. Genes with correlation scores higher than 0.8 in both datasets are highlighted. **b,** UMAP depicting GFP+ tumor cells from AKTP primary tumors colored by expression levels of EpiHR gene signature and Emp1 gene. **c,** Schematics of the IRES/iCaspase9/T2A/tdTomato/WPRE/BGHpA cassette introduced in the Emp1 locus of AKTP MTOs by CRISPR-Cas9. Confocal imaging of tdTomato and E-Cadherin immunostaining in Emp1-iCasp9-tdTomato (Emp1-iCT) MTOs. Scale bar, 50 µm. **d,** Representative flow cytometry plot of tdTomato expression in wild-type and Emp1-iCT MTOs. **e,** Relative mRNA expression in Emp1-high and Emp1-low sorted cell populations from Emp1-iCT MTOs in vitro. Two-sided t-test after normalizing by PPIA. n=3 technical replicates. Mean +/- SD. **f,** Boxplot showing normalized intensity of EpiHR and Lgr5 signature expression in Emp1-high and Emp1-low cells dissociated from primary tumors 4 weeks post-implantation. Box plots have whiskers of maximum 1.5 times the interquartile range; boxes represent first, second (median) and third quartiles. n=4 mice per condition. ROAST-GSA adjusted p-values are shown. **g,** Representative immunostaining for tdTomato and E-Cadherin in Emp1-iCT tumors implanted in the caecum 4 weeks post-implantation. Emp1-tdTomato fluorescence is shown with an mpl-inferno LUT. Dashed lines encompass tumor buds and invasion fronts in between the tumor core and the edge of the caecum. The inset shows an amplified view of tumor buds. NM: Normal Mucosa; ML: Muscle Layer. Scale bar: 500 µm; 100 µm (insert). **h,** Immunofluorescence staining of tdTomato, CD31 and DAPI in a tumor bud away from the primary tumor intermingled in muscle layer (ML) next to CD31+ blood vessels. Scale bar, 50 µm. **i-j,** tdTomato immunostainings of mice isografted with primary tumors where (**i**) a single tumor cell or (**j**) clusters both labelled in red extravasate into the liver from portal veins. Portal and central regions are distinguished by the zonated pattern of E-Cadherin staining (in grey). Scale bars, 50 µm (i) and 500 µm (j). **k,** Emp1-iCasp9-tdTomato and Lgr5-GFP alleles introduced in AKTP MTOs. Confocal imaging of tdTomato, GFP and E-Cadherin immunostaining in edited MTOs. Single z-plane. Scale bar, 10 µm. **l,** Immunostaining of Emp1-tdTomato, Lgr5-GFP and E-Cadherin in Emp1-iCT Lgr5-GFP MTO derived primary tumors 4 weeks post-implantation. Lgr5-high and Emp1-high cells show a mutually exclusive pattern (quantified in Extended Data Figure 11e). Dashed lines encompass invasion fronts and tumor buds. **m,** Quantification of the percentage of Lgr5-high and Emp1-high cells in the tumor core, invasive fronts and tumors buds. n= 855,330 cells from 18 different mice. Two-sided Paired Wilcoxon test. **n,** RNA ISH of EMP1 and LGR5 combined with E-Cadherin immunostaining in a primary tumor from a CRC human patient (ID= C21006_06). Scale bar, 250 µm. **o,** Immunofluorescence staining of tdTomato, GFP and E-Cadherin in liver metastases of increasing size (as indicated) generated from the CRC-relapse model. Scale bars, 50 µm (micro and small), 250 µm (medium), 500 µm (macro). Images shown in l and o belong to the same experiment, and were stained and scanned together. **p,** Percentage of Emp1-high and Lgr5-high cells per metastases, ordered by their size, measured in cell numbers. n= 318276 cells from 137 liver metastases from 17 different mice. Trend lines (bold) show a LOESS model with a 95% confidence interval (colored shades). NM= normal mucosa, ML= muscular layer, TB= tumor bud, PV= portal vein.

upregulation of *Emp1* expression whereas tdTomato-low cells were characterized by expression of intestinal stem cell (ISC)-specific genes such as *Lgr5* and *Smoc2* (C1-Figure 3e). Gene expression profiling confirmed elevated expression levels of the EpiHR-signa-

ture in Emp1-Tomato-high cells whereas the WNT/Lgr5+ ISC program was downregu-




**C1-Extended Data Figure 8 | HRCs are enriched in invasion fronts and micrometastases. a,** Primary tumor outlined by cyan line and colored in 4 different regions identified with HALO classifier (tumor-red, stroma-green, background-yellow, necrosis-blue). Scale bar, 1mm. **b,** tdTomato cell intensity analysis in the tumor area after segmentation into individual cells. B' and B'' show magnified regions corresponding to tumor core (B')

and invasion fronts + tumor buds (B''). Scale bars, 1mm (B), 100 µm (B' and B''). **c,** Representative immunostaining of tdTomato and E-Cadherin in the tumor core and in tumor buds of primary tumors derived from E-iCT MTOs 4 weeks post implantation in the caecum. Emp1-tdTomato fluorescence is shown with mpl-inferno LUT. The dashed line delimits the caecum edge. Arrows point to tumor buds. Scale bars, 100 µm (tumor core) 50 µm (tumor buds). **d,** Quantification of EMP1-high cells (defined as cells in percentile 90 for tdTomato expression) in the tumor core (submucosal area), invasion fronts (inside muscular layer) and isolated glands (over muscular layer). Box and whiskers from minimum to maximum values. Two-sided Wilcoxon test on percentages. n= 8 mice. **e,** Immunofluorescence of td-Tomato, CD31 and DAPI in primary tumors. Amplified insets show the tumor core (TC) and invasive glands intermingled in mucosal layers (ML) next to blood vessels. Dashed lines outline healthy intestinal epithelium. Scale bars, 250 µm, 100 µm (tumor core) and 50 µm (tumor buds). **f,** Representative images of Emp1-tdTomato and E-Cadherin staining in micro (left) and medium (right) size metastases. Scale bars: 50 µm and 250 µm. **g,** Percentage of tumor area containing Emp1-tdTomato high and low fluorescent pixels versus metastases size (in pixels). Each dot represents an individual metastasis.

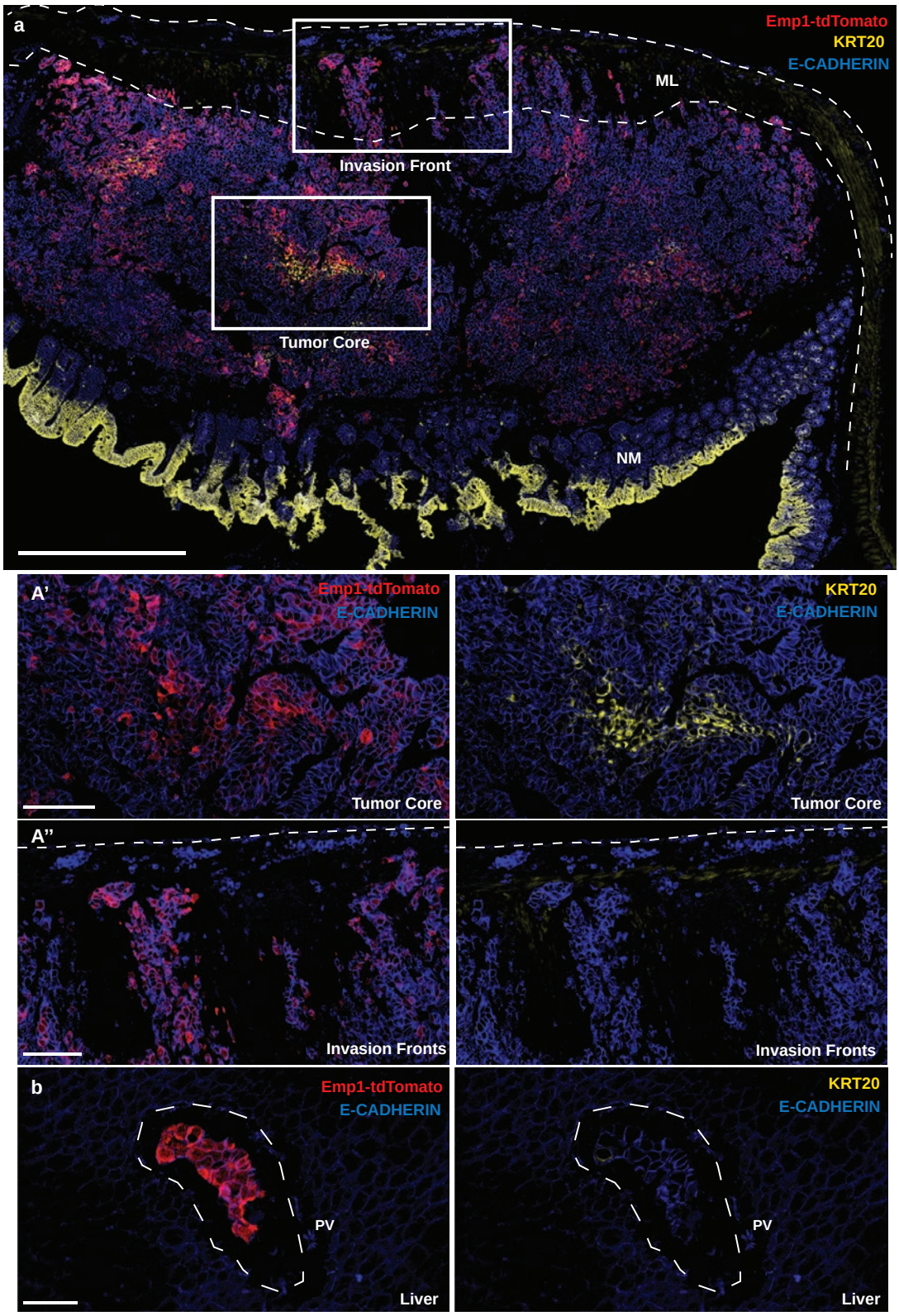
lated in these cells (C1-Figure 3f). These findings validated further the use of Emp1-td-Tomato reporter to track the HRC population.

Inspection of tissue sections evidenced that cancer cells invading the muscular layer were strongly labeled by the fluorescent reporter (C1-Figure 3g). In particular, isolated tumor buds and larger clusters in contact with the stroma at the edges on invasion fronts exhibited the highest Tomato expression (Inset in C1-Figure 3g, Extended Data Figure 8c, d).

Tumor budding predicts disease relapse and metastasis in CRC patients (Lugli et al., 2017; van Wyk et al., 2019). Consistent with this association, Emp1-TOM$^{high}$ isolated tumor buds were often found in proximity to peripheral blood vessels in mouse primary CRCs, suggesting a connection with hematogenous or lymphatic dissemination (C1-Figure 3h and Extended Data Figure 8e). We also examined livers of mice bearing AKTP tumors at various time points post orthotopic MTO implantation. In samples collected at early time points, individual DTCs and micrometastatic lesions trapped within portal veins and liver sinusoids were populated entirely by Emp1-TOM$^{high}$ cells (C1-Figure 3i,j and Extended Data Figure 8f).

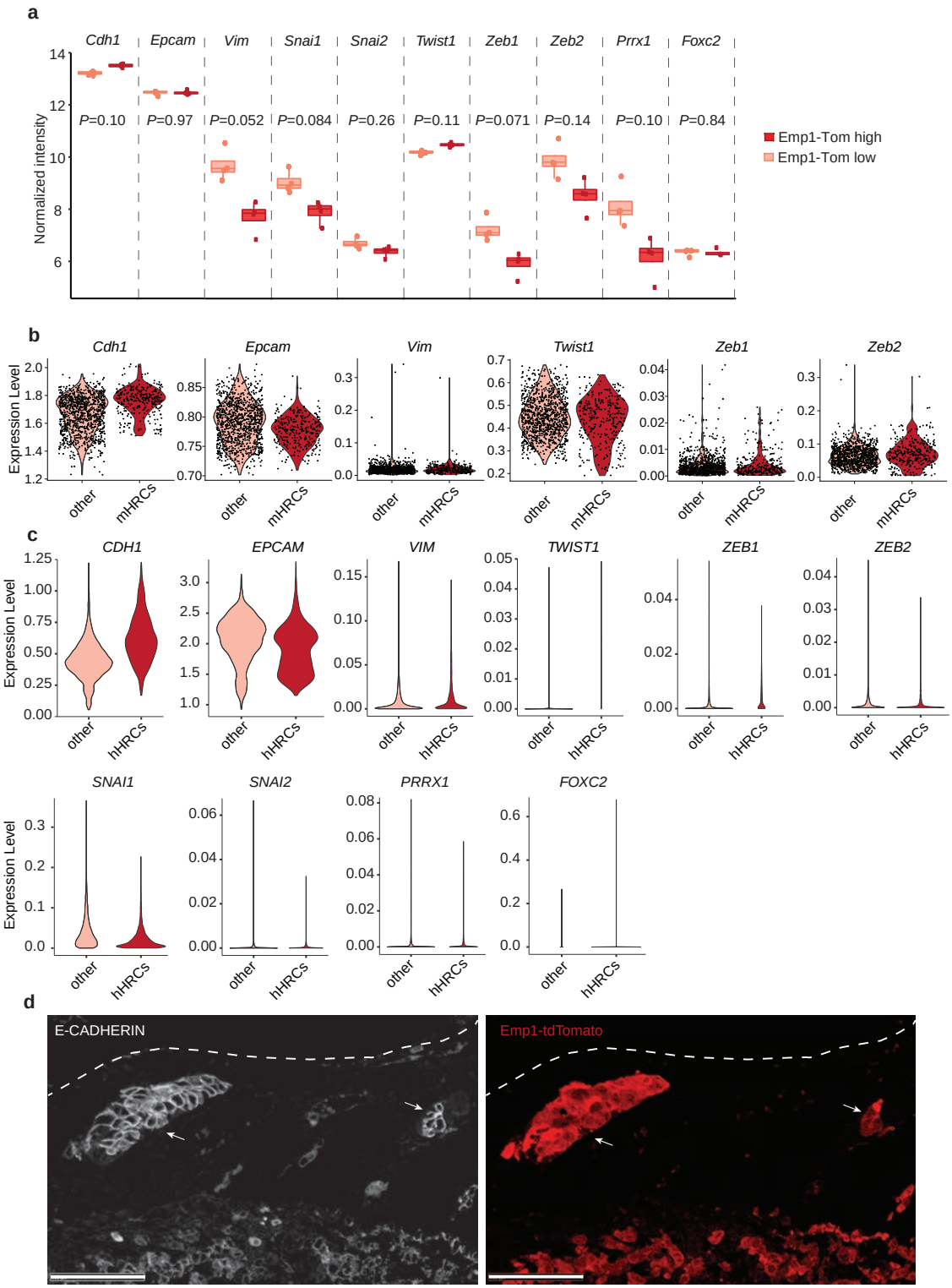
Fitting the single cell RNAseq analyses, we identified two Emp1-TOM+ subsets; one expressed KRT20 and was located mainly in the tumor cores whereas the other was positioned at invasion fronts, lacked KRT20 and expressed the highest levels of Tomato reporter (C1-Extended Data Figure 9a). Liver Emp1-TOM$^{high}$ micrometastases were also KRT20 negative (C1-Extended Data Figure 9b).

Of note, Emp1-TOM$^{high}$ tumor buds and micrometastases were labeled with Epcam and




**C1-Extended Data Figure 9 | KRT20+/EMP1+ cells are located in the tumor core. a,** Emp1-tdTomato, KRT20 and E-Cadherin staining in primary tumors generated by Emp1-iCasp9-tdTomato MTOs. Dashed lines encompass invasion fronts and tumor buds. KRT20 staining is observed in normal mucosa (NM) and to a lesser extent in the tumor core. Tumor cell clusters invading the muscular layer (ML) express high levels of tdTomato and no KRT20. Amplified insets show an example of tumor core (A') and invasion fronts (A") with

tdTomato (left) and KRT20 (right) stainings. Scale bars, 500 µm (a) and 100 µm (A' and A''). **b,** Immunofluorescence of tdTomato and E-Cadherin (left) and KRT20 and E-Cadherin (right) in a cluster of tumor cells that enter the liver through a portal vein (PV, delimited with dashed lines). Scale bar, 50 µm. NM= normal mucosa, ML= muscular layer, PV= portal vein.



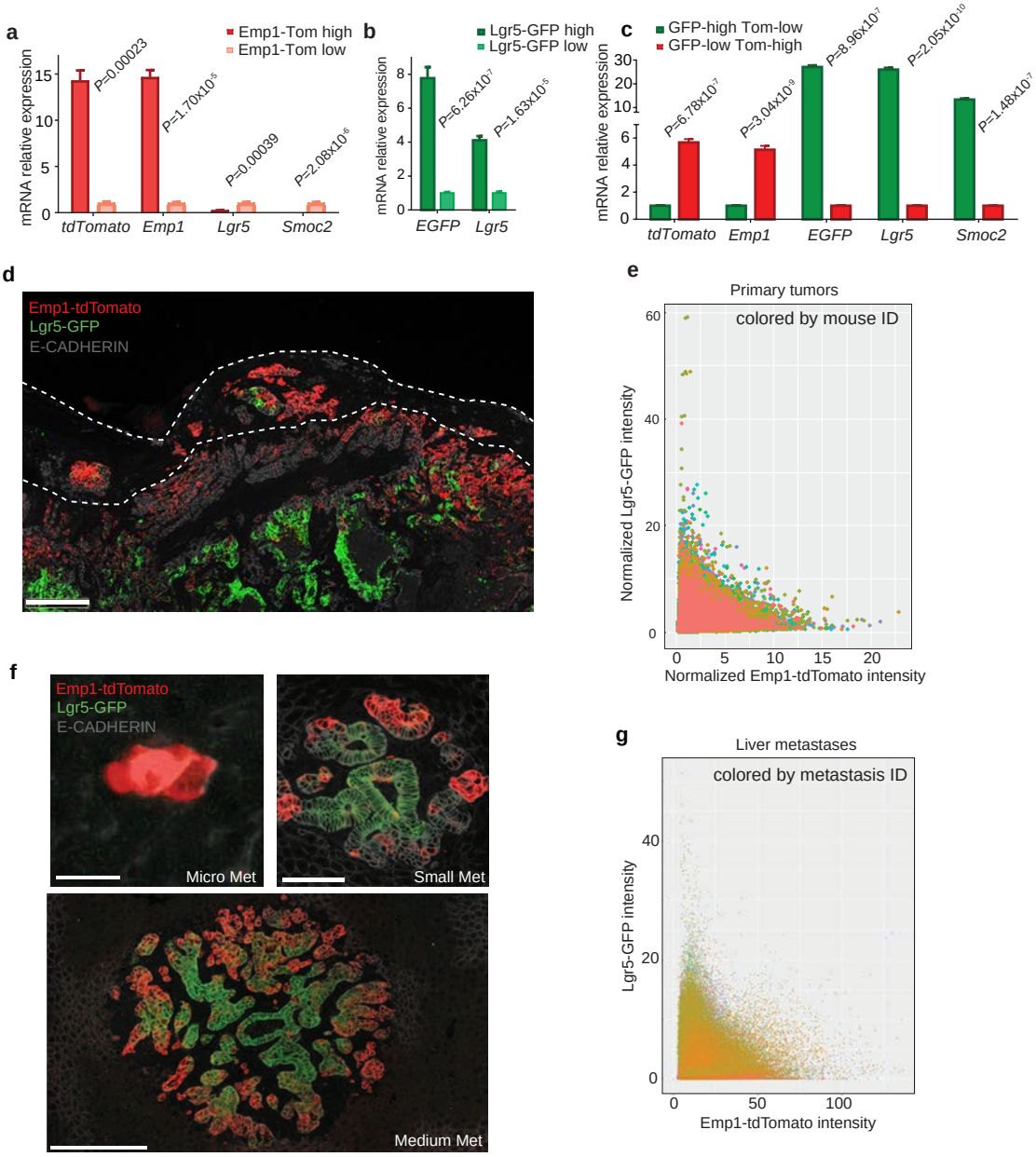

**C1-Extended Data Figure 10 | HRCs preserve an epithelial phenotype. a,** Boxplot showing the normalized expression of genes related to epithelial-to-mesenchymal transi-

tion (EMT) in Emp1-tdTomato-low vs –high cells. Box plots have whiskers of maximum 1.5 times the interquartile range; Boxes represent first, second (median) and third quartiles. Differential expression with Linear Model for Microarray Analysis (limma). n=4 biological replicates. Cdh1 (E-Cadherin); Vim (Vimentin); Snai1 (Snail); Snai2 (Slug); Twist1 (Twist Family BHLH Transcription Factor 1); Zeb1, Zeb2 (Zinc Finger E-Box Binding Homeobox 1 and 2); Prrx1 (Paired Related Homeobox 1); FoxC2 (Forkhead Box C2) **b-c,** Violin plot showing expression of selected EMT-related genes in HRCs versus the rest of other cells in mouse epithelial primary tumor cells (**b**) and human tumor cells from the SMC cohort (**c**). Genes present in a not shown (Snai1 and Snai2) were undetected in (**b**). **d,** Immunostaining of E-Cadherin and Emp1-tdTomato in Emp1-iCT primary tumors 4 weeks post-implantation of MTOs. Arrows point at examples of E-cadherin+ invasion fronts and tumor buds. Dashed lines show the caecum edge. Scale bars, 100 µm.

E-cadherin implying that they retained an epithelial organization (C1-Extended Data Figure 8f and Extended Data Figure 9). Analysis of EMT master transcription factors showed equivalent expression levels in Emp1-TOM$^{high}$ and Emp1-TOM$^{low}$ cells (C1-Extended Data Figure 10a) and scRNAseq also supported lack of EMT in human and mouse HRCs (C1-Extended Data Figure 10b-c).

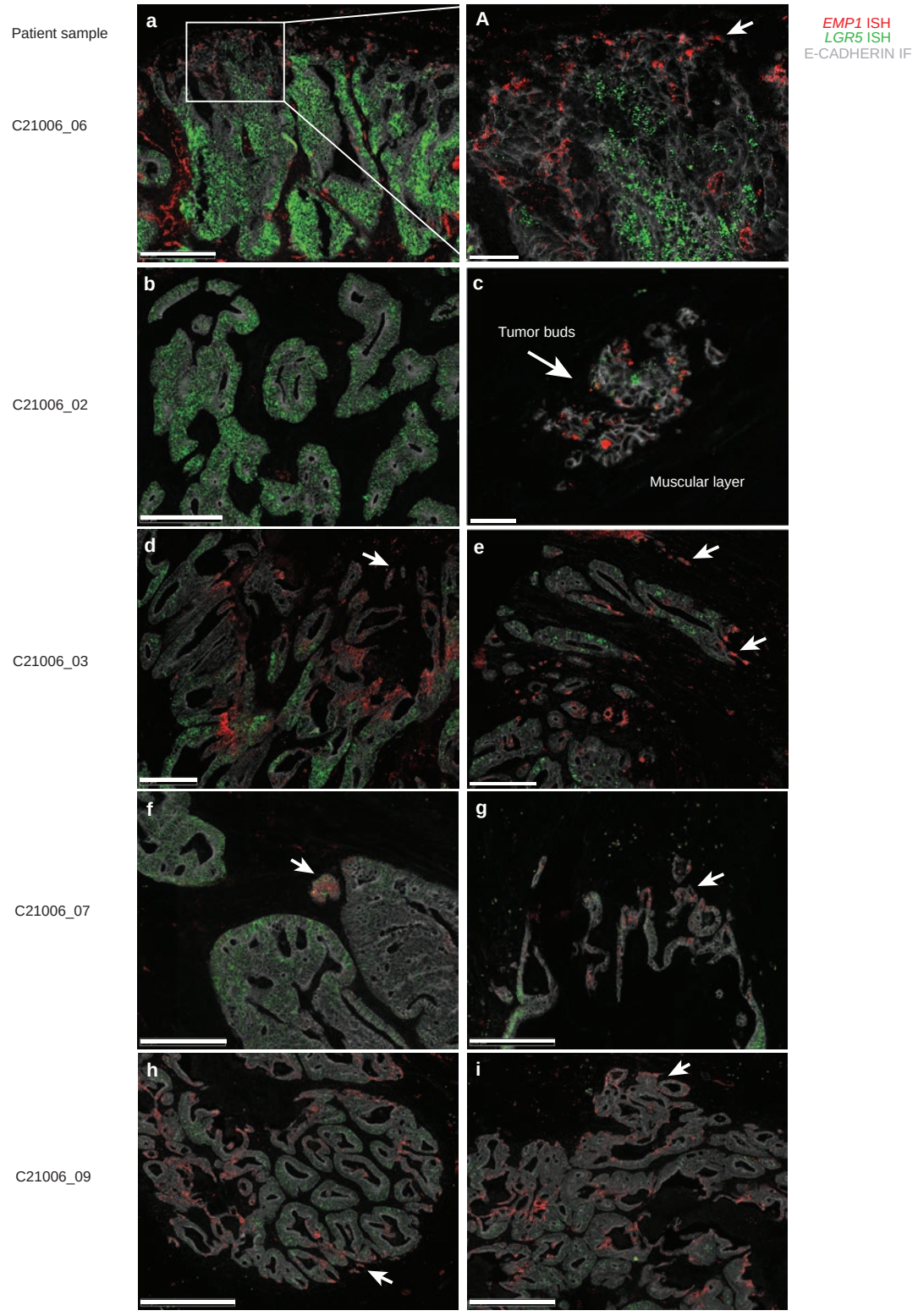
To further explore the relationship between *Emp1* and *Lgr5* expression, we engineered AKTP MTOs bearing both Emp1-iCT and Lgr5-GFP knock-in reporter cassettes (C1-Figure 3k and Extended Data Figure 11a-c). Confocal imaging of dual labelled MTOs showed a mutually exclusive pattern of expression of the two reporters (C1-Figure 3k) and RT-qPCR analysis confirmed upregulation of *Emp1* and *Lgr5* in sorted tdTomato+ (TOM) and GFP+ cells respectively (C1-Extended data Figure 11a-c). Primary CRCs generated from inoculation of the dual labelled MTOs in the caecum also exhibited a mutually exclusive expression pattern of Emp1-tdTomato and Lgr5-GFP reporters (C1-Figure 3l and Extended Data Figure 11d-e). Emp1-TOM$^{high}$ cells were largely enriched at tumor buds which in contrast contained few Lgr5-GFP+ cells (C1-Figure 3l-m). RNA fluorescence in situ hybridization (FISH) analysis on human CRC patient samples also showed that *EMP1* expression was elevated at tumor invasion fronts whereas *LGR5* marked the tumor cores in most cases (C1-Figure 3n and examples in Extended Data Figure 12).

Finally, we analyzed the livers of mice bearing primary CRCs. Lgr5-GFP fluorescence was absent in DTCs and micrometastases, but was progressively gained during metastatic outgrowth, in a marked antithetic pattern to Emp1-TOM expression (C1-Figure 3o-p and Extended data Figure 11f-g). Together with CellRank bioinformatic predictions, these observations suggest that HRCs are endowed with the ability to migrate and disseminate to foreign organs where they initiate metastatic outgrowth and subsequently give rise to non-HRC populations.




**C1-Extended Data Figure 11 | Emp1 and Lgr5 mark distinct tumor cell populations. a,** RT-qPCR analysis of tdTomato, Emp1 and ISC genes in sorted Emp1-high and -low cells dissociated from in vitro organoids. n=3 technical replicates. Mean +/- SD. Two-sided t-test after normalizing by PPIA. **b,** RT-qPCR analysis of EGFP and Lgr5 genes in GFP-high and –low sorted cell populations from in vitro organoids. n=3 technical replicates. Two-sided t-test after normalizing to B2M. Mean +/- SD. **c,** Relative mRNA expression in GFP-high/Tom-low and GFP-low/Tom-high sorted cells dissociated from subcutaneous AKTP Emp1-iCT Lgr5-GFP tumors. Two-sided t-test after normalizing by PPIA. Mean +/- SD. **d,** Immunostaining of Emp1-tdTomato, Lgr5-GFP and E-Cadherin in Emp1-iCT Lgr5-GFP primary tumors 4 weeks post-implantation of MTOs. Dashed lines encompass tumor buds. Scale bar, 250 µm. **e,** Scatter plot showing normalized Emp1-tdTomato intensity versus normalized Lgr5-GFP intensity in 855,330 cells from 18 different primary tumors. Note the absence of double positive cells (tdTomato and GFP high). **f,** Representative immunofluorescence staining of tdTomato, GFP and E-Cadherin in liver metastases of increasing size generated from the mouse CRC relapse model. Scale bars, 25 µm (micro) 100 µm (small) 250 µm (medium). **g,** Scatter plot showing tdTomato intensity versus GFP intensity in 318276 cells from

137 different liver metastases. Note the absence of double positive cells.



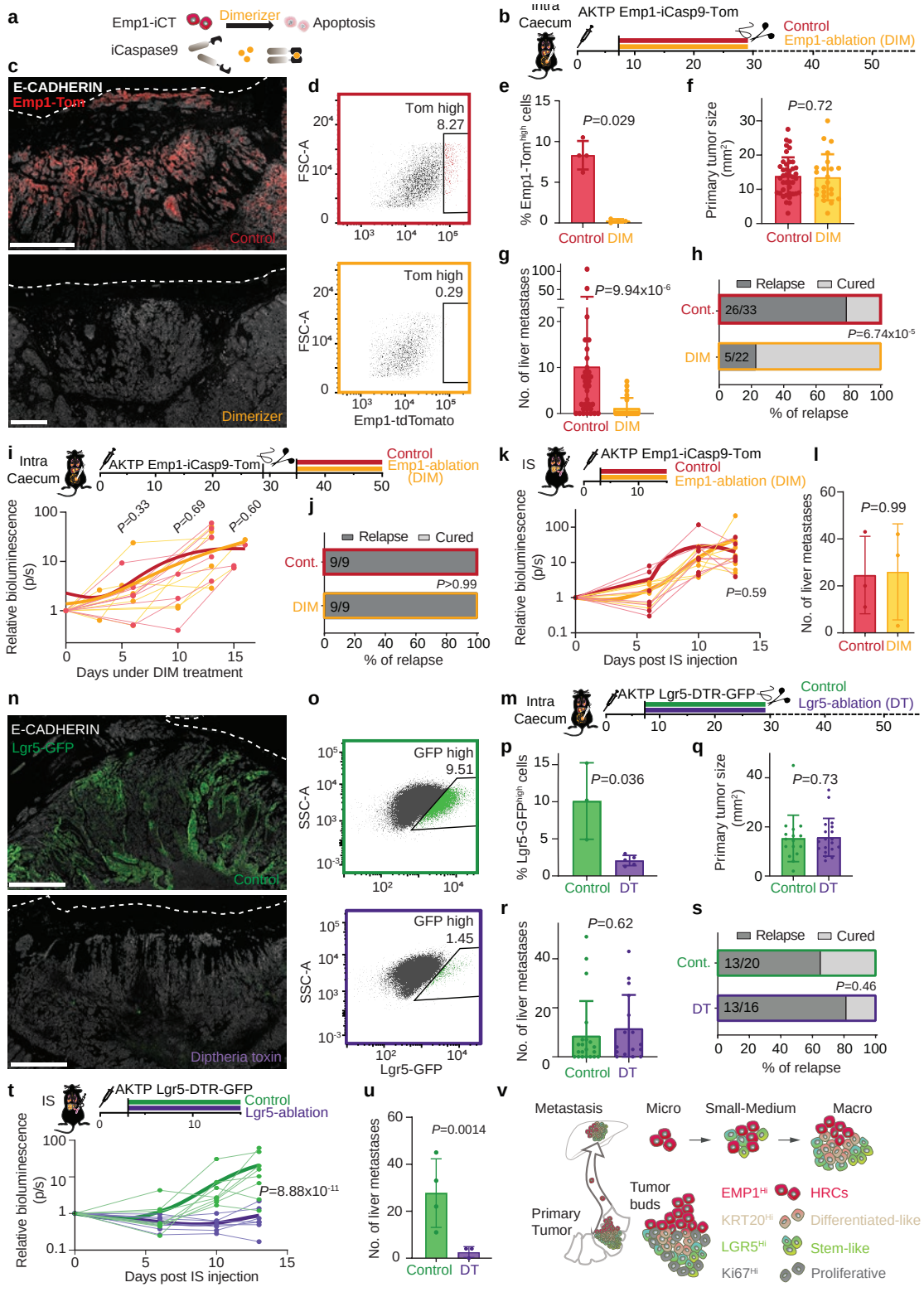

**C1-Extended Data Figure 12 | Dual EMP1 LGR5 mRNA ISH in human primary CRCs. a-j,** Dual EMP1 and LGR5 mRNA FISH combined with E-Cadherin immunofluorescence on human primary CRC tissue sections showing a mutually exclusive pattern of expression of EMP1 and LGR5. Each row corresponds to an individual patient. Note that EMP1 expression is high in invading tumor cell clusters (white arrows). Scale bars, 500 µm (h, i) 250 µm (a, b, d, e, f, g) 50 µm (A, c).

## HRCs are the cell of origin of metastatic relapse

To test functionally whether HRCs are responsible for metastatic relapse, we leveraged the inducible Caspase9 cassette inserted in the Emp1 locus to perform cell ablation experiments (C1-Figure 4a,b) (Morral et al., 2020; Shimokawa et al., 2017). Inoculation of mice with AP20187 dimerized the chimeric Caspase9 expressed under the Emp1 locus and specifically killed cells expressing the highest levels of Emp1-TOM reporter (C1-Figure 4c,d). Of note, dimerizer (DIM) treatment was only administered during primary tumor growth but was ceased the day before primary CRC resection (C1-Figure 4b). As macrometastases were not yet present when DIM treatment finished, this experimental setting aimed at ablating EMP1+ cells in the primary tumor and possibly in incipient metastatic lesions. Remarkably, while DIM treatment had no effect on primary tumor growth (Figure 4f), the vast majority of mice showed no signs of metastatic recurrence and were disease-free at experimental endpoints (C1-Figure 4g, h). On average, ablation of Emp1-high cells reduced liver metastases formation by 10-fold (C1-Figure 4g).

In contrast, when Emp1-high cell ablation started 1 week after primary CRC resection we observed no changes in metastatic progression and all mice suffered metastatic relapse (Figure 4i, j). Emp1high cell ablation neither halted metastasis formation when MTOs were directly inoculated in the liver through the spleen (C1-Figure 4k, l). We therefore conclude that EMP1+ HRCs drive metastatic relapse after primary tumor resection yet they are dispensable after metastatic seeding is completed.

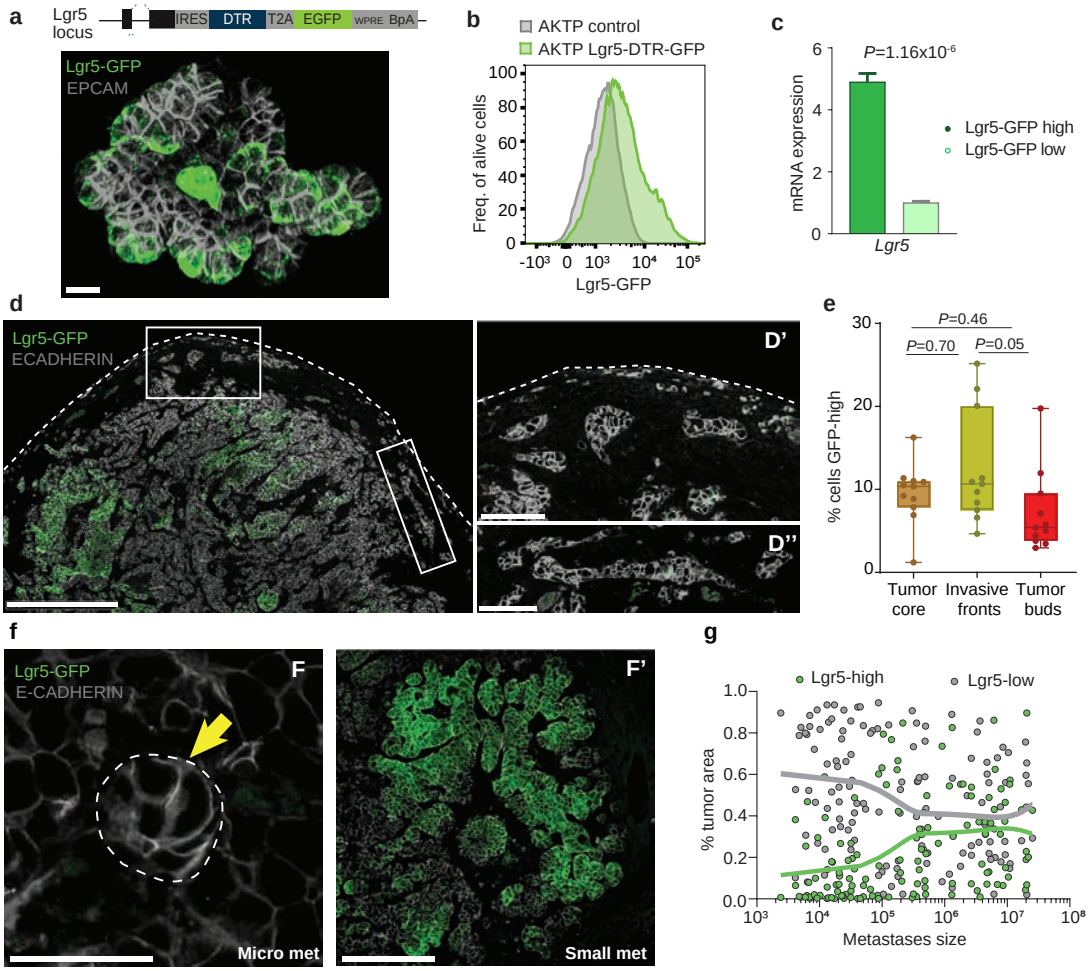
Using a diphtheria toxin receptor (DTR)-based ablation strategy in CRC models, it was previously shown that Lgr5+ CRC cells are necessary for liver metastasis formation (Melo et al., 2017). To assess the role of Lgr5+ tumor cells in our relapse models, we knocked-in a DTR cassette into the Lgr5 locus of AKTP MTOs (C1-Extended Data Figure 13a-c). Inoculation of this MTO line into the caecum further validated our previous observations that invasion fronts, tumor buds and micrometastases contain seldom Lgr5+ cells (C1-Extended Data Figure 13d-g). More importantly, treatment with Diphtheria toxin (DT) before surgical removal of the primary CRC effectively eliminated Lgr5+ cells (C1-Figure 4m-p) yet it did not prevent disease relapse and mice developed overt liver metastatic disease (C1-Figure 4q-s). Thus, Lgr5+ cells are dispensable for dissemination and metastatic colonization.




**C1-Figure 4 | Emp1-high cells are the origin of metastases relapse. a,** Scheme depicting iCaspase9-mediated Emp1-cell ablation upon dimerizer (DIM) treatment. **b,** Inducible ablation and surgery schedule of mice with AKTP E-iCT primary tumors. **c,** Immunostaining of Emp1-tdTomato and E-Cadherin showing effective ablation of Emp1-high cells in DIM-treated primary tumors compared to controls. Dashed lines delimitate the caecum edge. Scale bars, 500 µm (C), 250 µm (C'). **d,** Representative flow cytometry plot of tdTomato fluorescence in Epcam+ cells comparing control versus DIM-treated mice. **e,** Percentage

of Emp1-high tumor cells (defined as the top 10% of the tdTomato population, mean ± SD) in control and treated mice after three weeks of treatment, at the time of surgery. Every dot is an individual mice, n=4 in both groups. Two-sided Wilcoxon test. **f,** Primary tumor area (mean ± SD) measured after resection. Each dot is an individual mice, n= 33 (control) and 22 (DIM) mice. P-value for linear model after boxcox transformation. **g,** Liver metastases (mean ± SD) generated by MTO EiCT up to one month after primary tumor resection, treated with vehicle or DIM. Each dot is an individual mice, n as in f. P-value for generalized linear model with negative binomial family. **h,** Percentage of mice that developed liver metastases or remained metastases-free in control and Emp1-ablated tumors. Analyzed with a two-sided fisher test. **i,** Schematic of an experiment for late ablation of Emp1 cells. Untreated AKTP Emp1-iCT were left to grow and resected 4 weeks after injection. 1 week later we treated with DIM or vehicle for 13-16 days. Normalized bioluminescence (BLI) of liver metastases to the day DIM treatment started. Points and lines represent individual mice, trend lines (bold) show a LOESS model. n=9 mice per each group. P-values of mixed effects linear model with data normalized to time 0 and mouse as random effect. **j,** Percentage of mice that developed liver metastases or were cured in control versus late-ablated Emp1 tumors. Two-sided fisher test. **k,** BLI monitoring of the effect of the ablation of Emp1 cells 3 days after intrasplenic inoculation of Emp1-iCasp9-Tom organoids. Points and lines represent individual mice, trend lines (bold) show a LOESS model. n=7 mice per each group. P-values of mixed effects linear model with data normalized to time 0 and mouse as random effect. **l,** Number of liver metastases counted in the experiment described in k. Mean ± SD. n= 3 mice per group. Generalized linear model with negative binomial family. **m,** Inducible ablation and surgery schedule of mice orthotopically engrafted with AKTP Lgr5-DTR-GFP primary tumors. n, Representative immunofluorescence stainings of Lgr5-DTR-GFP primary tumors untreated or treated with diphtheria toxin to ablate Lgr5+ cells. Dashed lines outline the serosa. Scale bars, 100 µm. **o,** and Representative flow cytometry plot of Lgr5-GFP fluorescence in controls versus diphtheria toxin (DT)-treated mice. Epcam+ cells are shown. **p,** Percentage of Lgr5high tumor cells (defined as the top 10% of the GFP population, mean ± SD) in control and treated mice. n=3 and 5 mice from left to right. P-value for generalized linear model with binomial family. **q,** Primary tumor area (mean ± SD) measured after resection. n=20 (control) and 16 (DT) mice. Two-sided Wilcoxon test. **r,** Liver metastases (mean ± SD) generated by Lgr5-DTR-GFP MTOs up to one month after primary tumor resection, treated with vehicle or diphtheria toxin. Each dot is an individual mice, n as in r. p-value for generalized linear model with negative binomial family. **s,** Percentage of mice that developed liver metastases or remained metastases-free in control and Lgr5-ablated tumors. Two-sided fisher test. **t,** Bioluminescence monitoring of the effect of the ablation of Lgr5 cells 3 days after intrasplenic inoculation of Lgr5-DTR-EGFP MTOs. Points and lines represent individual mice, trend lines (bold) show a LOESS model. n=7 mice in control and n=8 in DT treated. P-values of mixed effects linear model with data normalized to time 0 and mouse as random effect. **u,** Number of liver metastases counted in the experiment detailed in t. Mean ± SD. n= as in t. Generalized linear model with negative binomial family. **v,** Proposed model for metastatic dissemination of CRC. HRCs in the invasion fronts of primary tumors disseminate to the livers where they de-differentiate into Lgr5+ cells that drive metastatic outgrowth.

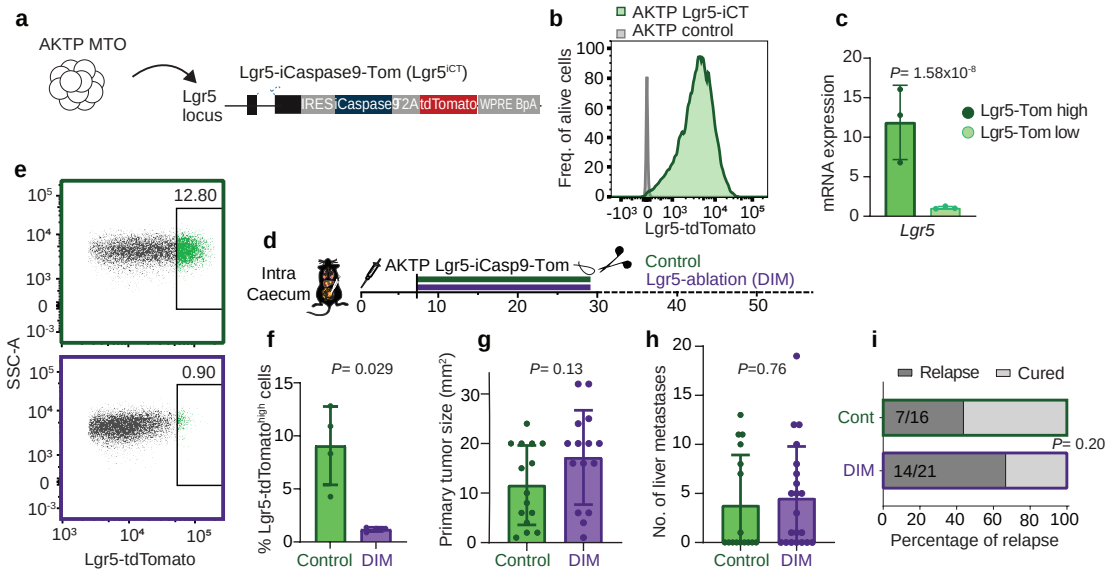
We validated these results using an independent AKTP MTO line engineered with an iCaspase9-tdTomato (iCT) cassette knocked-in in the Lgr5 locus (C1-Extended Data




**C1-Extended Data Figure 13 | Tumor cells at invasion fronts and micrometastases are Lgr5 negative. a,** CRISPR-Cas9 targeting strategy to introduce a IRES/DTR/T2A/EGFP/WPRE/BGHpolyA cassette into the Lgr5 locus of MTOs. Confocal imaging of immunostaining for GFP and Epcam in Lgr5-DTR-GFP organoids. Imaris 3D representation. Scale bar, 30µm. **b,** Representative flow cytometry plot of GFP expression in wild-type and Lgr5-GFP organoids. **c,** Relative Lgr5 mRNA expression of Lgr5-GFP-high over -low cells dissociated from Lgr5-DTR-GFP subcutaneous tumors. n=3 biological replicates. Two-sided t-test normalizing to B2M. **d,** Immunofluorescence of Lgr5-GFP and E-Cadherin in primary tumors. Insets show GFP- invasion fronts and tumor buds at higher magnification (D' and D''). Scale bars, 500 µm (D) and 100 µm (D' and D''). **e,** Quantification of Lgr5-GFP-high cells (defined as cells in percentile 90 for GFP expression) in the tumor core, invasion fronts and tumor buds. Paired two-sided Wilcoxon test on percentages. n= 11 mice. **f,** Representative images of Lgr5-GFP staining in micro (F) and small (F') metastases. Dashed lines and the yellow arrow surround a micrometastasis. Scale bars: (F) 50 µm; (F') 250 µm. **g,** Percentage of tumor area containing Lgr5-GFP high and low fluorescence versus metastases size in pixels. Each dot represents an individual metastasis.

Figure 14a-c). Again, effective ablation of Lgr5+ cells (C1-Extended Data Figure 13e,f) neither altered CRC tumor growth (C1-Extended Data Figure 14e-g) nor prevented metastatic recurrence (C1-Extended Data Figure 14h, i). On the other hand, ablation of Lgr5

cells after direct inoculation of MTOs in the liver through the portal vein halted metastasis formation (C1-Figure 4t, u), further supporting a requirement for Lgr5-cells during metastatic outgrowth (Melo et al., 2017).




**C1-Extended Data Figure 14 | Metastatic dissemination occurs in the absence of Lgr5+ cells. a,** CRISPR-Cas9 targeting strategy to introduce an IRES/iCaspase-9/T2A/tdTomato/WPRE/BGHpolyA cassette into the LGR5 locus of AKTP MTOs. **b,** Representative flow cytometry plot of tdTomato expression in Lgr5-iCasp9-tdTomato organoids vs non-transfected organoids. **c,** Quantification of Lgr5 mRNA by RT-qPCR in Lgr5-tdTomato-high and -low cells dissociated from primary tumors grown for 4 weeks. n=3 primary tumors. Analyzed with a mixed effects linear model. **d,** Timing of inducible ablation and surgery in mice implanted with AKTP Lgr5-iCasp9-TOM primary tumors. **e,** Representative flow cytometry plot of Lgr5-tdTomato fluorescence in controls versus dimerizer-treated mice. DAPI- Epcam+ cells are shown. **f,** Percentage of Lgr5high tumor cells (defined as the top 10% of the tdTomato population) in control and treated mice. n=4 mice each group. Two-sided Wilcoxon test. **g,** Primary tumor area measured after resection. n= 15 mice each group. Mean with SD, p-value of linear model after boxcox transformation. **h,** Liver metastases counted at experimental endpoints after primary tumor resection. n= 16 (control) and 21 (Lgr5-ablation) mice. Mean ± SD. Analyzed with a linear model with negative binomial family. **i,** Percentage of mice that developed liver metastases or remained cured in control and Lgr5-ablated tumors. n= 16 (control) and 21 (Lgr5-ablation) mice. Two-sided Fisher test.

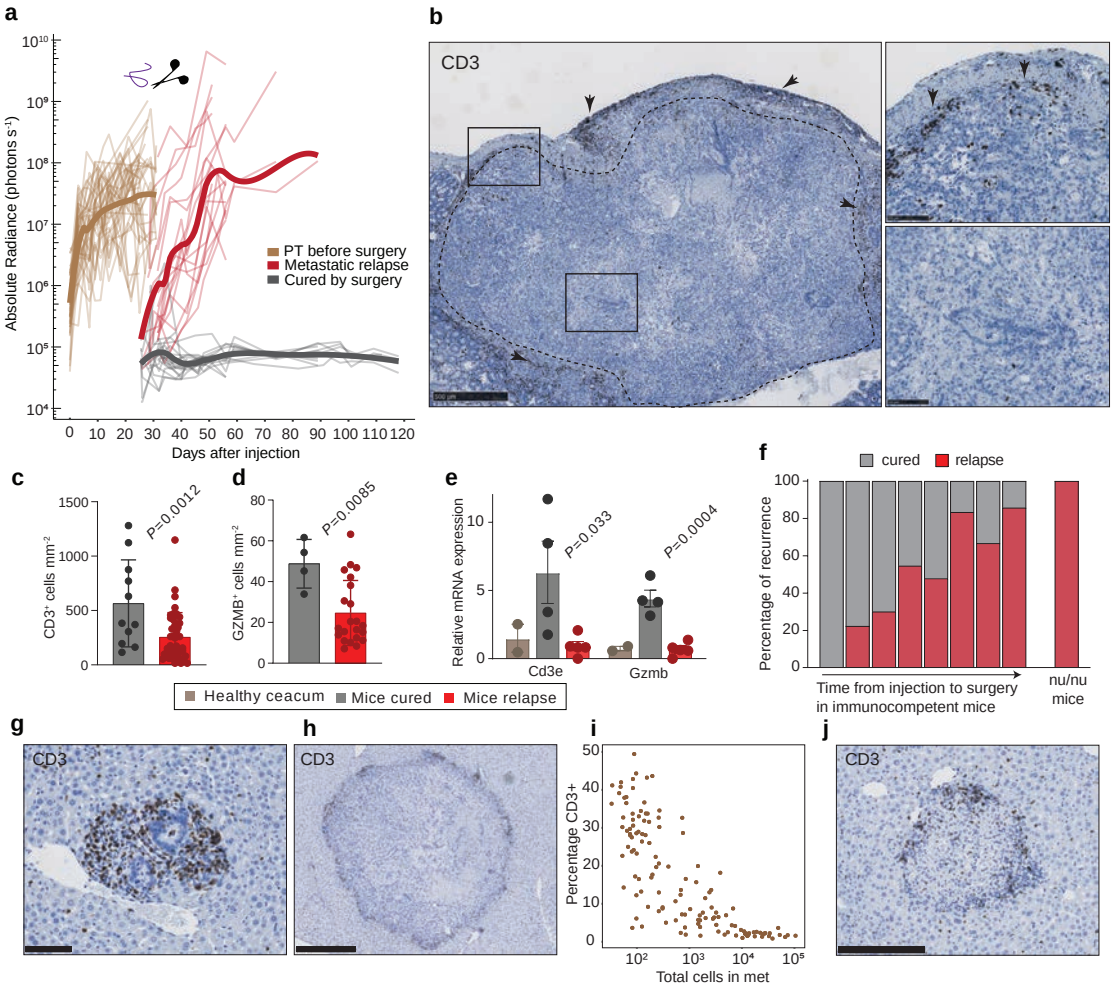
# Chapter 2 - Neoadjuvant immunotherapy prevents metastatic relapse in MSS CRCs

## Results

## Adaptive immunity restricts metastatic dissemination in MSS CRC

Seminal studies by the Galon lab revealed that poor T-cell infiltration in primary tumors predicts adverse outcomes after surgical intervention in patients with colorectal cancer (Galon et al., 2006). Subsequent research revealed that particular features of the tumor microenvironment, such as the presence of cancer-associated fibroblasts that secrete high levels of TGF-beta, were linked to poor prognosis in CRCs (Calon et al., 2015; Isella et al., 2015). Human-like mouse models of colorectal cancer bearing mutations in APC, KRAS, TP53 and TGFBRII (AKTP MTOs) reproduce key features of such poor-prognosis tumors, including both T cell exclusion and an immunosuppressive stromal-rich TGF-β-activated TME (Tauriello et al., 2018). Given that immune evasion in poor prognosis CRCs is particularly dependent on the TME, we decided to explore anti-tumor immunity during metastatic dissemination, where tumor cells abandon immune-privileged niches and land on unperturbed organs. For that, we used our recently developed CRC relapse model described in chapter 1 (Cañellas-Socias et al., *in revision*), in which luciferase-labelled MTOs generate primary tumors in the caecum of immunocompetent mice that can be surgically extirpated. After surgery, a fraction of mice develops metastases in the liver and other clinically relevant sites, while another subset remains relapse-free several months after surgery (C2-Figure 1a). Although the proportion of mice cured by surgery decreased as the interval between injection of MTOs and surgery increased, a subset of mice remained relapse-free even when primary tumors were resected after long periods of time (C2-Figure 1f). Akin to human patients, primary tumors from mice that relapsed after surgery contained fewer CD3+ and GZMB+ T cells than those that remained cured after surgery (C2-Figure 1c-e). To further test the role of the adaptive immune system, we implanted MTOs in nu/nu T cell-deficient mice and we observed full metastatic penetrance (C2-Figure 1f). Altogether, these data underscore the role of the immune system in restraining dissemination of MSS CRCs, and supports our preclinical model to study bona-fide adaptive immune responses during metastatic dissemination.

As a first approach to study the escape of tumor cells from adaptive immunity during metastatic dissemination, we collected livers at different timepoints after primary tumor implantation and analyzed T cell infiltration. Quantifying the percentage of CD3+ cells in liver metastases of increasing size revealed a marked downregulation of CD3+ cell infiltration as metastases grew (C2-Figure 1i). Similarly to primary tumors, CD3+ T cells in overt metastases were either excluded to the periphery or present in low abundance (C2-Figure 1h). In sharp contrast, tumor glands in liver micrometastases were heavily infiltrated by T cells (C2-Figure 1g). Moreover, we often observed groups of T cells surrounding dead cells, suggesting T cell-mediated killing of migratory tumor cells that reached the liver (C2-Figure 1j). Our data indicates that immune evasion decreases




**C2-Figure 1 | Adaptive immunity restricts metastatic dissemination in MSS CRC. a**, Longitudinal intravital bioluminescence (BLI) imaging quantification (photons s−1) of a representative experiment where AKTP MTOs were implanted in the caecum of c57BL/6J mice (in brown). 3-4 weeks post-implantation, primary tumors are excised resulting in reduction of bioluminescence to background levels. After surgery, bioluminescence reappears in the upper abdomen of some mice, indicating the presence of liver metastases (in red). Other mice are cured by surgery and remain free of bioluminescence until the end of the experiment (in grey). **b**, CD3 immunohistochemistry of a LAKTP T4 carcinoma (CA), arrowheads indicate CD3+ cells. Inset show magnified views of the periphery (upper panel) and the tumor core (bottom panel). Note that most CD3+ cells are excluded to the borders of tumors. Scale bar, 500 µm. **c-d**, Number of CD3+ (c) and GZMB+ (d) cells per mm2 (mean ± SD) in primary tumors from mice that were cured (grey) or relapsed (red) after surgery. Each dot is a primary tumor. Two-sided t test. **e**, Relative mRNA expression of in healthy intestine (brown) and primary tumors that relapsed (red) or were cured (grey) by surgery. Two-sided t-test after normalizing by PPIA. Dots represent biological replicates. Mean +/- SD. **f,** Percentage of metastatic recurrence depended on time elapsed from injection to primary tumor resection in immunocompetent mice and nu/nu immunodeficient mice. Time intervals are 5-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, 41-45. Nude mice could not be resected due to local spreading of primary tumors. Recurrence was determined 5 weeks post implantation. **g-h**, CD3 immunohistochemistry on a liver micrometastasis (g) and a grown metastases (h). Scale bars, 100 µm and 500 µm, respectively. **i,** CD3+ cell percentage versus individual metastases size (calculated as the number of cells present). **j**, CD3 immunohis-

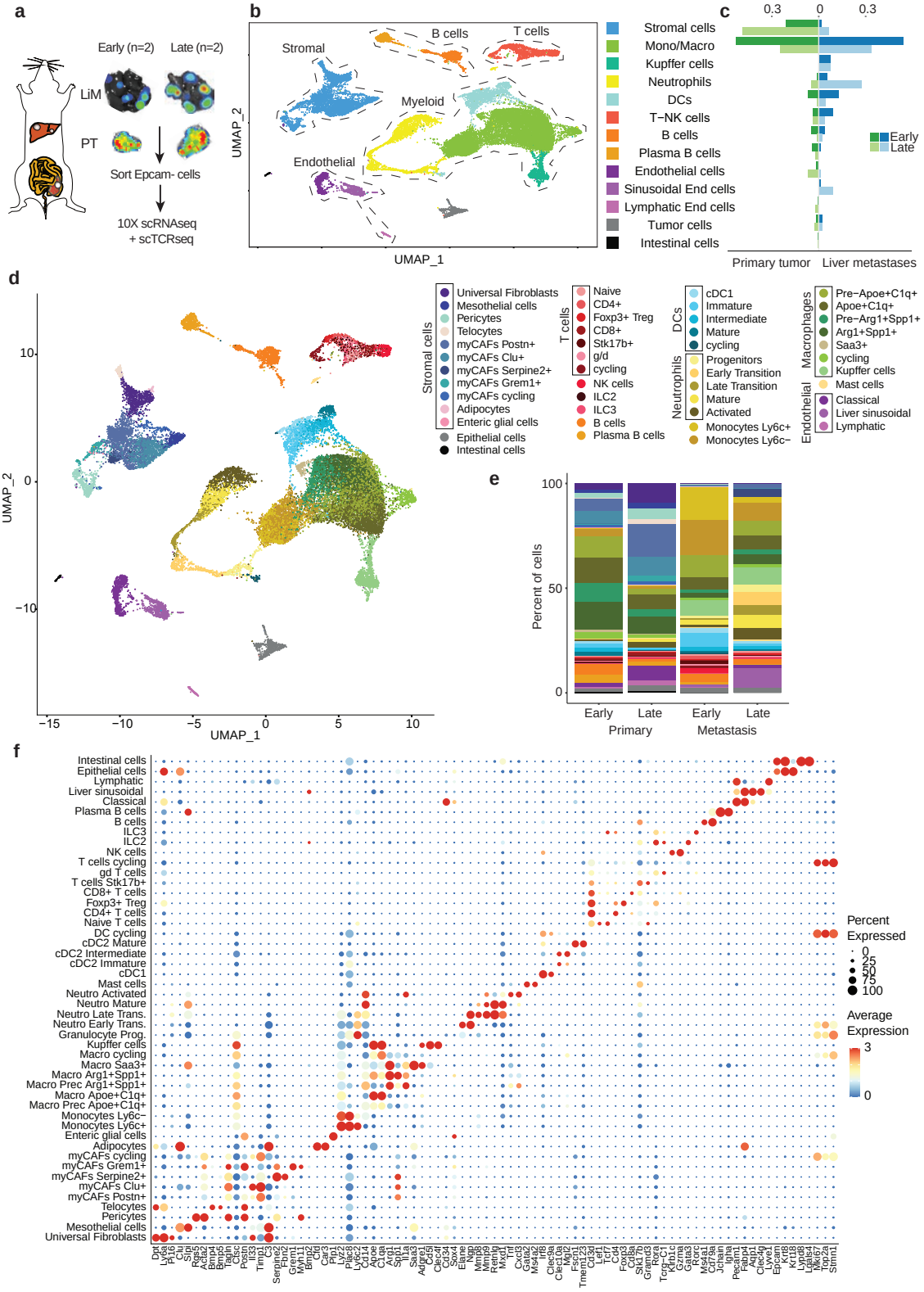
tochemistry showing groups of T cells surrounding dead cells in the liver of mice implanted with AKTP tumors, suggestive of T-cell mediated elimination of tumor cells. Scale bar, 250 µm.

during metastatic dissemination and suggests that micrometastases might be especially vulnerable to the immune system.

## Evolution of the TME during metastatic dissemination

We hypothesized that the absence of a mature TME in newly-disseminated lesions might favor immune infiltration. To investigate this question, we performed single-cell RNA sequencing on the TME of primary tumors and paired liver metastases at early and late timepoints (C2-Figure 2a). Briefly, we macro-dissected tissues and dissociated them into single cells, then Epcam- GFP- negative cells were sorted and their transcriptomes were analyzed using droplet-based 10X Genomic technologies (5' scRNA-seq + scTCR-seq). In this experiment we sequenced through the 5' of RNA, which also enables sequencing of variable regions of the T cell receptor in individual T cells. After quality control and removal of batch effects, 26999 single cells were included in the final dataset.

UMAP representation with cell type-specific marker gene-based annotations revealed 13 cell clusters that we grouped into 5 major cell types: myeloid, stromal, endothelial, T lymphocytes and B lymphocytes (C2-Figure 2b). Fibroblasts were highly abundant in primary tumors, whereas neutrophils were mostly present in metastases samples (C2-Figure 2c). In turn, neutrophils, stromal and endothelial cells were increased in late tumors, suggesting an association with disease-progression. On the other hand, T/NK lymphocytes, B lymphocytes, macrophages and dendritic cells were enriched in early samples. In agreement with our previous observations (C2-Figure 1i), small metastases contained the highest levels of T/NK lymphocytes. Notably, most clusters included cells from both primary and metastatic samples, with the exception of those corresponding to tissue-specific cells. Kupffer cells and liver sinusoidal endothelial cells were unique of liver metastases samples, and lymphatic endothelial cells were present only in primary tumor samples (C2-Figure 2c). Additionally, we identified contamination with intestinal cells and tumor cells albeit present at low abundance. Taken together, these results suggest that cell-types typically related to tumor-promotion in CRC – such as fibroblasts, endothelial cells and neutrophils- were enriched in primary tumors and large metastases, whereas cell types with anti-tumor functions – such as T/NK cells, B cells and DCs - were enriched in small metastases. Nonetheless, most cell types present in the tumor microenvironment can exert dual influence in cancer depending on their specific activation state, exerting either protumoral or antitumoral functions, thus increasing the complexity of understanding the TME. Therefore, and in order to better dissect the



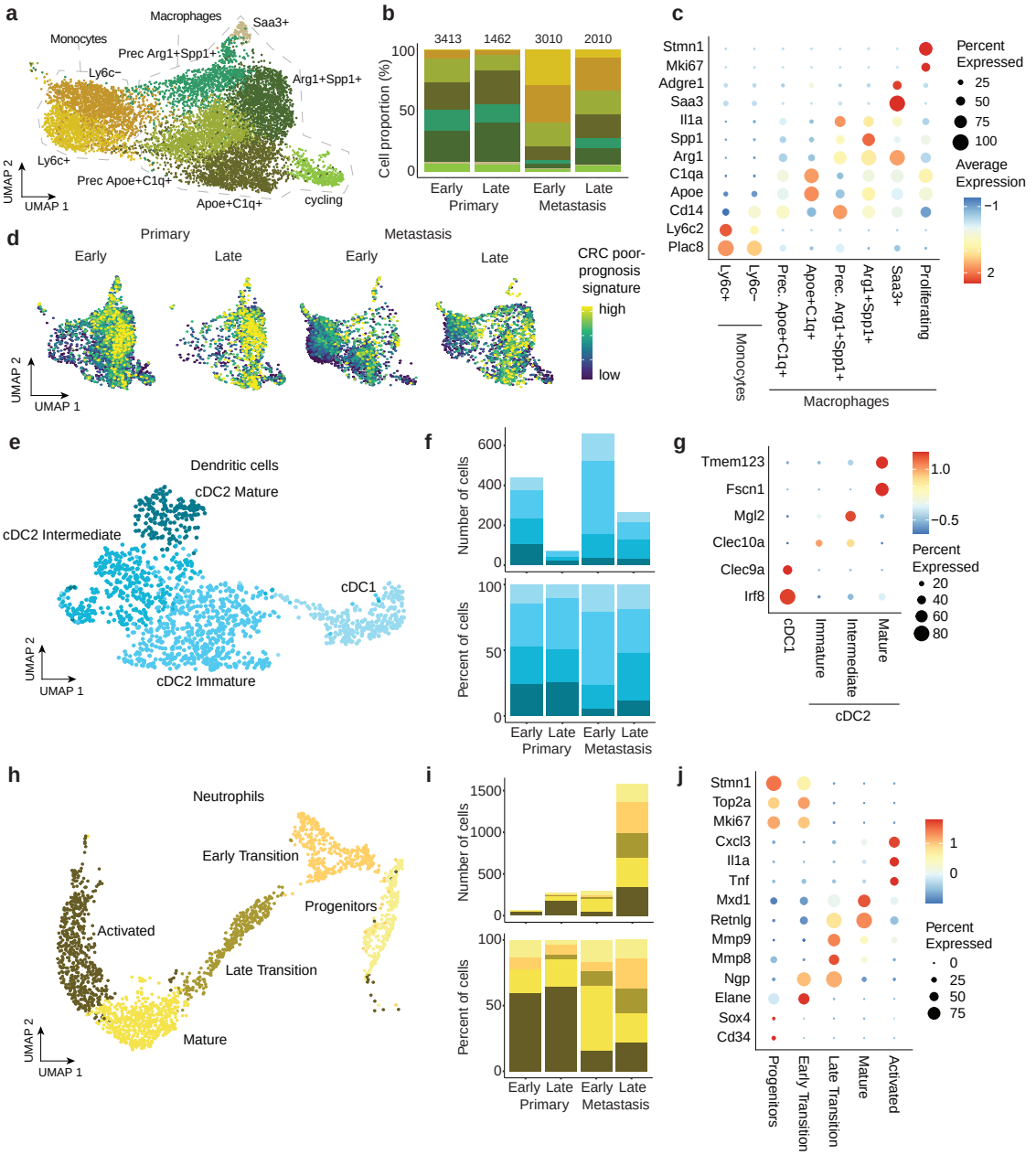

**C2-Figure 2 | Microenvironmental cell landscape in CRC primary tumors and metastases. a**, Schematics of the experiment to profile the tumor microenvironment of primary tumors and liver metastases at early and late timepoints. Samples from primary tumors and liver metastases are paired. **b**, UMAP layout of Epcam- cells isolated from 5 different mice colored by cell clusters. Dashed lines delimit 5 major cell-types identified (Myeloid, Stromal, Endothelial, T lymphocytes and B lymphocytes). **c**, Barplot showing cell-type composition by sample type. **d**, UMAP layout of Epcam- cells isolated from 5 different

mice colored by 48 cell clusters. **e**, Barplot showing Seurat cluster composition of all cells by sample type. Bars in each sample type shows the average proportion between 2 biological independent samples. **f**, Dotplot showing marker genes across 48 cell clusters from (d). Dot size indicates fraction of expressing cells, colored according to z-score normalized expression levels

evolution of the different cell types, further unsupervised clustering in each compartment gave rise to a total of 48 cell clusters, which we annotated based on literature described marker genes (C2-Figure 2d-f).

We first dissected the myeloid compartment (C2-Figure 3), a highly heterogeneous cell type that can contribute to malignancy through production of angiogenic factors, extracellular matrix remodeling and immunosuppression, but also play an important role in the anti-tumor immune response by phagocytosing cancer cells and presenting tumor antigens to CD4 T cells. Tumor-infiltrating monocytes were characterized by high expression of Lyz2 and Plac8 and were classified in classical and non-classical based on Ly6c expression (C2-Figure 3a,c). We found two main subsets of tumor-associated macrophages (TAMs), in agreement with a recent study (Zhang et al., 2020) (C2-Figure 3a,c). C1qc+ macrophages expressed high levels of complement pathway and antigen presentation genes, whereas Arg1, Spp1 and other immunomodulatory molecules defined a different TAM subset. Of note, analyses of genes associated with ‘‘classically activated’’ (M1) and ‘‘alternatively activated’’ (M2) macrophages in TAMs (Azizi et al., 2018) did not explain the dichotomy of C1qc+ and Spp1+ TAMs. In fact, both macrophage subsets expressed high levels of Cd274 and TGFβ signature genes, indicating an immunosuppressive role of both cell types (not shown). Two other clusters showed similar expression patterns to Spp1+ and C1qc+ macrophages but residual Cd14 expression (C3-Figure 3c). Given that a recent study suggested that both C1qc+ and Spp1+ TAMs develop from tumor-infiltrating monocytes (Zhang et al., 2020), we annotated these two clusters as a transitory state from monocytes to mature TAMs (precursors, prec.). In addition, we identified a cluster of proliferating macrophages and a small cluster of cells expressing high levels of Saa3 (C2-Figure 3c). Interestingly, the proportions of cell-types changed substantially during metastases evolution. While primary tumors contained mostly Spp1+ and C1qc+ TAMs, early metastases were highly enriched in monocytes and contained few Spp1+ macrophages (C3-Figure 3b). Later on, metastases increased the proportion of both Spp1+ and C1qc+ TAMs, which expressed high levels of genes associated with poor-prognosis in CRC compared to monocytes, further reinforcing their association with malignancy (C3-Figure 3d).

Additional myeloid cells in the TME included dendritic cells, neutrophils and mast cells. DCs were highly enriched in early-stage samples, whereas neutrophils were mostly



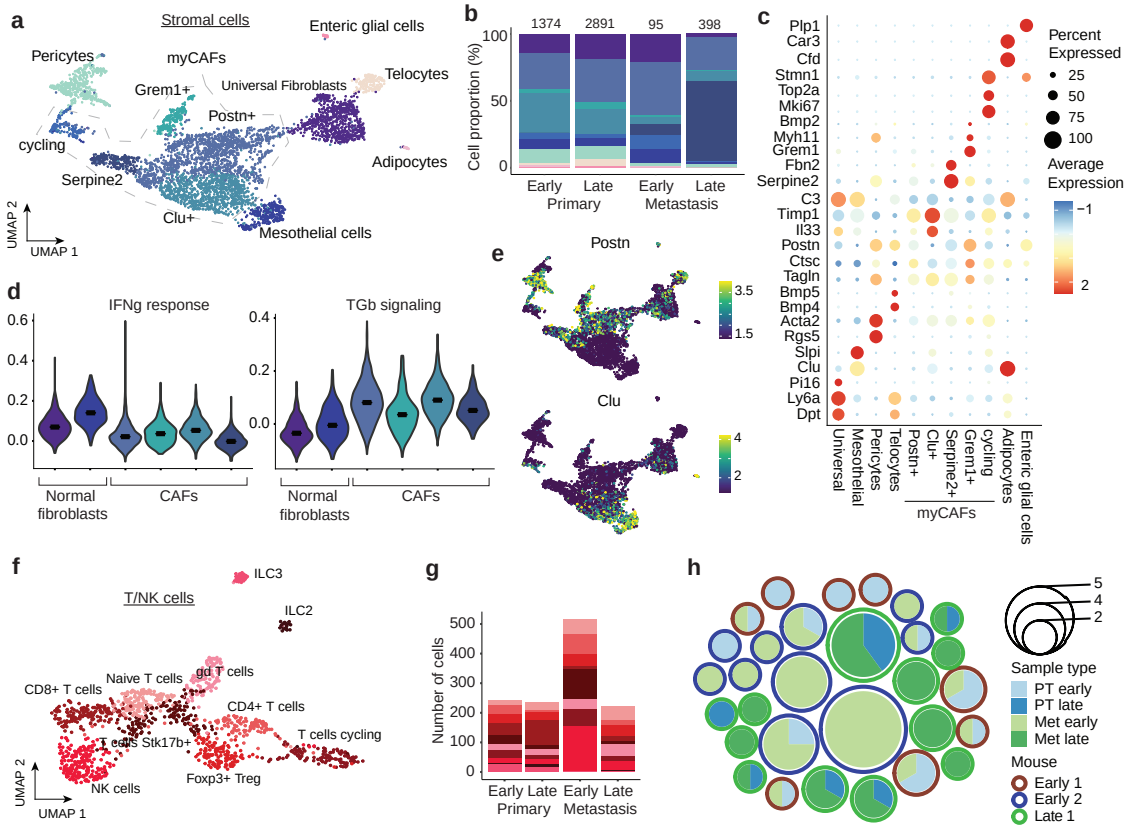

**C2-Figure 3 | Myeloid cell evolution during metastatic progression. a**, UMAPs depicting monocytes and macrophages colored by Seurat clusters. **b**, Barplot showing Seurat cluster composition in Monocytes and Macrophages by sample type. Bars in each sample type shows the average proportion between 2 biological independent samples, numbers on top of bars show the added absolute number of monocytes and macrophages. **c**, Dotplot showing marker genes across 8 myeloid clusters from (a). Dot size indicates fraction of expressing cells, colored according to z-score normalized expression levels. **d**, UMAP plots separated by sample type showing expression levels of the poor-prognosis signature (Cañellas-Socias et al., *in revision*) in monocytes and macrophages. **e**, UMAPs depicting dendritic cells colored by Seurat clusters. **f**, Barplot showing total cell numbers and percentage composition of Seurat clusters in dendritic cells by sample type. Bars in each sample type shows the average proportion between 2 biological independent samples. **g**, Dotplot showing marker genes across 4 dendritic cell clusters from (e). Dot size indicates fraction of expressing cells, colored according to z-score normalized expression levels. **h**, UMAPs depicting granulocytes

colored by Seurat clusters. **i**, Barplot showing total cell numbers and percentage composition of Seurat clusters in granulocytes cells by sample type. Bars in each sample type shows the average proportion between 2 biological independent samples. **j**, Dotplot showing marker genes across 5 granulocyte cell clusters from (h). Dot size indicates fraction of expressing cells, colored according to z-score normalized expression levels.

present in late-stage metastatic disease (C2-Figure 3e-j). Although the total amount of DCs and neutrophils varied, the distribution of immature and mature phenotypes was comparable between early and late tumors (C2-Figure 3f,i). Mast cells expressed a unique set of genes, such as Gata2 and Ms4a2, and were present in very low abundance (C2- Figure 2f).

UMAP representation of stromal cells revealed the presence of various cell clusters that resembled normal differentiated fibroblasts and cancer-associated myofibroblasts (C2-Figure 4a). Primary tumors contained pericytes, characterized by the expression of Rgs5 and contractile genes such as Acta2, Postn and Tagln (McCarthy et al., 2020b); universal fibroblasts, expressing Dpt, Ly6a and Pi16 (Buechler et al., 2021); telocytes, expressing high levels of Bmp4 and Bmp5 (McCarthy et al., 2020b); and mesothelial cells, which expressed high levels of Clu, Wt1 and Slpi (Fawkner-Corbett et al., 2021) (C2- Figure 4c). We identified 4 subtypes of CAFs, characterized by high levels of contractile genes but no Rgs5 expression (Lee et al., 2020)(C2-Figure 4a-c). Overall, CAFs were highly abundant in primary tumors and scarcely present in liver metastases (C2-Figure 4b). Different CAF subsets showed conserved expression programs, such as upregulation of TGFβ signaling and downregulation of IFN signaling compared to normal fibroblasts (C2-Figure 4d). Nonetheless, in contrast to TAMs, CAF subclusters were specialized for each anatomical location: while primary tumors contained three subtypes of CAFs, liver metastases contained almost exclusively Serpine2+ CAFs (C2-Figure 4b). A mutually exclusive pattern of expression of Postn and Clu distinguished universal fibroblasts and mesothelial cells, and, interestingly, it was preserved in two subsets of CAFs in primary tumors (C2-Figure 4e). This might suggest that different CAF subtypes emerge from distinct precursor cells. Two additional stromal cell clusters present at low abundance expressed high levels of Cfd and Car3, two markers of adipocytes; and Plp1, a marker of enteric glial cells (Lee et al., 2020) (Figure 4-c).

Finally, we dissected B and T lymphoid cells present in CRC primary tumors and liver metastases. Early metastases samples contained the highest proportion of B cells and T cells, in agreement with our previous observations (C2-Figure 2c,e). T-cell subclusters represented diverse subpopulations of CD4+, CD8+, γδ T cells and NK cells, as well as innate lymphoid cells (C2-Figure 4f). We then analyzed the immune profile of T cells, recovering TCR sequences for a total of 473 cells. Although we only captured TCR data for



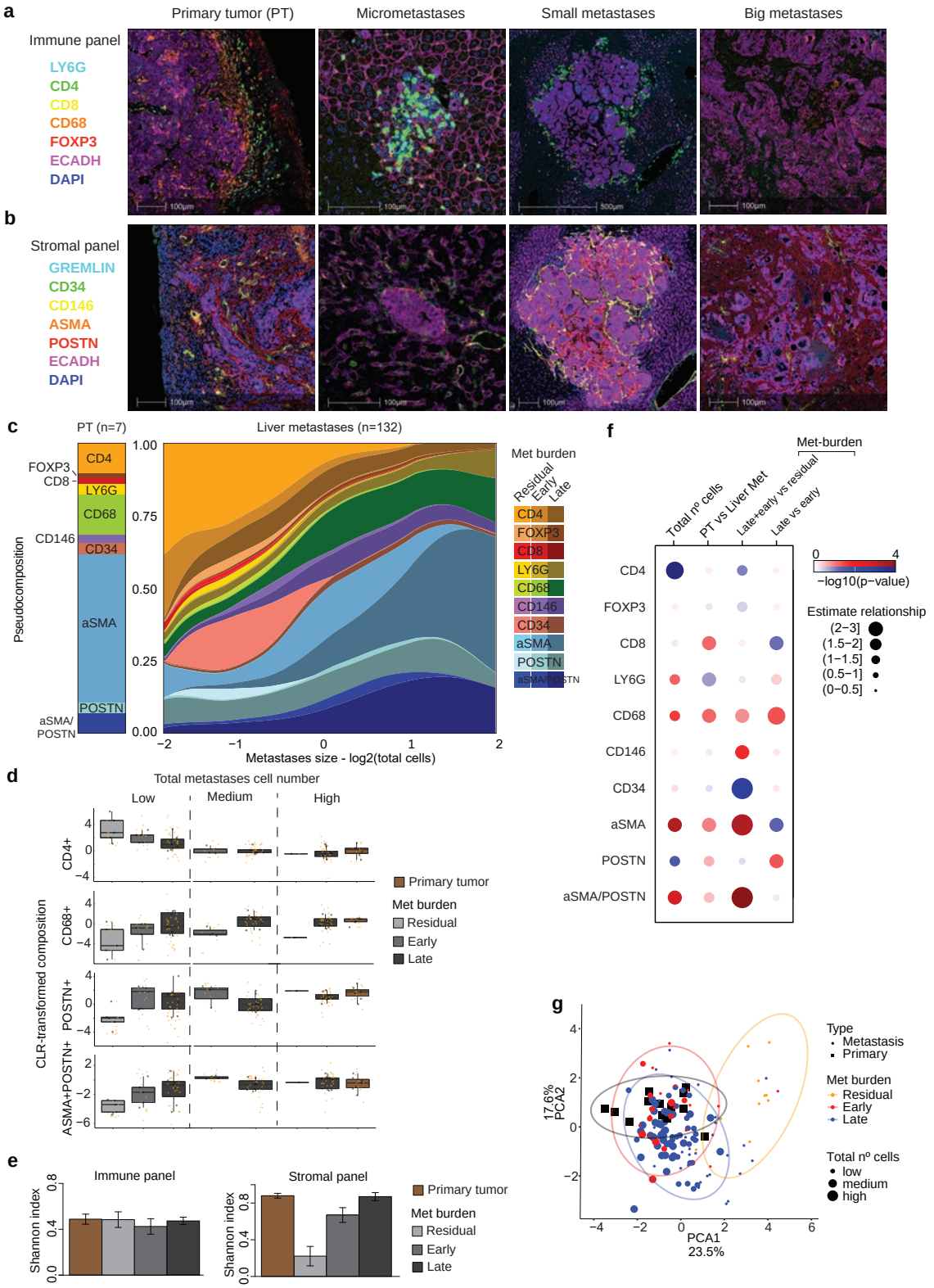

**C2-Figure 4 | Characterization of stromal cells and T/NK cells. a**, UMAPs depicting stromal cells colored by Seurat clusters. **b**, Barplot showing Seurat cluster composition in Stromal cells by sample type. Bars in each sample type shows the average proportion between 2 biological independent samples, numbers on top of bars show the added absolute number of stromal cells. **c**, Dotplot showing marker genes across 11 stromal clusters from (a). Dot size indicates fraction of expressing cells, colored according to z-score normalized expression levels. **d,** Violin plots showing expression of IFNg and TGFb gene signatures in 6 different Seurat clusters from (a). **e**, UMAPs showing expression of genes used to identify universal fibroblasts (Postn) and Mesothelial cells (Clu). Note that a subset of CAFs retains expression of Postn, whereas another expresses Clu. **f**, UMAPs depicting T cells and NK cells colored by Seurat clusters. **g**, Barplot showing Seurat cluster composition in T/NK cells by sample type. Bars in each sample type shows the average proportion between 2 biological independent samples, numbers on top of bars show the added absolute number of T/NK cells. **h**, Bubble plot showing expansion size (circle diameter), corresponding sample type (fill color) and mouse (outer circle color) in expanded TCRs.

35% of T cells, we could characterize 426 different clonotypes (cells having an identical α-chain and β-chain CDR3 nucleotide sequences), from which we detected expanded clones (>2 cells sharing identical clonotype) that represented around 20% of total. All expanded clones were private (mouse-specific), but expanded TCR clones were shared between primary tumors and liver metastases, suggesting that cell immunity against the primary CRC can also attack disseminated cancer cells (C2-Figure 4h).

## Residual micrometastases are a unique state of particular vulnerability

Although scRNAseq revealed consistent differences along metastases progression, tumor sampling was based on microdissection and therefore early-stage metastases were bigger than 1 mm. To analyze non-visible, residual micrometastases we tested a novel method developed by the Malanchi lab, in which a fluorescent protein (slp-mCherry) expressed by tumor cells diffuses through the cell membrane thus labelling neighboring cells (Ombrato et al., 2019). Unfortunately, the expression of this construct was suboptimal *in vivo*, probably owing to immune rejection of slp-mCherry derived neoantigens in immunocompetent mice (not shown). We thus leveraged multiplex immunofluorescence, a method that allowed us to quantify multiple stromal populations in primary tumors and livers metastases at multiple stages (C2-Figure 5). To obtain metastases at various stages, we implanted mice with primary tumors and sacrificed at several timepoints post-injection. Then, we measured bioluminescence *ex vivo* and counted the number of visible metastases for each liver, which allowed us to classify them based on metastatic burden. Residual disease livers looked apparently healthy but contained a small number of micro-lesions identified a posteriori through histology. Livers with early disease contained visible metastases that ranged from 0.5 to 1.5 mm in size, and late disease animals contained overt metastases bigger than 4 mm. Importantly, late disease livers contained a combination of large, small and micro-metastases, possibly owing to differential timing of dissemination. We then used an immune panel to stain for monocytes and macrophages (CD68), T helper cells (CD4), cytotoxic T cells (CD8), regulatory T cells (FOXP3), neutrophils (LY6G) and tumor cells (ECADHERIN); and a stromal panel to stain hematopoietic stem cells/vascular endothelial progenitors/mesenchymal stem cells (CD34), endothelial cells (CD146), different subsets of fibroblasts (POSTN and αSMA) and tumor cells (ECADHERIN). Livers and primary tumors were then scanned and we annotated 143 metastases and 7 primary tumors from 23 different mice.

Inspection of multiplex immunofluorescence revealed that primary tumors contained abundant αSMA+ fibroblasts intermingled with ECADH+ tumor glands (C2-Figure 5b). We also found invasive tumor buds close to CD146+ endothelial cells. At the tumor periphery, we also found many CD68+ macrophages that were located in between epithelial cells and CD4+ T cells (C2-Figure 5a). Micrometastases were heavily infiltrated with CD4+ T cells, from which only a minority expressed FOXP3, and contained very few stromal cells. In small metastases, CD4+ and CD8+ T cells were excluded to the periphery, and LY6G+, CD146+, αSMA+ and POSTN+ cells were intermingled with tumor cell glands. Bigger metastases contained even fewer T cells, while they accumulated a diverse set of stromal and immune cells.




**C2-Figure 5 | Evolution of the TME during metastatic dissemination. a**, Microphotographs of representative examples of multiplex immunofluorescence of immune cell markers in primary tumors and metastases at three different stages. Scale bars, 100 µm (PT, micro and big) 500 µm (small metastases). **b**, Microphotographs of representative examples of multiplex immunofluorescence of stromal cell markers in primary tumors and metastases at three different stages. Scale bars, 100 µm. **c**, Barplot showing the average populations percentages in primary tumors (left). Evolution plot showing the average population per-

centages over an increase in total cells distinguishing between residual, early and late metastatic burden (right). **d**, Boxplots showing CLR-transformed composition values in primary tumors and metastases divided in 3 groups according to their total cell number and further separated by metastatic burden. **e**, Barplot showing the average Shannon index diversity score of the immune and stromal populations in primary tumors and metastases from livers with different metastatic burden. **f**, Dotplot summarizing the results obtained from regression models. Effects of the total number of cells, tumor type (primary vs liver metastases) and metastases burden (late + early vs residual, and late vs early) on the composition of every cell population are represented by different point sizes (defining the magnitude of the effect) and colors (showing both the sign of the effect in blue(-)/red(+) and the statistical significance by color intensity). **g**, Principal components analysis (PCA) of primary tumors and metastases. Point size represent total number of cells, colors represent metastatic burden and shape differentiate between primary tumors and metastases. Ellipses are determined assuming a bivariate t-distribution on the PC scores, 95th percentile.

To analyze how the TME evolved during metastasis progression, we quantified the total number of each cell-type per tumor and plotted their relative frequencies in an evolution graph, ordering metastases by increasing size (C2-Figure 5c). Cell types are differently colored and attributed three different tonalities depending on mouse tumor burden: residual stage, early-stage or late-stage. In agreement with our previous observations, T lymphocytes were heavily enriched in micrometastases and their proportion decreased with metastases size. Most stromal cells — CD146+, αSMA+, POSTN+ and αSMA+POSTN+ — were largely absent in micrometastases, but were progressively gained during metastatic outgrowth, in a marked antithetic pattern to CD34+ cells. Of note, we observed that micrometastases from livers with high metastatic burden were not as infiltrated with T cells as those with residual disease (C2- Figure 5d). In addition, cell-types associated with disease-progression, such as CD68+, POSTN+ and αSMA+POSTN+ cells, were proportionally lower in animals with residual metastatic disease when comparing metastases of similar size (C2- Figure 5d). In summary, the immune excluded landscape of primary tumors was lost during the initial phases of liver colonization, but was progressively re-established as metastases expanded in size. In fact, stromal subpopulations scored low diversity in residual metastases compared to primary tumors and overt metastases, suggesting a premature (i.e. yet uncorrupted) state (C2- Figure 5e).

We next performed a multivariate analysis to statistically weight the effect of metastasis size, overall tumor burden and tumor location on cell type proportions in the TME (C2-Figure 5f). Results regarding tumor location were in agreement with scRNAseq data: neutrophils were significantly increased in liver metastases, while CD8+ T cells, CD68+ macrophages and αSMA+ fibroblasts were enriched in primary tumors. Interestingly, virtually all cell-types, in exception to FOXP3+ Tregs, showed significant compositional

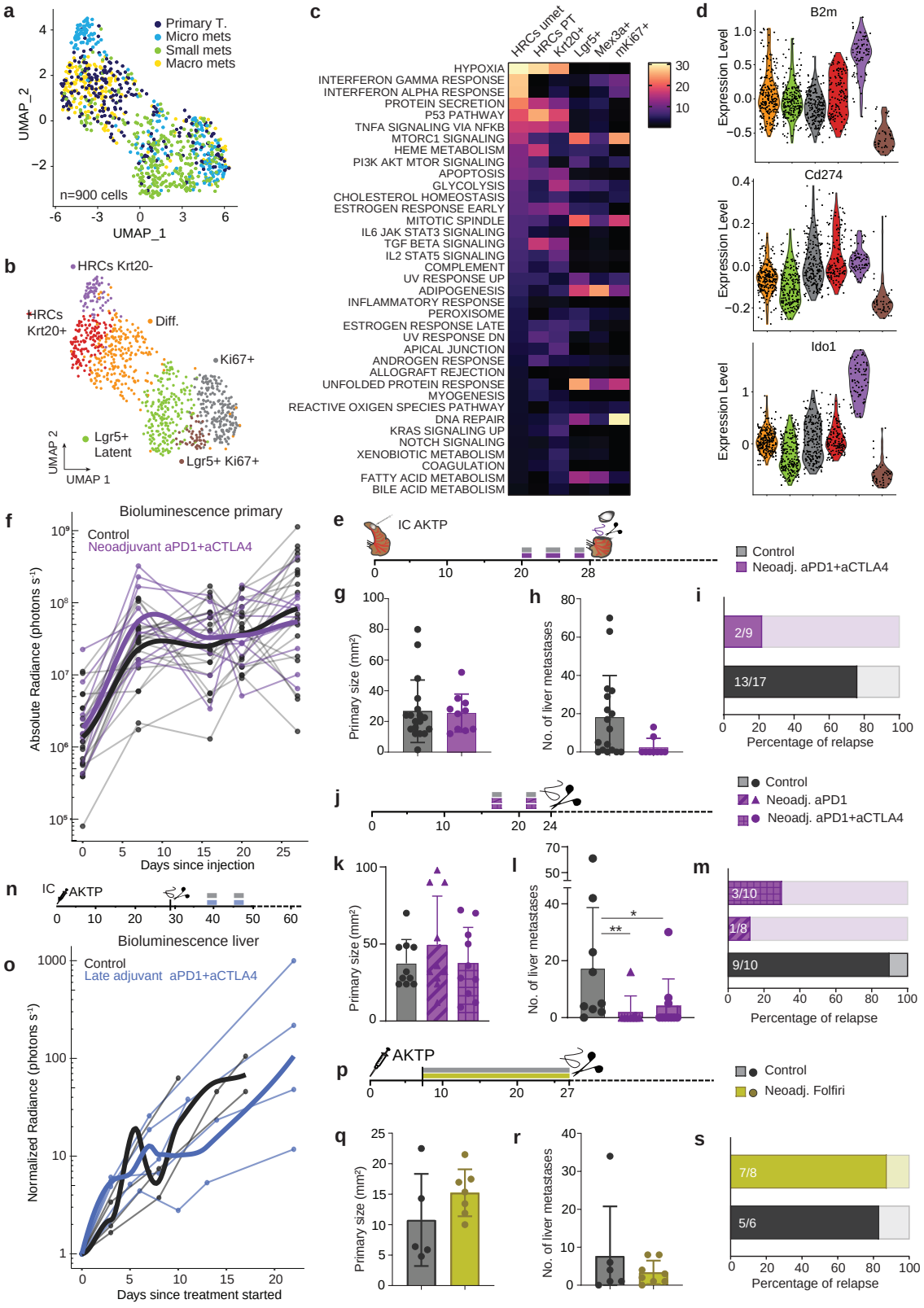
changes depending on metastases size and overall tumor burden. Both variables affected cell types, such as CD4+ T cells and CD68+ macrophages, in a similar manner, but showed a complex behavior in fibroblast subpopulations. Notably, CD4+ T cells were mostly affected by metastases size, whereas CD8+ T cells markedly decreased in mice with late-stage metastatic burden. Further supporting a combined effect of tumor burden and size, principal component analysis (PCA) evidenced that only metastases of small size in residual-stage livers clustered apart from the rest of samples (C2- Figure 5g). Collectively, our data support that residual micrometastases are a unique state of particular vulnerability to adaptive immunity.

## Neoadjuvant immunotherapy prevents metastatic relapse

Our data suggests that the ability to evade immunity fluctuates during metastatic dissemination, when immunosuppressive cell types of the TME need to be re-educated by tumor cells. We hypothesized that tumor cells that disseminate away from the immune-privileged environment of the primary CRC, may exploit additional strategies to evade the immune system while they recreate a protective TGF-β-driven TME in foreign organs. We have previously discovered the residual population of undifferentiated HRCs that originate metastatic relapse in CRC (Cañellas-Socias et al., *in revision*). To investigate the mechanism by which these tumor cells might escape immunosurveillance, we mined our single cell dataset on tumor cell evolution during metastases progression from an immune perspective (C2-Figure 6a-b). Gene set enrichment analysis on tumor cell clusters revealed that HRCs landing on the liver expressed high levels of Hypoxia, IFNγ and IFNα gene signatures, P53 pathway, TNFα, among others (C2-Figure 6c). Interestingly, IFNγ and IFNα signatures were uniquely enriched in HRCs from liver-micrometastases, but not in HRCs in primary tumors or big metastases. This suggests that the expression of these inflammatory signatures does not stem from tumor cell differentiation programs but rather from changes in the environment. We reasoned that HRCs in liver micrometastases may express high levels of IFN signature genes in response to IFN molecules secreted by large numbers of surrounding T cells. As part of the IFN program, HRCs show increased expression of antigen-presenting molecules such as B2m, yet, they also upregulate immunosuppressive molecules such as Cd274 and Ido1 (C2-Figure 6d). This suggests that early metastatic lesions are heavily attacked during the initial phases of colonization and upregulate checkpoint molecules to survive in the absence of a protective TME.

We have previously demonstrated that checkpoint blockade is ineffective in overt mCRC due to high levels of TGFβ in the TME (Tauriello et al., 2018). Yet, we hypothesized that at the onset of metastasis, low tumor cell numbers and an immature TME, would

render tumors vulnerable to ICB. To test so, we treated primary tumors with neoadjuvant anti-PD1 + anti-CTLA4 before surgical resection (C2-Figure 6e). The timing of ICB




**C2-Figure 6 | Neoadjuvant immunotherapy prevents metastatic recurrence in MSS CRC. a-b**, UMAP layout of 900 tumor cells colored by metastatic stage (a) and Seurat clusters (b). **c,** Heatmap showing Hallmark gene set enrichment analysis (GSEA) in tumor

cell clusters in (b). **d**, Violin plots showing expression of B2m, Cd274 and Ido1 in the 6 different Seurat clusters. **e**, Schematics of the experiment for neoadjuvant treatment with aPD1 and aCTLA-4 of primary tumors before resection. **f**, Bioluminescence monitoring of the effect neoadjuvant immunotherapy on primary tumor growth. Points and lines represent individual mice, trend lines (bold) show a LOESS model. **g**, Primary tumor area (mean ± SD) measured after resection. **h**, Liver metastases (mean ± SD) generated by AKTP primary tumors treated with vehicle or combination immunotherapy, up to one month after primary tumor resection. Each dot is an individual mice. **i**, Percentage of mice that developed liver metastases or remained metastases-free in control and immunotherapy-treated tumors. **j**, Schematics of the experiment for comparison of anti-PD1 monotherapy with anti-PD1+anti-CTLA4 combination neoadjuvant treatment. **k**, Primary tumor area (mean ± SD) measured after resection in the experiment described in (j). **l**, Liver metastases (mean ± SD) generated by AKTP primary tumors up to one month after primary tumor resection. Each dot is an individual mice. **m**, Percentage of mice that developed liver metastases or remained metastases-free in control and immunotherapy-treated tumors. **n**, Schematic of an experiment for late treatment with adjuvant immunotherapy. Untreated AKTP tumors were left to grow and resected 4 weeks after injection, 10 days later mice received two shots of anti-PD1 + anti-CTLA4. **o**, Normalized bioluminescence (BLI) of liver metastases to the day ICB treatment started. Points and lines represent individual mice, trend lines (bold) show a LOESS model. **p**, Schematics of the experiment testing therapeutic efficacy of neoadjuvant chemotherapy prior to primary tumor resection. **q**, Primary tumor area (mean ± SD) measured after resection. **r**, Liver metastases (mean ± SD) generated by AKTP primary tumors treated with vehicle or FOLFIRI chemotherapy, up to one month after primary tumor resection. Each dot is an individual mice. **s**, Percentage of mice that developed liver metastases or remained metastases-free in control and chemotherapy-treated tumors.

treatment was based on bioluminescence imaging of livers across different timepoints post-implantation, which suggested that metastases were still small in size around day 15-20. Primary tumor size and bioluminescence were unaltered by 3 shots of neoadjuvant immunotherapy (C2-Figure 6f-g). However, we observed that after resection most immunotherapy-treated mice did not show signs of metastatic recurrence, whereas the majority control mice were sacrificed due to extensive metastases (C2-Figure 6h-i). Counting of liver metastases revealed a 10-fold downregulation in metastasis formation in treated versus control mice. Importantly, we obtained similar results with anti-PD1 monotherapy, suggesting that combination with anti-CTLA-4 is not required for the curative effect (C2-Figure 6j-m).

Failed immunotherapy clinical trials in MSS CRC patients led to the conclusion that ICB is not a therapeutic option for these patients due to low neoantigen numbers. However, all such trials were performed in late-stage mCRC patients. In fact, when we mimicked the settings of failed human trials by treating metastases post-resection and at late-stages, we also observed a lack of response (C2-Figure 6n-o). Moreover, the current standard of care for stage III CRC patients – treatment with FOLFIRI chemotherapy – was ineffec-

tive in preventing relapse in our mouse model (C2-Figure 6p-s). Our data suggests that the timing of ICB treatment led to such disappointing results and strongly advocates for testing neoadjuvant ICB to prevent relapse in stage II and III CRC patients.

# Chapter 3 - Niche dependencies drive metastatic latency in colorectal cancer

## Results

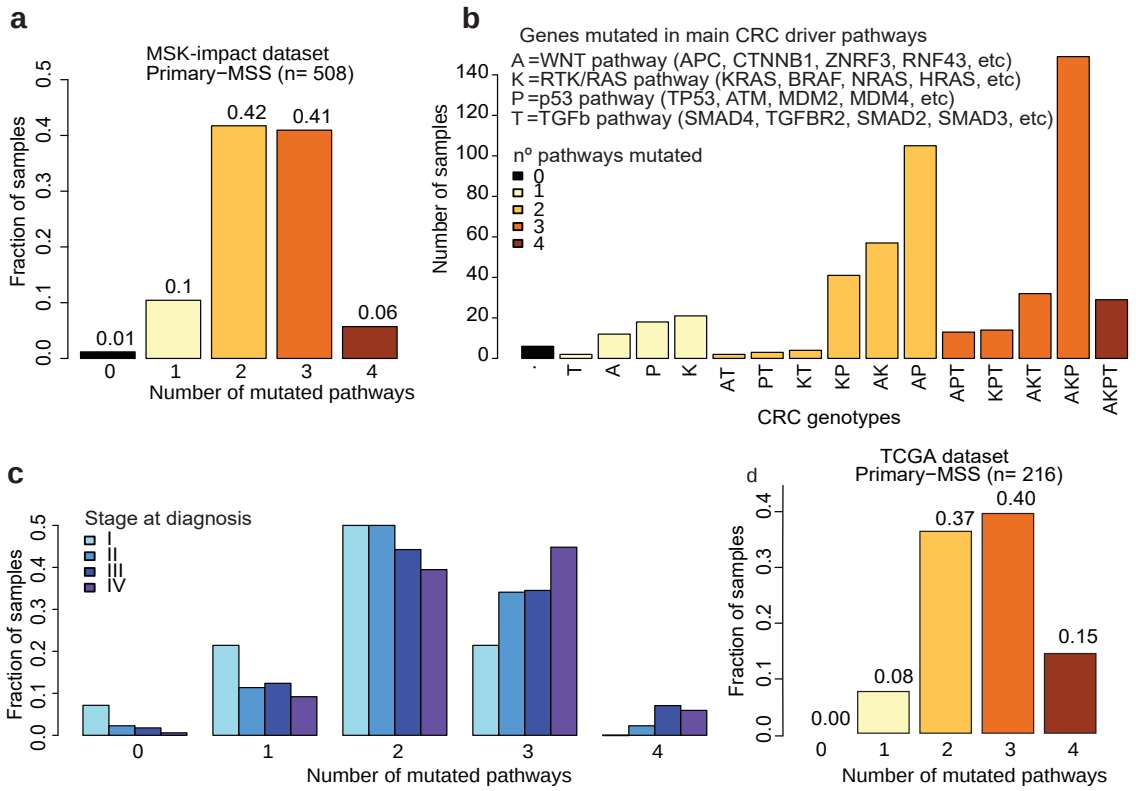
## Incomplete genetic progression leads to niche-dependence

Colorectal cancer has been classically viewed as a multistep mutational disease, where multiple mutations or 'hits' are required for the transformation of normal colon epithelium into invasive and metastatic colorectal cancer (mCRC). Large-scale sequencing studies have validated early observations indicating that the most common genetic alterations during CRC progression occur in signaling pathways that regulate ISC expansion. Mutations in WNT, EGFR, P53 and BMP/TGFβ pathways render tumor stem cells independent of the crypt niche, thereby facilitating growth in foreign environments environments (Batlle & Clevers, 2017; Fujii et al., 2016; Fumagalli et al., 2017). Supporting this view, we and others showed that colon organoids engineered with four driver mutations (Apc, Kras, P53 and Tgfbr2 or Smad4; AKPT/AKPS) can be cultured in the absence of stem cell factors and, upon inoculation in mice, they efficiently generate metastatic outgrowths in the liver and lungs (Fumagalli et al., 2017; Matano et al., 2015; O'Rourke et al., 2017; Tauriello et al., 2018).

However, we discovered that the vast majority of human primary mCRCs and metastases carry mutations in only two or three driver pathways (Figure R3.1 a, d). We downloaded data from 1,134 colorectal adenocarcinomas analyzed with MSK-IMPACT, a capture-based next-generation sequencing platform that detects mutations, copy-number alterations, and rearrangements in 341 cancer genes. The CRC cohort consisted of 1,011 tumors (478 primaries and 533 metastases) from 979 patients with mCRC and 123 tumors from 120 patients with early-stage CRC (Yaeger et al., 2018). We scored the mutational status in 4 canonical signaling pathways in CRC by assessing alterations in curated oncogenic pathway genes (Sanchez-Vega et al., 2018). For this analysis, we made the assumption that mutations in different components of the same pathway would have similar consequences in terms of niche dependencies. For example, mutations in either KRas or BRaf render CRC cells independent of EGF signals (Fujii et al., 2016). Remarkably, we found that dual and triple mutant CRC genotypes are the most frequent in CRC patients, whereas quadruple mutants are rare (C3-Figure 1a).

The number of mutated pathways related partially to the stage at diagnosis, and AKP was the most common genotype across CRC samples (C3-Figure 1b-c). We performed the same exercise with an independent cohort of exome-sequencing data from TCGA CRCs and obtained similar results (C3-Figure 1d) (Muzny et al., 2012). Our data is in agreement with patient-derived organoid (PDO) biobanks, where a large proportion of CRCs exhibited partial dependence on niche signals and required supplementation with complete stem cell media for expansion (Fujii et al., 2016; van de Wetering et al., 2015).

These analyses indicate that in the clinic, most mCRCs accumulate mutations in 2-3

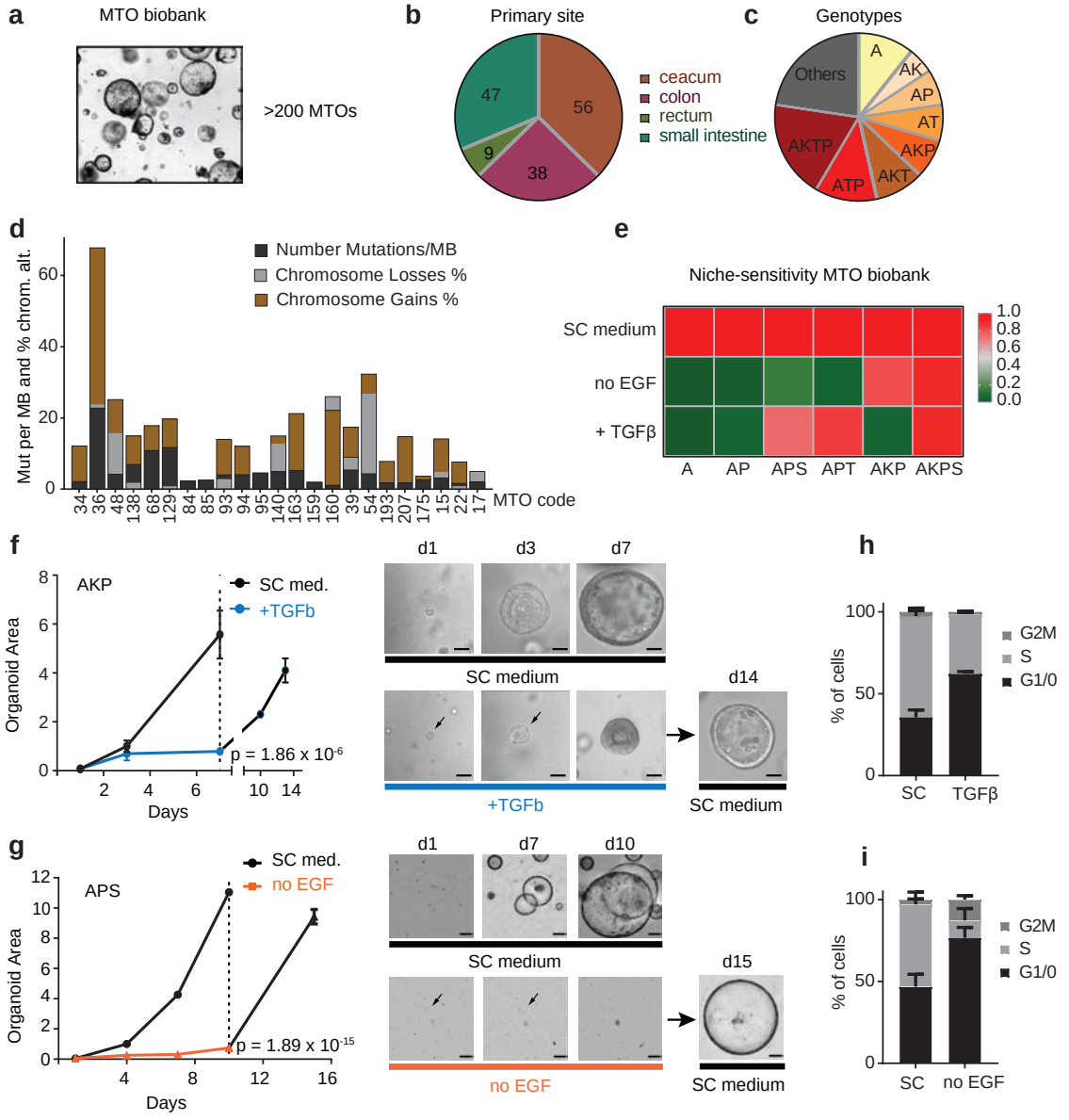



**C3-Figure 1 | Mutated driver pathways frequencies in CRC patients.** The mutational status of the 4 major driver signaling pathways in CRC (WNT, RAS, TP53 and TGFß) was scored by assessing alterations in components of oncogenic pathways according to the curated list signaling pathway component lists described in (Sanchez-Vega et al., 2018). We analyzed primary microsatellite stable primary CRCs and metastases of the MSK-impact dataset (Yaeger et al., 2018). **a**, Barplot showing the proportion of CRC primary tumor samples with the indicated combination of mutated pathways (n=508). **b**, Barplot showing the number of primary tumor metastatic samples with the indicated combination of mutated pathways (n=533). **c**, Frequency of primary CRC samples according to the number of mutated main driver pathways stratified by AJCC clinical staging at the time of diagnosis. d, Same analysis as in (a) but for TCGA dataset of primary-MSS CRCs (Muzny et al., 2012).

pathways, yet most laboratories, including ours, have thoroughly studied the ability of quadruple mutant tumors to metastasize. Thus, we decided to study the biology of triple mutant tumors, which remained overlooked. We had previously established a large collection (n> 200) of MTOs derived from primary CRCs and metastases arising in compound GEMMs (C3-Figure 2a). MTOs are 3D in vitro surrogates of the disease that, upon engraftment into syngeneic C57BL/6 mice, generate CRCs that mimic those developed by the GEMM of origin (Tauriello et al., 2018). Our MTO biobank comprises tumors arising from different intestinal sites that contain diverse combinations of mutant pathways and genetic alterations (C3-Figure 2b-d). We selected MTOs with three genetic alterations in driver pathways by subjecting them to niche-factor sensitivity experiments (C3-Figure 2e).

## Metastatic latency by niche-dependent triple mutant tumors

As expected, quadruple mutant tumors grew exponentially in the absence of stem cell niche factors. Supplementation with recombinant TGFβ arrested the growth of AKP MTOs cultures (C3-Figure 2f). Likewise, MTOs bearing wild-type KRAS alleles, i.e. APS (Apc -/- ; P53 -/-; Smad4-/-) or APT (Apc -/- ; P53 -/- ; Tgfbr2 -/-), did not expand in the absence of EGF (C3-Figure 2g). Interestingly, we found that AKP and APS MTOs remained viable but arrested in G1/G0 for more than a week in these suboptimal cul-




**C3-Figure 2 | In vitro modelling of latency with niche-dependent tumors**. **a**, Image of MTOs. **b-d**, Primary organ site (b), genotypes (c) and frequency of mutations and chromosomal alterations (d) in MTOs from our biobank. **e**, Heatmap summarizing the effect on organoid growth under suboptimal niche conditions (absence of EGF or presence of TGFβ1) vs growth in SC complete medium (with EGF and no TGFβ1) in MTOs with the indicated genotypes. **f-g**, Quantification of organoid growth (mean area ± s.e.m. relative to day

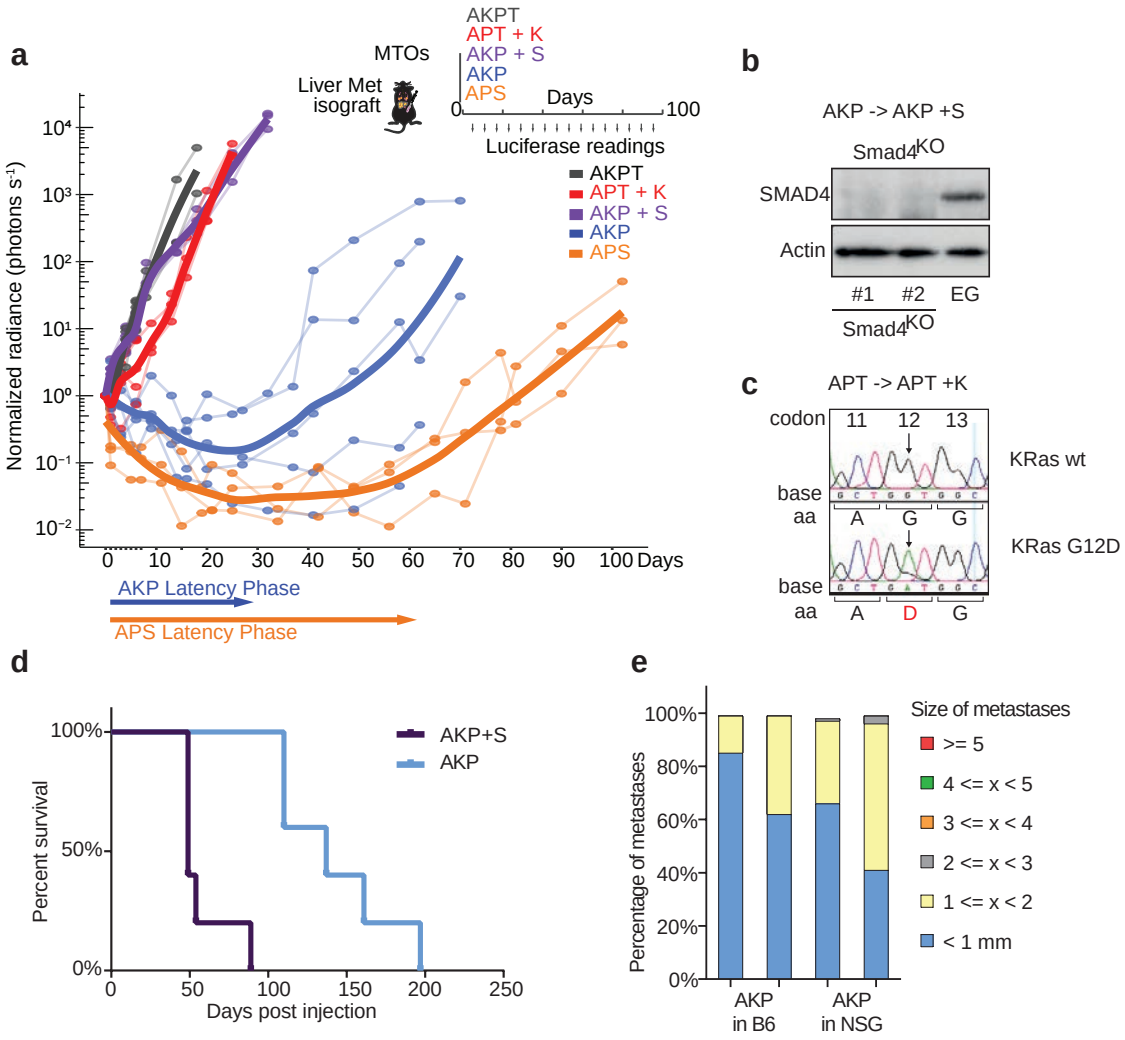
3) of (**f**) AKP MTOs (Apcmut, K-rasG12D, p53mut) in the presence of TGF-β, and (**g**), APS MTOs (Apcmut, Smad4mut, p53mut) in absence of EGF compared to complete stem cell medium (SC med.). Dashed lines indicate the day TGFβ was removed or EGF was re-added, respectively. n=4 independent wells per condition, linear model on log-transformed values. Scale bars: 50 µm. **h-i**, Barplots showing cell-cycle status in (h) AKP MTOs in the presence of control medium or TGF-β, and in (i) APS MTOs in absence of EGF compared to complete stem cell medium (SC med.).

ture conditions (C3-Figure 2f-i). AKP MTOs resumed expansion after removal of TGFβ, whereas supplementation of EGF after 10 days restored the growth rates of APS organoids (C3-Figure 2f-i).

The latent phenotype that we observed led us to hypothesize that an incomplete set of driver mutations might lead to delayed metastatic outgrowth in foreign environments. Cancer cells in primary tumors receive supporting niche signals from the surrounding TME (Medema & Vermeulen, 2011). However, during metastatic dissemination, tumor cells travel alone (or in small clusters) to distant organs, where they need to rebuild a favorable ecosystem to thrive. We hypothesized that tumors with incomplete niche independence enter a latent state until tumor-promoting signals are developed.

To test our hypothesis we investigated the ability of triple mutant tumors (AKP and APS/APT) to colonize the liver, the most common site of CRC metastases. In our previous study, we showed that intrasplenic inoculation of quadruple mutant (AKPT) MTOs in c57BL/6 mice produced rapidly growing liver metastases that killed the host in only 3-4 weeks (Tauriello et al., 2018). APS and AKP MTOs also colonized the liver, yet bioluminescence analysis showed that they entered a prolonged latency time, which was followed by exponential growth (C3-Figure 3a). The latent period between inoculation and the onset of rapid growth spanned 30-40 days for AKP MTOs and 60-80 days for APS MTOs. This delay is in sharp contrast to quadruple mutant tumors, which started exponential growth immediately after injection.

To formally demonstrate that the slower growth kinetics of 3x MTOs was due to the lack of the 4$^{th}$ driver mutation, we used CRISPR/Cas9 to knock-out Smad4 in AKP MTOs (C3-Figure 3b). We also introduced a knock-in base substitution that would result in a Glycine to Aspartic acid change in KRas codon 12 of APT MTOs (C3-Figure 3c). To select for mutant clones, we modulated culture conditions by adding TGFβ to select for Smad4 mutants and removing EGF to select for KRas G12D mutants (Drost et al., 2015; Matano et al., 2015). Smad4 knockout MTOs were generated through NHEJ by nucleofecting a guide disrupting exon 9. On the other hand, KRasG12D point mutation was obtained through HDR by nucleofecting a guide and a donor plasmid containing a G to A base substitution in codon 12. As expected, the modified MTOs (APT+K and AKP+S) gener-

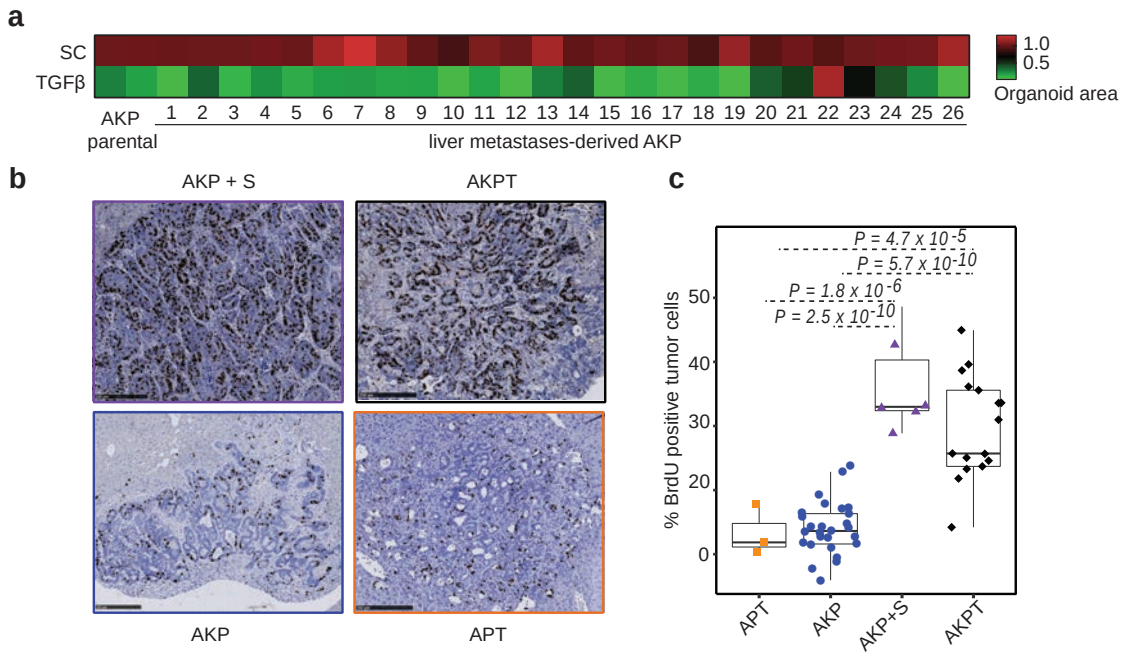



**C3-Figure 3 | Metastatic latency of triple mutant MTOs. a**, Quadruple mutant MTOs (AKPT, ATP+K or AKP+S) and triple mutant MTOs (AKP and APT) were labeled with a luciferase expression vector and inoculated through the spleen into in C57BL/6J mice. Liver metastatic growth was assessed by intravital bioluminescence imaging (BLI). Graph shows longitudinal BLI measurements (photons s−1), normalized to the day of injection. Points and lines represent individual mice, trend lines (bold) were generated using a LOESS model. For AKPT n=2, APT+K, APS and AKP+S n=3 and AKP n=5. **b**, Western blot for SMAD4 and actin of LAKP empty guide (EG)- or CRISPR-targeted (#1 and #2) Smad4KO organoids. **c**, Sanger sequencing results of APT organoids and CRISPR-targeted KRas G12D mutant organoids. Sequencing shows a guanosine to adenosine base substitution that results in a Glycine to Aspartic acid change. **d**, Kaplan–Meier survival curve of immunocompetent mice injected intrasplenic with AKP or AKP+S MTOs. **e**, Diameter composition of metastases generated by AKP organoids injected in immunocompetent (B6) or immunodeficient (NSG) mice.

ated rapidly progressing metastases *in vivo* that led to shorter overall-survival (C3-Figure 3d). To rule out that the latency we observed was immune-mediated (i.e. caused by an immune-equilibrium state followed by immune-escape), we injected triple mutant MTOs into fully immunodeficient NSG mice. We observed that liver metastases size was

comparable between B6 and NSG mice (C3-Figure 3e).

We next assessed the possibility that AKP metastases that expanded after a latency phase had acquired a 4th driver mutation that inactivated the TGFβ pathway. To this end, we



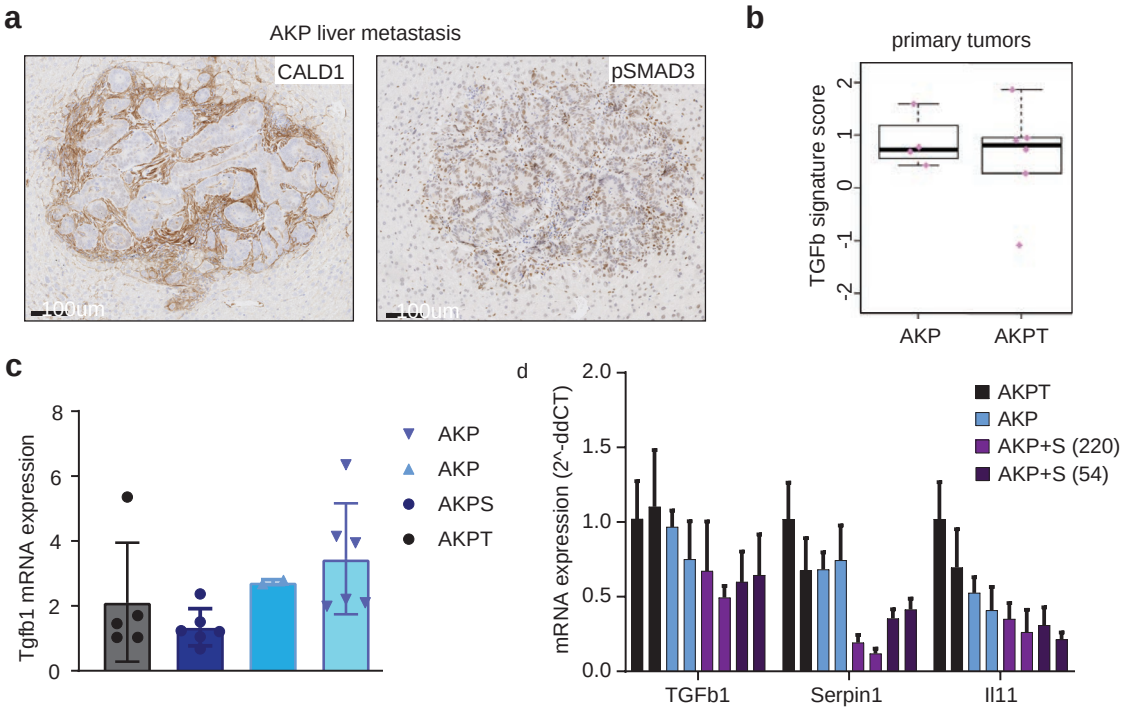

**C3-Figure 4 | Slow proliferation due to TGFβ-sensitivity drives tumor cell latency. a**, Heatmap showing the effect on organoid growth of TGFβ1 vs complete medium in AKP MTOs derived from liver metastases. **b**, Representative images of BrdU immunohistochemistry in liver metastasis generated by triple and quadruple mutant MTO, quantified in **c**, BrdU was injected into animals 2 hours before sacrifice. A mixed effects linear model was fitted to percentages after a boxcox transformation. Scale Bar 250 μm.

derived 26 new MTOs from AKP liver metastases and tested their sensitivity to TGFβ cytostasis. 24 out of 26 metastases-derived AKP organoids retained large sensitivity to TGFβ, pointing that they were wild-type for the the TGFβ pathway. This data suggests that acquisition of the 4th hit was not required for metastases formation (C3-Figure 4a). In agreement with this data, quantification of bromodeoxyuridine (BrdU) incorporation in tumor cells revealed lower rates of proliferation in triple mutant tumors compared to quadruple mutants (C3-Figure 4b-c).

## Regulation of the TME by KRas and TGFβ oncogenic mutations

Our results suggest that metastatic latency in triple mutant tumors owes to cell-autonomous mechanisms. Having said that, it is well-known that tumor cell oncogenes drive aberrant signaling not only in tumor cells but also in adjacent stromal cells. For example, it has been proposed that the ability to build a TGFβ-high immunosuppressive TME is

dependent on the previous loss of tumor cell sensitivity to TGFβ (Massagué, 2008). To investigate such long-lasting hypothesis, we compared TGFβ signaling in liver metastases and subcutaneous tumors generated by AKP, AKPT and AKP+S tumor cells. Using immunohistochemistry, we observed phosphorylation of the TGFβ signaling mediator SMAD3 in stromal cells surrounding AKP liver metastases, as well as high infiltration of TGFβ-activated CALD1+ CAFs (C3-Figure 5a). In agreement with this data, mRNA ex-

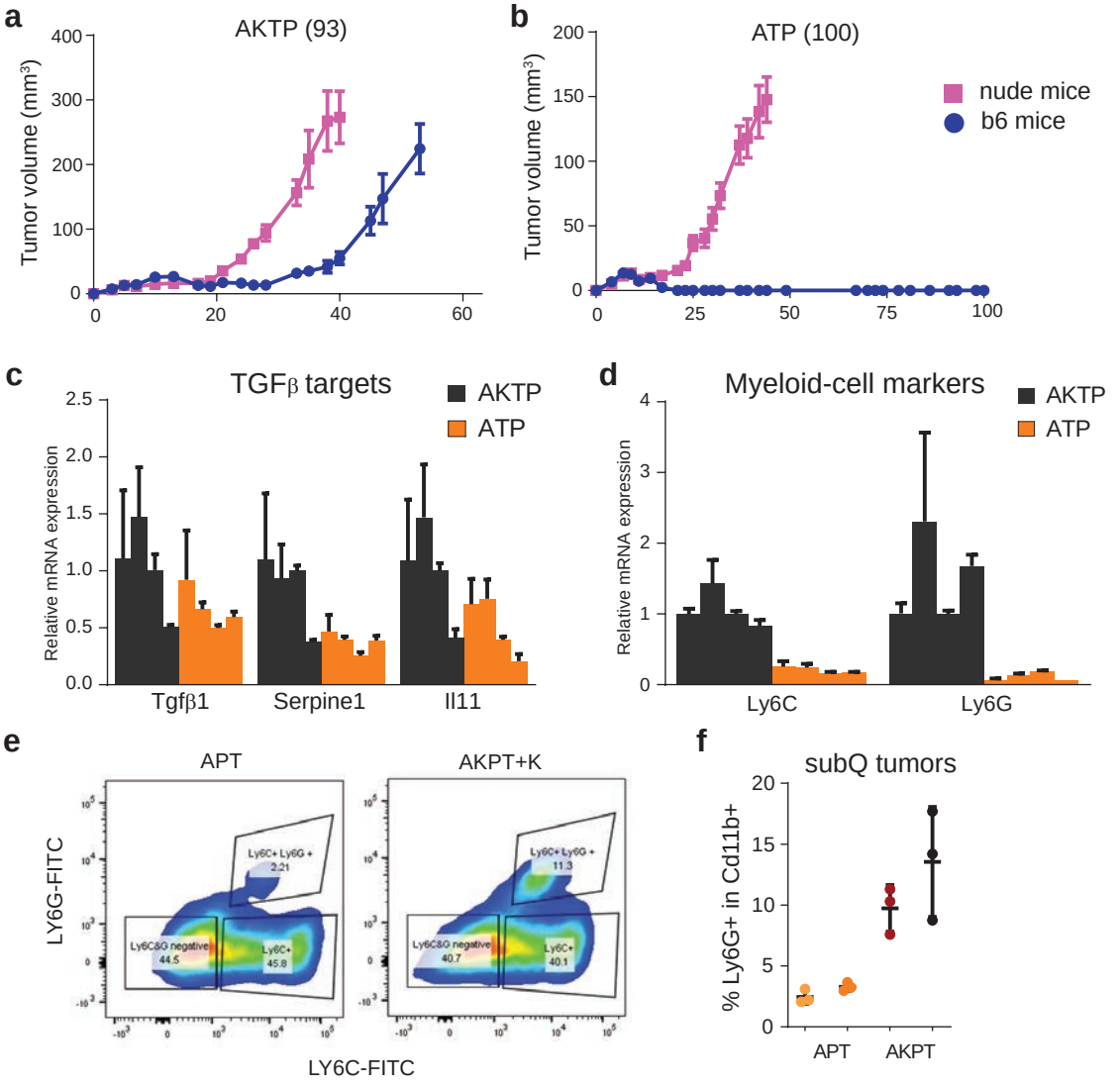



**C3-Figure 5 | TGFβ-sensitive tumors generate TGFβ-rich, poor prognosis-like metastases. a,** CALD1 and pSMAD3 immunohistochemistry of a liver metastases generated by AKP MTOs. Scale bars, 100 µm. **b**, TGFβ gene signature score in AKP and AKTP tumors analyzed through bulk RNA-sequencing. **c**, Relative Tgfβ1 mRNA expression in bulk RNA from liver metastases generated by MTOs with different driver mutations. **d**, Relative mRNA expression in bulk RNA from subcutaneous tumors generated by MTOs with differ-

pression of TGFβ signature and TGFβ1 was comparable in bulk RNA from TGFβR2 wild-type and knock-out primary tumors and liver metastases, respectively (C3-Figure 5b-c). Furthermore, TGFβ1 and the TGFβ target genes (Serpin1, Il11 and Postn) were similar in AKP subcutaneous tumors wild-type and knock-out for Smad4 or Tgfbr2 (C3-Figure 5d). We conclude that TGFβ levels are high in the TME regardless of the mutational status of the TGFβ pathway in tumor cells. Thus, we conclude that slow-growth and latency are a trade-off for TGFβ-sensitive tumors to evade anti-tumor immunity.

On the same lines, several studies demonstrated a role of oncogenic KRas signaling in the orchestration of tumor-promoting interactions in the TME (Canon et al., 2019; Car-

valho et al., 2018; Cheng et al., 2019; Cullis et al., 2018; Zdanov et al., 2016). For example, KRas G12D mutation induces the conversion of T cells into T regulatory cells in pancreatic cancer (Cheng et al., 2019). In fact, APT organoids engrafted poorly in the liver and subcutaneously compared to AKP or AKPT MTOs and this was dependent on the presence of T cells (C3-Figure 6a-b). To explore why APT tumors failed to evade immunesurveilance, we analyzed the expression of immunosuppressive molecules and cell-type marker genes in APT and AKPT subcutaneous tumors grown in nude mice. We found a two-fold decrease in the expression of TGFβ1 and the TGFβ-target genes Serpine1 and Il11 in triple APT mutants compared to quadruple mutant tumors (C3-Figure 6c). In addition, we observed a marked downregulation of Ly6g and Ly6c gene expression in APT tumors (C3-Figure 6d). These genes are known markers of MDSCs and




**C3-Figure 6 | KRas wild-type tumors show a deficiency in neutrophil recruitment. a-b**, Subcutaneous growth of AKTP quadruple-mutant tumors and ATP triple-mutant tumors in immunocompetent (b6) and T-cell deficient (nude) mice. **c-d,** Relative mRNA expression of (c) TGFβ target genes and (d) myeloid-cell markers in AKTP and ATP subcutaneous tumors implanted in T-cell deficient mice. **e-f,** Flow cytometry plots (e) and

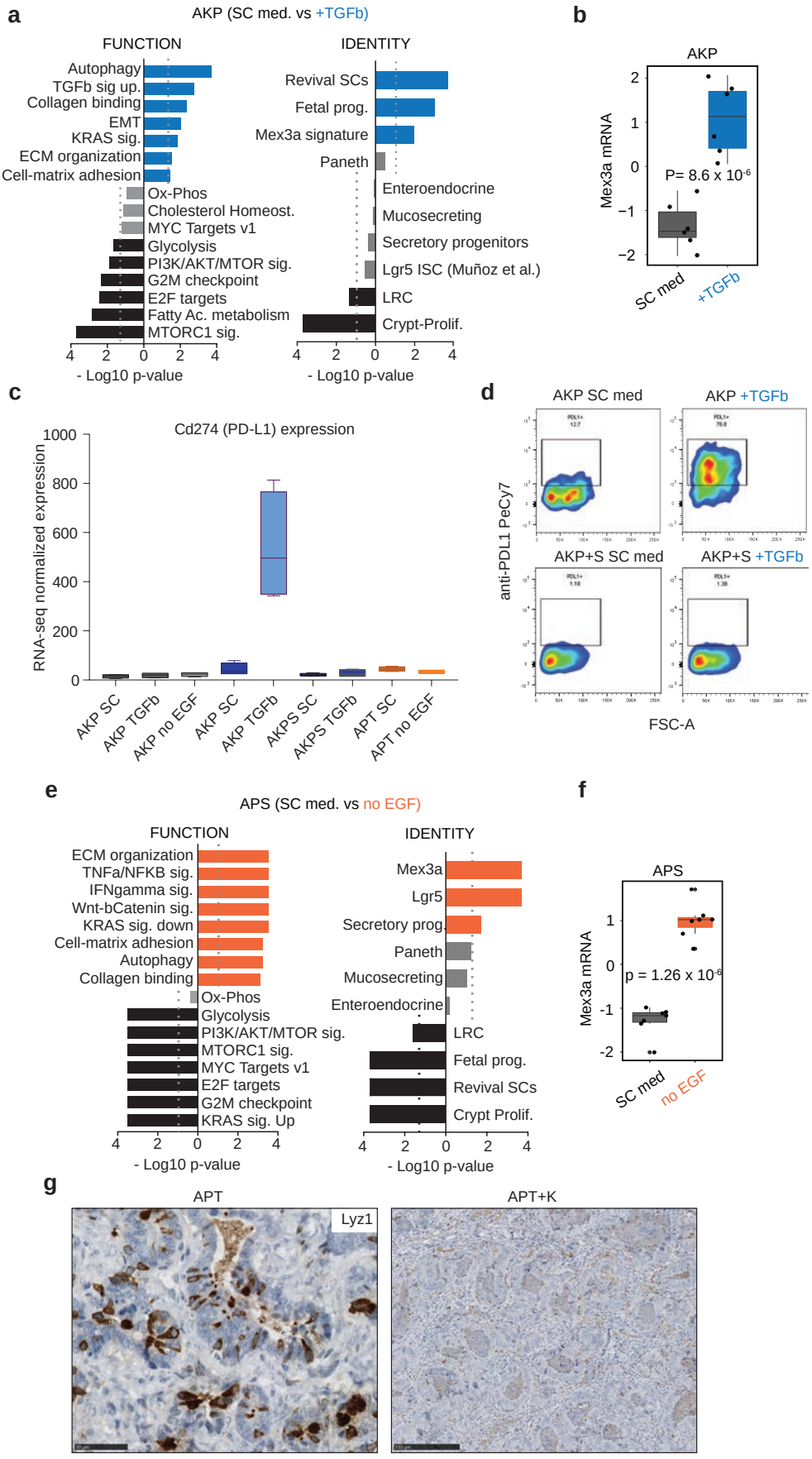
quantification (f, mean±SD) of Ly6c+ Ly6g+ neutrophils in subcutaneous tumors analyzed at day 10, before resection occurs in immunocompetent mice.

neutrophils, key mediators of immunosuppression in the TME. To validate these data, we used flow cytometry to analyze subcutaneous tumors grown in immunocompetent mice at day 10 – right before immune-mediated rejection –.The percentage of LY6C+ LY6G+ double positive cells within the myeloid cell compartment (CD11b+) was lower in ATP compared to AKTP tumors (C3-Figure 6e-f). Shortly after we obtained these results, Ronald DePinho and colleagues published similar observations using an inducible KRas G12D construct in CRC (Liao et al., 2019), thus, we decided not to follow these findings.

## Suboptimal stem cell niche induces a latent Mex3a+ state

Taken together, our data demonstrates that AKP and APS organoids experience a latency phase before they can efficiently colonize the liver. These growth kinetics may explain why a fraction of stage II and III patients, who show metastatic recurrence several months after removal of primary tumors. We hypothesize that late metastatic relapse in patients is caused by niche-dependencies of tumor cells with an incomplete set of driver mutations. Thus, we sought to characterize the latent state of triple mutant MTOs, with the aim to understand better the vulnerabilities of metastatic latency in patients. To gain insights into the latent state of triple mutant tumors, we performed RNA sequencing in suboptimal niche conditions compared to full SC media followed by gene set enrichment analysis (GSEA).

Although growth arrest triggered by the lack of the mitogen EGF and by activation of TGF-β signaling involved substantially different mechanisms and downstream gene programs, we identified some shared features. The latent phenotype in both MTO genotypes was characterized by silencing of proliferation gene programs and general downregulation of metabolic pathways including glycolysis, oxidative phosphorylation and fatty acid metabolism (C3-Figure 7a,e). Extracellular matrix remodeling and sensing genes were also upregulated in both AKP and APS MTOs (C3-Figure 7a,e). Of particular note, we observed that the Mex3a+ cell signature (Barriga et al., 2017) was amongst the common gene sets enriched both in latent AKP and APS/T MTO cultures (C3-Figure 7a,e). Mex3a marks a subset of slow proliferating Lgr5+ cells that show enhanced resistance to radiotherapy in the normal intestine (Barriga et al., 2017). Furthermore, Mex3a mRNA was upregulated in triple mutant CRC genotypes (C3-Figure 7b,f). This set of results is now part of an article that describes how triple mutant tumors are enriched in a slow-proliferating subset of cells, marked by the expression of Mex3a, that are resistant to chemotherapy and originate relapse after chemotherapy treatment (Álvarez-Varela et al. *in*




**C3-Figure 7 | Suboptimal stem cell niche induces a latent Mex3a+ state. a**, Relevant GSEAs obtained after RNA sequencing analysis of AKP MTOs grown in SC medium

versus +TGFβ1. Profiling was performed at day 3 for cells in control medium and day 7 for treated cells. Left panels in each case indicate Hallmarks or gene ontology biological processes (GOBPs) related to functional pathways (Function); right panels indicate GSEAs related to cell identity of different cell populations (Identity). Gene sets are ordered based on their -log10 p-value, dashed line indicates p-value = 0.05. **b,** Mex3a mRNA expression levels in AKP MTOs treated with TGF-β vs control SC medium (n=6 independent experiments). Boxplots indicate adjusted values of Mex3a mRNA corrected by experimental batch. **c**, Box and whiskers plot showing Cd274 expression in organoids with different genetic alteration and culturing mediums. **d**, Representative flow cytometry plots showing PD-L1 expression in AKP and AKP+S MTOs treated with TGF-β vs control SC medium. **e**, Relevant GSEAs obtained after RNA sequencing analysis of APS MTOs grown in SC medium versus medium with no EGF. Profiling was performed at day 3 for cells in control medium and day 7 for treated cells. Left panels in each case indicate Hallmarks or gene ontology biological processes (GOBPs) related to functional pathways (Function); right panels indicate GSEAs related to cell identity of different cell populations (Identity). Gene sets are ordered based on their -log10 p-value, dashed line indicates p-value = 0.05. **f,** Mex3a mRNA expression levels in APS MTOs treated with no EGF vs control SC medium (n=6 independent experiments). Boxplots indicate adjusted values of Mex3a mRNA corrected by experimental batch. **g**, LYZ1 immunohistochemistry showing the presence of Paneth-like tumor cells (LYZ+) in APT subcutaneous tumors, but not in APT+K tumors. Scale bars, 50 μm (left) and 250 μm (right).

*revision*).

In addition, TGF-β signaling in AKP MTOs induced the expression of the signature genes of revival stem cells (Ayyaz et al., 2019) and fetal progenitors (Nusse et al., 2018) (C3-Figure 7a), in agreement with previous studies (Han et al., 2020). TGF-β also induced the upregulation of pathways and genes related to epithelial-to-mesenchymal transition (C3-Figure 7a). Interestingly, we found a marked upregulation of Cd274 mRNA – the gene encoding for PD-L1- in AKP organoids treated with TGF-β (C3-Figure 7c), but not in TGFBR2 or SMAD4 mutant organoids. We validated the upregulation of PD-L1 at the protein level by flow cytometry (C3-Figure 7d).

Among the most upregulated genes in APT and APS organoids in non-EGF conditions, we found several Defensin genes (Defa24, Defa3, Defa31, Defa17 and others) (C3-Figure 7e). Defensins are known markers of Paneth cell; this finding thus suggested that these CRC tumors might retain this specialized secretory lineage. In fact, we observed an enrichment of signatures of secretory and Paneth cells in APS and APT MTOs without EGFR signaling. Accordingly, APT subcutaneous tumors contained cells stained with the Paneth cell-marker Lysozyme 1, whereas we could not find any LYZ1+ cells in APT+K tumors (C3-Figure 7g). This discovery is also in agreement with the upregulation of a Paneth cell signature in human KRas wild-type CRCs treated with EGFR inhibitors (Lupo et al., 2020).

## Modeling metastatic latency with triple mutant tumors

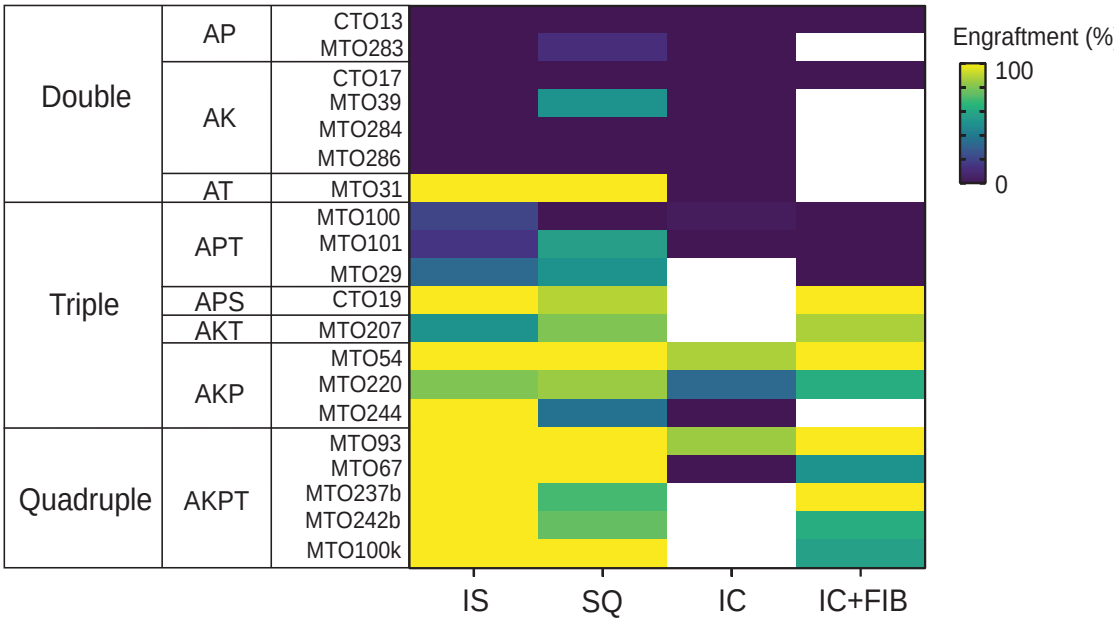
Stage II-III patients are treated with chemotherapeutic agents aimed to prevent metastatic relapse. However, these therapies only have a modest effect on the survival of these patients, suggesting that residual disease is largely refractory to such treatments. In previous chapters, we have characterized the nature of micrometastases formed by quadruple mutant tumors, from the perspective of tumor cells as well as from the TME. Our findings demonstrate that metastatic relapse is originated by a subset of cells with high Emp1 expression and that residual disease can be effectively treated with ICB immunotherapies. Nonetheless, the fast proliferating kinetics of 4x tumors imply that their residual-stage is only momentary and thus difficult to be exploited in the clinic.

In contrast, we discovered that niche-sensitive tumors – the most common type in mCRC patients - display a latent phenotype upon arrival to the liver, thus expanding the adjuvant therapeutic window. In fact, bioinformatic studies predict that metastatic seeding occurs, on average, 4.1 years before clinical detection (Hu et al., 2019, 2020). Thus, the residual-stage in CRC patients is most likely metastases that were seeded a considerable time before cancer diagnosis. These observations raise several important questions: we wonder whether the T cell-enriched phenotype that we observed in early metastases prevails in "old" but still small microlesions. Does residual disease differ depending on the timing of dissemination? Is there TME evolution in a context of slow-proliferating or latent tumor cells?

To shed some light on these problems, we explored the nature of tumor cells and their surrounding TME in latent residual micrometastases. For this objective, intrasplenic (IS) injections were not a suitable model, since the injection of cells directly into portal circulation, in the absence of a primary tumor, entails several caveats. First and foremost, in the absence of a primary tumor there is no pre-existing adaptive immunity that will control metastases upon arrival into the liver. Adaptive immunity takes at least seven days to build and therefore metastases initiation – the crucial step that we would like to characterize - will occur in the absence of a functional immune system. Moreover, direct transfer of tumor cells in circulation precludes the study of the extravasation and dissemination steps of the metastatic cascade and it bypasses the chronic inflammatory changes and immunosuppression that might be induced by primary tumors. On the other hand, tumor cell injection involves extensive cell death, which causes an artificial and exaggerated immunization of mice to tumor cell-antigens. Finally, the injection of hundreds of thousands of cells typically results in large numbers of liver metastases, which is in contrast to the metastatic burden of CRC patients, who typically develop few metastases. As we have seen in the previous chapter, the metastatic burden of the surrounding organ will have an influence on the local TME of a single metastases.

For all these reasons we used our recently developed model of disease relapse for CRC (Cañellas-Socias et al., *in revision*). Although this mouse model also relies on tumor cell transplantation (and therefore may not faithfully reproduce the priming of the immune system of spontaneous arising cancers), cell dissemination and metastases occur in the presence of fully functional adaptive immunity. Thus, the model is not best suited to study primary tumor initiation or engraftment, but it is especially useful to investigate metastases formation. Moreover, we have proved that triple mutant tumors are able to colonize the liver upon direct injection, but the question remains as to whether these tumors are successful in invading and disseminating.

Unfortunately, most MTOs tested showed poor engraftment upon transplantation into the caecum of immunocompetent mice (C3-Figure 8). We tested three different AKP MTOs (#54, #220 and #244), three APT MTOs (#100, #101 and #29), three AK MTOs (#39, #284, #286), one AP MTO(#283) and one AT (#31) MTO without success (C3-Figure 8). In general, quadruple mutant MTOs showed better engraftment than double or triple MTOs in the caecum, although engraftment in the caecum was consistently lower than in IS or in subcutaneous (SQ) injections for all MTOs (C3-Figure 8). Bioluminescence monitoring revealed that tumors grew during the first 7-10 days but then they abruptly disappeared (C3-Figure 9a). These kinetics suggested that again the adaptive immune system was responsible for the rejection of these tumors. In fact, depletion of



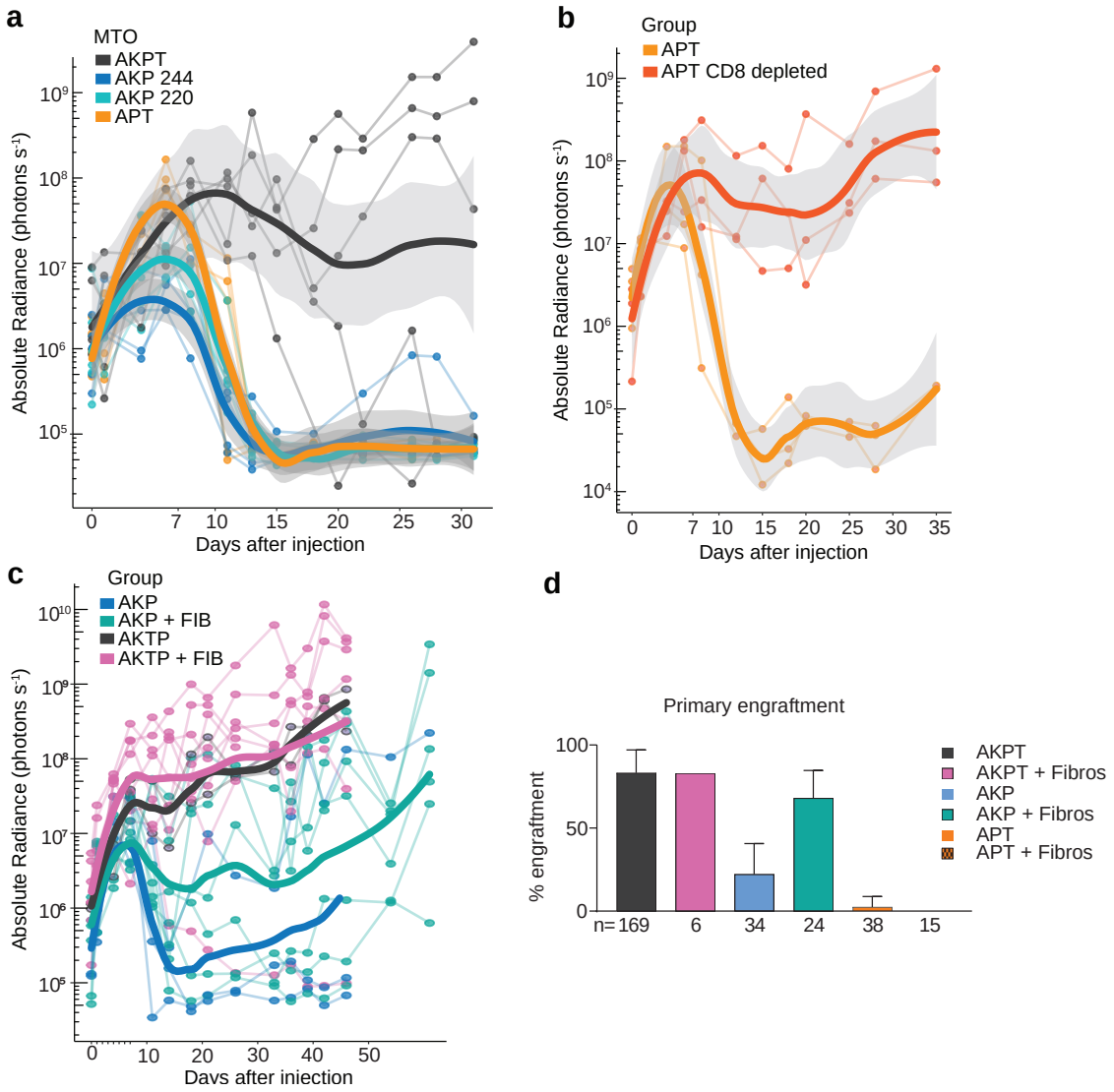

**C3-Figure 8 | Low engraftment of double and triple mutant MTOs implanted in the caecum.** Heatmap showing percentage of engraftment of MTOs with different genetic driver mutations implanted in different organs. MTOs are derived from GEMM, whereas mutations are introduced by CRISPR-Cas9 in vitro to generate CTOs IS= intrasplenic injections to the liver. SQ= subcutaneous injections skin. IC= intracaecum injection. IC + FIB= Coinjection of MTOs with colon fibroblasts in the caecum.

CD4/CD8 T cells enabled the growth of APT primary tumors (C3-Figure 9b). However, tumors grown in T cell-deficient mice spread locally, which impeded surgical excision and the study of metastasis formation in the presence of adaptive immunity (C3-Figure 10).

Importantly, all these MTOs were derived from full grown intestinal tumors, implying that the same tumor cells had been proficient in generating tumors in GEMM. As I previously discussed, this model system – akin to the intrasplenic injections - is limited by the artificial immunization of mice to cancer antigens due to the injection of cells. In injection models, tumor cells expressing neoantigens are implanted and have a very limited time (approximately seven to ten days) to build a TME that will protect them against adaptive immunity. The composition of tumors so early on is immature and therefore they are unable to face a well-primed immune system. In contrast, in patients (and up to some extent in GEMM) alterations in the genome evolve in parallel with the development of an immunosuppressive TME. Thus, immune recognition towards tumor cells is progressively acquired as tumor neoantigens accumulate, leaving time for tumor cells to defend themselves. Given these observations, we consider the differential primary tumor engraftment between genotypes artefactual and of little interest. Having said that, we considered that if we were successful in growing primary tumors, it would still be relevant to study the metastatic setting, in which there is a primary tumor.

In parallel studies, we found that the implantation of pieces of full grown *in vivo* subcutaneous tumors led to successful engraftment of AKP and ATP tumors (C3-Figure 10). This data further reinforced the notion that a mature TME enables immune escape and indicated that rejection of AKP and ATP tumors in needle-based injections is not physiologically relevant. However, we could not use this method of implantation (named nesting caecum) because it involves the use of surgical sutures that generate attachments to adjacent tissues thus preventing surgical excision of the primary tumor (C3- Figure 10). Given that cancer-associated fibroblasts (CAFs) are the major producers of poor prognosis-related genes in the stroma of CRC, we then attempted to inoculate tumor cells together with a cell line of colon fibroblasts. Co-injection of MTOs with fibroblasts significantly improved the engraftment of AKP tumors, yet it did not impede the rejection of ATP tumors (C3- Figure 9c).

Our previous experiments with subcutaneous tumors indicated that KRas wild-type tumors had a defective TME. Still, an important fraction of CRC tumors in patients are KRas wild-type, implying that they are able to expand in the context of the colonic mucosa. We hypothesized that the suboptimal TME of KRas *wild-type* tumors might be complemented by certain clinical conditions of the host associated with immunosuppression, such as infectious diseases, aging, obesity or diabetes. If this were the case, some tumors with




**C3-Figure 9 | Immune rejection of MTOs upon implantation in the caecum. a**, Quadruple mutant MTOs (AKPT) and triple mutant MTOs (AKP and APT) were labeled with a luciferase expression vector and inoculated through the spleen into in C57BL/6J mice. Primary tumor metastatic growth was assessed by intravital bioluminescence imaging (BLI). Graph shows longitudinal BLI measurements (photons s−1), normalized to the day of injection. Points and lines represent individual mice, trend lines (bold) were generated using a LOESS model. **b**, Bioluminescence monitoring of liver metastases generated by injection of APT MTOs in control or CD8-depleted mice. **c**, Bioluminescence monitoring of liver metastases generated by injection of AKP and AKTP MTOs in combination with colon-derived fibroblasts. **d**, Barplot (mean±SD) showing engraftment of AKTP, AKP and APT MTOs implanted in the caecum when injected alone or in combination with fibroblasts.

reduced ability to corrupt the TME might only thrive in certain circumstances. Obesity has been related to the accumulation of MDSCs and neutrophils (Quail et al., 2017). We hypothesized that obesity-induced systemic expansions of myeloid cells might counterbalance the reduced recruitment capability of such cell types by KRas wild-type tumors. Thus, we fed mice with high-fat diet for more than 14 weeks to induce obesity. Although

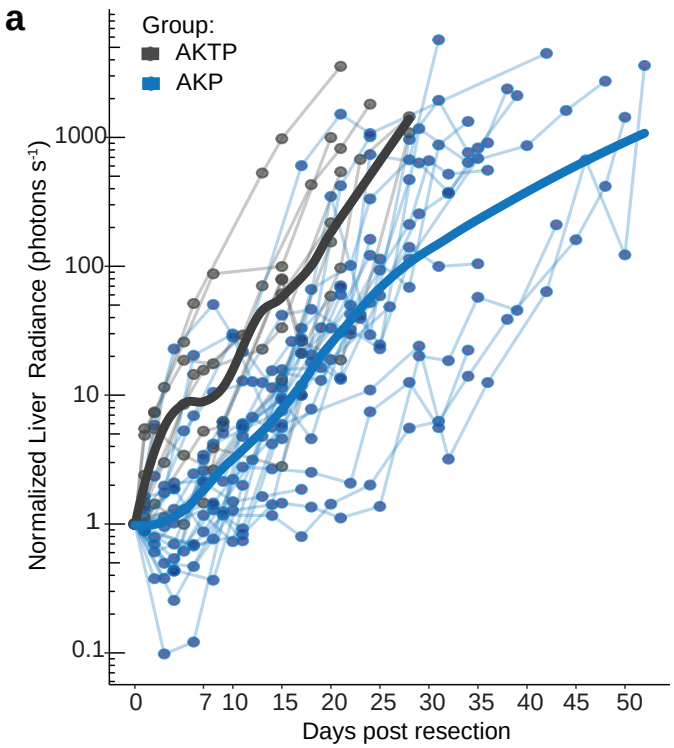
high fat-diet worsened metastases of AKPT tumors (not shown), it did not prevent the rejection of APT tumors in the caecum (C3- Figure 10). We obtained similar results with aged animals (20 weeks old) and animals treated with DSS to induce colonic inflammation (C3- Figure 10). We tested a large number of other different conditions that would circumvent this artifact of our model system (summarized in C3- Figure 10). For example, we tried injecting drops of Collagen I, instead of BME, and injecting organoids after niche-factor withdrawal. We also injected high and low MTO cell numbers, with or without xenoantigens such as luciferase and GFP. Unfortunately, none of these trials worked well and we could not monitor the growth kinetics of double and triple mutant tumors, which we expect might have even slower kinetics of metastasis progression.

Thus, we used fibroblast co-injection to follow the growth kinetics of AKP liver metastases after the resection of primary tumors. Mirroring our previous observations, we observed slower growth kinetics of AKP tumors compared to AKTP tumors (Figure 11a). Nonetheless, the latency phase was shorter than we had previously observed in intrasplenic injections, with the exception of a subset of mice with a latency phase that lasted

| Type perturbation | Perturbations in ATP | Engraftment ATP | Limitations |
|---|---|---|---|
| TME modification | CD4/CD8 ablation | 9/9 | Surgery impossible + immunodeficient |
| | Nesting cecum | 2/4 | Surgery impossible |
| | Digested TME | 0/5 | |
| | Fibroblast coinjection | 0/15 | |
| MTO modification | High/low cell numbers | 0/4 | |
| | Niche-factor withdrawal | 0/6 | |
| | Unlabelled (no xenoantigens) | 0/7 | |
| Host modification | High-fat diet | 0/5 | |
| | DSS | 0/4 | |
| | Aged mice | 0/4 | |
| | Antibiotics | 0/4 | |

**C3-Figure 10 | Strategies to avoid immune rejection of MTOs implanted in the caecum.** Table showing the different strategies we tested to increase the engraftment rate of ATP MTOs implanted in the caecum.

around 20-25 days. Of note, the growth kinetics of liver metastases were unaffected by the presence of fibroblasts at the time of injection (not shown).




**C3-Figure 11 | TGF-β sensitive tumors generate slower-progressing metastases.** a, Bioluminescence monitoring of liver metastases growth after AKTP or AKP primary tumor resection. Data is normalized to the first BLI measurement in the upper-thorax after primary tumor resection. Points and lines represent individual mice, trend lines (bold) were generated using a LOESS regression.

# Discussion

## Chapter I: High risk of metastatic relapse by residual EMP1+ tumor cells

The study of events underpinning metastatic relapse in most cancer types represents a major challenge due to the impossibility of sampling DTCs and micrometastases directly in patients. Instead, bulk transcriptomics have been used to assess gene programs associated with metastatic recurrence in primary tumors. By mapping recurrence-associated genes at the single cell level, we confirmed that most of the poor prognosis transcriptome is expressed by stromal cells (Calon et al., 2015; Isella et al., 2015). Yet, we also identified a subset of epithelial cancer cells, that we named High Relapse Cells (HRCs), associated with high risk of relapse after surgery of the primary CRC. HRCs represent a defined cell state present in a large proportion of samples, thus implying a common origin and mechanism of metastatic recurrence across CRC patients.

To study the contribution of this cell population to CRC relapse, we devised two new approaches. First, we established a preclinical CRC mouse model in which primary tumors can be surgically extirpated. These mice remain cured of the primary disease, but subsequently, as it occurs in patients, a fraction of them develop metastases in the liver and other relevant sites. The paucity of residual disseminated tumor cells compared to healthy tissue cells has limited attempts to profile directly this cell population from whole organs. The state-of-the-art approach involves an in vitro selection step of isolated residual cells that may distort their transcriptional state (Laughney et al., 2020). Here, we developed an isolation method from whole livers after resection of the primary CRC that results in an elevated enrichment of residual tumor cells. Coupled to single cell RNA sequencing, this methodology enabled us to capture the unperturbed cell states of residual tumor cells before metastatic disease is overt.

By leveraging these two approaches, we show that tumor cells occult in mouse livers after primary CRC surgery resembled the HRC population present in patient samples. Bioinformatic reconstruction from scRNAseq data predicts that HRCs are the cell of origin of disease relapse. Overtime, this cell population gives rise to other cell types, and regenerate the heterogenous cell populations present in macro-metastases. Our *in silico* characterization of HRCs allowed us to identify marker genes that allowed genetic manipulation of these cells. We chose the EMP1 (Epithelial morphogenic protein 1), a marker gene for mouse and human HRCs, to insert lineage tracing and suicidal cassettes through gene editing techniques. We tracked and selectively eliminated this cell population at different time points during disease progression. Ablation of Emp1+ cells before surgical excision of primary tumors prevented metastatic recurrence and cured mice.

The cell of origin of metastasis in CRCs has remained a matter of debate over the last years. Pioneering work by de Sauvage and colleagues revealed that Lgr5+ cancer stem cells are dispensable for primary CRC growth yet necessary for metastasis formation in experimental models (Melo et al., 2017). Subsequently, Van Rheenen and colleagues proposed that metastases are initiated by disseminated differentiated cells (Lgr5 negative) that through plasticity produce Lgr5+ cancer stem cells upon reaching the liver (Fumagalli et al., 2020). Massagué and colleagues also provided evidence that expression of the adhesion molecule L1CAM in Lgr5+ and Lgr5- cells is important for metastatic colonization (Ganesh et al., 2020). Our data unequivocally shows that HRCs are neither differentiated nor stem-like but rather represent a distinct state that enables migration and colonization of foreign organs. The similarity between HRCs and basal-like tumor cells present in pancreatic cancer is intriguing and warrants further exploration. Our data fit well with a model whereby HRCs disseminate out of the primary tumor prior to surgical resection and the HRC state is subsequently retained in residual tumor cells lodged in foreign organs (Figure 4v). Reacquisition of the Lgr5+ stem cell and proliferation programs occur at a later phase and are necessary for metastatic outgrowth (Figure 4v).

### *Features of HRCs*

We found no robust evidence supporting that HRCs undergo EMT but rather they appear to retain a complete epithelial program. Although loss of cell adhesion has been classically viewed as a pro-tumorigenic feature (Hanahan and Weinberg, 2000), normal cell adhesion through cell junctions is retained in many tumor types, enabling cancer cells to adopt a collective migration mode (Janiszewska et al., 2020a). The localization of HRCs at tumor buds further strengthens the well-established association of these anatomic structures with poor prognosis (Lugli et al., 2017; van Wyk et al., 2019). Invasion by collective migration is characterized by a fine balance between cell-cell and cell-ECM adhesion. In addition to E-Cadherin maintenance of adherens junctions, several genes related with desmosome formation —such as Keratins (Krt16, Krt17, Krt19 and Krt80), Periplakin (Ppl) and Junction Plakoglobin (Jup) — are upregulated in HRCs and have been associated with greater capacity to form tumor cell clusters (Li et al., 2021). Moreover, the HRC marker gene Epithelial Membrane Protein 1 (EMP1) encodes a four-transmembrane protein of 160 amino acids related to tight-junction formation. Proteins enhancing cell-cell junctions couple the actin cytoskeleton of groups of cells, thus transmitting forces that drive migration (Northey et al., 2017). HRCs also showed upregulated expression of ECM receptors that have been linked to tumor cell migration, such as Itga3, Itgb4 and Ddr1. Collective migration requires remodeling of the extracellular matrix, to facilitate the movement of a group of cells (Friedl and Gilmour, 2009). Fitting with this notion, we observed higher expression of several proteases — such as Kallikreins (Klk6, Klk7, Klk8,

Klk10), Ctse, Mmp13 and Plaur — as well as secreted components of the ECM, in particular the three genes encoding for laminin-5 (Lamc2, Lama3, Lamb3), a well-known marker of tumor buds in CRC(Pyke et al., 1995).

Stromal cell populations lying close to malignant cells have also been shown to assist ECM remodeling and invasion by creating tracks that are used by migrating cancer cells (Janiszewska et al., 2020b). In ongoing experiments, we have explored how HRCs modulate the surrounding non-tumoral cells by profiling the TME of primary tumors upon ablation of HRCs or Lgr5+ cells through scRNA-seq. Although preliminary, our results suggest that HRCs induce the differentiation of unspecialized fibroblasts into cancer-associated fibroblasts (CAFs). CAFs show high expression of proteins involved in extracellular proteolysis and are highly correlated with poor prognosis in CRC (Calon et al., 2015). The generation of organoid cultures that contain populations of tumor-associated stromal cells will be instrumental to study how CAFs modulate ECM remodeling and cancer cell invasion *in vitro*. Interestingly, a recent study suggests that CAF-mediated compression and mechanotransduction result in the formation of tumor budding structures in *in vitro* models (Barbazan et al., 2021).

### *Sources of HRCs*

The observation that CRC patients contain different numbers of HRCs suggest that certain signaling pathways or oncogenic mutations are necessary for this population to arise. By comparing the bulk transcriptomic profiles of patients with different levels of HRC-marker genes, we will explore which signaling cues are necessary to instruct this cell population. For example, the finding that CRC samples containing no Lgr5+ cells include abundant HRCs suggests that WNT signaling is dispensable for the specification of this population. In addition, early-stage CRCs (stage I-II) contain fewer HRCs, which suggests that features of advanced tumors induce the formation of this population. Such features might span from cancer cell-intrinsic properties, such as the accumulation of genomic alterations (driver mutations, copy number alterations, etc...) and epigenetic reprogramming, to changes in the TME—which consists of cellular interactions, physical parameters (oxygen levels, pH, mechanical pressure), and molecules including extracellular matrix components, growth factors, and cytokines—. In fact, scRNAseq profiling revealed substantial cancer cell heterogeneity in genetically stable organoids bearing 4 pathway mutations upon *in vivo* implantation. Together with a large body of literature, these observations suggest that environmental cues shape cancer cell phenotypes in vivo (Fuchs and Blau, 2020; Han et al., 2020; Lenos et al., 2018; Medema and Vermeulen, 2011). Moreover, the observation that HRCs were particularly enriched in invasion fronts further suggests that certain microenvironments induce this tumor cell subpopulation.

In preliminary experiments not included in this thesis, we have observed that increased collagen concentration, hypoxic conditions, and the presence of fibroblasts induce the expansion of HRCs in culture.

### *Plasticity of HRCs*

Cellular plasticity was first reported in healthy intestinal epithelial cells (Tian et al., 2011), and similar processes have been observed in primary CRC tumors, with various microenvironmental signals highlighted as being responsible for the switch of non-CSCs to CSCs (Medema and Vermeulen, 2011; de Sousa e Melo and de Sauvage, 2019). However, recent data suggested that plasticity in CRC cells can be also triggered independently of stemness-inducing factors provided by the mesenchyme (Fumagalli et al., 2020). The differences in the mechanisms governing stemness and plasticity in healthy and cancerous cells might arise from oncogenic mutations in signaling pathways that regulate stem cell hierarchies in the intestine, such as WNT signaling. Interestingly, although most CRC tumors harbor mutations in APC, which lead to constitutive WNT activation, not all cancer cells are stem cells. We hypothesize that, in the absence of environmental stimuli, all tumor cells harboring APC truncating mutations would behave like Lgr5+ cells. Nonetheless, in full-blown tumors, certain WNT-counteracting signals cooperate in specific niches to induce other cell states, such as HRCs. In fact, a recent article proposes a role of YAP/TAZ signaling in counteracting the stem cell phenotype and inducing a regenerative-like state in APC mutant tumors (Cheung et al., 2020). During cell dissemination, signaling of pathways counteracting WNT downstream of APC ceases and therefore the stem cell state is reacquired. However, as metastases regrow, an HRC-inducing niche is reestablished and HRCs are specified de novo. Thus, we propose that the stem cell phenotype is the equilibrium or predetermined state in WNT-mutant tumors, yet certain environmental cues induce cancer cell heterogeneity. In upcoming experiments we will explore which molecular changes are required for HRCs to transition to Lgr5+ cells. Theoretically, blocking such pathways could lock DTCs in an HRC-like non-proliferative state, which might have therapeutic potential.

### *Therapeutic targeting of HRCs*

The development of therapies targeting HRCs requires a deeper molecular and functional characterization of this cell population. Although HRCs overexpress several molecules that have been related to cancer progression, strategies blocking single molecules with neutralizing antibodies have repeatedly failed in a complex disease such as cancer. A more attractive alternative is to leverage antibody-dependent cellular cytotoxicity, antibody-drug conjugate formats, and chimeric-antigen receptor (CAR) T cells, which all

selectively deliver cytotoxic payloads to eradicate target-expressing cancer cells. Sparing of healthy tissue with these technologies strongly depends on the selection of target antigens. The establishment of large scRNAseq databases is likely to favor the identification of genes lowly expressed in healthy tissues that become upregulated in cancer cells, and more specifically in metastasis-initiating cells. An appealing candidate gene upregulated by HRCs is Mesothelin (Msln). In this regard, several CAR T cell therapies have already been developed against Msln, showing safe profiles in clinical trials (Adusumilli et al., 2019; Morello et al., 2016).

Our cell ablation experiments imply that therapies capable of eliminating HRCs may prevent disease relapse if applied before metastatic disease is overt. However, in our experiments, mice are treated shortly after the implantation of primary tumors, whereas untreated primary tumors in patients develop over many years until diagnosis (Hu et al., 2020). As a result, prevention of cell shedding in human primary tumors is unlikely to prevent dissemination, which most likely occurs before diagnosis as time from diagnosis to surgery is relatively short (Van Der Bij et al., 2009) compared to the estimated life of primary tumors (Hu et al., 2019). Thus, therapies to prevent metastatic relapse strongly depend on effectively eliminating residual cells lodged in distant organs that eventually regenerate tumors. In quadruple mutant tumors, where disease relapse occurs extremely rapidly, HRCs revert rapidly into Lgr5+ cells and therefore HRC-ablating therapies are bound to fail. Nonetheless, in Chapter III, we discovered that most CRCs harbor fewer oncogenic mutations, which results in longer disease-free survival intervals after surgery. It is highly relevant to understand whether the HRC state in these patients is retained for long periods or whether HRCs convert to another state, such as Mex3a+ slow-proliferating cells (Alvarez-Varela et al., under revision). We speculate that the timing of HRC-Lgr5+ state reversion may determine the success of therapies targeting HRCs.

Importantly, we observed that HRCs also located at the periphery of liver metastases, thereby suggesting a plausible role in secondary metastases seeding. Liver metastases have the most favorable prognosis of all metastases because they are potentially curable by resection (Yaeger et al., 2018). Unfortunately, over time many patients develop additional metastases in other tissues, including the lung, brain, peritoneum and omentum, which are associated with poor survival (Yaeger et al., 2018). If HRCs were found to be responsible for secondary metastases formation, HRC-ablating therapies would hold promising therapeutic value in the metastatic context. Unlike primary tumors, which develop and seed metastatic cells well before diagnosis, the development of asynchronous metastases can be predicted at diagnosis and patients receive timely monitoring. Therefore, stage II-III patients at high risk of metastatic relapse could be given anti-HRC therapies. While such treatment would probably not impede the formation of primary

metastases — given that seeding occurred before treatment commences — they would prevent further metastatic dissemination.

### *Technical advances to study residual disease*

Aside from the biological discoveries described, the technical advances in our study represent a breakthrough towards understanding CRC recurrence that will hopefully attract many researchers into investigating this neglected stage of the disease. The low proportion of tumor cells compared to healthy tissue makes the isolation of micrometastases a major bottleneck in cancer research. During tissue preparation for flow cytometry, we discovered that, upon mild dissociation of whole livers, most hepatocytes flowed through, whereas metastatic tumor cells were retained in 100 μm filters. This observation allowed us to develop a new protocol that massively enriches tumor cells in non-dissociated tissue, thereby allowing the profiling of DTCs for the first time. Although we cannot offer a solid explanation for this improvement, the differential sensitivity to enzymatic digestion might originate from distinct cell to cell and/or cell to matrix adhesions between tumor and healthy cells. Indeed, during *in vitro* culturing, tumor cells typically require harsher and longer treatments with dissociation agents than healthy cells, which supports the notion that cancer cells form stronger cell adhesions. Cell clusters retained in filters might be composed of tumor cells alone or attached to ECM proteins, other cell-types in the TME and/or hepatocytes. Thus, although we might lose some single cells by discarding the 100μm filter flow-through, this was a necessary compromise to enrich sufficiently for tumor cells and sort them in manageable times for scRNAseq. We speculate that this methodology might also enrich for tumor cells in metastases from other cancer types. Moreover, the implementation of the CRC relapse mouse model described herein may serve as a powerful pre-clinical platform for testing adjuvant therapies. In fact, it has been key to enable the study of the TME and response to immunotherapies in residual CRC, addressed in the second chapter of this thesis.

## Chapter II: Neoadjuvant immunotherapy prevents metastatic relapse in MSS CRCs

In sharp contrast to the exciting responses to immune checkpoint blockade (ICB) observed in patients with melanomas, renal cell cancer and lung tumors, the first clinical trial of PD-1 blockade in CRC led to disappointing results: only 1 out of 33 patients responded (Brahmer et al., 2012; Topalian et al., 2012). A follow-up study showed that only CRC patients with deficient mismatch-repair (MSI) and high TMB benefited from ICB with pembrolizumab (Le et al., 2015). Thereafter, all MSS CRCs were regarded as immunologically "cold", meaning scarcely T cell-infiltrated and non-immunogenic, due to the low number of mutations present in comparison with MSI CRCs, and they were therefore not considered candidates to benefit from ICB.

In 2018, however, we revealed that MSS CRCs are in fact recognized by the adaptive immune system, yet effector T cell activity is suppressed through high TGFβ levels in the TME (Tauriello et al., 2018). Mice bearing 4 oncogenic mutations (AKTP) - or MTOs derived from those mice and transplanted back into immunocompetent mice - generated metastatic intestinal tumors that display key hallmarks of human MSS CRC, including a low TMB, T-cell exclusion and TGFβ-activated stroma. Blocking TGFβ signaling with small molecule inhibitors enabled T cell infiltration and was sufficient to confer susceptibility to anti-PD-1–PD-L1 checkpoint- based therapies in advanced CRC mouse models. Consistent with the well-established role of TGFβ signaling in suppressing the differentiation and activity of T cells (Gorelik & Flavell, 2000; Laouar et al., 2002), we observed, both in mice and humans, that a TGFβ-activated TME antagonized a Th1-effector cell phenotype. Thus, we demonstrated that inhibition of TGFβ signaling could render poor-prognosis MSS patients susceptible to immunotherapy.

When analyzing the tumor stroma of residual metastases in our model, we observed striking infiltration by CD3+ T cells, in sharp contrast to the immune excluded phenotypes of primary tumors and overt metastases. Motivated by this finding, we treated mice with anti-PD1 + anti-CTLA4 before surgical resection of primary tumors, at a time point when metastatic disease is incipient. While 80% of control animals developed liver metastases after surgery, most ICB-treated mice remained disease-free after surgery, and the average number of metastases was reduced 10-fold. In contrast, primary tumors and overt metastases were unaffected by ICB treatment, in line with clinical data in patients. Our observations suggest that T cells readily detect tumor cells upon arrival at the liver, which become protected through PD1/PD-L1 signaling. ICB efficiently eliminates micrometastatic tumor deposits disseminated beyond the resected tumor, which are ultimately the source of postsurgical relapse.

We hypothesized that this curative effect is restricted to a narrow temporal window as a result of a rapidly evolving TME. By profiling at single-cell level metastatic lesions at different stages, we found that the TME of micrometastases is immature and poorly immunosuppressive, which, coupled with the low tumor burden, might render metastases sensitive to ICB monotherapy at this stage. On the other hand, the ecosystem of fully grown metastases is rich in cell types that produce TGFβ and other immunosuppressive molecules, which block T cell activity. In particular, monocytes are recruited to early metastatic lesions and are progressively converted to immunosuppressive macrophages, expressing checkpoint molecules, TGFβ and other immunomodulatory factors, such as Spp1, Arg1 and complement proteins. We validated these data using multiplex immunofluorescence, which evidenced that both metastases size and overall metastatic burden affect the composition of the TME. Residual disease metastases contain abundant T cells, whereas late stage metastases are heavily enriched in stromal subpopulations and T cells are scarce. As a result of increased tumor burden and the recruitment of stromal cell types, the TME of large tumors might possess multiple metabolic and cytokine- inhibitory mechanisms that render immunotherapies ineffective. Two concepts stem from these observations; first, the TME of metastases evolves and becomes complex over time. Thus, immunotherapeutic treatment must be tailored against different stages of this progression. Second, residual disease is a particular state that could benefit from immunotherapies when applied in a timely manner. These results could have enormous implications for the treatment of stage II-III CRC patients, since chemotherapy only mildly reduces the risk of relapse (André et al., 2015; Sinicrope et al., 2013) and TGFβ signaling inhibitors are not yet used in patients due to adverse cardiovascular effects (Tauriello et al., 2022).

### *Translation of neoadjuvant ICB to MSS CRC patients*

Given that, by definition, cancer is undetectable at the residual stage, human samples to validate our observations on the residual TME do not exist. Nonetheless, a recent clinical trial by Chalabi and colleagues showed, for the first time, pathological responses to ICB in early-stage MSS primary tumors (Chalabi et al., 2020). These responses further support our previous publication (Tauriello et al., 2018), in which we reported that the lack of success of ICB in MSS CRC was not due to inexistent neoantigens — as previously thought— but due to microenvironmental immunosuppression of T cell activity. In the exploratory NICHE study, patients with dMMR (mismatch repair deficient) or pMMR (mismatch repair- proficient) tumors received a single dose of ipilimumab and two doses of nivolumab before surgery. Although the primary objective of the study was safety and feasibility, pathological responses were observed after 4 weeks of therapy in 100% of dMMR tumors and more strikingly, 27% of pMMR tumors. Moreover, in non-respond-

ing pMMR tumors, increases in CD8+ T cell counts, TCR clonality and IFN-γ score expression reflected an underlying immune activation despite little or no tumor regression, in turn suggestive of tumor recognition. Studies with neoadjuvant ICB in melanoma, nonsmall cell lung cancer and bladder cancer had previously shown that, for tumor types in which activity is seen in stage IV disease, response rates go up when moving to earlier-stage disease. The NICHE study demonstrated, for the first time, that ICB activity can occur during early-stage disease in a tumor type that was thus far nonresponsive to ICB in stage IV. However, the study by Chalabi et al. has some limitations, namely the small number of MSS patients treated (n=15) and the short postoperative follow-up. Larger studies and at least a 3-year follow-up for disease-free survival are required to determine whether the responses observed translate into improved disease-free and overall survival.

Primary tumors that show a pathological response in neoadjuvant trials would have been removed by surgery alone and therefore the therapeutic benefit relies solely on the prevention of metastatic recurrence. The destruction of micrometastases is central to the notion that neoadjuvant PD-1 blockade should result in enhanced relapse-free and overall survival in operable patients who would otherwise relapse after surgery alone. The fact that most stage II-III CRC patients are cured by surgery alone calls for large cohorts of patients in order to accurately assess any therapeutic benefit of neoadjuvant therapies in terms of overall survival. The large number of patients, coupled to the long follow-up required (at least 3 years), make these clinical trials cumbersome. Therefore, we anticipate that the implementation of the CRC relapse mouse model described herein will serve as a powerful preclinical resource for testing neoadjuvant or adjuvant therapies. Such trials could also be facilitated by the recruitment of only stage II-III patients at high risk of relapse, which can be identified by clinical parameters and measurements of immune infiltration in primary tumors (Pagès et al., 2018). This strategy would alleviate the number of patients needed and, more importantly, would avoid treating patients that would be cured by surgery alone.

Ultimately, an important challenge is predicting which patients will respond to immunotherapeutic intervention before treatment. Although treatment-related adverse events are less frequent in immunotherapy compared to chemotherapy (Reck et al., 2016), blocking immune checkpoints unleashes powerful immune effector mechanisms that may not respect immune tolerance to self-tissues, with an estimated fatal incidence between 0.3 and 1.3% (Martins et al., 2019). It would also be advantageous to identify an early indicator of long-term benefit. Pathological response and immunological analyses of primary tumors have been used as early surrogate endpoints for OS and DFS in neoadjuvant therapies (Huang et al., 2019). However, the fact that we observed a major

effect on disease-free survival yet primary tumors did not show a pathological response, warrants special consideration when correlating responses between primary tumors and DTCs.

CTLA-4 restrains the initial TCR/CD28 mediated priming of conventional CD4+ T cells (Lenschow et al., 1996; Tivol et al., 1995) and PD-1 predominantly restrains CD8+ effector T cell responses in peripheral tissues (Nishimura et al., 2001). The combination of both antibodies has been shown to have a synergistic effect on the activation of anti-tumor immunity and has led to an increase in response rates in patients (Waldman et al., 2020). Our experiments comparing the two FDA-approved therapeutic antibodies showed comparable effects of aPD1 monotherapy to aPD1 + aCTLA4 combination therapy. Moreover, scTCRseq revealed that expanded T cell clonotypes were shared between primary tumors and metastases, thereby suggesting effective priming and recognition of tumor cell antigens before treatment. Thus, our data collectively support that systemic immunity is primed against primary tumors and consequently, T cells readily detect tumor cells upon arrival to the target organ. Residual lesions escape adaptive immunity through PD-L1/PD-1 signaling. However, in a lowly suppressive TME, anti-PD1 therapy is sufficient to unleash a curative response. Although we found that enhancing T cell priming through CTLA-4 inhibition is dispensable, it is worth mentioning that priming of adaptive immunity might be skewed in our model system. Upon injection of cancer cells *in vivo*, there is substantial cell death and inflammation that cause an exposure of tumor cell antigens before a mature TME has been developed, thus favoring priming. In contrast, acquisition of driver mutations and neoantigens in human tumors most likely occurs sequentially and therefore priming might be inhibited by a mature TME. Thus, we are cautious with respect to encouraging the use of monotherapy without anti-CTLA4.

### *Adjuvant vs Neoadjuvant Immunotherapy*

Compared with adjuvant (postsurgical) treatment, neoadjuvant (presurgical) ICI has been shown to induce a stronger and broader tumor-specific T cell response in breast cancer models (Blank et al., 2018; Liu et al., 2016). A key corollary to this hypothesis is that neoadjuvant PD-1 blockade while the primary tumor is in place, as opposed to adjuvant therapy directed only against micrometastatic disease after resection, will leverage the higher levels of endogenous tumor antigen present in the primary tumor to enhance T cell priming and expansion of tumor-reactive T cells (Topalian et al., 2020). On the other hand, the presence of an advanced tumor can cause systemic immunosuppression that blocks T cell responses systemically (Danna et al., 2004). The comparison of neoadjuvant and adjuvant treatments holds high clinical relevance but requires careful consideration, since the timing of surgery and treatments are asynchronous. In human

patients, such small differences in time (7 days) are insignificant, but, in experimental models where tumors disseminate in a matter of weeks, this point should be taken into account. At equal timing of surgery, neoadjuvant treatment will start earlier and therefore will fight against smaller lesions. On the other hand, at equal timing of treatments, primary tumors will be implanted for longer in neoadjuvant-treated mice, thus having longer time to grow and disseminate. Thus, to properly analyze whether the immune response is affected by the presence of a primary tumor, we will compare control, adjuvant and neoadjuvant treatments across several timepoints.

### *Additional immunotherapy targets in residual disease*

In addition to Cd274, we observed striking upregulation of Indoleamine 2,3-Dioxygenase 1 (Ido1) in HRCs that originate metastases. IDO1 can be expressed by DCs, MDSCs, and cancer cells, and it catabolizes tryptophan and generates kynurenine. Both the deprivation of tryptophan and the generation of its metabolic product inhibit T cell clonal expansion and promote the conversion of naïve T cells to Tregs (Joyce and Fearon, 2015). In future experiments we will test the effect of IDO1 antagonists on a neoadjuvant or adjuvant setting. Our preclinical model may become a key resource to test the effect of similar drugs on disease-free survival intervals after primary tumor resection. Considering that new drugs in phase III oncologic clinical trials are often combined with the standard of care (SOC), we will also verify that FOLFIRI treatment (the SOC in stage II-III CRC) does not diminish the therapeutic efficacy of ICB.

### *Evolution of the TME during metastatic progression*

Elegant studies in other types of cancer have shown that tumor cells migrate accompanied by other cell types, such as platelets and neutrophils (Anvari et al., 2021; Szczerba et al., 2019). Conversely, we found that tumor cell clusters travel alone and progressively rebuild the ecosystem of advanced tumors. As a result, the TME of small lesions is poorly infiltrated with stromal cells and heavily enriched in cell-types typically related to anti-tumor functions: DCs, B cells, NK cells, and CD4 and CD8 T cells. Given that stromal cells are a major source of immunosuppressive, poor prognosis-related products that enable immune escape in CRC, we hypothesize that their absence in micrometastases underlies their marked susceptibility to anti-PD1 therapy. The primary mechanism of action for PD-1 blockade is generally thought to be unleashing tumor-specific cytotoxic T cells that already reside in the TME before treatment (Chen and Mellman, 2013; Im et al., 2016). Interestingly, we detected low numbers of CD8+ T cells compared to CD4+ T cells by multiplex immunofluorescence. A straightforward possibility is that CD8+ tumor-infiltrating lymphocytes (TILs) massively expand upon PD-1 blockade and directly destroy

tumor cell glands. Nonetheless, CD4 T helper cells not only support CTLs by promoting effector functions and long-term survival, but also have direct cytotoxicity against cancer cells, as they have been shown to mediate CD8-independent clearing of tumors (Tran et al., 2014; Xie et al., 2010). CD4+ T cells with cytotoxic activity (CD4 CTL) have been reported in various immune responses and are characterized by their ability to secrete granzyme B and perforin and to kill the target cells in an MHC class II-restricted fashion (Hirata-Nozaki et al., 2019; Takeuchi and Saito, 2017). CD4+ T cells predominantly recognize neoantigens in melanoma patients (Linnemann et al., 2015) and PD-1 signaling also restrains their clonal expansion (Konkel et al., 2010).

Our scRNAseq profiling also evidenced an enrichment in B cells and NK cells in early metastatic lesions. Despite the fact that B-lineage cells have been shown to expand after PD-1 therapy, additional research is required to determine whether they have an active role in tumor rejection (Topalian et al., 2020). On the other hand, mounting evidence supports a role for NK cells in controlling metastatic outgrowth, especially in cancer types in which tumor cell dormancy upon dissemination leads to MHC-I downregulation (Malladi et al., 2016). DTCs in our experiments showed high expression of antigen-presenting molecules. Nevertheless, we are intrigued to establish whether innate immunity also participates in neoadjuvant-induced killing of residual metastases, given that anti-PDL1 therapy has been reported to activate PD-L1+ NK cells (Dong et al., 2019). Understanding which immune cell subsets mediate the clinical response is highly relevant. However, novel strategies to profile the TME of residual lesions are required. Thus, to assess which immune cell population is responsible for tumor clearance we will perform antibody-mediated depletions of CD4, CD8 and NK cells during neoadjuvant treatment. Moreover, in future experiments, we will explore whether aPD1 immunotherapy induces the proliferation of tumor-specific T cells, resulting in the clonal expansion of specific TCRs.

The coexistence of large numbers of CD4+ T cells with DCs in early lesions suggest active priming of tumor antigens within the TME, which might explain the futility of CTLA-4 pathway inhibition in our experiments. Although tumor-draining lymph nodes (TDLN) are a key site for tumor antigen presentation, a process enhanced by PD-1 pathway blockade, there is evidence that DCs can also present antigens to T cells within the tumor itself (Broz et al., 2014). The primary experimental approach to determine the contribution of intratumoral T cells vs. T cells egressing from TDLN is the use of sphingosine 1-phosphate (S1P) receptor inhibitor FYT720, which blocks the migration of T cells through the lymphatic system (Topalian et al., 2020). An expansion of T cell clones in blood after neoadjuvant treatment would be indicative of system response and could be used as an early marker of therapeutic efficacy.

Beyond the evidence of a poorly-immunosuppressive TME, the fact that tumor cell numbers *per se* are much lower in residual micrometastases, might partially explain their particular susceptibility to adaptive immunity. Cancer cells themselves can impair anti-tumor immunity and killing of fewer rival cells must be easier. This is, for example, reflected in T cell cytotoxic in vitro assays, where increasing tumor cell numbers correlate with their survival. In fact, clinical data indicate that tumor burden, defined as the total amount of cancer in the body, is a major predictor of immunotherapy success in a broad range of cancers (Dall'Olio et al., 2022; Kim et al., 2021). To date, we have been unable to discern the effects of an increasing number of tumor cells vs. changes in the TME on shaping anti-tumor immunity, since these two are heavily interconnected.

### *The tumor microenvironment of advanced CRCs*

Advanced primary and metastatic CRC tumors showed a paucity of T cells, which were retained in the stroma that surrounds tumor cell glands. These features suggest an existing anti-tumor response that is nonetheless rendered ineffective by the inactivation and retention of immune cells, a process largely mediated through TGFβ in CRC (Tauriello et al., 2018). Certain cell-types, such as immunosuppressive macrophages and endothelial cells, accumulated in grown tumors regardless of the organ of origin. Yet interestingly, CAFs were abundant in primary tumors, whereas neutrophils were enriched in liver metastases. We suggest that rewiring of the TME in advanced cancers is orchestrated by different cell-types depending on the organ site, which will have implications for the development of new drugs targeting the TME. The analyses of this scRNAseq dataset is ongoing and so far we have focused on macrophages and fibroblasts.

Macrophages were the most abundant non-cancerous cell-type in our samples, in agreement with recent data from CRC patients (Lee et al., 2020). Single-cell transcriptomic analysis recently identified two main subsets of TAMs in CRC that support angiogenesis, immune evasion and tumor progression (Zhang et al., 2020). We identified similar populations of TAMs, expressing inflammatory or angiogenic genes, in primary tumors and liver metastases. Interestingly, samples from small metastases contained a similar number of macrophages, but the vast majority were Cd14+ monocytes. Monocytes displayed low expression of genes associated with poor-prognosis in CRC compared to TAMs, suggesting that corruption of this cell-type occurs gradually. Moreover, our data suggest that TAMs in liver metastases derive from monocyte precursors and not tissue resident cells, although lineage tracing experiments would be required to confirm this. Intriguingly, a recent article suggested that liver metastases diminish immunotherapy efficacy systemically through macrophage-mediated elimination of T cells (Yu et al., 2021). In fact, we have observed that liver metastases burden skewed the TME into immunosuppression.

In the presence of a large primary tumor, micrometastases are highly infiltrated with T lymphocytes. However, when overt liver metastases are present, infiltration with T cells markedly decreases. These data suggest that certain secreted factors or cell populations rewire the TME of lesions elsewhere. A caveat of these results is that with 2D analyses metastases size cannot be accurately measured. For example, metastases classified as micro and small might be in fact bigger metastases that were cut in the edge. Thus, in livers with higher metastatic burden, the probability of misclassifying a big metastases as a smaller one is bigger. Analyzing T cell infiltration with 3D imaging would help to confirm whether microlesions in livers with high metastatic burden are in fact less infiltrated.

CAFs are a key component of the tumor microenvironment with diverse functions, including matrix deposition and remodelling, signalling interactions with endothelial and cancer cells and modulation of the immune system. The expression of fibroblast-marker genes is a hallmark of the poor-prognosis CMS4 subtype in CRC, thus highlighting the importance of this cell population in tumor progression (Guinney et al., 2015). In contrast to TAMs, CAFs from primary tumors and liver metastases formed distinct subpopulations in UMAPs, suggesting a diverse origin. Ongoing research in our laboratory is directed towards shedding light on the origin of CAFs in CRC. Our data points to the coexistence of two or three subsets of CAFs in primary tumors and hints that they might originate from different tissue-resident fibroblasts, given the conserved expression of marker genes for universal fibroblasts and mesothelial cells.

### *Dynamic classification of stromal subtypes*

Over the last decade, large sequencing efforts combined with bioinformatic analyses have led to the classification of tumor types into biological subtypes that differ in mutational background, metabolism, and immune activation (Guinney et al., 2015). Importantly, the classification of tumors based on bulk RNA expression is useful to predict outcome in patients and response to treatments (Dienstmann et al., 2017). In CRC, CMS4 tumors are characterized by an inflamed, but immune evasive, microenvironment that defines poor patient prognosis and resistance to immune checkpoint blockade (Lee et al., 2020). Here, we conceptualize that a single patient might belong to different biological subtypes during the course of cancer. The TME is a dynamic arena in which tumor cells interact with resident and recruited host cells. The outcome of these bidirectional interactions is a changing landscape that regulates, among a variety of processes, anti-tumor immunity. The microenvironment of early-stage tumors tends to exert anti-malignancy functions, whereas that of late-stage tumors tends to exert pro-malignancy functions. We hypothesize that this is caused by the necessity to rebuild and rewire *de novo* an immunosuppressive TME after arrival at the new soil. Moreover, our previous discoveries on the

dynamics of tumor cell phenotypes suggest that the crosstalk between tumor cells and the TME varies across space and time.

On the other hand, tumor mutational burden (TMB) is mostly a stable trait in human cancers (Sottoriva et al., 2015). CRC patients with mismatch repair deficiency or POLE mutations accumulate tens of mutations per megabase, whereas MSS CRC harbor only a handful. Nonetheless, even responses to ICB in MSI tumors vary depending on tumor stage and the TME (Chalabi et al., 2020; Mariathasan et al., 2018). This observation reflects that anti-tumor immunity is a balance between several parameters (Chen and Mellman, 2017). An important exception might be immune-desert tumors, which are considered unrecognized by adaptive immunity due to insufficient neoantigens and therefore should not be considered as candidates to benefit from immunotherapies. Importantly, Linneman et al. showed that tumor-specific immune responses occurred in melanoma patients with as few as ten somatic mutations per megabase of coding genome. These data support the notion that tumors with low TMB can also express neo-epitopes that can be recognized by T cells. Discerning which tumors have a low TMB but might benefit from ICB therapies once the immunosuppressive context is reverted, from those with even lower TMB that are uncapable of recognition is highly relevant for the design of clinical trials testing immunotherapies.

## Chapter III: Niche dependencies drive metastatic latency in colorectal cancer

Colorectal cancer has been classically viewed as a multistep mutational disease, where multiple mutations or 'hits' are required for the transformation of normal colon epithelium into invasive and metastatic colorectal cancer (mCRC). The most common genetic alterations during CRC progression occur in genes that regulate the key signaling pathways that allow colon stem cell expansion. Mutations in WNT, EGFR, P53 and BMP/TGFβ pathways render tumor stem cells independent of the crypt niche, a process that facilitates growth in foreign environments (Batlle & Clevers, 2017). Supporting this view, we and others showed that colon organoids engineered with four driver mutations (Apc, Kras, P53 and Tgfbr2 or Smad4; AKPT/AKPS) can be cultured in the absence of stem cell factors and, upon inoculation in mice, they generate metastatic outgrowths in the liver and lungs with high efficiency (Fumagalli et al., 2017; Matano et al., 2015; O'Rourke et al., 2017; Tauriello et al., 2018). However, we discovered that the vast majority of human primary mCRCs and metastases carry mutations in only two or three driver pathways. Many laboratories have focused on the study quadruple tumors; however, given that most mCRCs accumulate fewer mutations, we decided to study the biology of triple mutant tumors, which had remained overlooked.

During metastatic dissemination, tumor cells lose niche signals from the surrounding TME, which they gradually rebuild again in foreign organs. We modeled this scenario by removing specific niche-factors from the organoid culture medium. Interestingly, TGFβ supplementation or EGF withdrawal caused the growth-arrest of AKP and APT MTOs, respectively, yet organoids remained viable for several days and resumed growth upon reconstitution of the full Stem Cell (SC) medium. We hypothesized that tumors with an incomplete niche independence enter such a latent state during metastatic dissemination until tumor-promoting signals are developed, and this might explain the timing of disease-free survival intervals in patients. To test this hypothesis we investigated the ability of triple mutant tumors (AKP and APS/APT) to colonize the liver, the most common site of CRC metastases. Intrasplenic inoculation of quadruple mutant AKPT MTOs in c57BL/6 mice produces rapidly growing liver metastases that kill the host in only 3-4 weeks (Tauriello et al., 2018). In contrast, APS and AKP MTOs also colonized the liver, yet bioluminescence analysis showed that they entered a prolonged latency that was eventually followed by exponential growth. Importantly, we demonstrated that triple mutant tumors do not acquire a 4$^{th}$ hit *in vivo* that enables their growth.

To gain insights into the latent state of triple mutant tumors, we performed RNA sequencing in suboptimal niche conditions compared to full SC media followed by gene

set enrichment analysis (GSEA) (Subramanian et al., 2005). The latent phenotype in both MTO genotypes was characterized by silencing of proliferation gene programs and general downregulation of metabolic pathways including glycolysis, oxidative phosphorylation and fatty acid metabolism. Extracellular matrix remodeling and sensing genes were also upregulated in both AKP and APS MTOs. Our attention was particularly drawn by the observation that the Mex3a+ cell signature, a SC lineage showing enhanced resistance to radiotherapy in the normal intestine (Barriga et al., 2017), was amongst the common gene sets enriched both in latent AKP and APS/T MTO cultures. Furthermore, Mex3a mRNA was upregulated in triple mutant CRC genotypes. This set of results is now part of an article that describes how triple mutant tumors are enriched in a slow-proliferating subset of cells, which are marked by the expression of Mex3a and originate relapse after chemotherapy treatment (Alvarez-Varela et al., under revision).

### *Challenges to model slow-growing tumors in immunocompetent mice*

Due to time-constraints of diverse nature (journal and grant submissions, PhD thesis, etc...), the scientific field often favors the use of experimental models that provide timely results. Consequently, experimental models in oncology are heavily biased toward rapidly proliferating tumors, which may not faithfully mimic the great diversity of human tumors. An illustrative example is the widespread use of quadruple mutant tumors to model CRC over the last years, leaving behind other mutational backgrounds that are barely being studied. Here, we attempted to characterize models with fewer oncogenic mutations, yet we encountered a number of technical difficulties.

In our opinion, the characterization of DTCs should be performed in spontaneous metastases from primary tumors, given that intrasplenic inoculation of tumor cells misses several bottlenecks in the metastatic cascade and therefore, phenotype selection might be lost. In addition, the injection of thousands of cells results in increased metastatic burden and cell death, which generate substantial inflammation, which in turn may alter the microenvironmental landscape of small lesions. Unfortunately, the generation of primary tumors with most double and triple mutant MTOs failed due to immune rejection of transplanted cells. As we previously discussed, injection of cancer cells results in antigen-exposure under inflammatory conditions, and prior to the presence of a mature TME. By comparing engraftment rates across organs, we observed the lowest engraftment in the caecum compared to the skin or liver. The intestine is subject to continuous exposure to exogenous agents that probably result in higher levels of baseline activation of the immune system. We speculate that in slower growing tumors, tumor burden (and the TME) might not have had sufficient time to increase and therefore tumors succumb

against immunity. Importantly, we consider that these differential responses are artefactual, since during carcinogenesis anti-tumor immunity is most likely built as cancer cells accumulate mutations and recruit stromal cells. Therefore, we have focused on working to improve our model.

To circumvent immune rejection we tested a large number of alternatives. The most faithful method would be to achieve the generation of a single tumor in GEMMs, where mutations are induced in lower number of cells in their native environment. However, our attempts — which consisted of inducing such tumors by injecting 4-OH locally in the tip of the caecum to avoid the widespread development of tumors — were unsuccessful. Alternatively, transplantation of fully-grown tumors prevented immune rejection but subsequent surgery to remove the primary tumor became impossible due to attachments of sutures in other organs and connective and adipose tissue. Given that we had observed lower recruitment of neutrophils in Kras wild-type tumors, we also tested in vain injections in obese animals, which have been shown to have systemic expansions in neutrophils (Quail et al., 2017). Overall, we consider that further research into tumors with an incomplete set of driver mutations is necessary; however, development of more sophisticated models will be required.

### *Biology of DTCs across mutational backgrounds*

In future experiments, we would like to profile the latent state that triple tumor enter when they reach the liver. In the first chapter of this thesis, we discovered that Emp1+ HRCs migrate into the liver and rapidly revert to proliferative cells. We have data supporting a migratory role for HRCs in AKP tumors as well, but we would like to explore how the timing and nature of lineage reversion changes across mutational backgrounds. We speculate that niche-dependent tumors enter a quiescent state, HRC+ or Mex3a+, and the timing of reversion into proliferative states determines disease-free survival in patients.

### *Evolution of the TME across time and size*

The successful development of such experimental models would allow us to characterize the evolution of the TME in slower proliferating tumors. Based on our experience on the temporal and size-dependent evolution of the TME, special precautions should be taken when comparing tumor types with different growth kinetics. It is highly relevant to understand whether the T cell-rich environment observed in quadruple mutant tumors is due to short time after seeding or the small size of metastases. In other words, how is the TME of cancer cells that reach the liver and remain quiescent for long time?

Does it become immune excluded? Does immunity play a role in metastatic latency in CRC? Or are immune cells less attracted due to lower expression of MHC molecules in non-proliferative cells? We have also observed that, although TGFβ -sensitive tumors proliferate slower, they upregulate PD-L1 in response to TGFβ treatment, which might protect invasive cells that encounter effector T cells. The behavior of CRC that have certain niche dependencies might explain the timing from surgery to relapse. Those patients with longer latent intervals have a wider therapeutic window and it is therefore highly relevant to understand whether adjuvant immunotherapies would also eliminate these residual tumors.

# Conclusions

Our study unveils the biology of residual disease in CRC, both from a tumor-cell and a TME perspective. The main conclusions of this work are the following:

**Chapter I**

1. High Relapse Cells (HRCs) are a defined tumor cell subset in CRC primary tumor samples associated with poor-prognosis.

2. HRCs disseminate out of the primary tumor prior to surgical resection and colonize metastatic organs.

3. HRCs are neither differentiated nor stem-like, but rather represent a distinct state that, without undergoing epithelial to mesenchyme transition, enables migration to foreign organs.

4. The HRC state is reverted in foreign organs, reacquisition of the Lgr5+ stem cell and proliferation programs occur at a later phase and are necessary for metastatic outgrowth.

**Chapter II**

5. The capacity to evade adaptive immunity fluctuates during metastatic dissemination.

6. Residual metastases are heavily infiltrated with T cells and contain few stromal cells.

7. The tumor microenvironment landscape changes depending on the organ and the stage of tumors.

8. Recruitment and corruption of neighboring host cells occurs gradually, leading to a time and/or size-dependent evolution of the tumor microenvironment.

9. Disseminated HRCs express high levels of IFN-signature genes and more specifically upregulate Cd274 and Ido1 immunosuppressive molecules.

10. Neoadjuvant immunotherapy prevents metastatic recurrence in MSS CRCs.

11. Adjuvant immunotherapy is ineffective against overt metastases in MSS CRCs.

**Chapter III**

12. Most metastatic CRC accumulate driver mutations in two or three pathways controlling self-renewal and proliferation in the intestinal stem cells. Only a minority of patients contain four pathway mutations.

13. An incomplete set of driver mutations result in niche dependencies that induce a slow-growing phenotype with metabolic shutdown.

14. Instruction of a TGFβ-activated TME is not dependent on loss of TGFβ-sensitivity by tumor cells.

15. KRas wild-type tumors accumulate fewer neutrophils in the stroma.

16. TGFβ-sensitive tumors generate liver metastases that progress with slower kinetics.

# Materials and Methods

### Mouse tumor organoid (MTO) culture

We described previously (Tauriello et al., 2018) the establishment of MTOs from primary tumors arising in GEMMs with combined distinct genetic alterations. In brief, triple mutant AKP MTOs were established from CRCs arising in Lgr5-creERT2; Apc$^{fl/fl}$; Kras$^{LSL-G12D}$; Trp53$^{fl/fl}$ mice; ATP MTOs from Lgr5-creERT2; Apc$^{fl/fl}$; Tgfbr2$^{fl/fl}$; Trp53$^{fl/fl}$ mice and AKPT in Apc$^{fl/fl}$; Kras$^{LSL-G12D}$; Trp53$^{fl/fl}$ ; Tgfbr2$^{fl/fl}$ mice. APK+ Smad4 were generated by introducing a mutation in the SMAD4 locus in APK MTOs using CRISPR/Cas9 as previously described (Tauriello et al., 2018). MTOs were cultured as detailed by (Tauriello et al., 2018) and checked bimonthly for mycoplasma contamination.

### MTO latency

Complete stem cell medium corresponds to standard medium used for CRC organoid cultures(Tauriello et al., 2018). It is supplemented with 10 mM HEPES, Glutamax, B-27 without retinoid acid (all Life Technologies), 50 ng/ml recombinant human EGF (Peprotech), 100 ng/ml recombinant human NOGGIN and 1 µM galunisertib (LY2157299) in advanced DMEM/F12. For no EGF conditions, we used the same medium without supplementing EGF. In +TGFβ conditions, TGFBR1 inhibitor galunisertib was removed from the medium and 5 ng/ml of active TGFβ1 (Peprotech, 100-21-B) were added.

### CRISPR genome editing of MTOs

We used CRISPR-Cas9 technology to insert reporter and ablation cassettes into the locus of Emp1 and Lgr5. Small guide RNAs (20 bp) were designed to cut 9-11 base pairs after the STOP codons using the http://crispr.mit.edu web tool and were cloned into a pX330-IRFP hSp-enhanced-Cas9 plasmid (Cortina et al., 2017). The sgRNA sequence for Emp1 is “AAATAAGCCGAATACGCTCA” and for Lgr5 “GTCTCTAGTGACTATGAGAG”. 1kb 5’ and 3’ homology arms were amplified by PCR from MTO gDNA or synthesized by gene synthesis (Thermo Fisher) and sequentially cloned into pShuttle vectors containing the inducible Caspase9-tdTomato cassette (Morral et al. 2020), which was a kind gift from Toshiro Sato (Shimokawa et al., 2017). IRES-DTR-T2A-EGFP-WPRE-BGHpA sequence was ordered by gene synthesis at Thermofisher and cloned between Lgr5 homology arms flanking the gene stop codon. EGFP was amplified by PCR and cloned after IRES to generate the LGR5-IRES-EGFP-WPRE-BGHpA donor. CRISPR-Cas9 knock-in editing was carried out as described previously(Cortina et al., 2017). Briefly, organoids were nucleofected using a Nucleofector 2b (Lonza) in combination with the Cell Line Nucleofector Kit V (Lonza). Trypsinized organoids (1.0–1.5 × $10^6$ cells per guide) were resuspended in 100 µl nucleofection buffer mix containing 3 µg sgRNA and 7 µg donor plasmid, and nucleofected using program A32. 2 days after nucleofection with guide-

eCas9-IRFP+ and donor plasmids, iRFP+ single cell clones were FACS-sorted and after expansion screened by PCR for correct integration. Genotyping primers are detailed in Supplementary Table 2. AKTP-MTO93 Emp1-iCT was generated from a single cell tdTomato+ (clone#14). MTO93 Emp1-iCT Lgr5-EGFP was first generated by a single cell tdTomato+ (clone#49), then nucleofected again with the Lgr5 construct and a pool of GFP+ cells were sorted. MTO93 Lgr5-iCT was generated by a single cell tdTomato+ (clone#2). MTO93 Lgr5-DTR-GFP was generated by sorting a pool of GFP+ cells.

### Introduction of driver mutations in organoids using CRISPR

Guides were designed and cloned into px330-U6-Chimeric BB-CBh-hSpCas9 (Addgene: #42230), which was modified by introducing a SV40 promoter–IRFP expression cassette downstream of hSpCas9. Guide sequences: mo_Smad4#1:AGACAGGCATCGTTACTTGT; mo_trp53#16:AGTGAAGCCCTCCGAGTGTC, mo_kras g4:GACTGAGTATAAACTTGTGG. For donor plasmid construction, 750bp of 5´ and 3´ MmKRAS G12D homology arm (HA) flanking the knock-in insertion cassettes were synthetized by gene synthesis (Thermo Fisher) and cloned into pShuttle or pDONOR vectors. Organoids were nucleofected as explained in (Tauriello et al., 2018) and (Cortina et al., 2017). For Smad4 knockout mutants, three days after nucleofection, growth medium was exchanged for selection medium (+ TGFβ1 10 ng/ml and removal of Galunisertib). For p53 KO, organoids were selected using nutlin3 (MedChem, HY-50696) at 20 μM. For Kras G12D mutations, organoids were selected by seeding without EGF, and supplemented with Gefinitib (1 μM, Santa Cruz Biotechnology).

### Lentivirus production and organoid infection

For bioluminescent tracking, MTOs were infected with a lentivirus encoding an eGFPwwfirefly luciferase fusion reporter construct, followed by a WPRE sequence to enhance expression, cloned under the control of the PGK promoter. AKTP-MTO93 used for residual disease profiling was further infected with lentivirus encoding an eGFP construct controlled by an SV40 promoter to enhance fluorescence. Double-labelled Emp1-iCasp9-Tom, Lgr5-GFP organoids were infected with lentivirus encoding firefly-luciferase followed by an IRES-PURO resistance cassette, cloned under control of the CMV promoter. Virus were produced using packaging constructs in HEK293T cells in DMEM 10% FBS medium, filtered through 0.45μm, and concentrated with Lenti-X-concentrator (Clontech Takara Bio 631231) according to manufacturer specifications. Trypsinized organoids were suspended in ultra low attachment plates (Corning) in MTO medium and treated with two successive rounds of virus-containing medium in the presence of 8 μg/ml polybrene. After lentiviral infection, MTOs were seeded back in BME drops and selected with 0.5-1 μg/ml puromycin (InvivoGen) or by flow-cytometry sorting of GFP+ cells.

### Ethics and animal maintenance

All experiments with mouse models were approved by the Animal Care and Use Committee of Barcelona Science Park (CEEA-PCB) and the Catalan government. Mice were maintained in a specific-pathogen-free (SPF) facility with a 12-h light–dark cycle and given ad libitum access to standard diet and water.

### Mouse injections

For all injections, c57BL/6J mice were purchased from Janvier Labs at six weeks of age and injected at 7 to 9 weeks of age. Sex always matched the origin of the tumor. Intra-caecum injections were used for the generation of primary tumors. Organoids were harvested and incubated for 30 minutes with cold HBSS to break down BME, without disrupting their structure. Cells were then counted and resuspended in 70% BME in HBSS for injection at a concentration of 0.1 X $10^6$ cells in 10uL per mouse. Full organoids were injected with a 30G syringe into the submucosal wall of the distal caecum while looking through binocular lens. We introduced a significant modification to previous protocols (Céspedes et al., 2007) by moving the injection site to the apex of the caecum, which allowed posterior surgical resection. For liver colonization studies we used intrasplenic injections of organoids as described before (Tauriello et al., 2018).

### Bioluminescence imaging

Growth kinetics of luciferase-expressing MTOs were tracked by in vivo bioluminescence, using an IVIS Spectrum (Perkin Elmer) imager as previously described (Tauriello et al., 2018). For quantification, radiance per minute in the area of interest- lower thorax (primary tumor) or lower thorax and upper abdomen (IS) - was calculated using Living Image software (Perkin Elmer). Data were processed and visualized with R/RStudio and ggplot2. Depicted are longitudinal curves, connecting measurements of individual mice, and the group smooth (LOESS, span = 0.5).

### Primary tumor resection

Mice were anesthetized with isoflurane and placed in dorsal recumbency. The abdomen was shaved and sterilized with povidone-iodine surgical solution. A small midline incision - slightly to the left- was performed to open the skin and peritoneum and expose the abdominal area. We placed a sterile surgical drape on top of the abdomen with a circular hole above the incision and sprawled the caecum over the drape using cotton swabs and saline to keep it hydrated. After confirming the presence of a primary tumor, Kelly forceps were used to first knot the surgical suture into the caecum wall, in between the ileocecal junction and the primary tumor. This provided a grip for subsequent caecum

ligation. After ligation, the apical caecum containing the primary tumor was excised and any remaining cecal tissue was trimmed. After resection and organ fixation, we measured primary tumor size using a caliper. We provide a video of the surgery, which usually lasted from 5 to 10 minutes (Supplementary video 1). Fitness of mice was monitored bi-weekly throughout the experiment. Mice were euthanized four weeks after resection and metastasis were scored macroscopically.

### Mouse treatments

For iCasp9-inducible ablation experiments, animals were treated with dimerizer (AP20187, Medchem express, #HY-13992) via IP injection 2.5 mg/kg 3 times per week (Emp1-iCasp9) or 5 mg/kg every day (Lgr5-iCasp9). For DTR-inducible ablation mice were treated with 16.7 µg/kg of diphtheria toxin (Sigma-Aldrich, # 322326) three times per week. 200 µg of Rat αCD8α (YTS 169.4; BioXCell BE0117) or Rat αCD4α (GK1.5; BioXCell BE0003-1) or Rat IgG2b (LTF-2, BioXCell BE0090) isotype control antibodies were injected intraperitoneal for in vivo T cell depletion. All routinely validated and confirmed by flow cytometry analysis on peripheral blood (by us). 250 µg of anti-mouse CTLA-4 (C2444, Leinco) and anti-mouse PD-1 (P372, Leinco) via intraperitoneal injection were used for immunotherapy.

### Tumor dissociation for flow cytometry

Primary tumors and micro-dissected liver metastases were chopped with razor blades. Subsequent enzymatic digestion was performed with 200 U/ml collagenase IV (Sigma Aldrich) in HBSS (Lenovo) for 30 min at 37 °C, in a shaking water bath or a gentleMACS Dissociator (Miltenyi Biotec). Digested tissue fragments were then filtered through 100- and 40-µm meshes, washed, and treated for 5 minutes with ammonium chloride. Single-cell preparations were first blocked with anti-CD16/32 (clone 93; eBioscience) and then stained with APC anti-Epcam (118214; Biolegend) and BV605 anti-CD45 (103155, Biolegend) antibodies. Finally, cells were resuspended in HBSS with 0.5% FBS and DAPI (Sigma Aldrich).

### Isolation of residual DTCs

For isolation of low numbers of tumor cells from mice without visible metastases, whole livers were thoroughly minced with razor blades. After an initial 30-minute digestion with 200 U/ml collagenase IV, samples were filtered through 100 µm meshes. Although most cells flowed through, a small fraction of the sample –highly enriched for tumor cells- was retained in the filter (Supplementary Figure 5c). Filters were then laid into a 6-well plate and covered with HBSS containing collagenase IV, Dispase II and DNase I.

After re-digestion in a water bath for 30 minutes al 37 °C, most cells now seeped through the filters. The protocol continued with washing, ammonium chloride treatment and antibody stainings as described above. DAPI- GFP+ CD45- cells were gated to sort tumor cells.

### Histology and tissue stainings

Standard hematoxylin eosin and antibody staining were performed on 4-µm tissue sections using standard procedures, as described previously (Tauriello et al., 2018). Briefly, after antigen retrieval, samples were blocked with Peroxidase-Blocking Solution (Dako, S202386) for 10 min at room temperature, and primary antibodies were then incubated with samples at corresponding times and temperatures (see Supplementary Table 3). Slides were washed with EnVision FLEX Wash Buffer (Dako, K800721), and incubated at RT 1:400 for 1h with the corresponding secondary antibodies: donkey anti-goat conjugated to Alexa 488/568/647 (Life Technologies A11055, A11057, A21447), donkey anti-rabbit conjugated to Alexa 488/568/647 (Life Technologies A21206, A10042, A31573) and donkey anti-mouse conjugated to Alexa 488/568/647 (Life Technologies A-21202, A10037, A31571) at RT. DAPI was added at 1:2500 after secondary antibody incubation and slides were mounted with Fluorescent mounting medium (Dako, 53023). Digital scanned bright-field and fluorescent images were acquired with a NanoZoomer-2.0 HT C9600 scanner (Hamamatsu, Photonics, France) equipped with a 20X objective and using NDP.scan2.5 software U10074-03 (Hamamatsu, Photonics, France). Fluorescent images were acquired with a mercury lamp unit L11600-05 coupled to the NanoZoomer. All images were visualized with the NDP.view 2 U123888-01 software (Hamamatsu, Photonics, France) with a gamma correction set at 1.8 in the image control panel.

### In Situ Hybridization (ISH)

Samples from patients were provided by the HCB-IDIBAPS Biobank (B.0000575), integrated in the Spanish National Biobanks Network and they were processed following standard operating procedures with the appropriate approval of the Ethics and Scientific Committees. Paraffin-embedded tissue sections (2-3 µm in thickness) of human CRC primary tumors were air dried and further dried at 60 °C over-night prior any staining. Sections were hybridized with RNAscope® Probe Hs-LGR5 (ref: 311021, Bio-Techne R&D Systems) in C1 channel, a custom-made RNAscope® Probe Hs-EMP1 (ref: 895051-C2, Bio-Techne R&D Systems) in C2 channel and an Alexa568-conjugated antibody against E-Cadherin. FISH probe were detected using the RNAscope® Multiplex Fluorescent Detection Reagents Kit v2 (ref: 323110, Bio-Techne R&D Systems). Epitope retrieval was performed at 100°C for 15 min with Target Retrieval Reagent followed by protease Plus treatment (30 min at 40°C). Probe hybridization, signal amplification and colorimetric

detection were subsequently performed using manufacturer recommendations. For colorimetric detection, Opal™ 650 (FP1496001KT, Akoya Bioscience) was used at 1:500 for C1 channel and Opal™ 520 (FP1487001KT, Akoya Bioscience) was used at 1:500 for C2 channel. Fluorescent images were acquired with a mercury lamp unit L11600-05 coupled to the NanoZoomer.

### Multiplex immunofluorescence

Multiplex immunofluorescence staining was performed on 4-µm-thick formalin-fixed paraffin embedded sections using the OPAL protocol (Akoya Biosciences, Marlborough, MA) on the Leica BOND RXm autostainer (Leica Microsystems, Wetzlar, Germany). Six consecutive staining cycles were performed using the following primary antibody-Opal fluorophore pairings. Stroma panel: CD34 (1:3000, ab81289; Abcam)–Opal 520; CD146 (1:500, ab75769; Abcam)–Opal 570; SMA (1:1000, ab5694; Abcam)–Opal 620; Periostin (1:1000, ab227049; Abcam)–Opal 690; and E-cadherin (1:500, 3195; Cell Signaling)–Opal 650. Immune panel: (1) Ly6G (1:300, 551459; BD Pharmingen)–Opal 540; (2) CD4 (1:500, ab183685; Abcam)–Opal 520; (3) CD8 (1:800, 98941; Cell Signaling)–Opal 570; (4) CD68 (1:1200, ab125212; Abcam,)–Opal 620; (5) FoxP3 (1:400, 126553; Cell Signaling)–Opal 650; and (6) E-cadherin (1:500, 3195; Cell Signaling)–Opal 690. Primary antibodies were incubated for 60 minutes and detected using the BOND Polymer Refine Detection System (DS9800; Leica Biosystems, Buffalo Grove, IL) according to the manufacturer's instructions, substituting DAB for the Opal fluorophores, with a 10-minute incubation time and without adding hematoxylin. Antigen retrieval was performed at 100°C for 20 minutes, in accordance with standard Leica protocols, with Epitope Retrieval (ER) Solution 1 or 2 was performed before each primary antibody was applied. Sections were then incubated for 10 minutes with spectral DAPI (FP1490, Akoya Biosciences) and the slides mounted with VECTASHIELD Vibrance Antifade Mounting Medium (H-1700-10; Vector Laboratories). Whole-slide scans and multispectral images (MSI) were obtained on the Akoya Biosciences Vectra Polaris. Batch analysis of the MSIs from each case was performed with the inForm 2.4.8 software provided. Finally, batched analysed MSIs were fused in HALO (Indica Labs, Albuquerque, NM) to produce a spectrally unmixed reconstructed whole-tissue image, ready for analysis.

### Statistical analyses multiplex immunofluorescence

HALO software was used for quantification of cell phenotypes within metastases and primary tumors. To obtain metastases at various stages, mice were implanted with primary tumors and sacrificed at several timepoints post-injection. Then, bioluminescence was measured *ex vivo* and the number of visible metastases were counted for each liver, which allowed us to classify them based on metastatic burden. Residual disease livers

looked apparently healthy but contained a small number of micro-lesions identified a posteriori through histology. Livers with early-stage disease contained visible metastases that ranged from 0.5 to 1.5 mm in size, and late-stage disease animals contained overt metastases bigger than 4 mm. importantly, late-stage disease livers contained a combination of large, small and micro-metastases, owing to differential timing of dissemination. As initial pre-processed data, two counts matrices (distinguishing immune and stromal panels) with the total number of cells per metastasis/primary (in rows) assigned to each population (in columns) were obtained. For each metastases information about metastasis size (nº of total cells), metastatic burden and organ site were used. Multiple positives were present in both immune and stromal panels. These cases observed in the immune panel only represented the 0.7% of the total assignations and were removed from the analysis. For the stromal panel, aSMA and POSTN double positives were kept and labeled in a distinct category aSMA/POSTN. The rest of multiple positives, that accumulated the 7% of the total cells, were amalgamated and labeled as stromal others. Even though the stromal others counts were used for the normalization of the whole matrix of counts, they were not considered into the graphical representation and interpretation of the results. An evolution plot showing the average population percentages over an increase in total cells distinguishing between micro, small and big metastasis was generated using the ggstream R package. We considered the percentages of Thelp, Treg, Tctx, Neutro and Macro for the immune panel, and the percentages of CD146, CD34, aSMA, POSTN and aSMA/POSTN for the stromal panel, thus ignoring Epithelial and unassigned cells. These percentages were averaged out by the interaction of covariables Definition (identifying micro, small and big metastasis) and Total cells (in logarithm scale), and were shown using the geom_stream function with parameter type = "proportional". This option forced the percentages to be rescaled to 1 for the whole Total cells range. To account for the compositional nature of the data, the zero values of the counts matrices were replaced by pseudo-counts using the cmultRepl function of the zCompositions R package (Palarea-Albaladejo and Martín-Fernández, 2015). Such imputed matrices were then CLR transformed using the compositions R package (van den Boogaart and Tolosana-Delgado, 2008). In this way, classical statistical analysis to the transformed observations could be applied (Aitchison, 1986). Linear mixed effects models were fitted independently for every population using the CLR-transformed values as response variable, the Definition and the Total cells in log scale as fixed effects, and the tumor Id (the tumor identifier) as random effect to take into account the dependence between metastasis for the same tumor.

### Liver Optical Clearing

Whole livers containing GFP-expressing tumor cells were optically cleared for 3D imag-

ing following the "CUBIC protocol I" described by Tainaka et al. (Tainaka et al., 2018). Mice were sacrificed by an overdose of pentobarbital, and transcardially perfused first with 15 mL of PBS, followed by 20 mL of 4% PFA-PBS and an additional 15 mL of PBS to rinse the PFA. Once fixation was completed, livers were cleared with the perfusion of 30 ml CUBIC-P (a mixture of 5 wt% 1-methylimidazole (Tokyo Chemical Industry, M0508), 10 wt% CU#0414 (N-Butyldiethanolamine, Tokyo Chemical Industry, B0725) and 5 wt% Triton X-100 (Nacalai Tesque, 12967-45) during 10 minutes. The livers were then dissected, cut into small pieces of approximately 3-4 $cm^3$ and immersed in CUBIC-L (a mixture of 10 wt% CU#0414 and 10 wt% Triton X-100), to be incubated with shaking at 37°C for 6-7 days. Since the immersion period was longer than 4 days, CUBIC-L was refreshed at least once. After delipidation, the pieces of liver were washed in PBS with gentle shaking at room temperature for 6 hours. The pieces were then immersed in 1:1 water-diluted CUBIC-RA (a mixture of 45 wt% CU#0640 (antipyrine, Tokyo Chemical Industry, D1876) and 30 wt% CU#1283 (N-methylnicotinamide, Tokyo Chemical Industry, M0374) recommended for organs with fluorescence) and incubated with gentle shaking at room temperature overnight. The organs were finally immersed in CUBIC-RA with gentle shaking at room temperature for at least 2 days before imaging.

### Lightsheet Imaging

Liver samples were imaged on a custom lightsheet microscope, MacroSPIM (used previously in (Huang et al., 2015)). In brief, cleared samples were mounted on 0.2mm stainless steel Austerlitz insect pins (Entomoravia, Czech Republic) held by a 1cm silicon pad (made with double component silicone glue, Twinsil by Picodent, Wipperfürth, Germany), positioned inside a custom quartz cuvette (WxDxH; 2x2x3.5cm). Lightsheet imaging was performed horizontally, and samples were illuminated with a pivot-scanned lightsheet of waist ranging from 4 to 10µm, adjustable to the field of view, formed by a 50mm focal length cylindrical lens (ACY254-050-A, Thorlabs, Newton, NJ, USA). GFP fluorescence was excited by a 488nm laser, autofluorescence by a 561nm DPSS laser, and detection was made through 525/50 and 609/54 bandpass filters, using a 2x air lens with a AZ100M macroscope (Nikon, Japan), set to image at magnifications of 2.4x or 9.6x, and recorded on a Flash4.0 v2 camera (Hamamatsu, Japan).

### Organoid immunofluorescence

Organoids were dissociated into single cells and plated in 50 µl drops in µ-Slide 8 Well Ibidi plates (Ibidi Cat.No: 80826). Organoid were fixed after 3 days, stained following a published protocol(O'Rourke et al., 2016) and mounted in DAKO fluorescence mounting medium. After immunostaining, organoids were imaged with an Inverted SP5 confocal microscope (Leica), using HCX PL APO CS 40x/1.25 objective, at format of 1024x1024

pixels. Excitation wavelengths were 488 nm (Argon laser), 561 nm (DPSS laser) and 633 nm (HeNe laser). Detection windows were set at 495-557 nm, 576-625 nm and 645-695 nm respectively (PMT detectors). The images were taken in a sequential mode with 45 μm z-stack, at 1-1.5 μm intervals. Images were processed using Image J and Imaris. Brightness and contrast were adjusted for each channel for maximal visibility.

### Quantification of tdTomato GFP fluorescence intensity

For the quantification of percentages of Emp1-tdTomato high and Lgr5-GFP-high cells across different zones of primary tumors and liver metastases of different sizes HALO® IMAGE ANALYSIS PLATFORM was used. Briefly, the epithelial tumor area was classified apart from the stroma, background and necrosis using a random forest algorithm. Single cells were detected and Tomato/GFP fluorescence intensity was measured for every cell. In primary tumors, Tomato-high and GFP-high tumor cells were defined as the cells in the 90$^{th}$ percentile of each sample. In liver metastases, Tomato-high and GFP-high tumors cells were defined as the cells in the 90$^{th}$ percentile for all metastases measured. For Emp1-iCT liver metastases (in Supplementary Figure 8g) and Lgr5-DTR-GFP liver metastases (in Supplementary Figure 13g), tumor cell fluorescence-intensity was analyzed using ImageJ with a custom-made macro. Tumor cells were first detected and isolated using E-Cadherin to create a mask. Tomato or GFP intensity was calculated for every pixel inside the masked area. Then, we plotted the percentage of fluorescence high and low pixels as a function of the area of the metastases (measured in pixels).

### Protein extraction and western blot analysis

Cells were washed twice and collected with cold PBS and lysed in cold RIPA buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.4, 1% NP-40, 0.5% Sodium deoxycolate, 0.1% SDS and 5mM EDTA and 50 mM NaF) supplemented with protease and phosphatase inhibitors (Sigma Aldrich). Cell extracts were centrifuged at 14000 rpm for 10 min at 4°C. The supernatant was kept as the protein extract. Protein content was quantified with the Pierce BCA Protein Assay Kit (23227, Thermo Scientific). Equal amounts of protein per sample were separated by standard SDS-PAGE and transferred to PVDF membranes (Millipore). The membranes were incubated in TBS-T (20mM Tris, 150mMNaCl, 0.1% Tween 20) supplemented with 5% milk for 30 min at RT to block unspecific antibody binding. Primary antibodies SMAD4 (B-8, Santa Cruz, ref: sc7966), or ACTIN (Abcam, ref: ab20272),) were incubated overnight at 4°C. Secondary antibodies were diluted 1:5000 and incubated for 1h at RT with the PVDF membranes. Membranes were washed 3 times with TBS-T 0.1% for 10 min, incubated for 5 min with ECL Prime Western Blotting Detection Reagent (RPN 2232, Amersham Cytiva) and visualized using Hyper Processor (Amersham pharmacia Biotech)

### RNA purification and RT-qPCR

RT-qPCR and Microarrays were used to compare subpopulations of Emp1-high and -low, Lgr5-high and -low cells dissociated from MTO93 organoids grown *in vitro* or *in vivo*. Subpopulations were defined in flow cytometry as the top/bottom 10% in fluorescence expression. RNA from 2000 cells was extracted and retrotranscribed to cDNA as described previously(Gonzalez-Roca et al., 2010). cDNA was purified using a commercial kit (PureLink Quick PCR Purification Kit, Invitrogen). To analyze gene expression changes RT-qPCR was performed using 5 ng of cDNA per each real-time qPCR well. Real-time qPCRs were performed with TaqMan Universal PCR Master Mix (Applied Biosystems 4369016) in triplicates following manufacturer's instructions. Gene expression levels were normalized using the housekeeping genes PPIA or B2M. The following TaqMan assays were used: Tomato-BGHPA (custom made probe; F: GGGCATGGCACCGGCAGCACC, R: CCTACTTGTACAGCTCGTCCATGCC), MmPPIA (Mm002342430_g1), MmB2m (Mm00437762_m1), EGFP (Mr04097229mr), MmEmp1 (Mm00515678_m1), MmLgr5 (Mm0043889_m1), MmSmoc2 (Mm00491553_m1).

### Microarray library preparation

Microarrays were used to compare Emp1-high and Emp1-low cells dissociated from MTO93 EiCT primary tumors grown for 4 weeks. 8 µg of cDNA were fragmented and labelled (GeneChip Mapping 250K Nsp Assay Kit, Affymetrix) according to manufacturer instructions. Array hybridization was performed on using the GeneChip Hybridization, Wash, and Stain Kit (Applied Biosystems). cDNA was hybridized to the microarray for 16 hours at 45°C, washing and staining of microarrays was performed using a GeneAtlas Fluidics Station (Affymetrix, Santa Clara, CA). Finally, arrays were scanned with a GeneChip Scanner GCS3000 (Affymetrix/ThermoFisher Scientific) using default parameters. The CEL files containing the microarray data were generated with the Command Console software (Affymetrix/ThermoFisher Scientific), and were used for downstream bioinformatics analysis.

### Microarray analyses

Samples were processed with the R package oligo (v1.46.0) (Carvalho and Irizarry, 2010). Raw cel files were normalized applying RMA background correction and summarization(Bolstad et al., 2005), by oligo functions fitProbeLevelModel (background = TRUE, normalize = TRUE, target = "core", method = "plm") and rma (default paramaters). Probesets were annotated using the Affymetrix databases (version Clariom_S_Mouse_HT-na36-mm10-transcript). Standard quality controls were considered to identify abnormal samples(Irizarry et al., 2003) regarding: a) spatial artifacts in the hybridization

process (scan images and pseudo-images from probe level models); b) intensity dependences of differences between chips (MvA plots); c) RNA quality (RNA digest plot); d) global intensity levels (boxplot of perfect match log-intensity distributions before and after normalization and RLE plots); e) anomalous intensity profile compared to the rest of the samples (NUSE plots, Principal Component Analyses); f) impact of quality metrics (Eklund and Szallasi, 2008) on expression measures. No samples were excluded from the studies due to quality issues.

Differential expression analysis was performed using a linear model with empirical shrinkage as implemented in the limma R package (Ritchie et al., 2015), taking into account the paired data setting. Benjamini and Hochberg False Discovery Rate (FDR) was used for multiple comparisons correction. Gene set analysis was used to explore the enrichment in custom gene sets. The rotation-based approach for enrichment(Wu et al., 2010) implemented in the R package limma was considered to represent the null distribution. The maxmean enrichment statistic proposed in(Efron and Tibshirani, 2007), under restandardization, was used for competitive testing. To get a measure of the pathway activity in the transcriptomic data, we summarized some signatures (Supplementary Table 4) of interest by a z-score(Lee et al., 2008). For doing so, normalized expression values were adjusted for biological replicate, then centered and scaled gene-wise according to the mean and the standard deviation computed across samples, which were then averaged across all genes included in the signature. In addition, a global signature was computed using all the genes in the expression matrix and used for a-priori centering of the signature scores. This strategy has been proved to be useful to avoid systematic biases due to the gene-correlation structure present in the data, and to adjust by the expectation under gene randomization, i.e., the association expected for a signature whose genes have been chosen at random(Efron and Tibshirani, 2007; Mestres et al., 2018). Only the expression levels of the most variable probe sets per gene were considered for these gene set analyses.

**RNA-Seq data**

RNAseq was used to compare AKP, APS and APT organoids in complete stem cell media compared to +TGFβ or media without EGF respectively. Single end reads were aligned to the mouse genome version mm10 using STAR (2.3.0e) (Dobin et al., 2013). SAM files were converted to BAM and sorted using sambamba (Tarasov et al., 2015) The count matrix was generated with Rsubread (Liao et al., 2019) with the built-in annotation for mm10. DESEq2 (Love et al., 2014) was used for differential expression with fold change shrinkage as implemented in the lfcShrink function.

### Genotype distributions in mCRC patients

Clinical data including MSS status and staging was downloaded from (Yaeger et al., 2018) while mutational status was obtained from the cBioPortal (Cerami et al., 2012) version "crc_msk_2018". Pathway definitions were downloaded from (Sanchez-Vega et al., 2018). Only homozygous deletions and high amplifications were considered as copy number alterations.

### 10X single-cell 3' sequencing

From FACS analysis, single cells were sorted into 1.5 mL tubes (Eppendorf). Cell concentration was adjusted to approximately 1,000 cells/µl and approximately 8,300 sorted cells were used for 3' single-cell gene expression profiling. Cell partition into GEMs was carried out on the Chromium platform (10x Genomics), according to the manufacturer instructions. Briefly, 5,000 cells were targeted for partitioning into GEMS and the reverse-transcribed cDNA was amplified using 12 cycles. The resulting cDNA was quality controlled using a high sensitivity DNA assay on the Bioanalyzer 2100 (Agilent) system. The purified libraries were quality controlled, quantified and used to prepare a 20 nM equimolar sequencing pool. NGS Sequencing was performed at the Centre for Genomic Regulation using two lanes of the HiSeq2500 (Illumina) system and a paired-end 125 nt strategy. In total, 116,819,313 Million paired-end reads were obtained.

### 10X mouse single cell RNA sequencing analysis (Epithelial cells)

Chromium single cell RNA sequencing reads were aligned to a custom refdata-gex-mm10-2020-A transcriptome including the EGFP and Luciferase reporter genes with CellRanger (Zheng et al., 2017) version 4.0.0. The count utility was used with default options to quantify gene expression. The subsequent processing steps and analysis were performed with Seurat package version v4.0.3 (Butler et al., 2018; Hao et al., 2021; Satija et al., 2015; Stuart et al., 2019). A total of 1,330 cells having <20% mitochondrial read content and >3,000 detected genes were considered for downstream analyses. Ribosomal reads, which accounted for a 17% of the total, were removed. The proportion of mitochondrial reads was regressed out during the normalization and variance stabilization of raw counts, which was performed with the sctransform method (Hafemeister and Satija, 2019).

The first 19 principal components were used to obtain the Uniform Manifold Approximation and Projection (UMAP) for visualization purposes. Cells were assigned to clusters using FindClusters Seurat function (resolution = 1.2). SCT transformed counts were further imputed and smoothed with magic (v.2.0.3) (van Dijk et al., 2018a). The expression of gene signatures was summarized by taking the average magic score of its constituent

genes. HRCs and Lgr5+ cell populations were defined by identifying those cells having a score above the 75th percentile. FindAllMarkers was used to identify differentially expressed genes from the raw counts stored in the RNA slot of Seurat object. The HRCs cell population was compared to the rest of cells in order to identify HRC markers in mice. Testing was limited to genes that were detected in a minimum of 10% of cells and showed, on average, at least 0.25-fold difference (log-scale). The 100 strongest markers were selected for functional enrichment analysis with a hypergeometric test. The enrichment of terms with a Benjamini and Hochberg adjusted p-value smaller than 0.05 were considered significant. For those genes with a human orthologue, we also considered their correlation with EpiHR signature in SMC and KUL datasets. Candidate genes were prioritized by selecting those that were independently identified as marker genes of the EpiHR cell population in >10 human samples from the SMC dataset. The resulting list of genes was ranked by the average correlation across mice and human datasets.

**10X mouse single cell RNA sequencing analysis (Epcam- cells)**

Sequencing reads were processed with the 10X Genomics Inc. software package CellRanger (v.5.0.0) against the mouse mm10 reference genome and VDJ reference provided by 10X with v5 release. To simultaneously profile the transcriptomic profile and the TCR repertoire, we followed the "cellranger multi" pipeline [https://support.10xgenomics.com/single-cell-vdj/software/pipelines/5.0/using/multi]. Chemistry was set to "SC5P-R2" and expect-cells to "6000". Quality control (QC) and normalization were performed considering all 8 libraries together after ensuring there were no remarkable differences on the three main QC metrics (library size, library complexity and mitochondrial expression) among the different samples, as described in the guidelines from (Luecken and Theis, 2019). Low quality cells were removed by removing those barcodes with a very low number of UMI (< 500) and genes (< 175), or with a percentage of expression from mitochondrial expression higher than 15% as it is indicative of lysed cells. Additionally, barcodes with a large library size and complexity (> 15000 UMIs and > 4000 genes) and genes detected in less than 5 cells were removed. Finally, data was normalized and log transformed.

To combine data from all 8 samples, batch-effect associated with the confounder variable were removed using Seurat's standard integration pipeline (Butler et al., 2018). This harmonization approach is based on the identification of "anchor" correspondences between pairs of datasets. Prior to integration, we obtained condition-specific highly variable features; only 2000 features common among all libraries were used to compute the integration anchors, thereby reducing noise. In order to speed up the anchor identification, we used the alternative reciprocal PCA (RPCA) as suggested by the authors. To explore our data in a two-dimensional embedding we applied the Uniform Manifold

Approximation and Projection (UMAP) algorithm. After integration, we used the first 30 Principal Components to create a k-nearest neighbors graph with the "FindNeighbors" function followed by the cell clustering using the Louvain clustering algorithm with the "FindClusters" function, and setting the resolution parameter to 0.4. We performed a Differential Expression Analysis (DEA) for all clusters to determine their marker genes using the normalized RNA counts instead of the integrated data slot. The clusters were annotated by expression of canonical markers of immune cell types and comparing it with the results of the DEA analysis, and grouped into major cell groups (Stromal cells, Monocytes/Macrophages, Kupffer cells, Granulocytes, DCs, T-NK cells, B cells, Plasma B cells, Endothelial cells, Liver sinusoidal Endothelial cells, Lymphatic Endothelial cells, Epithelial cells, Intestinal cells). Subsequently, the major cell groups (Stromal cells, Monocytes/Macrophages, Granulocytes, DCs, and T-NK cells) were independently re-processed following the previously described steps to obtain a fine-grained resolution of clusters and to annotate them into specific cell types and states. At this point, doublets and low quality cells were automatically removed.

In order to identify shared clonotypes across samples, we re-analyzed the CDR3 nucleotide sequences obtained by CellRanger using scirpy. A clonotype was defined by all T cells having identical α-chain and β-chain CDR3 nucleotide sequences; for the cells where multiple alpha or beta chains were captured, only the most abundant pair was considered. Clonal expansion was considered when a specific clonotype was detected in more than 2 cells.

**SMART-seq single-cell sequencing**

Each 96-well plate contained cells from all 4 experimental conditions to avoid batch effects. Full-length single-cell RNA-seq libraries were prepared using the SMART-Seq v5 Ultra Low Input RNA Kit for Sequencing (Takara Bio). All reactions were downscaled to one quarter of the original protocol and performed following thermal cycling manufacturer's conditions. Briefly, freshly harvested single cells were sorted into 96-well plates containing 2.5 µl of the Reaction buffer (1x Lysis Buffer, RNase Inhibitor 1U/µl). Reverse transcription was performed using 2.5 µl of the RT MasterMix (SMART-Seq v5 Ultra Low Input RNA Kit for Sequencing, Takara Bio). cDNA was amplified using 8 µl of the PCR MasterMix (SMART-Seq v5 Ultra Low Input RNA Kit for Sequencing, Takara Bio) with 25 cycles of amplification . Following purification with Agencourt Ampure XP beads (Beckmann Coulter), product size distribution and quantity were assessed on a Bioanalyzer using a High Sensitivity DNA Kit (Agilent Technologies). A total of 140 pg of the amplified cDNA was fragmented using Nextera XT (Illumina) and amplified with double indexed Nextera PCR primers (IDT). Products of each well of the 96-well plate were pooled and purified twice with Agencourt Ampure XP beads (Beckmann Coulter). Final

libraries were quantified and checked for fragment size distribution using a Bioanalyzer High Sensitivity DNA Kit (Agilent Technologies). Pooled sequencing of Nextera libraries was carried out using a HiSeq4000 (Illumina) to an average sequencing depth of 0.5 million reads per cell. Sequencing was carried out as paired-end (PE75) reads with library indexes corresponding to cell barcodes.

### Mouse SMART-seqv2 single cell RNA sequencing analysis

The raw output of Smart-seq2 single-cell RNA sequencing was analyzed with the zUMIs pipeline (Parekh et al., 2018). Reads were aligned to the UCSC_GRCm38.mm10 genome. The subsequent processing steps and analysis were performed with Seurat package (Butler et al., 2018; Hao et al., 2021; Satija et al., 2015; Stuart et al., 2019) (version v4.0.3). The four technical batches were merged into a single Seurat object. A total of 1,057 cells having <20% mitochondrial read content and >20,000 detected UMIs were considered for downstream analyses. Ribosomal reads, which accounted for a 6.78% of the total, were removed. The proportion of mitochondrial reads was regressed out during the normalization and variance stabilization of raw counts, which was performed with the sctransform method. The first 22 principal components were used to obtain a Uniform Manifold Approximation and Projection (UMAP) for visualization purposes. SCT transformed counts were further imputed and smoothed with magic (v.2.0.3) (van Dijk et al., 2018a). The expression of gene signatures (Supplementary Table 4) was summarized by taking the average magic score of its constituent genes. Non-epithelial cells (Epcam signature < 2.2) were removed and normalized again. Substantial batch effect was observed in the data. In order to improve the integration of the four batches, the IntegrateData function was used with pre-computed Anchors that were obtained from 3,000 integration features. The first 18 principal components of the integrated dataset, which was SCT normalized, were used to obtain the UMAP and the clustering of cells with FindClusters Seurat function (resolution = 1). SCT integrated counts were imputed and smoothed with magic (v.2.0.3) (van Dijk et al., 2018a). The expression of gene signatures was summarized by taking the average magic score of its constituent genes. FindMarkers was used to identify differentially expressed genes from the raw counts (RNA slot of Seurat object). Velocyto(La Manno et al., 2018) and CellRank(Lange et al., 2020)were used to uncover the cell-state dynamics of CRC metastasis from RNA velocity estimates. Gene expression or signature expression was represented as a function of latent time using the smoothing function geom_smoth in R(R Core Team, 2020). In addition to the overall analysis, the integration, normalization, imputation, and trajectory analysis was also performed independently for the subset of cells harvested from primary tumors, incipient metastasis, and macrometastases.

### Creation of CRC transcriptomic Meta-cohort

Public CRC transcriptomic datasets were downloaded from GEO (Barrett and Edgar, 2006) and NCI GDC commons (Grossman et al., 2016) and were pre-processed and homogenized in order to create a unique Meta-cohort for analysis including 1.830 tumor samples from seven different sources: TCGA (Muzny et al., 2012), GSE38832 (Tripathi et al., 2014), GSE44076 (Sanz-Pamplona et al., 2014), GSE33113 (Kemper et al., 2012), GSE14333 (Jorissen et al., 2009), GSE39582(Marisa et al., 2013) and GSE37892 (Laibe et al., 2012) (Supplementary Table 1). Importantly, the four latest datasets include disease-free survival information with a median follow-up of 3.7 years. Other clinical information commonly available across datasets is gender, age, stage and location of primary tumor (Supplementary Table 1). When not available, Microsatellite Instability (MSI) status was imputed using the transcriptomic signature reported in (Jorissen et al., 2008) by means of a density-based non-parametric clustering method (Azzalini and Menardi, 2014; Azzalini and Torelli, 2007). For doing so, gene expressions were centered and scaled to one standard deviation to create a z-score, which were then averaged across all genes included in the signature. The resulting score was centered and scaled across samples, which was used in the clustering algorithm.

Microarray data were processed serparately using RMA (Irizarry et al., 2003) and probesets were annotated using the information provided by Affymetrix. Technical information concerning sample processing and hybridization was retrieved from the original CEL files. Regarding TCGA data, the Legacy version was used for expression measures while data updated up to October, 2016 was retrieved for clinical and sample annotation. Ensembl Biomart database (GRCh37) (Drost and Paszkowski, 2017; Smedley et al., 2015) was used to annotate genes in TCGA. Duplicated samples across platforms were removed from the Genome Analyzer dataset, as well as those patients reporting malignant lesions in locations other than colon or rectum. Expression measures were expressed in RSEM (Li and Dewey, 2011) in this TCGA version, which were log2-transformed and quantile normalized. Samples TCGA-A6-2679-01A and TCGA-AA-A004-01A were excluded as their gene expression showed an anomalous distribution compared to the rest of samples in their dataset, even after quantile normalization.

Each microarray series was also corrected a-priori by Eklund metrics (Eklund and Szallasi, 2008), center and date of scanning using a mixed-effect linear model (Bates et al., 2015). Briefly, occurences of center-scan date combinations were fitted as a random effect, while Eklund metrics were included as fixed effects. Available information about age, gender, stage, site and MSI were also included in these models. Similarly, TCGA sets were corrected for ocurrences of combinations of center and plate identifier (random effect). For microarrays, probesets were summarized at the gene-level (entrez) using the

first principal component from probesets mapping to the same gene. This component was then centered and scaled to the weighted mean of the means and standard deviations of the probesets using the corresponding contribution to the component as weight. The sign of the component was eventually changed to be congruent to the sign of the probeset contributing the most to the component. Finally, all datasets were merged after gene-wise standardization (mean and standard deviation) to the GSE39582 series according to the distribution of gender, age, MSI and stage using undersampling. For doing so, a sub-sample of the same number of patients and the same distribution according to these clinical variables was selected from the GSE39582 series for each of the rest of datasets. Expression values were truncated to the maximum and minimum values observed in the reference dataset.

### Gene screening for association with relapse

Each gene in the Meta-cohort was evaluated for linear association of its expression with recurrence using a frailty Cox proportional hazards model as described in (Therneau et al., 2003) and implemented in (Therneau, 2020), in which dataset and technical variables were included as covariates. Statistical significance was assessed by means of a Wald test. Hazard Rations (HR) and their corresponding 95% confidence intervals were computed as a measure of association. 2530 genes passed this selection criterion (all HR signature)

We then assessed the expression of each gene with this relapse geneset at the cell population level using the GSE39397 dataset (REF) which include expression profiles of cancer cells (EPCAM+), CAFs (FAP+), leukocytes (CD45+) and endothelial cells (CD31+) isolated by FACs from dissociated primary CRCs (n=14). The Epi-HR signature contains those genes within the relapse geneset which expression is upregulated (Fold change >1; $p<0,05$) in Epcam+ cells compared to Cd31+, Cd45+ and Fap+ cells. The relapse-associated genes that did not pass this cut-off comprised the TME-HR signature.

### Patient 10X single-cell analysis

Count matrices were downloaded from arrayExpress (accession number E-MTAB-8107 for samples EXT001, EXT002, EXT003, EXT009, EXT010, EXT011, EXT012, EXT013, EXT014, EXT018, EXT019, EXT020, EXT021, EXT022, EXT023, EXT024, EXT025, EXT026, EXT027, EXT028) and GEO (accession number GSE132465 for all SMC..-T samples) (Lee et al., 2020). The remaining EXT samples were processed as referred in E-MTAB-8107 and deposited in ArrayExpress under accession number E-MTAB-9934. Cells with mitochondrial content higher than 20%, less than 1000 counts, more than 6000 or less than 200 genes detected were discarded. Ribosomal genes were also re-

moved from the matrix to avoid technical biases during the normalization step. Samples with less than 500 cells and not classified as core tumor were discarded from further analyses. The Korea (SMC samples) and Leuven (EXT samples) cohorts were processed independently following the R [1] package Seurat V3 recommendations(Stuart et al., 2019) : samples were merged into a single Seurat object and normalized using the SCTransform function regressing mitochondrial percentage, with the method "glmGamPoi" and setting the min.cells parameter to 1 and return.only.genes to FALSE in order to maintain the maximum number of genes. Dimensionality reduction and visualization were performed using the RunPCA and RunUMAP functions, using 26 principal components. Expression was imputed and smoothened using the MAGIC algorithm(van Dijk et al., 2018b). The expression of the EPCAM gene was used to define the connected components corresponding to epithelial cells.

Once defined, the epithelial component of each cohort was processed as follows: cells with less than 1000/3000 (SMC/KUL) counts and less than 200/1250 genes detected were removed from the dataset. No further normalization was applied. RunPCA, RunUMAP, FindNeighbors and FindClusters were applied, with 5/7 principal components and resolution of 0.7. Expression was imputed and smoothened using the MAGIC algorithm. Expression of gene sets was computed as the mean value of the MAGIC expression per cell for all genes in the signature. The EpiHR and LGR5 cell populations were defined as those cells with the corresponding signature expression above the 75 percentile. Population markers were found using the FindMarkers function comparing each population with all remaining cells. A signature was defined as the top 100 genes ordered by the fold change. Functional enrichment was computed through the Gene Set Enrichment Analysis implementation of (Mootha et al., 2003) with the GseaPreranked function.

**Association of gene expression and survival data**

A multivariate Cox model was fitted to the expression signature of EpiHR and AllHR genesets, with technical and clinical covariates. Signature scores were computed as the scaled mean values of the genes' expression. The technical variables included were global signature (mean expression of all genes) and dataset; while clinical variables were age, stage, gender and MSS/MSI status. The model was fitted using the *coxph* function from the R survival package. The likelihood ratio test p-value was computed with the *drop1* R function. Kaplan Meier plots were generated using the survfit and plot functions of the R survival package. A Cox proportional hazards model with dataset and global signature as covariates was used to compute significance of differences between groups of samples (lower tercile and upper two terciles). P-values and hazard ratios were obtained with the coxph function from the survival package.

**Sample sizing and collection.**

All samples and animals were assigned randomly to experimental conditions, as well as the sample collection. Automated blind quantifications and blind data analysis were done whenever possible. A minimum of three mice were quantified in each experiment and each condition.

# References

Adusumilli, P.S., Zauderer, M.G., Rusch, V.W., O'Cearbhaill, R., Zhu, A., Ngai, D., McGee, E., Chintala, N., Messinger, J., Cheema, W., et al. (2019). Regional delivery of mesothelin-targeted CAR T cells for pleural cancers: Safety and preliminary efficacy in combination with anti-PD-1 agent. J. Clin. Oncol.

Agudo, J., Park, E.S., Rose, S.A., Alibo, E., Sweeney, R., Dhainaut, M., Kobayashi, K.S., Sachidanandam, R., Baccarini, A., Merad, M., et al. (2018). Quiescent Tissue Stem Cells Evade Immune Surveillance. Immunity *48*, 271-285.e5.

André, T., Boni, C., Navarro, M., Tabernero, J., Hickish, T., Topham, C., Bonetti, A., Clingan, P., Bridgewater, J., Rivera, F., et al. (2009). Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. J. Clin. Oncol. *27*, 3109–3116.

André, T., De Gramont, A., Vernerey, D., Chibaudel, B., Bonnetain, F., Tijeras-Raballand, A., Scriva, A., Hickish, T., Tabernero, J., Van Laethem, J.L., et al. (2015). Adjuvant fluorouracil, leucovorin, and oxaliplatin in stage II to III colon cancer: Updated 10-year survival and outcomes according to BRAF mutation and mismatch repair status of the MOSAIC study. J. Clin. Oncol. *33*, 4176–4187.

Anvari, S., Osei, E., and Maftoon, N. (2021). Interactions of platelets with circulating tumor cells contribute to cancer metastasis. Sci. Rep.

Ayyaz, A., Kumar, S., Sangiorgi, B., Ghoshal, B., Gosio, J., Ouladan, S., Fink, M., Barutcu, S., Trcka, D., Shen, J., et al. (2019). Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. Nature *569*, 121–125.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371-385.e18.

Barbazan, J., Pérez-González, C., Gómez-González, M., Dedenon, M., Richon, S., Latorre, E., Serra, M., Mariani, P., Descroix, S., Sens, P., et al. (2021). Cancer-associated fibroblasts actively compress cancer cells and modulate mechanotransduction. BioRxiv 2021.04.05.438443.

Bardelli, A., and Siena, S. (2010). Molecular mechanisms of resistance to cetuximab and panitumumab in colorectal cancer. J. Clin. Oncol. *28*, 1254–1261.

Barker, N., Van Es, J.H., Kuipers, J., Kujala, P., Van Den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., et al. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003–1007.

Barriga, F.M., Montagni, E., Mana, M., Mendez-Lago, M., Hernando-Momblona, X., Sevillano, M., Guillaumet-Adkins, A., Rodriguez-Esteban, G., Buczacki, S.J.A., Gut, M., et al. (2017). Mex3a Marks a Slowly Dividing Subpopulation of Lgr5+ Intestinal Stem Cells. Cell Stem Cell *20*, 801-816.e7.

Barry, E.R., Morikawa, T., Butler, B.L., Shrestha, K., De La Rosa, R., Yan, K.S., Fuchs, C.S., Magness, S.T., Smits, R., Ogino, S., et al. (2013). Restriction of intestinal stem cell expansion and the regenerative response by YAP. Nature *493*, 106–110.

Basak, O., Beumer, J., Wiebrands, K., Seno, H., van Oudenaarden, A., and Clevers, H. (2017). Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells. Cell Stem Cell *20*, 177-190.e4.

Batlle, E., and Clevers, H. (2017). Cancer stem cells revisited. Nat. Med. *23*, 1124–1134.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414.

Beyaz, S., Mana, M.D., Roper, J., Kedrin, D., Saadatpour, A., Hong, S.-J., Bauer-Rowe, K.E., Xifaras, M.E., Akkad, A., Arias, E., et al. (2016). High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. Nature *531*, 53–58.

Van Der Bij, G.J., Oosterling, S.J., Beelen, R.H.J., Meijer, S., Coffey, J.C., and Van Egmond, M. (2009). The perioperative period is an underutilized window of therapeutic opportunity in patients with colorectal cancer. Ann. Surg. *249*, 727–734.

Bjerknes, M., and Cheng, H. (1999). Clonal analysis of mouse intestinal epithelial progenitors. Gastroenterology *116*, 7–14.

Blank, C.U., Rozeman, E.A., Fanchi, L.F., Sikorska, K., van de Wiel, B., Kvistborg, P., Krijgsman, O., van den Braber, M., Philips, D., Broeks, A., et al. (2018). Neoadjuvant versus adjuvant ipilimumab plus nivolumab in macroscopic stage III melanoma. Nat. Med.

Brahmer, J.R., Tykodi, S.S., Chow, L.Q.M., Hwu, W.-J., Topalian, S.L., Hwu, P., Drake, C.G., Camacho, L.H., Kauh, J., Odunsi, K., et al. (2012). Safety and Activity of Anti–PD-L1 Antibody in Patients with Advanced Cancer. N. Engl. J. Med. *366*, 2455–2465.

Brannon, A.R., Vakiani, E., Sylvester, B.E., Scott, S.N., McDermott, G., Shah, R.H., Kania, K., Viale, A., Oschwald, D.M., Vacic, V., et al. (2014). Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. Genome Biol. *15*, 454.

Braun, D.A., Street, K., Burke, K.P., Cookmeyer, D.L., Denize, T., Pedersen, C.B., Gohil, S.H., Schindler, N., Pomerance, L., Hirsch, L., et al. (2021). Progressive immune dysfunction with advancing disease stage in renal cell carcinoma. Cancer Cell *39*, 632-648.e8.

Broz, M.L., Binnewies, M., Boldajipour, B., Nelson, A.E., Pollack, J.L., Erle, D.J., Barczak, A., Rosenblum, M.D., Daud, A., Barber, D.L., et al. (2014). Dissecting the Tumor Myeloid Compartment Reveals Rare Activating Antigen-Presenting Cells Critical for T Cell Immunity. Cancer Cell.

Buczacki, S.J.A., Zecchini, H.I., Nicholson, A.M., Russell, R., Vermeulen, L., Kemp, R., and Winton, D.J. (2013). Intestinal label-retaining cells are secretory precursors expressing Lgr5. Nature

*495*, 65–69.

Buechler, M.B., Pradhan, R.N., Krishnamurty, A.T., Cox, C., Calviello, A.K., Wang, A.W., Yang, Y.A., Tam, L., Caothien, R., Roose-Girma, M., et al. (2021). Cross-tissue organization of the fibroblast lineage. Nature *593*, 575–579.

Calon, A., Espinet, E., Palomo-Ponce, S., Tauriello, D.V.F., Iglesias, M., Céspedes, M.V., Sevillano, M., Nadal, C., Jung, P., Zhang, X.H.F., et al. (2012). Dependency of Colorectal Cancer on a TGF-β-Driven Program in Stromal Cells for Metastasis Initiation. Cancer Cell *22*, 571–584.

Calon, A., Lonardo, E., Berenguer-llergo, A., Espinet, E., Hernando-momblona, X., Iglesias, M., Sevillano, M., Palomo-ponce, S., Tauriello, D.V.F., Byrom, D., et al. (2015). Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat. Genet. *47*, 320–329.

Campbell, P.J., Getz, G., Korbel, J.O., Stuart, J.M., Jennings, J.L., Stein, L.D., Perry, M.D., Nahal-Bose, H.K., Ouellette, B.F.F., Li, C.H., et al. (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93.

Canale, F.P., Basso, C., Antonini, G., Perotti, M., Li, N., Sokolovska, A., Neumann, J., James, M.J., Geiger, S., Jin, W., et al. (2021). Metabolic modulation of tumours with engineered bacteria for immunotherapy. Nature *598*.

Castle, J.C., Kreiter, S., Diekmann, J., Löwer, M., Van De Roemer, N., De Graaf, J., Selmi, A., Diken, M., Boegel, S., Paret, C., et al. (2012). Exploiting the mutanome for tumor vaccination. Cancer Res. *72*, 1081–1091.

Cerezo-Wallis, D., Contreras-Alcalde, M., Troulé, K., Catena, X., Mucientes, C., Calvo, T.G., Cañón, E., Tejedo, C., Pennacchi, P.C., Hogan, S., et al. (2020). Midkine rewires the melanoma microenvironment toward a tolerogenic and immune-resistant state. Nat. Med.

Chalabi, M., Fanchi, L.F., Dijkstra, K.K., Van den Berg, J.G., Aalbers, A.G., Sikorska, K., Lopez-Yurda, M., Grootscholten, C., Beets, G.L., Snaebjornsson, P., et al. (2020). Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. Nat. Med. *26*, 566–576.

Chen, D.S., and Mellman, I. (2013). Oncology meets immunology: The cancer-immunity cycle. Immunity *39*, 1–10.

Chen, D.S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. Nature *541*, 321–330.

Cheng, H., and Leblond, C.P. (1974). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine V. Unitarian theory of the origin of the four epithelial cell types. Am. J. Anat. *141*, 537–561.

Cheung, P., Xiol, J., Dill, M.T., Yuan, W.C., Panero, R., Roper, J., Osorio, F.G., Maglic, D., Li, Q., Gurung, B., et al. (2020). Regenerative Reprogramming of the Intestinal Stem Cell State via Hip-

po Signaling Suppresses Metastatic Colorectal Cancer. Cell Stem Cell *27*, 590-604.e9.

Chowell, D., Yoo, S.-K., Valero, C., Pastore, A., Krishna, C., Lee, M., Hoen, D., Shi, H., Kelly, D.;, Patel, N., et al. (2021). An integrated clinical-genetic model improves prediction of immune checkpoint blockade efficacy across multiple cancer types. Nat. Biotechnol. *In Press*.

Clevers, H. (2013). The Intestinal Crypt, A Prototype Stem Cell Compartment. Cell *154*, 274–284.

Clevers, H., and Nusse, R. (2012). Wnt/β-catenin signaling and disease. Cell *149*, 1192–1205.

Coffelt, S.B., Kersten, K., Doornebal, C.W., Weiden, J., Vrijland, K., Hau, C.S., Verstegen, N.J.M., Ciampricotti, M., Hawinkels, L.J.A.C., Jonkers, J., et al. (2015). IL-17-producing γδ T cells and neutrophils conspire to promote breast cancer metastasis. Nature *522*, 345–348.

Coussens, L.M., Zitvogel, L., and Palucka, A.K. (2013). Neutralizing Tumor-Promoting Chronic Inflammation: A Magic Bullet? Science (80-. ). *339*, 286–291.

Crosnier, C., Stamataki, D., and Lewis, J. (2006). Organizing cell renewal in the intestine: Stem cells, signals and combinatorial control. Nat. Rev. Genet. *7*, 349–359.

Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A. a, Sim, S., Okamoto, J., Johnston, D.M., Qian, D., et al. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat. Biotechnol. *29*, 1120–1127.

Dall'Olio, F.G., Marabelle, A., Caramella, C., Garcia, C., Aldea, M., Chaput, N., Robert, C., and Besse, B. (2022). Tumour burden and efficacy of immune-checkpoint inhibitors. Nat. Rev. Clin. Oncol.

Danna, E.A., Sinha, P., Gilbert, M., Clements, V.K., Pulaski, B.A., and Ostrand-Rosenberg, S. (2004). Surgical Removal of Primary Tumor Reverses Tumor-Induced Immunosuppression Despite the Presence of Metastatic Disease. Cancer Res. *64*, 2205–2211.

Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M., and Wallace, M.B. (2019). Colorectal cancer. Lancet *394*, 1467–1480.

Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., and Tabernero, J. (2017). Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. Nat. Rev. Cancer *17*, 79–92.

Dong, W., Wu, X., Ma, S., Wang, Y., Nalin, A.P., Zhu, Z., Zhang, J., Benson, D.M., He, K., Caligiuri, M.A., et al. (2019). The mechanism of anti–pd-l1 antibody efficacy against pd-l1–negative tumors identifies nk cells expressing pd-l1 as a cytolytic effector. Cancer Discov.

Douillard, J.-Y., Oliner, K.S., Siena, S., Tabernero, J., Burkes, R., Barugel, M., Humblet, Y., Bodoky, G., Cunningham, D., Jassem, J., et al. (2013). Panitumumab–FOLFOX4 Treatment and RAS Mutations in Colorectal Cancer. N. Engl. J. Med. *369*, 1023–1034.

Drost, J., van Jaarsveld, R.H., Ponsioen, B., Zimberlin, C., van Boxtel, R., Buijs, A., Sachs, N., Overmeer, R.M., Offerhaus, G.J., Begthel, H., et al. (2015). Sequential cancer mutations in cultured human intestinal stem cells. Nature *521*, 43–47.

Drost, J., Van Boxtel, R., Blokzijl, F., Mizutani, T., Sasaki, N., Sasselli, V., De Ligt, J., Behjati, S., Grolleman, J.E., Van Wezel, T., et al. (2017). Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science (80-. ). *358*, 234–238.

Dunn, G.P., Old, L.J., and Schreiber, R.D. (2004). The immunobiology of cancer immunosurveillance and immunoediting. Immunity *21*, 137–148.

Dupage, M., Mazumdar, C., Schmidt, L.M., Cheung, A.F., and Jacks, T. (2012). Expression of tumour-specific antigens underlies cancer immunoediting. Nature *482*, 405–409.

Durand, A., Donahue, B., Peignon, G., Letourneur, F., Cagnard, N., Slomianny, C., Perret, C., Shroyer, N.F., and Romagnolo, B. (2012). Functional intestinal stem cells after Paneth cell ablation induced by the loss of transcription factor Math1 (Atoh1). Proc. Natl. Acad. Sci. U. S. A. *109*, 8965–8970.

Dvorak, H.F. (1986). Tumors: Wounds That Do Not Heal. N. Engl. J. Med. *315*, 1650–1659.

Elamin, Y.Y., Rafee, S., Toomey, S., and Hennessy, B.T. (2015). Immune Effects of Bevacizumab: Killing Two Birds with One Stone. Cancer Microenviron. *8*, 15–21.

Fawkner-Corbett, D., Antanaviciute, A., Parikh, K., Jagielowicz, M., Gerós, A.S., Gupta, T., Ashley, N., Khamis, D., Fowler, D., Morrissey, E., et al. (2021). Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell *184*, 810-826.e23.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759–767.

Friedl, P., and Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. Nat. Rev. Mol. Cell Biol.

Fuchs, E. (2009). The Tortoise and the Hair: Slow-Cycling Cells in the Stem Cell Race. Cell *137*, 811–819.

Fuchs, E., and Blau, H.M. (2020). Tissue Stem Cells: Architects of Their Niches. Cell Stem Cell *27*, 532–556.

Fujii, M., Shimokawa, M., Date, S., Takano, A., Matano, M., Nanki, K., Ohta, Y., Toshimitsu, K., Nakazato, Y., Kawasaki, K., et al. (2016). A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. Cell Stem Cell *18*, 827–838.

Fumagalli, A., Drost, J., Suijkerbuijk, S.J.E., Van Boxtel, R., De Ligt, J., Offerhaus, G.J., Begthel, H., Beerling, E., Tan, E.H., Sansom, O.J., et al. (2017). Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. Proc. Natl. Acad. Sci.

U. S. A. *114*, E2357–E2364.

Fumagalli, A., Suijkerbuijk, S.J.E., Begthel, H., Beerling, E., Oost, K.C., Snippert, H.J., Van Rheenen, J., and Drost, J. (2018). A surgical orthotopic organoid transplantation approach in mice to visualize and study colorectal cancer progression. Nat. Protoc. *13*, 235–247.

Fumagalli, A., Oost, K.C., Kester, L., Morgner, J., Bornes, L., Bruens, L., Spaargaren, L., Azkanaz, M., Schelfhorst, T., Beerling, E., et al. (2020). Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. Cell Stem Cell *26*, 569-578.e7.

Galon, J.J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., et al. (2006). Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. Science (80-. ). *313*, 1960–1964.

Ganesh, K., Basnet, H., Kaygusuz, Y., Laughney, A.M., He, L., Sharma, R., O'Rourke, K.P., Reuter, V.P., Huang, Y.-H., Turkekul, M., et al. (2020). L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer. Nat. Cancer *1*, 28–45.

Gardner, A., and Ruffell, B. (2016). Dendritic Cells and Cancer Immunity. Trends Immunol. *37*, 855–865.

Garner, H., and de Visser, K.E. (2020). Immune crosstalk in cancer progression and metastatic spread: a complex conversation. Nat. Rev. Immunol. *20*, 483–497.

GBD (2017). Global Burden of Disease Study 2017. Lancet 1–7.

Gehart, H., and Clevers, H. (2019). Tales from the crypt: new insights into intestinal stem cells. Nat. Rev. Gastroenterol. Hepatol. *16*, 19–34.

Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S. V, Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med. *21*, 938–945.

Gorelik, L., and Flavell, R.A. (2000). Abrogation of TGFβ signaling in T cells leads to spontaneous T cell differentiation and autoimmune disease. Immunity *12*, 171–181.

Gregorieff, A., Liu, Y., Inanlou, M.R., Khomchuk, Y., and Wrana, J.L. (2015). Yap-dependent reprogramming of Lgr5+ stem cells drives intestinal regeneration and cancer. Nature *526*, 715–718.

Guinney, J., Dienstmann, R., Wang, X., De Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular subtypes of colorectal cancer. Nat. Med. *21*, 1350–1356.

Han, T., Goswami, S., Hu, Y., Tang, F., Zafra, M.P., Murphy, C., Cao, Z., Poirier, J.T., Khurana, E., Elemento, O., et al. (2020). Lineage Reversion Drives WNT Independence in Intestinal Cancer. Cancer Discov. *10*, 1590–1609.

Hanahan, D., and Coussens, L.M. (2012). Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. Cancer Cell *21*, 309–322.

Hanahan, D., and Weinberg, R.A. (2000). The Hallmarks of Cancer. Cell *100*, 57–70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation. Cell *144*, 646–674.

Haramis, A.-P.G., Begthel, H., van den Born, M., van Es, J., Jonkheer, S., Offerhaus, G.J.A., and Clevers, H. (2004). De Novo Crypt Formation and Juvenile Polyposis on BMP Inhibition in Mouse Intestine. Science (80-. ). *303*, 1684–1686.

Heijstek, M.W., Kranenburg, O., Borel Rinkes, I.H.M., Kranenburg, M.W.H.O., Rinkes, I.H.M.B., Heijstek, M.W., Kranenburg, O., and Borel Rinkes, I.H.M. (2005). Mouse Models of Colorectal Cancer and Liver Metastases. Dig. Surg. *22*, 16–25.

De Henau, O., Rausch, M., Winkler, D., Campesato, L.F., Liu, C., Cymerman, D.H., Budhu, S., Ghosh, A., Pink, M., Tchaicha, J., et al. (2016). Overcoming resistance to checkpoint blockade therapy by targeting PI3Kγ in myeloid cells. Nature *539*, 443–447.

Hirata-Nozaki, Y., Ohkuri, T., Ohara, K., Kumai, T., Nagata, M., Harabuchi, S., Kosaka, A., Nagato, T., Ishibashi, K., Oikawa, K., et al. (2019). PD-L1-specific helper T-cells exhibit effective antitumor responses: New strategy of cancer immunotherapy targeting PD-L1 in head and neck squamous cell carcinoma. J. Transl. Med.

Hodi, F.S., O'Day, S.J., McDermott, D.F., Weber, R.W., Sosman, J.A., Haanen, J.B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J.C., et al. (2010). Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. N. Engl. J. Med. *363*, 711–723.

Hu, Z., Ding, J., Ma, Z., Sun, R., Seoane, J.A., Scott Shaffer, J., Suarez, C.J., Berghoff, A.S., Cremolini, C., Falcone, A., et al. (2019a). Quantitative evidence for early metastatic seeding in colorectal cancer. Nat. Genet. *51*.

Hu, Z., Ding, J., Ma, Z., Sun, R., Seoane, J.A., Scott Shaffer, J., Suarez, C.J., Berghoff, A.S., Cremolini, C., Falcone, A., et al. (2019b). Quantitative evidence for early metastatic seeding in colorectal cancer. Nat. Genet. *51*, 1113–1122.

Hu, Z., Li, Z., Ma, Z., and Curtis, C. (2020a). Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. Nat. Genet.

Hu, Z., Li, Z., Ma, Z., and Curtis, C. (2020b). Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. Nat. Genet.

Huang, A.C., Orlowski, R.J., Xu, X., Mick, R., George, S.M., Yan, P.K., Manne, S., Kraya, A.A., Wubbenhorst, B., Dorfman, L., et al. (2019). A single dose of neoadjuvant PD-1 blockade predicts clinical outcomes in resectable melanoma. Nat. Med. *25*, 454–461.

Huch, M., Dorrell, C., Boj, S.F., Van Es, J.H., Li, V.S.W., Van De Wetering, M., Sato, T., Hamer, K., Sasaki, N., Finegold, M.J., et al. (2013a). In vitro expansion of single Lgr5 + liver stem cells induced by Wnt-driven regeneration. Nature *494*, 247–250.

Huch, M., Bonfanti, P., Boj, S.F., Sato, T., Loomans, C.J.M., Van De Wetering, M., Sojoodi, M., Li, V.S.W., Schuijers, J., Gracanin, A., et al. (2013b). Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. EMBO J. *32*, 2708–2721.

Huels, D.J., and Sansom, O.J. (2017). R-spondin Is More Than Just Wnt's Sidekick. Dev. Cell *41*, 456–458.

Hurwitz, H., Fehrenbacher, L., Novotny, W., Cartwright, T., Hainsworth, J., Heim, W., Berlin, J., Baron, A., Griffing, S., Holmgren, E., et al. (2004). Bevacizumab plus Irinotecan, Fluorouracil, and Leucovorin for Metastatic Colorectal Cancer. N. Engl. J. Med. *350*, 2335–2342.

Hwang, W.L., Pike, L.R.G., Royce, T.J., Mahal, B.A., and Loeffler, J.S. (2018). Safety of combining radiotherapy with immune-checkpoint inhibition. Nat. Rev. Clin. Oncol. *15*, 477–494.

Im, S.J., Hashimoto, M., Gerner, M.Y., Lee, J., Kissick, H.T., Burger, M.C., Shan, Q., Hale, J.S., Lee, J., Nasti, T.H., et al. (2016). Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. Nature.

Isella, C., Terrasi, A., Bellomo, S.E., Petti, C., Galatola, G., Muratore, A., Mellano, A., Senetta, R., Cassenti, A., Sonetto, C., et al. (2015). Stromal contribution to the colorectal cancer transcriptome. Nat. Genet. *47*, 312–319.

Jackstadt, R., and Sansom, O.J. (2016). Mouse models of intestinal cancer. J. Pathol. *238*, 141–151.

Janiszewska, M., Primi, M.C., and Izard, T. (2020a). Cell adhesion in cancer: Beyond the migration of single cells. J. Biol. Chem.

Janiszewska, M., Primi, M.C., and Izard, T. (2020b). Cell adhesion in cancer: Beyond the migration of single cells. J. Biol. Chem. *295*, 2495–2505.

Janssen, E., Subtil, B., Ortiz, F. de la J., Verheul, H.M.W., and Tauriello, D.V.F. (2020). Combinatorial immunotherapies for metastatic colorectal cancer. Cancers (Basel). *12*, 1–29.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (80-. ). *337*, 816–821.

Jones, S., Chen, W.D., Parmigiani, G., Diehl, F., Beerenwinkel, N., Antal, T., Traulsen, A., Nowak, M.A., Siegel, C., Velculescu, V.E., et al. (2008). Comparative lesion sequencing provides insights into tumor evolution. Proc. Natl. Acad. Sci. U. S. A. *105*, 4283–4288.

Joyce, J.A., and Fearon, D.T. (2015). T cell exclusion, immune privilege, and the tumor microen-

vironment. Science (80-. ). *348*, 74–80.

Joyce, J.A., and Pollard, J.W. (2009). Microenvironmental regulation of metastasis. Nat. Rev. Cancer *9*, 239–252.

June, C.H., O'Connor, R.S., Kawalekar, O.U., Ghassemi, S., and Milone, M.C. (2018). CAR T cell immunotherapy for human cancer. Science (80-. ). *359*, 1361–1365.

Jung, P., Sato, T., Merlos-Suárez, A., Barriga, F.M., Iglesias, M., Rossell, D., Auer, H., Gallardo, M., Blasco, M. a, Sancho, E., et al. (2011). Isolation and in vitro expansion of human colonic stem cells. Nat. Med. *17*, 1225–1227.

Kalbasi, A., and Ribas, A. (2020). Tumour-intrinsic resistance to immune checkpoint blockade. Nat. Rev. Immunol. *20*, 25–39.

Kalluri, R. (2016). The biology and function of fibroblasts in cancer. Nat. Rev. Cancer *16*, 582–598.

Kaneda, M.M., Messer, K.S., Ralainirina, N., Li, H., Leem, C.J., Gorjestani, S., Woo, G., Nguyen, A. V., Figueiredo, C.C., Foubert, P., et al. (2016). PI3Kγ 3 is a molecular switch that controls immune suppression. Nature *539*, 437–442.

Karthaus, W.R., Iaquinta, P.J., Drost, J., Gracanin, A., Van Boxtel, R., Wongvipat, J., Dowling, C.M., Gao, D., Begthel, H., Sachs, N., et al. (2014). Identification of multipotent luminal progenitor cells in human prostate organoid cultures. Cell *159*, 163–175.

Kim, S.I., Cassella, C.R., and Byrne, K.T. (2021). Tumor Burden and Immunotherapy: Impact on Immune Infiltration and Therapeutic Outcomes. Front. Immunol.

Kim, T.H., Li, F., Ferreiro-Neira, I., Ho, L.L., Luyten, A., Nalapareddy, K., Long, H., Verzi, M., and Shivdasani, R.A. (2014). Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity. Nature *506*, 511–515.

Koebel, C.M., Vermi, W., Swann, J.B., Zerafa, N., Rodig, S.J., Old, L.J., Smyth, M.J., and Schreiber, R.D. (2007). Adaptive immunity maintains occult cancer in an equilibrium state. Nature *450*, 903–907.

Konkel, J.E., Frommer, F., Leech, M.D., Yagita, H., Waisman, A., and Anderton, S.M. (2010). PD-1 signalling in CD4+ T cells restrains their clonal expansion to an immunogenic stimulus, but is not critically required for peptide-induced tolerance. Immunology.

Korinek, V., Barker, N., Moerer, P., Van Donselaar, E., Huls, G., Peters, P.J., and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat. Genet. *19*, 379–383.

Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2020). CellRank for directed single-cell fate mapping. BioRxiv 2020.10.19.345983.

Laouar, Y., Gorelik, L., and Flavell, R. a (2002). Immune-Mediated Eradication of Tumors through the Blockade of Transforming Growth Factor-beta Signaling in T Cells. Dir. Sci. *1*, 113–115.

De Lau, W., Barker, N., Low, T.Y., Koo, B.K., Li, V.S.W., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P.J., Van De Wetering, M., et al. (2011). Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. Nature *476*, 293–297.

Laughney, A.M., Hu, J., Campbell, N.R., Bakhoum, S.F., Setty, M., Lavallée, V.P., Xie, Y., Masilionis, I., Carr, A.J., Kottapalli, S., et al. (2020). Regenerative lineages and immune-mediated pruning in lung cancer metastasis (Springer US).

Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N. Engl. J. Med. *372*, 2509–2520.

Lee, H.O., Hong, Y., Etlioglu, H.E., Cho, Y.B., Pomella, V., Van den Bosch, B., Vanhecke, J., Verbandt, S., Hong, H., Min, J.W., et al. (2020). Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat. Genet. *52*, 594–603.

Lenos, K.J., Miedema, D.M., Lodestijn, S.C., Nijman, L.E., van den Bosch, T., Romero Ros, X., Lourenço, F.C., Lecca, M.C., van der Heijden, M., van Neerven, S.M., et al. (2018). Stem cell functionality is microenvironmentally defined during tumour expansion and therapy response in colon cancer. Nat. Cell Biol. *20*, 1193–1202.

Lenschow, D.J., Walunas, T.L., and Bluestone, J.A. (1996). CD28/B7 system of T cell costimulation. Annu. Rev. Immunol.

Li, K., Wu, R., Zhou, M., Tong, H., and Luo, K.Q. (2021). Desmosomal proteins of DSC2 and PKP1 promote cancer cells survival and metastasis by increasing cluster formation in circulatory system. Sci. Adv.

Linnemann, C., van Buuren, M.M., Bies, L., Verdegaal, E.M., Schotte, R., Calis, J.J., Behjati, S., Velds, A., Hilkmann, H., Atmioui, D.E., et al. (2015). High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. Nat Med *21*, 81–85.

Liu, J., Blake, S.J., Yong, M.C.R., Harjunpää, H., Ngiow, S.F., Takeda, K., Young, A., O'Donnell, J.S., Allen, S., Smyth, M.J., et al. (2016). Improved efficacy of neoadjuvant compared to adjuvant immunotherapy to eradicate metastatic disease. Cancer Discov.

Lugli, A., Kirsch, R., Ajioka, Y., Bosman, F., Cathomas, G., Dawson, H., El Zimaity, H., Fléjou, J.F., Hansen, T.P., Hartmann, A., et al. (2017). Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. Mod. Pathol. *30*, 1299–1311.

Lukonin, I., Serra, D., Challet Meylan, L., Volkmann, K., Baaten, J., Zhao, R., Meeusen, S., Colman, K., Maurer, F., Stadler, M.B., et al. (2020). Phenotypic landscape of intestinal organoid regeneration. Nature *586*, 275–280.

Luksza, M., Riaz, N., Makarov, V., Balachandran, V.P., Hellmann, M.D., Solovyov, A., Rizvi, N.A., Merghoub, T., Levine, A.J., Chan, T.A., et al. (2017). A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. Nature *551*, 517–520.

Malladi, S., MacAlinao, D.G., Jin, X., He, L., Basnet, H., Zou, Y., De Stanchina, E., and Massagué, J. (2016). Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. Cell *165*, 45–60.

Maman, S., and Witz, I.P. (2018). A history of exploring cancer in context. Nat. Rev. Cancer *18*, 359–376.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Mariathasan, S., Turley, S.J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, E.E., Koeppen, H., Astarita, J.L., Cubas, R., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature *554*, 544–548.

Marisa, L., de Reyniès, A., Duval, A., Selves, J., Gaub, M.P., Vescovo, L., Etienne-Grimaldi, M.C., Schiappa, R., Guenot, D., Ayadi, M., et al. (2013). Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value. PLoS Med. *10*.

Martins, F., Sofiya, L., Sykiotis, G.P., Lamine, F., Maillard, M., Fraga, M., Shabafrouz, K., Ribi, C., Cairoli, A., Guex-Crosier, Y., et al. (2019). Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. Nat. Rev. Clin. Oncol. *16*, 563–580.

Massagué, J. (2008). TGFβ in Cancer. Cell *134*, 215–230.

Massagué, J. (2012). TGFβ signalling in context. Nat. Rev. Mol. Cell Biol. *13*, 616–630.

Matano, M., Date, S., Shimokawa, M., Takano, A., Fujii, M., Ohta, Y., Watanabe, T., Kanai, T., and Sato, T. (2015). Modeling colorectal cancer using CRISPR-Cas9–mediated engineering of human intestinal organoids. Nat. Med. *21*, 256–262.

Matsushita, H., Vesely, M.D., Koboldt, D.C., Rickert, C.G., Uppaluri, R., Magrini, V.J., Arthur, C.D., White, J.M., Chen, Y.-S., Shea, L.K., et al. (2012). Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. Nature *482*, 400–404.

McCarthy, N., Kraiczy, J., and Shivdasani, R.A. (2020a). Cellular and molecular architecture of the intestinal stem cell niche. Nat. Cell Biol. *22*, 1033–1041.

McCarthy, N., Manieri, E., Storm, E.E., Saadatpour, A., Luoma, A.M., Kapoor, V.N., Madha, S., Gaynor, L.T., Cox, C., Keerthivasan, S., et al. (2020b). Distinct Mesenchymal Cell Populations Generate the Essential Intestinal BMP Signaling Gradient. Cell Stem Cell *26*, 391-402.e5.

Mcgranahan, N., Furness, A.J.S., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-han-

jani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science (80-. ). *351*, 1463–1469.

Medema, J.P., and Vermeulen, L. (2011). Microenvironmental regulation of stem cells in intestinal homeostasis and cancer. Nature *474*, 318–326.

Mellman, I., Coukos, G., and Dranoff, G. (2011). Cancer immunotherapy comes of age. Nature *480*, 480–489.

Melo, F. de S. e, Kurtova, A. V., Harnoss, J.M., Kljavin, N., Hoeck, J.D., Hung, J., Anderson, J.E., Storm, E.E., Modrusan, Z., Koeppen, H., et al. (2017). A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer. Nature *543*, 676–680.

Merlos-Suárez, A., Barriga, F.M., Jung, P., Iglesias, M., Céspedes, M.V., Rossell, D., Sevillano, M., Hernando-Momblona, X., Da Silva-Diz, V., Muñoz, P., et al. (2011). The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell Stem Cell *8*, 511–524.

Metcalfe, C., Kljavin, N.M., Ybarra, R., and De Sauvage, F.J. (2014). Lgr5+ stem cells are indispensable for radiation-induced intestinal regeneration. Cell Stem Cell *14*, 149–159.

Moertel, C.G., Fleming, T.R., Macdonald, J.S., Haller, D.G., Laurie, J.A., Goodman, P.J., Ungerleider, J.S., Emerson, W.A., Tormey, D.C., Glick, J.H., et al. (1990). Levamisole and Fluorouracil for Adjuvant Therapy of Resected Colon Carcinoma. N. Engl. J. Med. *322*, 352–358.

Morello, A., Sadelain, M., and Adusumilli, P.S. (2016). Mesothelin-targeted CARs: Driving T cells to solid Tumors. Cancer Discov.

Morral, C., Stanisavljevic, J., Hernando-Momblona, X., Mereu, E., Álvarez-Varela, A., Cortina, C., Stork, D., Slebe, F., Turon, G., Whissell, G., et al. (2020). Zonation of Ribosomal DNA Transcription Defines a Stem Cell Hierarchy in Colorectal Cancer. Cell Stem Cell 1–17.

Moynihan, K.D., Opel, C.F., Szeto, G.L., Tzeng, A., Zhu, E.F., Engreitz, J.M., Williams, R.T., Rakhra, K., Zhang, M.H., Rothschilds, A.M., et al. (2016). Eradication of large established tumors in mice by combination immunotherapy that engages innate and adaptive immune responses. Nat. Med. *22*, 1402–1410.

Murata, K., Jadhav, U., Madha, S., van Es, J., Dean, J., Cavazza, A., Wucherpfennig, K., Michor, F., Clevers, H., and Shivdasani, R.A. (2020). Ascl2-Dependent Cell Dedifferentiation Drives Regeneration of Ablated Intestinal Stem Cells. Cell Stem Cell *26*, 377-390.e6.

Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., Newsham, I.F., et al. (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

Nishimura, H., Okazaki, T., Tanaka, Y., Nakatani, K., Hara, M., Matsumori, A., Sasayama, S., Mizoguchi, A., Hiai, H., Minato, N., et al. (2001). Autoimmune dilated cardiomyopathy in PD-1 receptor-deficient mice. Science (80-. ).

Noguchi, T., Ward, J.P., Gubin, M.M., Arthur, C.D., Lee, S.H., Hundal, J., Selby, M.J., Graziano, R.F., Mardis, E.R., Korman, A.J., et al. (2017). Temporally distinct PD-L1 expression by tumor and host cells contributes to immune escape. Cancer Immunol. Res. *5*, 106–117.

Northey, J.J., Przybyla, L., and Weaver, V.M. (2017). Tissue force programs cell fate and tumor aggression. Cancer Discov. *7*, 1224–1237.

Nusse, R., and Clevers, H. (2017). Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. Cell *169*, 985–999.

O'Rourke, K.P., Loizou, E., Livshits, G., Schatoff, E.M., Baslan, T., Manchado, E., Simon, J., Romesser, P.B., Leach, B., Han, T., et al. (2017). Transplantation of engineered organoids enables rapid generation of metastatic mouse models of colorectal cancer. Nat. Biotechnol. *35*, 577–582.

Ombrato, L., Nolan, E., Kurelac, I., Mavousian, A., Bridgeman, V.L., Heinze, I., Chakravarty, P., Horswell, S., Gonzalez-Gualda, E., Matacchione, G., et al. (2019). Metastatic-niche labelling reveals parenchymal cells with stem features. Nature *572*, 603–608.

Pagès, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F.S., Bifulco, C., Lugli, A., Zlobec, I., Rau, T.T., Berger, M.D., et al. (2018). International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet *391*, 2128–2139.

Palucka, A.K., and Coussens, L.M. (2016). The Basis of Oncoimmunology. Cell *164*, 1233–1247.

Palucka, K., and Banchereau, J. (2012). Cancer immunotherapy via dendritic cells. Nat. Rev. Cancer *12*, 265–277.

Parikh, A.R., Szabolcs, A., Allen, J.N., Clark, J.W., Wo, J.Y., Raabe, M., Thel, H., Hoyos, D., Mehta, A., Arshad, S., et al. (2021). Radiation therapy enhances immunotherapy response in microsatellite stable colorectal and pancreatic adenocarcinoma in a phase II trial. Nat. Cancer *2*, 1124–1135.

Pascual, G., Avgustinova, A., Mejetta, S., Martín, M., Castellanos, A., Attolini, C.S.O., Berenguer, A., Prats, N., Toll, A., Hueto, J.A., et al. (2017). Targeting metastasis-initiating cells through the fatty acid receptor CD36. Nature *541*, 41–45.

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. Nature *556*.

Pérez-González, C., Ceada, G., Greco, F., Matejcic, M., Gómez-González, M., Castro, N., Kale, S., Álvarez-Varela, A., Roca-Cusachs, P., Batlle, E., et al. (2020). Mechanical compartmentalization of the intestinal organoid enables crypt folding and collective cell migration. BioRxiv 2020.09.20.299552.

Pyke, C., Rømer, J., Danø, K., Ralfkiær, E., Salo, S., and Tryggvason, K. (1995). Laminin-5 Is a Marker of Invading Cancer Cells in Some Human Carcinomas and Is Coexpressed with the Receptor for Urokinase Plasminogen Activator in Budding Cancer Cells in Colon Adenocarcinomas.

Cancer Res. *55*, 4132–4139.

Qi, Z., Li, Y., Zhao, B., Xu, C., Liu, Y., Li, H., Zhang, B., Wang, X., Yang, X., Xie, W., et al. (2017). BMP restricts stemness of intestinal Lgr5+ stem cells by directly suppressing their signature genes. Nat. Commun. *8*, 13824.

Qin, X., Sufi, J., Vlckova, P., Kyriakidou, P., Acton, S.E., Li, V.S.W., Nitz, M., and Tape, C.J. (2020). Cell-type-specific signaling networks in heterocellular organoids. Nat. Methods *17*, 335–342.

Qu, M., Xiong, L., Lyu, Y., Zhang, X., Shen, J., Guan, J., Chai, P., Lin, Z., Nie, B., Li, C., et al. (2021). Establishment of intestinal organoid cultures modeling injury-associated epithelial regeneration. Cell Res. *31*, 259–271.

Quail, D.F., Olson, O.C., Bhardwaj, P., Walsh, L.A., Akkari, L., Quick, M.L., Chen, I.C., Wendel, N., Ben-Chetrit, N., Walker, J., et al. (2017). Obesity alters the lung myeloid cell landscape to enhance breast cancer metastasis through IL5 and GM-CSF. Nat. Cell Biol. *19*, 974–987.

Radtke, F., and Clevers, H. (2005). Self-renewal and cancer of the gut: Two sides of a coin. Science (80-. ). *307*, 1904–1909.

Rahbari, N.N., Carr, P.R., Jansen, L., Chang-Claude, J., Weitz, J., Hoffmeister, M., and Brenner, H. (2019). Time of Metastasis and Outcome in Colorectal Cancer. Ann. Surg. *269*, 494–502.

Reck, M., Rodríguez-Abreu, D., Robinson, A.G., Hui, R., Csőszi, T., Fülöp, A., Gottfried, M., Peled, N., Tafreshi, A., Cuffe, S., et al. (2016). Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. N. Engl. J. Med. *375*, 1823–1833.

Ries, C.H., Cannarile, M.A., Hoves, S., Benz, J., Wartha, K., Runza, V., Rey-Giraud, F., Pradel, L.P., Feuerhake, F., Klaman, I., et al. (2014). Targeting tumor-associated macrophages with anti-CSF-1R antibody reveals a strategy for cancer therapy. Cancer Cell *25*, 846–859.

Ringel, A.E., Drijvers, J.M., Baker, G.J., Catozzi, A., García-Cañaveras, J.C., Gassaway, B.M., Miller, B.C., Juneja, V.R., Nguyen, T.H., Joshi, S., et al. (2020). Obesity Shapes Metabolism in the Tumor Microenvironment to Suppress Anti-Tumor Immunity. Cell *183*, 1848-1866.e26.

Ritsma, L., Ellenbroek, S.I.J., Zomer, A., Snippert, H.J., De Sauvage, F.J., Simons, B.D., Clevers, H., and Van Rheenen, J. (2014). Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live imaging. Nature *507*, 362–365.

Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S., et al. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science (80-. ). *348*, 124–128.

Roerink, S.F., Sasaki, N., Lee-Six, H., Young, M.D., Alexandrov, L.B., Behjati, S., Mitchell, T.J., Grossmann, S., Lightfoot, H., Egan, D.A., et al. (2018). Intra-tumour diversification in colorectal cancer at the single-cell level. Nature.

Rosenberg, S.A., and Restifo, N.P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. Science (80-. ). *348*, 62–68.

Routy, B., Le Chatelier, E., Derosa, L., Duong, C.P.M., Alou, M.T., Daillère, R., Fluckiger, A., Messaoudene, M., Rauber, C., Roberti, M.P., et al. (2018). Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science (80-. ). *359*, 91–97.

Ruffell, B., Chang-Strachan, D., Chan, V., Rosenbusch, A., Ho, C.M.T., Pryer, N., Daniel, D., Hwang, E.S., Rugo, H.S., and Coussens, L.M. (2014). Macrophage IL-10 Blocks CD8+ T Cell-Dependent Responses to Chemotherapy by Suppressing IL-12 Expression in Intratumoral Dendritic Cells. Cancer Cell *26*, 623–637.

Russo, M., Crisafulli, G., Sogari, A., Reilly, N.M., Arena, S., Lamba, S., Bartolini, A., Amodio, V., Magrì, A., Novara, L., et al. (2019). Adaptive mutability of colorectal cancers in response to targeted therapies. Science (80-. ). *366*, 1473–1480.

Sadanandam, A., Lyssiotis, C.A., Homicsko, K., Collisson, E.A., Gibb, W.J., Wullschleger, S., Ostos, L.C.G., Lannon, W.A., Grotzinger, C., Del Rio, M., et al. (2013). A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat. Med. *19*, 619–625.

Saha, D., Martuza, R.L., and Rabkin, S.D. (2017). Macrophage Polarization Contributes to Glioblastoma Eradication by Combination Immunovirotherapy and Immune Checkpoint Blockade. Cancer Cell *32*, 253-267.e5.

Sasaki, N., Sachs, N., Wiebrands, K., Ellenbroek, S.I.J., Fumagalli, A., Lyubimova, A., Begthel, H., Van Born, M. Den, Van Es, J.H., Karthaus, W.R., et al. (2016). Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. Proc. Natl. Acad. Sci. U. S. A. *113*, E5399–E5407.

Sato, T., Vries, R.G., Snippert, H.J., Van De Wetering, M., Barker, N., Stange, D.E., Van Es, J.H., Abo, A., Kujala, P., Peters, P.J., et al. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262–265.

Sato, T., Van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., Van Den Born, M., Barker, N., Shroyer, N.F., Van De Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature *469*, 415–418.

Sato, T., Clevers, H., Flier, L.G. van der, Clevers, H., Cheng, H., Leblond, C.P., Barker, N., Potten, C.S., Sangiorgi, E., Capecchi, M.R., et al. (2013). Growing Self-Organizing Mini-Guts from a Single Intestinal Stem Cell: Mechanism and Applications. Science (80-. ). *340*, 241–260.

Schepers, A.G., Snippert, H.J., Stange, D.E., Van Den Born, M., Van Es, J.H., Van De Wetering, M., and Clevers, H. (2012). Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. Science (80-. ). *337*, 730–735.

Schumacher, T.N., and Schreiber, R.D. (2015). Neoantigens in cancer immunotherapy. Science (80-. ). *348*, 69–74.

Schutgens, F., Rookmaaker, M.B., Margaritis, T., Rios, A., Ammerlaan, C., Jansen, J., Gijzen, L., Vormann, M., Vonk, A., Viveen, M., et al. (2019). Tubuloids derived from human adult kidney and urine for personalized disease modeling. Nat. Biotechnol. *37*, 303–313.

Serra, D., Mayr, U., Boni, A., Lukonin, I., Rempfler, M., Challet Meylan, L., Stadler, M.B., Strnad, P., Papasaikas, P., Vischi, D., et al. (2019). Self-organization and symmetry breaking in intestinal organoid development. Nature *569*, 66–72.

Shankaran, V., Ikeda, H., Bruce, A.T., White, J.M., Swanson, P.E., Old, L.J., and Schreiber, R.D. (2001). IFNγ, and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature *410*, 1107–1111.

Shimokawa, M., Ohta, Y., Nishikori, S., Matano, M., Takano, A., Fujii, M., Date, S., Sugimoto, S., Kanai, T., and Sato, T. (2017). Visualization and targeting of LGR5+ human colon cancer stem cells. Nature 1–21.

Simoni, Y., Becht, E., Fehlings, M., Loh, C.Y., Koo, S.L., Teng, K.W.W., Yeong, J.P.S., Nahar, R., Zhang, T., Kared, H., et al. (2018). Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. Nature *557*, 575–579.

Sinicrope, F.A., Mahoney, M.R., Smyrk, T.C., Thibodeau, S.N., Warren, R.S., Bertagnolli, M.M., Nelson, G.D., Goldberg, R.M., Sargent, D.J., Alberts Frank A Sinicrope, S.R., et al. (2013). Prognostic Impact of Deficient DNA Mismatch Repair in Patients With Stage III Colon Cancer From a Randomized Trial of FOLFOX-Based Adjuvant Chemotherapy Written on behalf of the Alliance for Clinical Trials in Oncology.

Siravegna, G., Mussolin, B., Buscarino, M., Corti, G., Cassingena, A., Crisafulli, G., Ponzetti, A., Cremolini, C., Amatu, A., Lauricella, C., et al. (2015). Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat. Med. *21*, 795–801.

Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., et al. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134–144.

Sottoriva, A., Kang, H., Ma, Z., Graham, T.A., Salomon, M.P., Zhao, J., Marjoram, P., Siegmund, K., Press, M.F., Shibata, D., et al. (2015). A Big Bang model of human colorectal tumor growth. Nat. Genet. *47*, 209–216.

de Sousa e Melo, F., and de Sauvage, F.J. (2019). Cellular Plasticity in Intestinal Homeostasis and Disease. Cell Stem Cell *24*, 54–64.

De Sousa E Melo, F., Wang, X., Jansen, M., Fessler, E., Trinh, A., De Rooij, L.P.M.H., De Jong, J.H., De Boer, O.J., Van Leersum, R., Bijlsma, M.F., et al. (2013). Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. Nat. Med. *19*, 614–618.

Sprangers, J., Zaalberg, I.C., and Maurice, M.M. (2021). Organoid-based modeling of intestinal development, regeneration, and repair. Cell Death Differ. *28*, 95–107.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. *71*, 209–249.

Szczerba, B.M., Castro-Giner, F., Vetter, M., Krol, I., Gkountela, S., Landin, J., Scheidmann, M.C., Donato, C., Scherrer, R., Singer, J., et al. (2019). Neutrophils escort circulating tumour cells to enable cell cycle progression. Nature *566*, 553–557.

Takeuchi, A., and Saito, T. (2017). CD4 CTL, a cytotoxic subset of CD4+ T cells, their differentiation and function. Front. Immunol.

Taniguchi, S., Elhance, A., van Duzer, A., Kumar, S., Leitenberger, J.J., and Oshimori, N. (2020). Tumor-initiating cells establish an IL-33–TGF-b niche signaling loop to promote cancer progression. Science (80-. ). *369*.

Tauriello, D.V.F., Sancho, E., and Batlle, E. (2022). Overcoming TGFβ-mediated immune evasion in cancer. Nat. Rev. Cancer *22*, 25–44.

Tauriello, D.V.F.F., Palomo-Ponce, S., Stork, D., Berenguer-Llergo, A., Badia-Ramentol, J., Iglesias, M., Sevillano, M., Ibiza, S., Cañellas, A., Hernando-momblona, X., et al. (2018). TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis. Nature *554*, 538–543.

Tetteh, P.W., Basak, O., Farin, H.F., Wiebrands, K., Kretzschmar, K., Begthel, H., Van Den Born, M., Korving, J., De Sauvage, F., Van Es, J.H., et al. (2016). Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. Cell Stem Cell *18*, 203–213.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The Immune Landscape of Cancer. Immunity 1–19.

Tian, H., Biehs, B., Warming, S., Leong, K.G., Rangell, L., Klein, O.D., and de Sauvage, F.J. (2011). A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. Nature *478*, 255–259.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science (80-. ). *352*, 189–196.

Tivol, E.A., Borriello, F., Schweitzer, A.N., Lynch, W.P., Bluestone, J.A., and Sharpe, A.H. (1995). Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. Immunity.

Tomasetti, C., and Vogelstein, B. (2015). Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science (80-. ). *347*, 78–81.

Tomasetti, C., Li, L., and Vogelstein, B. (2017). Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science (80-. ). *355*, 1330–1334.

Topalian, S.L., Hodi, F.S., Brahmer, J.R., Gettinger, S.N., Smith, D.C., McDermott, D.F., Powderly, J.D., Carvajal, R.D., Sosman, J.A., Atkins, M.B., et al. (2012). Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer. N. Engl. J. Med. *366*, 2443–2454.

Topalian, S.L., Taube, J.M., and Pardoll, D.M. (2020). Neoadjuvant checkpoint blockade for cancer immunotherapy. Science (80-. ). *367*.

Tran, E., Turcotte, S., Gros, A., Robbins, P.F., Lu, Y.C., Dudley, M.E., Wunderlich, J.R., Somerville, R.P., Hogan, K., Hinrichs, C.S., et al. (2014). Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. Science (80-. ).

Tüting, T., and De Visser, K.E. (2016). How neutrophils promote metastasis. Science (80-. ). *352*, 145–146.

Varghese, A. (2015). Chemotherapy for Stage II Colon Cancer. Clin. Colon Rectal Surg. *28*, 256–261.

Waldman, A.D., Fritz, J.M., and Lenardo, M.J. (2020). A guide to cancer immunotherapy: from T cell basic science to clinical practice. Nat. Rev. Immunol. *20*, 651–668.

Wellenstein, M.D., Coffelt, S.B., Duits, D.E.M., van Miltenburg, M.H., Slagter, M., de Rink, I., Henneman, L., Kas, S.M., Prekovic, S., Hau, C.S., et al. (2019). Loss of p53 triggers WNT-dependent systemic inflammation to drive breast cancer metastasis. Nature *572*, 538–542.

Van de Wetering, M., Sancho, E., Verweij, C., De Lau, W., Oving, I., Hurlstone, A., Van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A.P., et al. (2002). The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell *111*, 241–250.

Winton, D.J., Blount, M.A., and Ponder, B.A.J. (1988). A clonal marker induced by mutation in mouse intestinal epithelium. Nature *333*, 463–466.

Wu, P., Wu, D., Ni, C., Ye, J., Chen, W., Hu, G., Wang, Z., Wang, C., Zhang, Z., Xia, W., et al. (2014). γδT17 cells promote the accumulation and expansion of myeloid-derived suppressor cells in human colorectal cancer. Immunity *40*, 785–800.

van Wyk, H.C., Roseweir, A., Alexander, P., Park, J.H., Horgan, P.G., McMillan, D.C., and Edwards, J. (2019). The Relationship Between Tumor Budding, Tumor Microenvironment, and Survival in Patients with Primary Operable Colorectal Cancer. Ann. Surg. Oncol. *26*, 4397–4404.

Xie, Y., Akpinarli, A., Maris, C., Hipkiss, E.L., Lane, M., Kwon, E.K.M., Muranski, P., Restifo, N.P., and Antony, P.A. (2010). Naive tumor-specific CD4+ T cells differentiated in vivo eradicate established melanoma. J. Exp. Med.

Yaeger, R., Chatila, W.K., Lipsyc, M.D., Hechtman, J.F., Cercek, A., Sanchez-Vega, F., Jayakuma-

ran, G., Middha, S., Zehir, A., Donoghue, M.T.A., et al. (2018). Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. Cancer Cell *33*, 125-136.e3.

Yamamoto, K., Venida, A., Yano, J., Biancur, D.E., Kakiuchi, M., Gupta, S., Sohn, A.S.W., Mukhopadhyay, S., Lin, E.Y., Parker, S.J., et al. (2020). Autophagy promotes immune evasion of pancreatic cancer by degrading MHC-I (Springer US).

Yang, Q., Bermingham, N.A., Finegold, M.J., and Zoghbi, H.Y. (2001). Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. Science (80-. ). *294*, 2155–2158.

Yao, Y., Xu, X., Yang, L., Zhu, J., Wan, J., Shen, L., Xia, F., Fu, G., Deng, Y., Pan, M., et al. (2020). Patient-Derived Organoids Predict Chemoradiation Responses of Locally Advanced Rectal Cancer. Cell Stem Cell *26*, 17-26.e6.

Yu, J., Green, M.D., Li, S., Sun, Y., Journey, S.N., Choi, J.E., Rizvi, S.M., Qin, A., Waninger, J.J., Lang, X., et al. (2021). Liver metastasis restrains immunotherapy efficacy via macrophage-mediated T cell elimination. Nat. Med. *27*, 152–164.

Yui, S., Nakamura, T., Sato, T., Nemoto, Y., Mizutani, T., Zheng, X., Ichinose, S., Nagaishi, T., Okamoto, R., Tsuchiya, K., et al. (2012). Functional engraftment of colon epithelium expanded in vitro from a single adult Lgr5 + stem cell. Nat. Med. *18*, 618–623.

Yui, S., Azzolin, L., Maimets, M., Pedersen, M.T., Fordham, R.P., Hansen, S.L., Larsen, H.L., Guiu, J., Alves, M.R.P., Rundsten, C.F., et al. (2018). YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. Cell Stem Cell *22*, 35-49.e7.

Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S.M., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat. Med. *23*, 703–713.

Zhang, L., Li, Z., Skrzypczynska, K.M., Fang, Q., Zhang, W., O'Brien, S.A., He, Y., Wang, L., Zhang, Q., Kim, A., et al. (2020). Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. Cell *181*, 442-459.e29.

(2017). The problem with neoantigen prediction. Nat. Biotechnol. *35*, 97–97.

# Appendix

# Appendix 1. Supplementary tables

**Supplementary Table 1 | Descriptive table of CRC metacohort**

Datasets used for the metacohort of 1830 patients and clinical information for each one.

| Dataset (GSE) | | TCGA | 38832 | 44076 | 33113 | 14333 | 39582 | 37892 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|
| | | (538) | (122) | (98) | (90) | (290) | (562) | (130) | (1830) |
| Gender | Male | 280 | | 71 | 42 | 164 | 307 | 69 | 933 |
| | | (52.0%) | | (72.4%) | (46.7%) | (56.6%) | (54.6%) | (53.1%) | (54.6%) |
| | Female | 258 | | 27 | 48 | 126 | 255 | 61 | 775 |
| | | (48.0%) | | (27.6%) | (53.3%) | (43.4%) | (45.4%) | (46.9%) | (45.4%) |
| | Miss. | 0 | 122 | 0 | 0 | 0 | 0 | 0 | 122 |
| | | (0.0%) | (100.0%) | (0.0%) | (0.0%) | (0.0%) | (0.0%) | (0.0%) | (6.7%) |
| Age at diagnosis | Mean (Sd) | 65.36 (12.87) | | 70.54 (9.02) | 70.39 (12.95) | 65.96 (12.52) | 66.84 (13.28) | 68.25 (12.69) | 66.73 (12.83) |
| | Miss. | 0 (0.0%) | 122 (100.0%) | 0 (0.0%) | 1 (1.1%) | 0 (0.0%) | 1 (0.2%) | 0 (0.0%) | 124 (6.8%) |
| Tumor location | Right | 191 | | 38 | | 125 | 222 | 57 | 633 |
| | | (41.9%) | | (38.8%) | | (43.7%) | (39.5%) | (44.2%) | (41.3%) |
| | Left | 117 | | 60 | | 122 | 340 | 72 | 711 |
| | | (25.7%) | | (61.2%) | | (42.7%) | (60.5%) | (55.8%) | (46.4%) |
| | Rectum | 148 | | 0 | | 39 | 0 | 0 | 187 |
| | | (32.5%) | | (0.0%) | | (13.6%) | (0.0%) | (0.0%) | (12.2%) |
| | Miss. | 82 | 122 | 0 | 90 | 4 | 0 | 1 | 299 |
| | | (15.2%) | (100.0%) | (0.0%) | (100.0%) | (1.4%) | (0.0%) | (0.8%) | (16.3%) |
| Stage | I | 93 | 18 | 0 | 0 | 44 | 33 | 0 | 188 |
| | | (17.9%) | (14.8%) | (0.0%) | (0.0%) | (15.2%) | (5.9%) | (0.0%) | (10.4%) |
| | II | 198 | 35 | 98 | 90 | 94 | 264 | 73 | 852 |
| | | (38.1%) | (28.7%) | (100.0%) | (100.0%) | (32.4%) | (47.0%) | (56.2%) | (47.0%) |
| | III | 157 | 39 | 0 | 0 | 91 | 205 | 57 | 549 |
| | | (30.2%) | (32.0%) | (0.0%) | (0.0%) | (31.4%) | (36.5%) | (43.8%) | (30.3%) |
| | IV | 72 | 30 | 0 | 0 | 61 | 60 | 0 | 223 |
| | | (13.8%) | (24.6%) | (0.0%) | (0.0%) | (21.0%) | (10.7%) | (0.0%) | (12.3%) |
| | Miss. | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 18 |
| | | (3.3%) | (0.0%) | (0.0%) | (0.0%) | (0.0%) | (0.0%) | (0.0%) | (1.0%) |
| MSI status | MSS | 462 | 99 | 98 | 61 | 240 | 483 | 130 | 1573 |
| | | (85.9%) | (81.1%) | (100.0%) | (67.8%) | (82.8%) | (85.9%) | (100.0%) | (86.0%) |
| | MSI | 76 | 23 | 0 | 29 | 50 | 79 | 0 | 257 |
| | | (14.1%) | (18.9%) | (0.0%) | (32.2%) | (17.2%) | (14.1%) | (0.0%) | (14.0%) |
| Time of follow-up (years, not events) | Median (MAD) | 1.63 (1.28) | 3.34 (2.44) | | 3.64 (1.62) | 3.96 (2.22) | 5.33 (2.72) | 4.17 (1.45) | 3.83 (2.85) |

**Supplementary Table 2 | Immunofluorescence protocol**

This table shows the antibodies that were used for immunofluorescence stainings, as

| Antibody | Company | Reference | Antigen retrieval | Dilution | Time inc. | Host |
|---|---|---|---|---|---|---|
| tdTomato | Rockland | 600-401-379 | Tris-EDTA | 1/100 | ON 4ºC | Rabbit |
| tdTomato | Sicgen | AB8181-200 | Tris-EDTA | 1/100 | ON 4ºC | Goat |
| GFP | Abcam | ab6673 | Tris-EDTA | 1/200 | 2h RT | Goat |
| E-Cadherin | BD | 610182 | Tris-EDTA | 1/300 | ON 4ºC | Mouse |
| CD31 | Abcam | ab28364 | Tris-EDTA | 1/300 | ON 4ºC | Rabbit |
| KRT20 | Sigma | HPA024309 | Tris-EDTA | 1/100 | ON 4ºC | Rabbit |
| CD3 | Abcam | ab16669 | Tris-EDTA | 1/500 | ON 4ºC | Rabbit |

**Supplementary Table 3 | qRT-PCR primers**

This table shows the Taqman primers used for the RT-qPCRs.

| Gene Symbol | Taqman probe |
|---|---|
| EMP1 | Mm00515678_m1 |
| KRT20 | Mm01306857_m1 |
| LAMC2 | Mm00500494_m1 |
| LGR5 | Mm0043889_m1 |
| PPIA | Mm002342430_g1 |
| MUC2 | Mm01276696_m1 |
| SMOC2 | Mm00491553_m1 |
| MKI67 | Mm01278671_m1 |
| CD3 | Mm00445553_m1 |
| GZMB | Mm00442837_m1 |

# Appendix 2. Publications

Tauriello, D. V. F., Palomo-Ponce, S., Stork, D., Berenguer-Llergo, A., Badia-Ramentol, J., Iglesias, M., Sevillano, M., Ibiza, S., **Cañellas, A.**, Hernando-Momblona, X., Byrom, D., Matarin, J. A., Calon, A., Rivas, E. I., Nebreda, A. R., Riera, A., Attolini, C. S. O., & Batlle, E. (2018). TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature*, 554(7693), 538–543.

Urosevic, J., Blasco, M. T., Llorente, A., Bellmunt, A., Berenguer-Llergo, A., Guiu, M., **Cañellas, A.**, Fernandez, E., Burkov, I., Clapes, M., Cartana, M., Figueras-Puig, C., Batlle, E., Nebreda, A. R., & Gomis, R. R. (2021). ERK1/2 signaling induces upregulation of ANGPT2 and CXCR4 to mediate liver metastasis in colon cancer. *Cancer Research*, 80, 4668–4680.

Bram Herpers, Berina Eppink, Mark I. James, Carme Cortina3, **Adrià Cañellas-Socias**, Sylvia F. Boj, Xavier Hernando-Momblona, Dominik Glodzik, Rob C. Roover, Marc van de Wetering, Carina Bartelink Clements, Vanessa Zondag van der Zand, Jara Garcia Mateos, Kuan Yan , Lucia Salinaro, Abdul Basmeleh, Szabolcs Fatrai, David Mausssang, Jeroen J Lammerts van Bueren, Irene Chicote, Garazi Serna, Laia Cabellos, Lorena Ramirez, Paolo Nuciforo , Ramon Salazar, Cristina Santos, Alberto Villanueva, Camille Stephan-Otto-Attolini, Elena Sancho, Hector G. Palmer, Josep Tabernero, Michael R. Stratton, John de Kruif, Ton Logtenberg, Hans Clevers, Leo S. Price, Robert Vries, Eduard Batlle* & Mark Throsby*. (2022) Functional screening on patient-derived 1 organoids identifies a therapeutic bispecific antibody that triggers EGFR degradation in LGR5+ tumor cells. *Nature Cancer*, *in press*.

**Adrià Cañellas-Socias**, Carme Cortina, Xavier Hernando-Momblona, Gemma Turon, Lidia Mateo, Felipe Slebe, Sergio Palomo-Ponce, Tamara Sipka, Marta Sevillano, Adrià Caballé-Mestres, Antonio Berenguer-Llergo, Laura Jiménez-Gracia, Sefora Conti, Adrian Álvarez-Varela, Diana Stork, Nicola Fenderico, Lidia Bardia, Julien Colombelli, Xavier Trepat, Holger Heyn, Daniele VF Tauriello, Sabine Tejpar, Elena Sancho, Camille Stephan-Otto Attolini, Eduard Batlle. (-) High risk of metastatic recurrence in colorectal cancer by residual EMP1+ tumor cells. *Nature, in revision*.

Adrián Álvarez-Varela, Laura Novellasdemunt, Francisco M. Barriga, Xavier Hernando-Momblona, **Adrià Cañellas-Socias**, Sara Cano-Crespo, Marta Sevillano, Carme Cortina, Diana Stork, Clara Morral, Gemma Turon, Laura Jiménez-Gracia, Ginevra Caratù, Holger Heyn, Daniele VF Tauriello, Lidia Mateo, Sabine Tejpar, Elena Sancho, Camille Stephan-Otto Attolini, Eduard Batlle. (-) Identification of colorectal cancer cells that mediate relapse after chemotherapy by marker gene Mex3a. *Nature, in revision*.

Marcos Fernández-Alfara, Annarita Sibilio, Judit Martin, Victor Alcalde, Veronica Chanes, **Adrià Cañellas-Socias**, Eduard Batlle and Raúl Méndez. (-) A CPEB4-mediated response to chronic endoplasmic reticulum stress is required for the antitumor effector function of CD8 T lymphocytes. *Science Advances, submitted*.

## Appendix 3. Author contributions

Chapter I

Adrià Cañellas Socias (AC-S) and Eduard Batlle (EB) conceived the study, coordinated experiments and wrote the manuscript. AC-S designed and performed key experiments including the profiling of residual disease, the analyses of tumor buds and micrometastases and genetic ablation studies in the CRC relapse model. AC-S generated the figures for the manuscript. Carme Cortina, Gemma Turon, Felipe Slebe and Diana Stork generated and characterized MTO knock-in lines (Emp1-iCasp9-tdTom, Lgr5-DTR-GFP and Lgr5-iCasp9-tdTom). AC-S, Xavier Hernando-Momblona and Sergio Palomo-Ponce developed the CRC relapse model and performed all mice work. Lidia Mateo (LM) and Camille Stephan-Otto Attolini (CS-OA) analyzed mouse Smart-seq scRNAseq. LM and Laura Jiménez-Gracia analyzed mouse 10X scRNAseq. CS-OA analyzed human scRNA-seq and performed statistical analyses. Tamara Sipka imaged organoids in vitro. Sefora Conti and Xavier Trepat further characterized Emp1 cells in vitro. AC-S and Adrián Álvarez-Varela quantified immunofluorescence stainings. AC-S and Nicola Fenderico developed the method to purify residual tumor cells in whole livers. Marta Sevillano performed IF. Adrià Caballé Mestres analyzed expression arrays. Lidia Bardia and Julien Colombelli performed organ clearing and 3D lightsheet imaging of mice livers with metastases. Patricia Lorden and Holger Heyn provided support with scRNAseq experiments. Sabine Tejpar generated single cell RNA sequencing data from CRC patient samples. Daniele Tauriello generated MTOs and provided support. Elena Sancho provided strategic support and helped with figures and manuscript writing. EB supervised the study.

This manuscript is currently in revision at Nature.

Chapter II

Adrià Cañellas Socias (AC-S) and Eduard Batlle (EB) conceived the study and coordinated experiments. AC-S generated figures and wrote the manuscript drat. AC-S designed and performed key experiments including the analyses of residual microlesions, scRNAseq profiling of the TME and immunotherapy experiments in the CRC relapse model. Laura Jiménez Gracia analyzed 10x scRNAseq. AC-S and Eoghan Mullholland analyzed multiplex immunofluorescence. AC-S, Sergio Palomo-Ponce and Xavier Hernando-Momblona performed all mice work. Adrià Caballé Mestres performed statistical analyses of multiplex immunofluorescence. Lidia Mateo and Camille Stephan-Otto Attolini analyzed mouse Smart-seq scRNAseq. Marta Sevillano performed immunohistochemistry stainings. Maria Salvany, Ana Henriques, Alejandro Prados and Paula Nieto provided help to annotate scRNAseq clusters. Holger Heyn provided support with scRNAseq experi-

ments. Simon Leedham provided support with multiplex immunofluorescence.

This manuscript is in preparation.

Chapter III

Adrià Cañellas Socias (AC-S) and Eduard Batlle (EB) conceived the study and coordinated experiments. AC-S generated figures and wrote this chapter. AC-S designed and performed key experiments including metastatic latency of triple mutant tumors, analysis of their TME and the profiling of triple mutant tumors in suboptimal conditions. AC-S and Sergio Palomo-Ponce performed all mice work. Adrián Álvarez Varela analyzed Mex3a in triple mutant tumors in suboptimal growth conditions. Daniele Tauriello generated MTOs and provided advice during the project. Marta Sevillano performed immunohistochemistry. CS-OA analyzed RNAseq and exome-sequencing data.

Part of the data in chapter III are included in an article (Álvarez-Varela et al) currently in revision at Nature.

# Acknowledgements

Miro enrere i me n'adono que cada dia d'aquests últims 5 anys m'he despertat amb ganes de venir al laboratori. Vull aprofitar aquest apartat per donar les gràcies a les persones que han fet de la tesi una experiència brutal.

Gràcies Eduard per confiar amb mi, però sobretot per aquesta interacció que hem mantingut i que m'ha fet disfrutar al màxim de la recerca. La teva motivació és contagiosa i a mi, m'has fet creure que el que fem al laboratori és important. Acabo la tesi pensant que hem fet coses noves, que farem una bona aportació al camp i espero que algun dia, tota aquesta feina serveixi per ajudar a algú i a la seva família a sofrir una mica menys. Gràcies Elena per la teva inesgotable ajuda i per donar el teu toc d'espontaneïtat, crits i diversió al lab. Formeu un equip invencible!

Thank you Daniele. I learnt so much from you... I realize that, still nowadays, I keep doing so many things just as you taught me. I was extremely lucky to have you as a mentor. Stormborn, I have truly missed you these last years, it felt incredibly good to chat with you. I will come to visit soon.

Gracias Emmesito, no podría haber tenido un mejor compañero de tesis. Juntos nos lo hemos pasado en grande y a la vez disfrutado muchísimo de la ciencia. A los de las parties... Gracias Dron por aceptarme como nuevo user, por toda tu currazo, por soltar tacos constantemente, por todas las risas... sin ti esto no hubiera sido lo mismo. Gracis Pons, per tota sa feinada que hem fet junts, per ses teves filosofades, ses teves invencions i per compartir es teu punt de vista únic sobre sa vida amb mi. Gracis Confit, no ho sé posar en paraules però quan estic amb tu me sent molt bé. Gracias al Ahueonao, que ya no está en el lab pero es un amigo para toda la vida. Cuando la Gioia este un poco más mayorcita te espero de nuevo para venir a correr a pleno sol. Gracis Gremmis i Bardiens per fernos sentir tan acollits quan vam arribar al lab i tots els plans que hem fet junts. Gracis senyora Calrissien per aportar tanta energia i bon rotllo. Gracias Tiiiika por tu curro, tu alegria, pero sobretodo, por aguantar mi pesadez a diario. I parlant de molestar... gràcies Kefa Nouvilles of the World, per ser tan bona gent i aguantar ses meves locures. Thank you Oroman for coping with such a bad teacher. Thank you Nicola, I had a marvelous time with you. Thanks to every member of the Batlle lab that helped to generate such a friendly and healthy atmosphere during all these years. Messi, Geri, Henrietta, Alex, Tamara, Jelena, Angela, Sara, Chang, Doris, Gerhard, Cris, Clara, Francisco, Enza, Elisa... It has been a pleasure to share this adventure with you.

Más allá del lab, ha sido un privilegio poder hacer la tesis en el IRB. Mil gracias al increíble equipo de bioinformática que tantísimo nos ha ayudado: Camille, Toni, Lidia y Adrià. Camille, gracias por aguantarme. La interacción contigo es de las mejores experiencias que me llevo del doctorado. Gracias Nacho y David, por ser tan buena gente y es-

tar siempre dispuestos a ayudar. Mil gracias a las chicas de Histopatología, al equipo de Microscopia, el equipo de Citómetria y a la gente del estabulario por la constante ayuda durante todos estos años.

Gracias Marcos por ser un oasis de inspiración, de diversión, de chorradas, de conversaciones serias y de “ohs”. Gracis Moki, estic convençut que aquesta tesi no hagués estat la mateixa sense els ciments que em vas donar. Viure un any entre niguls va valdré la pena per haver-te conegut. Gracies Laura JG per tota sa teva ajuda i energia positiva (els meetings online no m’inspiren... et falta un mote!). Y un gracias general a toda la gente del IRB que genera este ambiente tan guay.

A més a més voldria agrair el suport de tota la gent de fora del laboratori. Quan et sents envoltat de gent que t’estima i que t’estimes, tots els entrebancs de la vida resulten més fàcils.

A nes millor grup d’amics i amigues que se pot tenir. Gracis per ser com sou. Gracis a sa meva estimada familia. Mama, Victor, ho donaría tot perquè estessiu aquí. Vos enyor immensament però vos tenc ben guardats dins meu. Victor, cada any que passa ets més es meu referent. Mama, aquesta tesi hagués estat impossible sense tu. Papa, Joana, Carlos, Maria, Joaquin, Padri, Fatima i tota sa familia, me sent molt afortunat de tenir-vos a nes meu costat, vos estim molt. Me gustaría terminar esta tesis de la forma más alegre, y eso me lleva a ti Leyre. Gracias por estos fantásticos años juntos. Gracias por soportarme y darme otro ángulo para ver la vida. Ya sabes que para mi eres una persona con algo muy especial.