



UNIVERSITAT DE  
BARCELONA

## A genome editing based approach to study tumor cell heterogeneity

Gemma Turon Rodrigo

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT DE  
BARCELONA

Programa de Doctorat en Biomedicina de la Universitat de Barcelona

Facultat de Biologia

A GENOME EDITING BASED APPROACH TO STUDY TUMOR  
CELL HETEROGENEITY

Memòria presentada per  
GEMMA TURON RODRIGO  
per optar al grau de  
**Doctora per a la Universitat de Barcelona**

Tesi realitzada sota la direcció del Dr. Eduard Batlle i la Dra. Carme Cortina al  
grup de Colorectal Cancer de l'IRB Barcelona i sota la tutela de la Dra.  
Montserrat Corominas del Departament de Genètica de la Facultat de Biologia  
de la UB.

Universitat de Barcelona

2014-2019

EDUARD BATLLE	CARME CORTINA	MONTSERRAT COROMINAS	GEMMA TURON
Co-Director	Co-Directora	Tutora	Candidata



Als pares.



## AGRAÏMENTS

---

Aquest apartat podria ser tan llarg com la tesi mateixa, intentaré ser breu. Primer, agrair a l'Edu la oportunitat de formar part d'aquest grup i la confiança que ha dipositat en mi des del primer dia, i a la Carme per haver-me ensenyat pràcticament tot el que sé, i que ha portat a una nova dimensió el treball en equip. A la resta de Batlle lab - Elena, Sergio, Xavi, Marta, Jelena, Diana, Daniele, Felipe, Jordi, Clara, Adrià, Adrián, Maria - moltes gràcies per ser-hi sempre, per estar disposats a ajudar, col·laborar, rentar histologies els caps de setmana. Especialment gràcies al Xavi, per la paciència infinita en ensenyar-me i, sobretot, per convertir les hores d'estabulari en les millors del dia a dia, i acompanyar-me fins a Holanda i a on faci falta. I clarament, ja ho sabeu, Young Batlles, m'heu donat molt més del que m'atreviria a demanar, i sé que això no acaba aquí. Clara, companya de despatx i primeres hores, ja sigui per parlar d'experiments, esquiades, esport o plans de futur sempre hi ets. Jordi, m'has donat sempre la tranquil·litat de no estar sola, i la teva confiança de que tot anirà bé m'ha animat en els dies més difícils. Adrià i Adrián, sou un pack inseparable i, des del primer dia, indispensables. Us culpo d'haver de passar caps de setmana viatjant, esquiant, veient sortir el sol massa dies. Maria, la incorporació més recent i amb més energia, la teva motivació i ganes de provar coses alegres a qualsevol, gràcies per no dubtar ni quan et dic que et pengis cap per avall. I finalment vull agrair als membres temporals del grup que ja en semblen un més, Angela i Chang, i als antics EBatlle que em veu rebre tan bé – Enza, Alex, Peter, Elisa, Mark, Francisco -.

Aquesta etapa no hagués sigut el mateix sense la resta de PhDs, encara em sorprèn el ràpid que ens vam entendre i tot el que hem aconseguit fer malgrat la feina que teníem. Us heu convertit en gent molt especial per a mi i espero que ens trobem per aquests móns sovint. Laura, HelenaM i Cris, em feu massa feliç, a la meua tenda sempre hi tindreu un lloc. HelenaR, millor veci, espero que es multipliquin els nostres sopars. Ernest, Anna i Craig, crec que el Toni es pensa que treballa en el seu lab, gràcies per acollir-me sempre. Jurgen, Alex i Ricardo, no fallais nunca, sobre todo si hay cerveza de por medio! Martí, les muntanyes seran sempre nostres, gràcies per trobar moments entre les teves mil activitats. Ela, Bushra, Gianmarco, Albert, amics de promoció i passadís. I la llista pot seguir infinitament, moltes gràcies Freixo, Jessi, Carlos, Salva, Enric, Marcos, Sandra, Berta, Clara, Alicia, Irene, Núria, Víctor, Marc(s), Joana, Rosa.

Tampoc em vull oblidar dels que des de fora m'heu acompanyat aquests anys, sense dubtar mai de mi. Roger i Andrea, imprescindibles. Esther, Ana, Blanca, Rosa i Marta, gràcies per donar-me un espai on ser jo. I gràcies, gràcies a totes aquelles que, irònicament, heu fet que toqui més de peus a terra durant aquests anys: Anna, Joan, Yoli, Rebe, Alba, Clares, Yaiza, Biel, Bea, Elis, Juana, Kat, Thais, Jordi, Tommy, Judit, Ingrid, Míriam, Anna, Miquel, Guada, Lucho, César, Numa i un llarg etc. Evidentment, gràcies a la família, en el seu sentit més extens, aquells "que sempre hi seran", tiets/es, cosins/es, i també padrins/es, quasi-germanes, amics, amigues... Als meus

pares, que han fet sempre el possible perquè sigui feliç, i a la meva germana, que em recorda la bellesa de la ciència, espero que en segueixis gaudint sempre.

I només em queda donar les gràcies al Miquel, per deixar-me descobrir el món a través dels seus ulls i recordar-me què és el més important. No falles mai, amb tu, tot és possible.

## CONTENTS

---

<b>i</b>	<b>INTRODUCTION</b>	1
1	INTRODUCTION	3
1.1	Adult stem cells . . . . .	3
1.1.1	Properties of adult stem cells . . . . .	3
1.1.2	Division modes of adult stem cells . . . . .	4
1.1.3	Plasticity of adult stem cells . . . . .	5
1.1.4	Examples of human adult stem cells . . . . .	6
1.1.5	Identification of adult stem cells . . . . .	7
1.2	Cancer stem cells . . . . .	10
1.2.1	Properties of CSCs . . . . .	10
1.3	The mammalian intestine . . . . .	12
1.3.1	Intestinal structure and function . . . . .	12
1.3.2	Signaling pathways involved in ISC renewal and differentiation . . . . .	13
1.3.3	Differentiated intestinal cell types . . . . .	16
1.3.4	Intestinal stem cells . . . . .	17
1.4	Colorectal cancer . . . . .	22
1.4.1	CRC staging . . . . .	22
1.4.2	Types of CRC . . . . .	23
1.4.3	Colon cancer stem cells . . . . .	27
1.4.4	Tumor microenvironment and the CSC theory . . . . .	29
1.4.5	Metastasis in CRC . . . . .	30
1.5	Genome-editing tools . . . . .	32
1.5.1	Zinc finger nucleases (ZNF) . . . . .	33
1.5.2	Transcription activation like elements (TALE) . . . . .	33
1.5.3	CRISPR/Cas9 . . . . .	33
<b>ii</b>	<b>OBJECTIVES</b>	35
2	OBJECTIVES	37
<b>iii</b>	<b>RESULTS</b>	39
3	RESULTS	41
3.1	Characterization of LGR5+ cells in human CRC organoids . . . . .	43
3.1.1	CRISPR/Cas9 labelling of <i>LGR5</i> gene . . . . .	43
3.1.2	EGFP knock-in reports LGR5 levels . . . . .	44
3.1.3	CRISPR/Cas9 genome editing does not introduce random mutations in PDO7 DNA) . . . . .	45
3.1.4	LGR5 marks a stem-like population in human CRC xenografts . . . . .	46
3.1.5	LGR5+ cell population fuels tumor growth . . . . .	52
3.2	Labelling proliferative cells in human CRCs . . . . .	58
3.2.1	Labelling endogenous KI67 protein by CRISPR/Cas9 . . . . .	58
3.2.2	One third of the LGR5+ cells are in a slow cycling state . . . . .	61
3.3	Defining the differentiated compartment of CRC tumors . . . . .	65

3.3.1	KRT20 marks differentiated cells in human tumors . . . . .	65
3.3.2	Characterization of EMP1+ cells . . . . .	70
3.4	EMP1 defines a subset of CRC cells with invasive characteristics . . . . .	77
3.4.1	Tagging EMP1 with reporter and suicide cassettes using CRISPR/- Cas9 technology . . . . .	77
3.4.2	EMP1 marks a proliferating differentiated-like population <i>in vivo</i> . . . . .	78
3.4.3	EMP1 expression overlaps with invasive markers . . . . .	82
3.4.4	Innate immunity cells are recruited to the tumor by EMP1+ cells . . . . .	86
3.4.5	Establishment of a conditional EMP1+ cell ablation system . . . . .	87
3.4.6	Failure of PDO6 and PDO7 to generate metastases precluded analysis of the role of EMP1+ CRC cells in the dissemination of the disease . . . . .	89
3.5	Mouse tumor organoids as a model system . . . . .	95
3.5.1	EMP1+ cells retain invasive behavior in MTO-derived tumors . . . . .	97
3.5.2	LGR5 and EMP1 populations overlap in MTO93 . . . . .	104
3.5.3	Study of EMP1+ cells in MTO-derived CRCs . . . . .	108
3.5.4	Primary resection allows liver metastasis seeding . . . . .	113
3.5.5	EMP1+ cells do not have higher metastatic potential when in- jected via intraspleen . . . . .	116
<b>iv</b>	<b>DISCUSSION</b> . . . . .	119
<b>4</b>	<b>DISCUSSION</b> . . . . .	121
<b>v</b>	<b>CONCLUSIONS</b> . . . . .	137
<b>5</b>	<b>CONCLUSIONS</b> . . . . .	139
<b>vi</b>	<b>MATERIALS AND METHODS</b> . . . . .	141
<b>6</b>	<b>MATERIALS AND METHODS</b> . . . . .	143
6.1	Organoid cultures . . . . .	143
6.1.1	Patient Derived Organoids . . . . .	143
6.1.2	Mouse Tumor Organoids . . . . .	143
6.2	CRISPR/Cas9 knock-in design . . . . .	144
6.2.1	Donor plasmid construction . . . . .	144
6.2.2	Guide RNA design . . . . .	144
6.2.3	px330-iRFP Cas9 plasmid construction . . . . .	144
6.2.4	CRISPR/Cas9 knock-in generation in PDOs and MTOs . . . . .	145
6.3	Lentiviral infections . . . . .	147
6.4	Mouse studies . . . . .	147
6.4.1	Subcutaneous xenografts . . . . .	147
6.4.2	Orthotopic transplantations . . . . .	149
6.4.3	Intraspleen injections . . . . .	150
6.5	Quantitative real-time PCR . . . . .	150
6.6	Microarray analysis . . . . .	151
6.7	Whole exome sequencing . . . . .	152
6.8	Immunostaining and quantifications . . . . .	152
6.8.1	Immunostaining and confocal imaging of <i>in vitro</i> cultured PDOs . . . . .	152
6.8.2	Immunostaining on paraffin sections . . . . .	153

6.8.3	Quantification of lineage tracing clone number and area . . . . .	154
6.8.4	Quantification of LGR5-EGFP+ cells within the xenografts . . . . .	155
6.8.5	Visualization of 1 month grown <i>LGR5</i> -derived clones . . . . .	155
vii	<b>BIBLIOGRAPHY</b>	157
7	<b>BIBLIOGRAPHY</b>	159
viii	<b>APPENDIX</b>	183
A	<b>APPENDIX</b>	185
A.1	Supplementary Tables . . . . .	185
A.2	Supplementary Figures . . . . .	188
A.3	Article: A genome editing approach to study cancer stem cells in human tumor . . . . .	203

## ACRONYMS

---

**4-OHTam** 4-hydroxytamoxifen

**AAVS1** Adeno-Associated Virus Integration Site 1

**AAVS1-BFPTOM** AAVS1-LoxP-Ubiquitin-BFP-STOP-LoxP-TOM knock-in

**AJCC** American Joint Committee on Cancer

**ALPI** Alkaline Phosphatase, Intestinal

**AML** Acute Myeloid Leukemia

**ANXA** Annexin

**AOM** Azoxymethane

**Apc** Adenomatous Polyposis Syndrome

**APCDD1** Adenomatosis Polyposis Coli Down-Regulated 1

**ASC** Adult Stem Cell

**ASCL2** Achaete-Scute Family BHLH Transcription Factor 2

**AURKB** Aurora Kinase B

**BFP** mTagBFP2 fluorescent protein

**BIRC3** Baculoviral IAP Repeat Containing 3

**BMP** Bone Morphogenic Protein

**BMPR** Bone Morphogenic Protein Receptor

**bp** Base pairs

**BUB1** BUB1 Mitotic Checkpoint Serine/Threonine Kinase

**CABLES** CDK5 and ABL Enzyme Substrate

**CAIX** Carbonic Anhydrase IX

**CAV2** Caveolin2

**CBC** Crypt Base Columnar Cell

**CCLx** C-C motif Chemokine Ligand x

**CCNF** Cyclin F

**CCRx** C-C motif Chemokine Receptor x

CDC45 Cell Division Cycle 45  
CDCA2 Cell Division Cycle Associated 2  
CDCA5 Cell Division Cycle Associated 5  
CDK1 Cyclin Dependent Kinase 1  
CDKN1A Cyclin Dependent Kinase Inhibitor 1A  
CDX2 Caudal Type Homeobox 2  
CEACAM5 Carcinoembryonic Antigen Related Cell Adhesion Molecule 5  
CFP Cyan Fluorescent Protein  
CHEK1 Checkpoint Kinase 1  
CHGA Chromogranin A  
CID Chronic Inducer of Dimerization  
CIMP CpG Island Methilator Phenotype  
CIN Chromosomal Instability  
CMS Consensus Molecular Subtypes  
COCH1 Cochlin  
CRC Colorectal Cancer  
CRISPR Clustered Regularly Interspaced Short Palindromic Repeats  
CSC Cancer Stem Cell  
CTSE Cathepsin E  
CXCL<sub>x</sub> C-X-C motif Chemokine Ligand x  
DACH1 Dachshund Family Transcription Factor 1  
DAPI 4'6'-Diamidino-2-phenylindole  
DCLK1 DoubleCortin Like Kinase I  
DLL<sub>x</sub> Delta Like canonical notch Ligand x  
DMBT1 Deleted In Malignant Brain Tumors 1  
DSB Double Strand Break  
DSS Dextran Sodium Sulphate  
dv1 Disheveled  
ECM Extracellular Matrix

EdU Ethynyl Deoxyuridine  
EFNB1 Ephrin-B1  
EFNB2 Ephrin-B2  
EGFP Enhanced Green Fluorescent Protein  
EGFR EGF Receptor  
EGFR Epidermal Growth Factor  
EMP1 Epithelial Membrane Protein 1  
EMP1-EGFP/Cre EMP1-LF2A-EGFP / EMP1-LF2A-CreERT2 double knock-in  
EMP1-iCT EMP1-inducibleCaspase9-tdTOMATO knock-in  
EMT Epithelial to Mesenchymal Transition  
ENU N-ethyl-N-nitrosurea  
EPCAM Epithelial Cell Adhesion Molecule  
EPHB2 Ephrin Receptor B2  
EPHB3 Ephrin Receptor B3  
ESC Embryonic Stem Cell  
FABP1 Fatty Acid Binding Protein 1  
FACS Fluorescent Activated Cell Sorting  
FAP Familial Adenomatous Polyposis  
FOXM1 Forkhead Box M1  
FS Frameshit Mutation  
GABBR1 Gamma-Aminobutyric Acid Type B Receptor Subunit 1  
GEMM Genetic Engineered Mouse Model  
GFP Green Fluorescent Protein  
GO Gene Ontology  
gRNA guide RNA for Cas9 (tracrRNA+crRNA)  
GSEA Gene-Set Enrichment Analysis  
HE Hematoxilin-Eosin  
HNPCC Hereditary Non Polyposis Colorectal Cancer  
HOPX HOP Homeobox

HR Homologous Recombination  
HSC Hematopoietic Stem Cell  
iCASP9 Inducible Caspase 9  
IF Intermediate Filaments  
IFI27 Interferon Alpha Inducible Protein 27  
IFI44 Interferon Induced Protein 44  
IL- x Interleukin x  
INSM1 INSM Transcriptional Repressor 1  
IRFP Near Infrared Fluorescent Protein  
ISC Intestinal Stem Cell  
kb kilobases  
KI67 Marker of Proliferation Ki67  
KI67-RFP KI67-mTagRFP2 (fusion protein) knock-in  
KLK10 Kallikrein Related Peptidase 10  
KRT20 Keratin 20  
KRT20-Cre KRT20-LF2A-CreERT2 knock-in  
KRT20-IRFP KRT20-IRFP (fusion protein) knock-in  
KRT6A Keratin 6A  
L1CAM L1 Cell Adhesion Molecule  
LAMC2 Laminin Subunit Gamma 2  
LGR5 Leucine Rich Repeat Containing G Protein-Coupled Receptor 5  
LGR5-Cre LGR5-LF2A-CreERT2 knock-in  
LGR5-EGFP LGR5-LF2A-EGFP knock-in  
LOH Loss of Heterozygosity  
LRC Label Retaining Cell  
LRIG Leucine Rich Repeats And Immunoglobulin Like Domains 1  
Luc Luciferase  
LY TGFβ inhibitor LY2157299  
M cells Microfold cells

MAPK Mitogen-Activated Protein Kinase Patwhay  
MDSCs Myeloid Derived Supressor Cells  
MEK Mitogen-activated Protein Kinase Kinase  
MET Mesenchymal to Epithelial Transition  
MIC Metastasis Initiation Capacity  
Min Multiple Intestinal Neoplasia  
MLH1 MutL Homolog 1  
MMP7 Matrix Metallopeptidase 7  
MMR Mismatch Repair  
MSH2 MutS Homolog 2  
MSI microsatellite instability  
MSI1 Musashi RNA Binding Protein 1  
mTERT Telomerase Reverse Transcriptase  
MTO Mouse tumor organoid  
MUC2 Mucin 2  
MYB MYB Proto-Oncogene, Transcription Factor  
MYC MYC Proto-Oncogene, BHLH Transcription Factor  
NEK2 NIMA Related Kinase 2  
NHEJ Non Homologous End Joining  
NS Non Significant  
NT Non treated  
NTHL1 Nth Like DNA Glycosylase 1  
OLFM4 Olfactomedin 4  
OR51E1 Olfactory Receptor Family 51 Subfamily E Member 1  
PAM protospacer adjacent motif  
PCR polymerase chain reaction  
PCR 3'specific PCR to determine if knock-in is correctly inserted (3' end)  
PCR 5'specific PCR to determine if knock-in is correctly inserted (5' end)  
PCR H/H PCR to determine if knock-in is homozigous or heterozigous

PD1 Programmed Cell Death 1  
PD-L1 Programmed Cell Death 1 Ligand  
PDO Patient Derived Organoid  
PDX Patient Derived Xenograft  
PGC1alpha Peroxisome Proliferator Activated Receptor Gamma, Coactivator 1 Alpha  
PI3 Peptidase Inhibitor 3  
PIK3CA Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha  
PPBP / CXCL7 Pro-Platelet Basic Protein  
PPP1R12C Protein Phosphatase 1 Regulatory Subunit 12C  
PRUNE2 Prune Homolog 2 or KIAA0367  
qPCR Quantitative or Real-time PCR  
R26R Rosa-26 mouse safe genomic locus  
RACGAP1 Rac GTPase Activating Protein 1  
RAD51 RAD51 recombinase  
RARRES3 Retinoic Acid Receptor Responder 3  
RFP mTagRFP2 Fluorescent Protein  
RNF43 Ring Finger Protein 43  
ROS Reactive Oxygen Species  
Rspo Rspodin  
SC Stem Cell  
SDCBP2 Syndecan Binding Protein 2  
SI Sucrase-Isomaltase  
SLC35F6 Solute Carrier Family 35 Member F6  
SMOC2 SPARC Related Modular Calcium Binding 2  
SPIM Selective Plane Illumination Microscopy  
SSA Sessile Serrated Adenomas  
TA Transient Amplifying  
TALE Transcription Activation Like Elements  
Tam Tamoxifen

TCGA The Cancer Genome Atlas  
TGFB Transforming Growth Factor Beta  
TIC Tumor Initiation Capacity  
TM4SF20 Transmembrane 4 L Six Family Member 20  
TNM Tumor invasiveness, lymphatic node dissemination and metastases grade  
TNNC1 Troponin C1, Slow Skeletal And Cardiac Type  
TOM TdTOMATO Fluorescent Protein  
TOP2A DNA Topoisomerase II Alpha  
TP53 Tumor Protein p53  
UBE2C Ubiquitin Conjugating Enzyme E2 C  
VEGF Vascular Endothelial Growth Factor  
WT Wild Type  
YFP Yellow Fluorescent Protein  
ZNF Zinc-Finger Nucleases

Part I  
INTRODUCTION



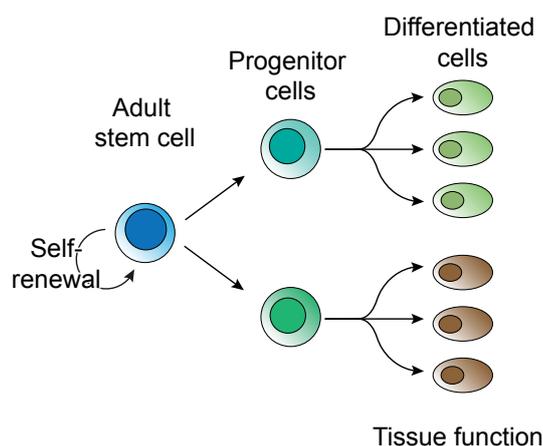
## INTRODUCTION

---

### 1.1 ADULT STEM CELLS

Stem cells (SCs) are the cells of origin of all the tissues in our body. Embryonic-SCs (ESCs) conform the embryo and, as it grows, they differentiate towards specific fates, restricting their potential. This exquisitely calibrated system enables the formation of the distinct body organs, and relies on a strict hierarchy fueled by ESC division (Zernicka-Goetz et al., 2009). This hierarchy is essential not only during development but also for the proper maintenance of adult tissues, as most tissues require cell replacement and turnover for their well-functioning.

SCs found in mature organs are called Adult-SCs (ASCs). ASCs are located in specific niches in the body, maintained in an undifferentiated state yet they are already committed to certain lineages within their resident organ or tissue, a property called multipotency. ASCs are at the apex of the hierarchy, as their division ensures both the maintenance of the stem cell pool and the generation of progenitor cells, their immediate descendants (Clevers, 2015). Progenitors are characterized by a high proliferation rate and a more restricted lineage commitment, but do not retain the self-renewal capacity. As they occupy their correspondent positions in the organs, they become post-mitotic, fully differentiated cells. ASCs are thus in charge of maintaining tissue homeostasis whereas differentiated cells ensure its well-functioning (Figure 1.1).



**Figure 1.1: Adult stem cells are at the apex of the hierarchy.** ASCs maintain the stem cell pool intact through-self renewal and generate proliferating progeny already committed to specific differentiation lineages. Mature tissues are composed of post-mitotic differentiated cells that ensure their function and are constantly replenished by ASC division.

#### 1.1.1 Properties of adult stem cells

Stem cells must fulfill two properties in order to be considered as such:

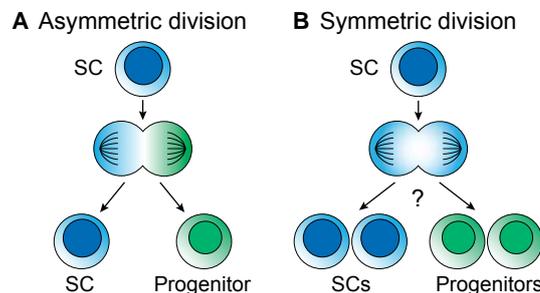
- Self-renewal: ability to proliferate and give rise to a cell with identical properties.

- Multipotency: capacity to specify into more than one differentiation program, which mainly codifies for specific proteins necessary for cell function, such as enzyme production in the gastrointestinal tract. In contrast to the ESC, that are able to give rise to all the embryonic layers (pluripotent), the ASC potential is restricted to a specific tissue (multipotent).

These two properties do not imply that all ASCs behave identically in terms of proliferation. The first described stem cells in adult body were the hematopoietic stem cells (HSC), and they were found to be a pool of quiescent, seldom-dividing cells. In the bone marrow, progenitor cells are the ones with highest proliferative burden (Till et al., 1964). As this was the first system studied, corroborated preconceived ideas such as that the stem cells should not replicate often in order to protect the genetic material prevailed. For years, researchers had long been long trying to find other cells with similar abilities. Instead, they found ASCs with a complete different behavior. For example, intestinal stem cells (ISCs) are highly proliferative, and capable of renewing the whole intestinal epithelium in 5 days (Leblond and Messier, 1958). Therefore, each tissue has different needs and thus different organization.

### 1.1.2 Division modes of adult stem cells

ASCs not only proliferate at different rates, but they may also replicate through different modes in each tissue. Self-renewal can be achieved by two different means (Figure 1.2):



**Figure 1.2: Division modes of adult stem cells.** Asymmetric SC division gives rise to one SC and a progenitor cell in each cell cycle. **B.** Symmetric divisions stochastically give rise either to two adult stem cells or two progenitors.

On one hand, upon division each ASC may give rise to an identical ASC and to a progenitor daughter, in what is called **asymmetrical division**. This process often requires unequal division of cell contents between daughter cells, as is the case of the Numb protein in mouse muscular satellite cells. Numb is a Notch-1 inhibitor, the inheritance of which allows the commitment of the daughter cell to the differentiation path (Conboy and Rando, 2002). Additionally, asymmetric stem cell division can be guided by microenvironment factors. The division of the ASC oriented in a specific plane by the mitotic spindle leads to loss of contact of one of the cells with the stem cell niche, inducing its differentiation. The daughter that remains in closer contact with the niche is kept in a stem state. This process has been well studied in *Drosophila melanogaster* germline cells (Yamashita et al., 2003).

On the other hand, self-renewal can be also achieved at the population level: each ASC stochastically gives rise to two ASCs or two progenitor cells in a **symmetrical division** fashion. By these means, the total number of ASCs remains constant as the

loss of one ASC is compensated by the division of another into two new ASCs. Moreover, symmetric division is a powerful tool to increase stem cell numbers both during development and in response to tissue injury (Morrison and Kimble, 2006). For example, following a brain stroke the mouse subventricular zone neurogenesis is increased via symmetric cell division (Zhang et al., 2004).

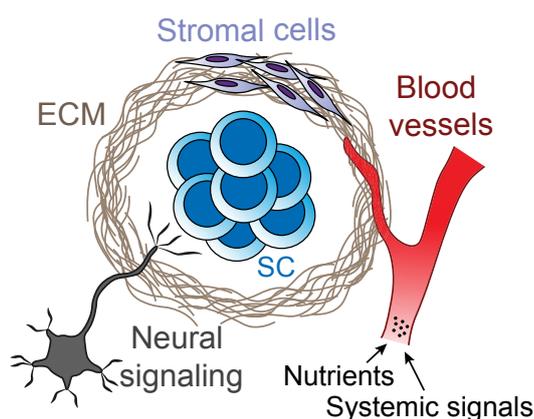
In summary, we have to understand ASCs not as a single population but as particular groups of cells best adapted to fulfill each tissue self-renewal needs. Importantly, their behavior is not hardwired, but rather plastic and adaptable.

### 1.1.3 Plasticity of adult stem cells

Plasticity is defined as the ability of a cell down the hierarchy to revert its differentiation program and give rise to another cell type or even a stem cell. This is a property of essential importance in the stem cell field (Merrel and Stagner, 2016). Upon certain conditions, committed progenitor cells can de-differentiate and replenish a damaged ASC pool, as is the case for the skin (Blanpain and Fuchs, 2014). This ensures tissue integrity and function. *In vitro*, even terminally differentiated cells have been reprogrammed into full pluripotency (Jaenisch and Young, 2008). However, the experimental settings used to study this dramatic change in cell fate are generally far from being physiological.

#### *Stem cell niche*

Stem cell properties do not only depend upon intrinsic factors, but also on environmental cues that are essential to maintain their phenotype. The stem cell niche is the area where the microenvironment produces these signals, and it is thought to play an essential role during plasticity since certain cells de-differentiate when receiving specific niche factors.



**Figure 1.3: Stem cell niche composition.** SC niches are constituted of various elements (extracellular matrix, stromal cells, blood vessels and neurons) that maintain SC function and integrate signals from the whole body.

Each ASC type has a particular niche requirement, but some conserved elements can be found across different niches: an extracellular matrix (ECM) scaffold to provide structural support, blood vessels that maintain a constant supply of nutrients and systemic signals, stromal cells that provide necessary factors and cell-to-cell interactions

and neural inputs that integrate signals from different organs (Figure 1.3) (Ferraro et al., 2010). The niche, then, modulates SC function via both short-range signals (cell-to-cell interactions) and long-range secreted factors (such as WNT proteins secreted in the intestine (Koch, 2017)). A more detailed review of the intestinal stem cell niche factors will be discussed in sections below.

#### 1.1.4 *Examples of human adult stem cells*

The abovementioned diversity of ASCs is well represented within the most studied self-renewing tissues of the body:

##### *Hematopoietic system*

Hematopoietic stem cells (HSCs) were first described in the 1950's thanks to bone marrow transplantation experiments in irradiated mice. Their discovery opened the door to studies on stemness. HSCs are conformed by a pool of rare and quiescent cells (Long Term-HSCs) that seldom divide and have full self-renewal capacity, and a Short Term-HSC population that sustains hematopoiesis in the adult. These Short Term-HSC do not retain full self-renewal ability. (Cheshier et al., 1999, Benveniste et al., 2010, Busch et al., 2015) (Figure 1.4A).

##### *Epidermis*

The epidermis can be divided in two layers: a basal stratum, consistent of a single-cell layer of highly proliferative cells and a spinous or suprabasal stratum, with terminally differentiated cells that express high levels of KRT1 and KRT10 to form the intermediate filaments (IF). IF are essential to keep the skin barrier intact. Basal cells conform the pool of epidermal SC (reviewed in Banplain and Fuchs, 2009). More recent reports suggest the presence of two different SC pools, located in specific areas within the skin basal layer. One is a slow-cycling, label-retaining population, and the other is a highly proliferative population, yet both retain SC properties. Of note, each SC type preferentially gives rise to a certain differentiated cell type, but upon injury they can interconvert (Gómez et al., 2013, Sada et al., 2016) (Figure 1.4B).

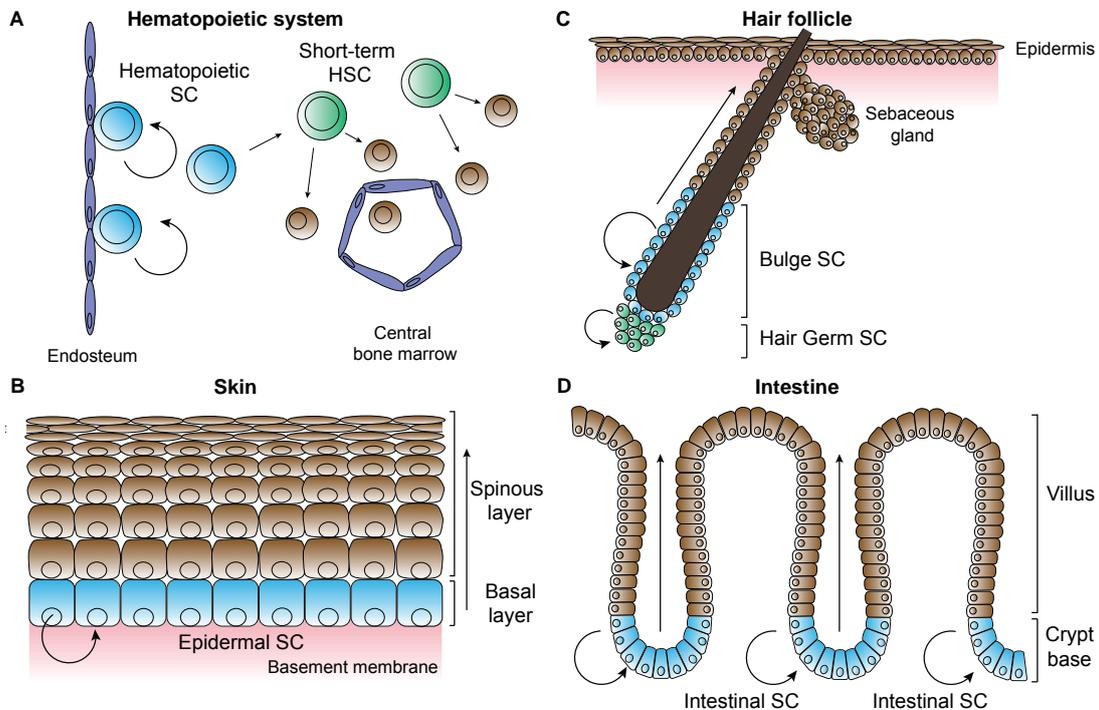
##### *Hair follicle*

The hair follicle is a case of interest due to its cycling activity. In homeostatic setting, it does not renew at a stable rate but rather goes through expansion and regression episodes during adult life. The stem cell compartment can be divided in two elements, the bulge cells, in a quiescent, slow cycling state, and the hair germ cells, situated at the base of the follicle and more prone to proliferate and differentiate. Whether the hair germ cells are bonafide stem cells or already committed to differentiation remains unclear to date (Rompolas et al., 2012, Krieger and Simons, 2015) (Figure 1.4C).

##### *Intestine*

The intestinal epithelium is of special interest in stem cell studies, as is the fastest self-renewing organ of the body. The engine of this turnover is the stem cell pool

residing at the base of the crypts, which proliferates and pushes the progeny out of the niche. As the cells move upwards, they differentiate and are finally shed into the lumen (Clevers, 2013)(Figure 1.4D). The maintenance of intestinal homeostasis will be detailed in **section 1.3**.



**Figure 1.4: Examples of adult self-renewing tissues.** **A.** The hematopoietic system relies on rarely dividing HSC (blue), located at the perimeter of the bone marrow. These HSC divide once every 4-5 months and give rise to amplifying progenitors (green), that proliferate and differentiate into the all blood cell lineages (brown). **B.** The epidermis is sustained by basal cell division (blue), which functions as a stem cell pool. **C.** Hair follicles contain two different types of SC: highly proliferative hair germ cells (green) and bulge cells (blue), more quiescent. **D.** Intestinal self-renewal is fueled by SC located at the crypt base (blue). Their progeny transits up towards the top of the villi where they are shed. Based on Fuchs and Chen, 2013.

### 1.1.5 Identification of adult stem cells

The study of stem cell behavior poses many challenges, as they are not a static population *per se*, but rather adapt and respond to a dynamic environment and are sensitive to both intrinsic and extracellular signals. Moreover, the genes used to follow specific SC populations are often just markers, not proved to be linked to SC functions. All these factors difficult the discrimination between real SC and their progeny. Consequently, a series of functional assays have been developed to facilitate SC studies:

#### *Label-retaining assays*

Quiescence has long been proposed as a hallmark of stemness, based on the findings in the hematopoietic system (**section 1.1.4**). These assays consist in the incorporation

of a DNA label that is diluted with cell divisions, hence a slow-cycling cell will maintain the mark for long periods of time (Figure 1.5A). The discovery of label retaining cells (LRC) in the hair follicle led to the definition of the bulge stem cell pool (Cotsarelis et al., 1990). This type of assays, however, do not take into consideration the possibility of having highly proliferative stem cells in homeostasis, as is the case for the intestine. A detailed discussion of the existence of quiescent and proliferating stem cells in the crypt can be found in **section 1.3.4**.

#### *Clonogenic assays*

Putative stem cells can be isolated from the tissue and grown *in vitro* in a culture system that replicates niche conditions and factors. The colony formation assay measures the ability of single cells to give rise to colonies, as a surrogate of their unlimited division capacity (Franken et al., 2006). The progenitor and differentiated cell types, with limited proliferation capacity, will extinguish with time, whereas the true stem cells will be able to grow and persist over time (Figure 1.5B). The major drawbacks of this assay is that the cells are removed from their native microenvironment and that it does not rule out the possibility of plasticity.

#### *Transplantation assays*

Similarly to clonogenic assays but more physiological, cells can be transplanted *in vivo* into new recipients to observe if they can regenerate the tissue of origin. This method was key to discover the HSCs by bone marrow transplantation (Lorenz et al., 1951), and it has been very useful in a number of other studies, such as to determine the proportion of stem cells in a given organ by limiting dilution experiments. This assay consists in transplanting very few number of cells to determine the relative abundance of SC within a pool (Illa-Bochaca et al., 2010) (Figure 1.5C). Nevertheless, it also requires the isolation of the cells from their original niche, and furthermore, not all tissues from the body can be engrafted into secondary hosts. In addition, cells have to be transplanted to immunosuppressed hosts, and the level of mice immunosuppression has been shown to alter the assay outcome (McDermott et al., 2010).

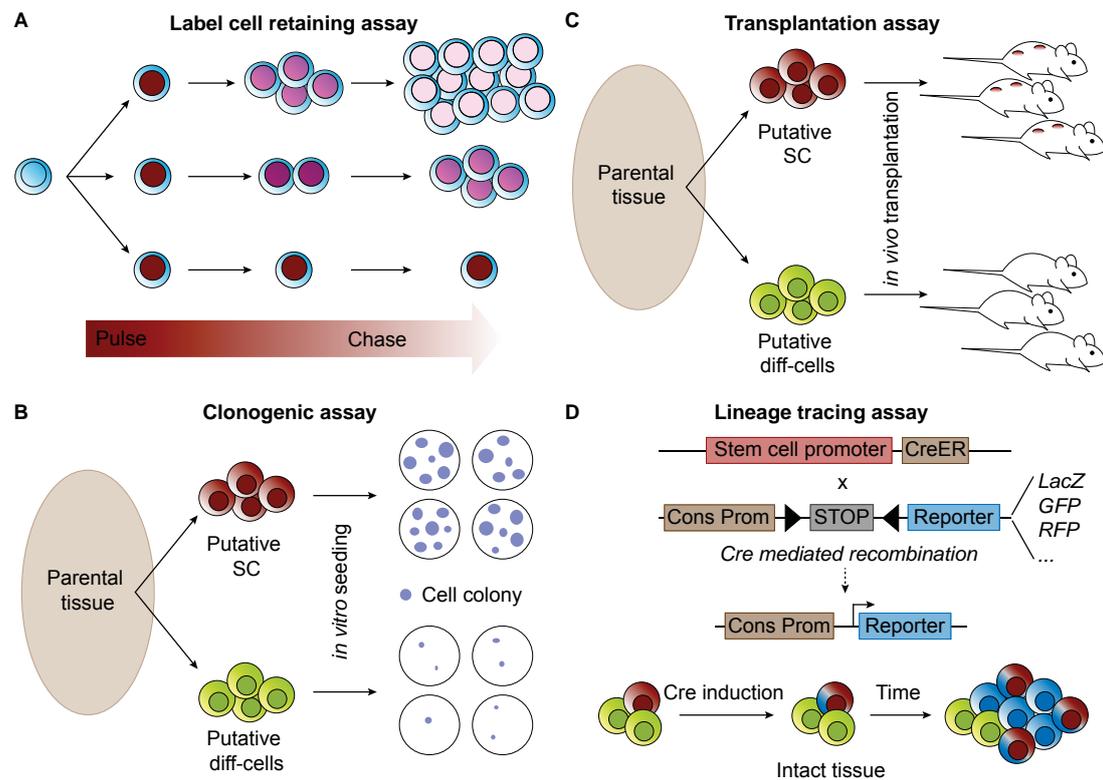
#### *Lineage tracing*

Lineage tracing consists in following the progeny of a specific subpopulation in its native environment. It is considered the golden stem cell assay, as you can visualize how the descendants of a specific cell proliferate by clonal analysis. The tracing strategies are based on genetic modification of the cells to incorporate an inducible reporter gene that upon extrinsic cues will mark the cell of interest and all its descendants. The Cre-LoxP system is widely used due to its simplicity: the Cre enzyme recognizes short DNA sequences named LoxP and induces their recombination, excising from the genome anything encompassed between them (Sauer and Henderson, 1988) (Figure 1.5D).

Generally, the Cre recombinase is placed under the control of the gene of interest, and the reporter system (usually a fluorescent protein) is inserted in a safe locus, with its expression is blocked by a STOP sequence flanked by LoxP sites. When the gene of

interest is expressed, the Cre will remove the transcriptional blockade and the cell will acquire the fluorescent marking. Moreover, the system can be refined to induce the tracing at desired time points by using a modified version of the enzyme, the CreERT2. This enzyme contains a modified fragment of the ligand-binding domain of the estrogen receptor that sequesters the protein in the cytoplasm and translocates to the nucleus only upon Tamoxifen addition (Feil et al., 1997).

This assay is a very powerful tool to observe the behavior of cell populations. Nevertheless, it relies on finding genes that are expressed only in the desired population, and it involves prior genetic manipulation of the cells.



**Figure 1.5: Assays to determine the stemness potential of given a cell population.** **A.** Low proliferation activity is considered a putative SC feature. By label-retaining assays, slow-cycling cells are identified due to the maintenance of a dye over time. The dye is rapidly diluted in actively dividing cells. **B.** Clonogenic assays determine the ability of different cell types to grow *in vitro*. **C.** Transplantation assays consist on reinjecting of the cells in their original environments to observe if they can rebuild the tissue. **D.** Lineage tracing assay is the only method that allows the analysis of growth dynamics of a given population in its native environment. It requires the insertion of an inducible recombinase (i.e CreER) under the control of a marker gene of interest, and a second knock-in containing a reporter protein with a roadblock. Recombination is induced only in the cells that express the stem cell gene, but the label is maintained in all descendants, regardless of whether they still express the stem cell gene or not.

## 1.2 CANCER STEM CELLS

The discovery of how homeostatic tissues organize, self-renew and cope with damage opened the door for a better understanding of cancer, a disease provoked by deregulation of these mechanisms. The last decades have seen the development of the cancer stem cell (CSC) theory. It postulates that, analogous to the organization of healthy tissue, a pool of stem cell-like cells, termed the CSCs, fuels tumor growth. This concept would explain long-known effects such as tumor relapse, tumor dormancy and metastasis.

The idea that not all the tumor cells, despite sharing the same mutational background, displayed similar proliferative behavior was groundbreaking. Up to then, the main current of thought was focused on elucidating the properties that enabled the tumor cells to grow. It was known that each tumor is composed of several clones, understanding clone as a population of cells with the same mutational background. The intratumor heterogeneity was thought to be due to differential clonal growth, i.e. environmental and stochastic events that led to the evolution of some of them and the extinction of others, but the tumorigenic potential of cells with identical phenotype could not be predicted (Hanahan and Weinberg, 2000).

The discovery of cells with heterogeneous behavior despite sharing the same genetic alterations was first made in hematopoietic tumors. By separating a CD34+CD38- population from acute myeloid leukemia (AML), researchers were able to engraft the disease in immunosuppressed mice (Lapidot et al., 1994). As in normal tissue, these tumor-initiating cells were giving rise to all the cell types within the tumor, thus proving that a certain degree of hierarchy was maintained in AML cancers (Bonnet and Dick, 1997). Soon, similar findings related to other cancer cell types followed: CD44+CD24- cells were the tumor initiating cells of breast cancer (Al-Hajj et al., 2003), CD133+ cells corresponded to brain tumor propagating cells (Singh et al., 2004) and CD44, CD117 markers separated ovarian putative CSC (Zhang et al., 2008) (Figure 1.6A).

The CSC field, as well as homeostatic stem cell research, arose in the first place from a hematopoietic-CSC based point of view, where CSCs were thought to be rare and quiescent, and little plasticity was ascribed to tumor bulk cells. Slowly, the view has evolved to a more dynamic concept partly due to techniques such as lineage tracing (Batlle and Clevers, 2017).

### 1.2.1 *Properties of CSCs*

#### *Plasticity*

As explained in **section 1.1.3**, injury in homeostatic tissues has been clearly associated with a reversion of committed progenitors -and even differentiated cells- to a stem cell-like phenotype, as is the case for the trachea (Tata et al., 2013) and kidney tubules, among others (Kusaba et al., 2014). This backs up the idea that plasticity might be more common in tumors than expected, as it occurs in normal tissues. Indeed, a

nice work recently demonstrated that isolation and culture of different breast cancer phenotypes led to a re-equilibrium of the original tumor composition regardless of the phenotype of origin (Gupta et al., 2011) (Figure 1.6B).

#### *Epithelial to mesenchymal transition*

Epithelial to mesenchymal transition (EMT) is the acquisition of a mesenchymal gene program necessary for migration and invasion (Yang and Weinberg, 2008). The major mediators of EMT are the transcription factors *SNAI1*, *SLUG*, *ZEB* and *TWIST-1*, as well as miRNA and epigenetic and post-translational regulators (Nieto et al., 2016). EMT has been linked to the CSC phenotype in many reports (Mani et al., 2008, Morel et al., 2008), but it is important to bear in mind that metastatic outgrowth requires a reversion to an epithelial state (Celià-Terrassa et al., 2012). Moreover, some reports suggest there is no need of EMT for tumor cells to migrate and invade (Fischer et al., 2015, Zheng et al., 2015). Taken together, it is likely that EMT is uncoupled from the stemness program, and that a plastic phenotype comprising intermediate reversible mesenchymal states present in the tumor borders allows cell dissemination (Brabletz et al., 2005) (Figure 1.6C).

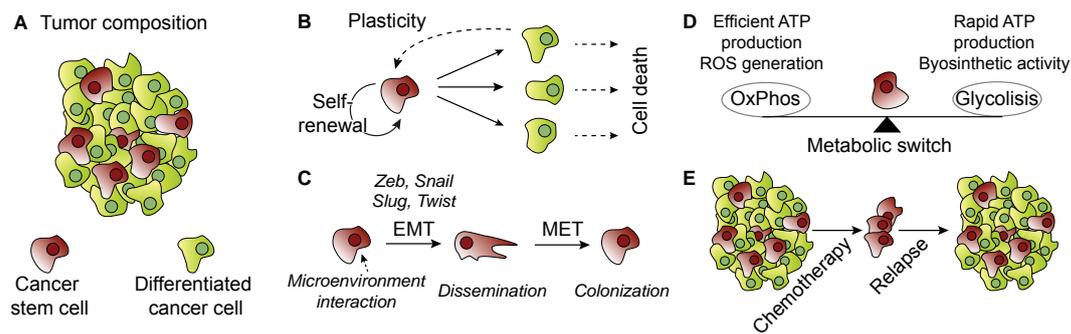
#### *Metabolism*

Tissues rely on mitochondrial oxidative phosphorylation (OxPhos) for energy obtainment. The major drawback of the process is the generation of reactive oxygen species (ROS), which can cause stem cell dysfunction, as is the case for HSCs (Suda et al., 2011). On the contrary, other adult stem cells, such as LGR5+ cells in the intestine, heavily rely on OxPhos to maintain their proliferative capacity (Rodriguez-Colman et al., 2017) (Figure 1.6D).

A hallmark of tumor development is an altered metabolism, yet not all the cells in the tumors have the same metabolic pattern (Hanahan and Weinberg, 2011). It has been shown that CSC of different tissues rely either on OxPhos or glycolysis for their growth (Sancho et al., 2016), but they are able to switch between metabolic states in response to environmental cues, such as nutrient restriction (Flavahan et al., 2013).

#### *Resistance to treatments*

Quiescent cancer stem cells are, by definition, resistant to radiation and chemotherapy (Borst, 2012). Their slow-cycling rates make them a protected target from first line chemotherapies, which are mostly directed to block proliferation (i.e 5-fluorouracil, irinotecan and oxaliplatin). In addition, they present an increase in pumps for drug efflux (ATP binding cassette transporters), higher DNA repair activity and high levels of antiapoptotic proteins (Holoohan et al., 2013). It is therefore increasingly evident that quiescent CSCs are the mediators of disease relapse after successful treatment. Conversely, in cancers with highly proliferative CSCs, such as intestinal tumors, differentiated non-cycling cells could be the chemotherapy resistant population, replenishing the CSC pool after treatment (Figure 1.6E).



**Figure 1.6: Cancer stem cell features.** **A.** Tumors are heterogeneous entities composed of differentiated-like cells and stem-like cells, termed CSCs, to which the following features are attributed. **B.** Plasticity: CSCs self-renew and give rise to a differentiated progeny which in turn may have the ability to revert and regain CSC traits. **C.** CSCs are suspected to harbor Epithelial to Mesenchymal Transition (EMT) capability, which is required for cell migration. Upon new organ colonization, cells must revert to an epithelial phenotype (Mesenchymal to Epithelial Transition, MET). **D.** CSCs can either obtain energy via glycolysis, which rapidly generates ATP to sustain cell growth and maintains a reduced cell state, or by OxPhos, which produces ATP more efficiently. **E.** CSCs harbor chemotherapy resistant traits and could be the mediators of disease relapse.

### 1.3 THE MAMMALIAN INTESTINE

The intestine is a tubular structure that conforms most of the gastrointestinal tract, achieving a length of 3 to 6 meters in adult humans. It is essential as most of the chemical digestion and nutrient and water absorption takes place in it. In addition, it represents a very interesting model to study ASC identity due to its peculiar organization, where stem and differentiated cells are located in well defined compartments.

#### 1.3.1 Intestinal structure and function

The intestine is divided in two anatomical parts, small (duodenum, jejunum and ileum) and large (caecum, colon and rectum) (Figure 1.7A). They share the same layers, concentrically organized around the lumen: the first layer is a monostratified epithelium with invaginations called crypts of Lieberkühn (from here on, referred as crypts). The second layer, the lamina propia, consists of connective tissue and stromal cells, enveloped by a submucosa layer where most of the blood vessels are found. Finally, the outermost layer is made of smooth muscle which controls the peristaltic movements. This muscular layer is covered by the serosa, a protective sheet of squamous epithelial cells (Rao and Wang, 2010).

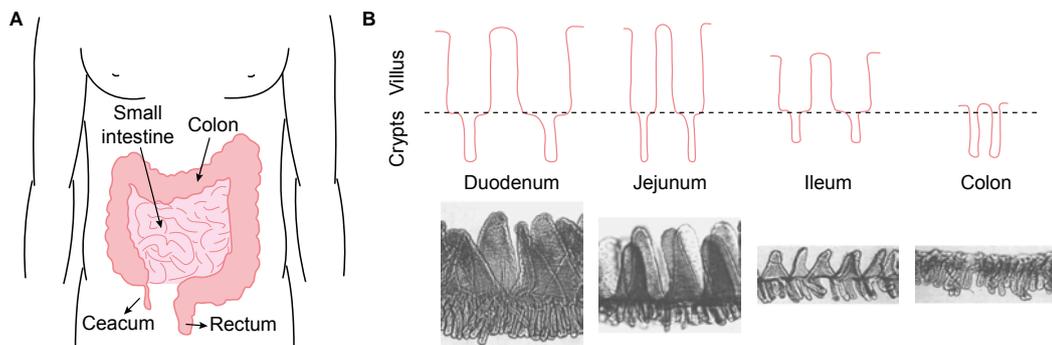
##### *Small intestine*

The small intestine functions to finish the chemical digestion of lipids, proteins and sugars and enhance the mechanical disaggregation of the stomach content, the chyme, through peristaltic movements. It receives digestive enzymes via the pancreatic main duct and the bile duct (coming from the liver), which empty their contents in the duodenum. When the food breaks down into its essential components (aminoacids, fatty

acids, glycerol and monosaccharides), they are absorbed across all the epithelium. The absorptive surface of the small intestine has evolved specifically for this purpose: the intestinal epithelium is folded in hundreds of invaginations, generating crypt-villus complexes that are both the structural and functional units of the gut. The villus are finger-like and protrude into the lumen, increasing the absorptive surface by 100-fold (Figure 1.7B).

### *Large intestine*

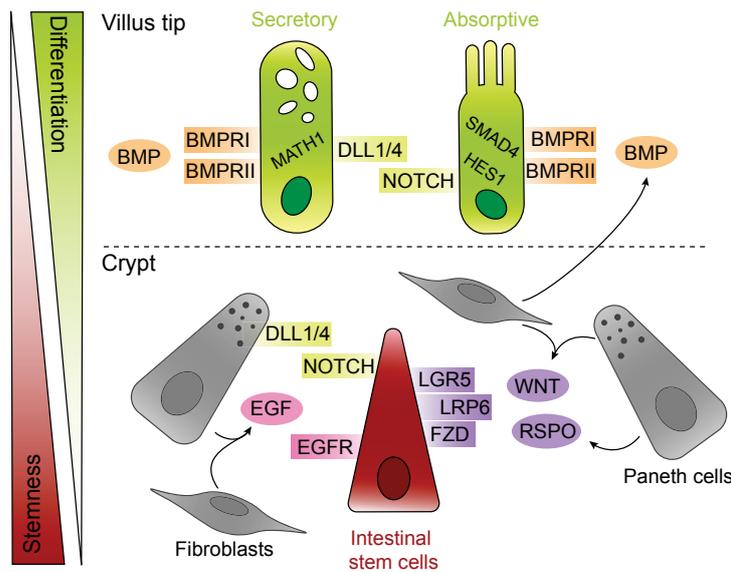
The large intestine is mostly dedicated to completing nutrient absorption, especially of water, and creating the stool. The caecum is a very short fragment that unites the ileum with the colon, the major unit of the large intestine and where most of the absorption takes place. In the rectum the stool is finally compacted for expulsion. Interestingly, the histology of the colon epithelium is similar to that the small intestine but without the villus structures, just columnar invaginations covered by a monolayer of absorptive cells (Figure 1.7B).



**Figure 1.7: Intestinal epithelium organization.** **A.** Representation of the small and large intestine. The small intestine is divided in duodenum, jejunum and ileum and the large intestine is composed of the colon, in its major part, the caecum and the rectum. **B.** Representation of the crypt-villus units of the small and large intestine and representative images of the structures in each intestinal segment. Images extracted from Bjercknes and Cheng, 1981d.

### 1.3.2 *Signaling pathways involved in ISC renewal and differentiation*

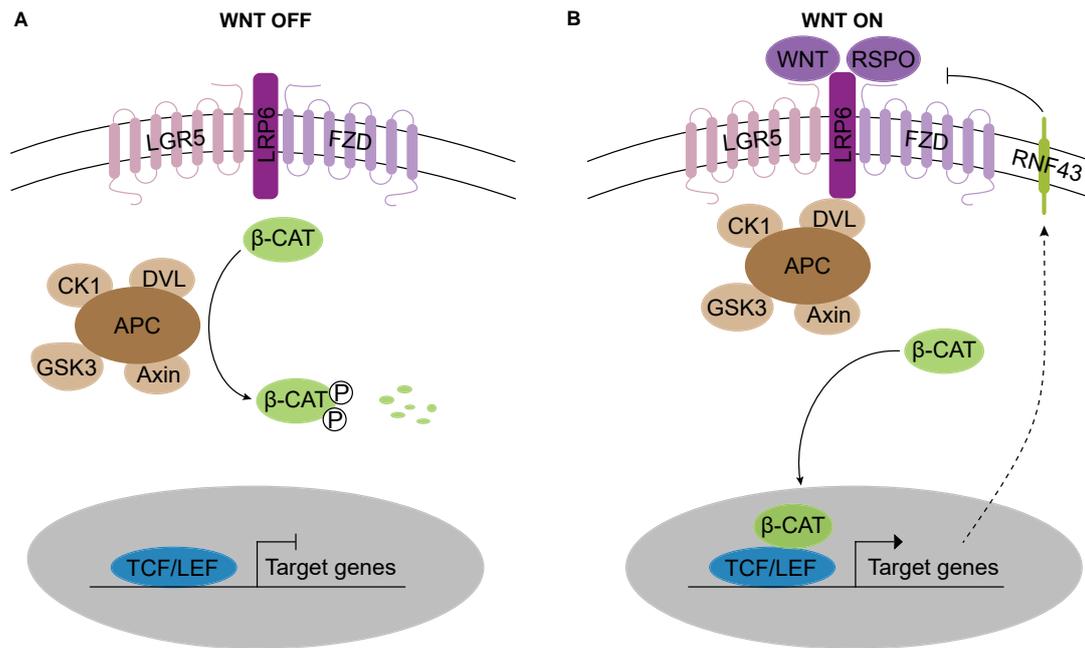
The crypt-villus units represent a structural scaffold for the cell hierarchy in the intestine. At the base of the crypt we find the ISCs, supported by Paneth cells (terminally differentiated). As the daughter cells of ISCs move up the crypt, cells differentiate into the secretory and absorptive lineages, and transit to the top of the villus where they eventually die and are shed into the lumen (Clevers, 2013). The intestinal epithelium, thus, relies on a heavy replacement system mediated by ISC proliferation and differentiation. In order to maintain homeostasis, the system must be finely tuned. Four major signaling pathways control this process: WNT, EGF, NOTCH and BMP (Figure 1.8).



**Figure 1.8: Intestinal signaling pathways that regulate stemness and differentiation potential.** The crypt stem cell niche provides pro-proliferative and stem signals via EGF, Notch activators and WNT signaling. In the upper parts of villi, BMPs induce differentiation. Lineage decision between secretory of absorptive phenotype is mediated by Notch signaling.

### WNT signaling

WNT is a major player in promoting ISC proliferation, its blockade resulting in crypt loss and differentiation (Pinto et al., 2003).  $\beta$ -catenin is the principal mediator of WNT signaling; in basal conditions, the APC complex (named after the Adenomatous Polyposis Syndrome) phosphorylates  $\beta$ -catenin inducing its ubiquitination and subsequent degradation by proteasome. Upon WNT addition to the media, Frizzled+LRP5/6 receptor complexes block the APC kinase activity via Disheveled (DVL) recruitment to the membrane, resulting in  $\beta$ -catenin accumulation and nuclear translocation. In the nucleus,  $\beta$ -catenin turns the TCF4/LEF transcriptional repressors into activators, inducing a WNT target gene program (Cadigan, 2008) (Figure 1.9). Some of the most important WNT target genes in the intestine are, for example, ASCL2, necessary for ISC specification, and the proliferation driver MYC (Sansom et al., 2007). Moreover, genes specific of the intestinal stem cell pool, such as EPHB2 and LGR5, are also responsive to WNT signals (Batlle et al., 2002, Barker et al., 2007). A comprehensive summary of WNT target genes is kindly provided by the Nusse lab ([https://web.stanford.edu/group/nusselab/cgi-bin/wnt/target\\_genes](https://web.stanford.edu/group/nusselab/cgi-bin/wnt/target_genes)). The importance of WNT pathway for intestinal homeostasis becomes evident during tumorigenesis: mice carrying a single mutated allele of APC develop adenomas at an early age (Moser et al., 1990).



**Figure 1.9: WNT signaling pathway components.** **A.** In the absence of WNT ligands, the APC complex mediates β-catenin degradation. TCF/LEF nuclear transcription factors act as gene suppressors, inhibiting the expression of WNT target genes. **B.** WNT activation of LRP6 and its co-receptors Frizzled (FZD) and LGR5 recruits the APC complex to the membrane, and β-catenin translocates into the nucleus, activating the expression of WNT target genes via TCF/LEF. In a negative feedback loop, RNF43 is produced upon TCF/LEF activation and blocks LRP6 activation.

### *Epidermal growth factor*

EGF is one of the best described mitogens. They exist as pro-proteins that are activated/secreted upon cleavage, inducing dimerization of their receptors (EGFRs), which in turn transactivate their kinase domains and transduce pro-survival signals to the nucleus (Yarden and Shilo, 2007). ISCs express high levels of EGFRs, and the ligand is provided by Paneth cells (Sato et al., 2011b)(Figure 1.8). The importance of EGF pathway in ISC homeostasis is highlighted by *in vitro* studies: EGF addition to the media is essential to maintain intestine-derived organoids in stem cell conditions (Sato et al, 2009, Jung et al, 2011). Moreover, most colorectal tumors harbor activation mutations downstream the EGF pathway, either *KRAS* or *BRAF* (Vogelstein et al., 1988, Kedrin and Gala, 2015) that render tumor cells independent of niche signals.

### *Notch signaling*

Notch pathway is well known for controlling cell fate decisions in many organisms. In the intestine, Notch signaling is active in the stem cell zone and in the transient amplifying (TA) compartment, where it determines the fate between absorptive and secretory lineages. Notch downstream effector, *HES1*, specifies the cells towards an enterocytic fate, whereas the transcription factor *MATH-1* induces secretory differentiation. *MATH-1* also increases *DLL1/4* NOTCH ligand abundance in secretory-specified cells, inducing then an absorptive phenotype on the lateral cells (Fre et al., 2005, Tian et al., 2015) (Figure 1.8).

### *BMP pathway*

Bone morphogenic proteins (BMP) are part of the TGF- $\beta$  superfamily. They act via dimerization of their receptors (BMPRI and 2) in a similar fashion as EGF. Active BMPRI phosphorylate SMAD1, 5 and 8 that in turn interact with SMAD4. SMAD4 mediates a transcriptional response that suppresses the proliferation and stemness in the villus (Sancho et al., 2004). They are essential for homeostatic maintenance, as blockade of the pathway results in the outgrowth of ectopic stem cell niches (Haramis et al., 2004) (Figure 1.8).

### 1.3.3 *Differentiated intestinal cell types*

There are six well-described differentiated epithelial cell subtypes (Figure 1.10)

#### *Absorptive lineage*

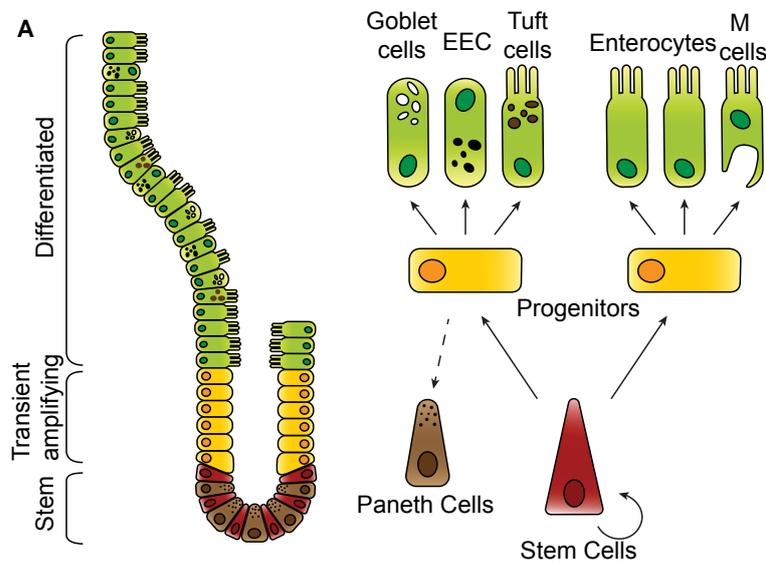
- Enterocytes: enterocytes, also called absorptive cells, conform the most abundant lineage of the intestinal epithelium (80% of total cells). They are highly polarized, with a typical columnar shape, and a brush border of microvilli that facilitates nutrient uptake. They require of HES1 signaling and ELF3 activation of the TGF- $\beta$  pathway to differentiate (Jensen et al., 2000 and Flentjar et al., 2007).
- M cells: microfold (M) cells are a specialized type of cell found in Peyer's patches (lymphoid tissue of the intestine) and have the essential function of translocating antigens from the lumen towards immunitary cells (de Lau et al., 2012). Their specification requires the SPIB transcription factor, a downstream effector of RANK signaling, actively induced by stromal cells of the Peyer patches (Clevers and Batlle, 2013).

#### *Secretory lineage*

- Goblet cells: goblet cells secrete mucins (such as MUC2) and trefoil proteins, to ease movement of chyme and stool through the intestine. Accordingly, their numbers increase from the duodenum (4%) to the colon (16%). Their formation requires acute NOTCH blockade and SPDEF transcription factor activation (Crosnier et al., 2005, Gregorieff et al., 2009).
- Enteroendocrine cells: the hormone-producing cells of the intestine occupy up to 1% of the epithelial surface, and can be subclassified according to the type of hormones they produce (Schonhoff et al., 2004). The differentiation of this lineage requires NOTCH inactivation and Neurogenin3 signaling (Jenny et al., 2002).
- Paneth cells: Paneth cells are the only epithelial differentiated subtype that migrates downwards, occupying their position within the stem cell niche. They are filled with secretory granules containing lysozymes, antimicrobial peptides and defensins, creating thus an innate immunity defense against the gastrointestinal flora. As mentioned above, they also function as the ISC niche, providing

essential EGF, WNT and NOTCH signals (Sato et al, 2011b). Interestingly, they are longer-lived than the rest of terminally differentiated cells, with a half-life of around four weeks (Bjerknes and Cheng, 1981b). They need active WNT signaling to differentiate (Farin et al., 2012). Paneth cells are not present in the colon, although CD24<sup>+</sup> cells with analogous function have been identified (Sato et al., 2011b, Sasaki et al., 2016).

- Tuft cells: tuft cells are a minority of the intestinal compartment (0.4%). They have a characteristic brush border and are identified by Doublecortin and CaM kinase-like-1 expression (DCLK1) (Gerbe et al., 2011). As has been recently described (Gerbe et al., 2016), they initiate type 2 immunity responses during parasitic infections.



**Figure 1.10: Intestinal hierarchy.** Intestinal stem cells reside at the base of the crypt, supported by niche-providing Paneth cells. Cells that leave the crypt enter the transient amplifying region, where progenitor cells specify their fate. The upper part of the villus is composed of terminally differentiated cells of the secretory (goblet, EEC and tuft cells) and absorptive lineages (enterocytes and M cells).

#### 1.3.4 Intestinal stem cells

##### *Identification of the intestinal stem cells*

Crypt-villus units are self-renewing structures with hierarchical organization. The extreme proliferative activity of the intestinal epithelium has been a matter of intense research over the last century, with the aim to find the true intestinal stem cell. The first reports of an actively cycling population residing near the crypt bottom date from the 1950's (Leblond and Stevens, 1948). With a very simple mitosis-blocking experiment with colchicine, they demonstrated that around 3% of the total intestinal epithelium was undergoing mitosis at a given time, but the process was strictly restricted to the crypts. Moreover, they postulated the model of balanced cell proliferation – cell loss as a conveyor belt that guided the differentiating cells from the

crypt base to the top of the villi where they were shed to the lumen. Since then, two currents of thought have developed about the nature of the ISC population able to maintain such proliferative status:

#### *Crypt base columnar cells*

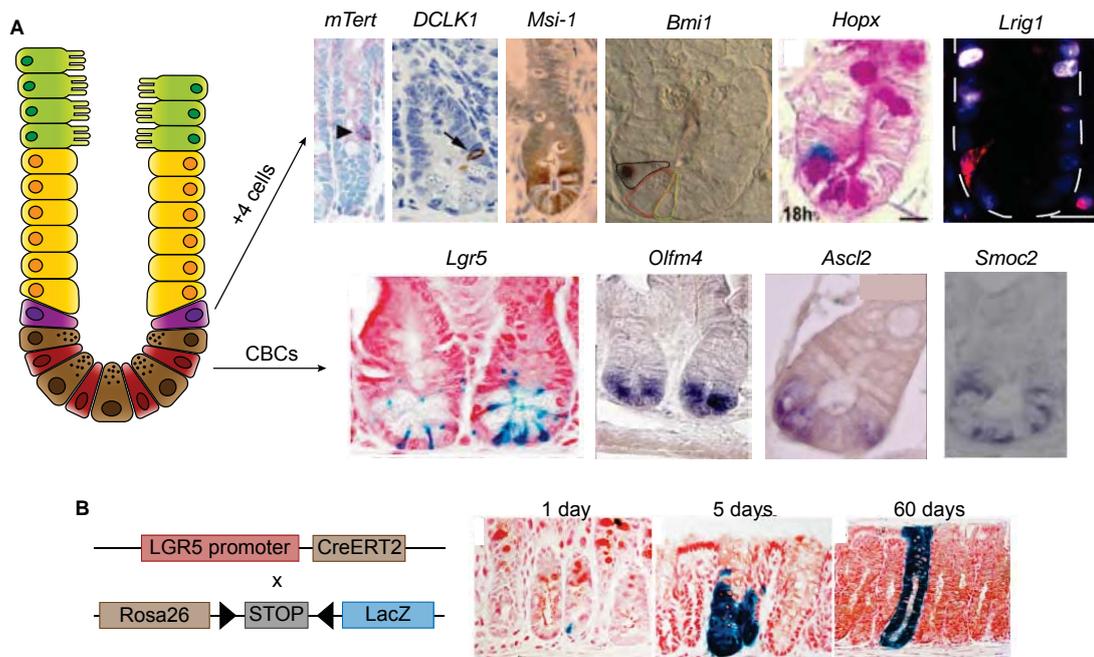
On one hand, Leblond was again the first one to report the presence of small slender cells intermingled in between the Paneth cells that could act as ISC, proposed after the first lineage tracing experiments with 3H-Thymidine and mutation accumulation (Cheng and Leblond, 1974, Bjerknes and Cheng, 1981c). Due to their shape, they named these cells crypt base columnar cells (CBC). The Clevers group pursued this hypothesis, starting with the notion that WNT signals are essential for intestinal homeostasis (Korinek et al., 1998). By generating an inducible dominant negative TCF4 construct, they were able to generate a WNT responsive gene program of the intestine (van de Wetering et al., 2002). Within this signature, they observed that the Ephrin receptors EPHB2 and EPHB3 were downregulated upon TCF4/ $\beta$ -catenin inhibition, whereas the expression of the ligand ephrin-B1 (EFNB1) was increased. Indeed, EPHB2 was expressed in a decreasing gradient from the crypt bottom to the top, inversely correlating with EFNB1 levels (Batlle et al., 2002, Cortina et al., 2007). Since then EPHB2 has been widely used as a marker of stemness both in normal and tumor settings (Merlos-Suárez et al., 2011).

Another top gene in the intestinal WNT signature list was *LGR5*, a WNT co-receptor that perfectly matched the CBC expression domain visualized in *Lgr5*-GFP-CreERT2/R26R-LacZ transgenic mouse (Barker et al., 2007) (Figure 1.11A). By GFP labelling, they counted an average of 15 cells per crypt that divided every day, and, taking advantage of the Cre system, they were able to trace the progeny of *Lgr5*-derived cells. In solely 5 days, blue ribbons occupying the whole villus length and containing all epithelial lineages appeared, validating the notion that CBCs were indeed bonafide ISCs (Figure 1.11B). Further studies comparing the *Lgr5*-GFP-High vs the *Lgr5*-GFP-Low populations allowed the identification of *Olfm4* and *Ascl2* as other CBC markers (van der Flier et al., 2009a, van der Flier et al., 2009b) (Figure 1.11A). Moreover, by generating a knock-in analogous to that of *Lgr5*, a later study identified SPARC related modular calcium binding 2 (*Smoc2*) as a stem cell gene (Muñoz et al., 2012) (Figure 1.11A).

#### *+4 reserve stem cell pool*

On the other hand, Potten and colleagues found out a group of LRC cells located at the +4 position, immediately above the Paneth cells, rather than at the crypt base (Potten et al., 1974, Potten et al., 2002). +4 cells were proposed as the putative ISC, and many devoted their efforts to finding specific genes to visualize these slow-cycling cells. *mTert*, necessary to maintain the telomere length, was signaled as a marker for this LRC cells (Breault et al., 2008 and Montgomery et al., 2011), as well as *Dclk1* (May et al., 2008), *Mushashi-1* (*Msi1*) (Potten et al., 2003 and Kayahara et al., 2003) and *Bmi1* (Sangiorgi and Capecchi, 2008) (Figure 1.11A). Most of the studies were based on the localization of these cells within the crypt, all in the +4 position, although DCLK1+ cells were shown to give rise to spheroids *in vitro* that, when injected in immunodeficient

cient mice, formed small nodules resembling poorly formed crypts (May et al., 2008). Moreover, the Capecchi group generated a *Bmi1*-CreER mouse combined with an inducible LacZ reporter. By lineage tracing, they demonstrated that *Bmi1* progeny was able to colonize the villus, with 10% of them fully LacZ+ one month post-induction (Sangiorgi and Cappecchi, 2008). Later on, two more genes, *Hopx* and *Lrig1*, were also proposed as specific markers of a quiescent ISC pool able to repopulate the whole villus (Takeda et al., 2011 and Powell et al., 2012) (Figure 1.11A).



**Figure 1.11: Specific markers for the +4 position and crypt base columnar cells. A.** Schematic representation of the CBCs (red) and +4 cells (magenta) position within the crypt, and representative stainings of markers for each population respectively in mouse small intestines. Images are extracted from Breault et al., 2008 (*mTert*), May et al., 2008 (*Dclk1*), Potten et al., 2003 (*Msi-1*), Sangiorgi et al., 2008 (*Bmi1*), Takeda et al., 2011 (*Hopx*), Powell et al., 2012 (*Lrig1*), Barker et al., 2007 (*Lgr5*), Van der Flier et al., 2009 (*Ascl2* and *Olfm4*) and Muñoz et al., 2012 (*Smoc2*). **B.** Schematic representation of the first *Lgr5* lineage tracing system and LacZ staining in mouse small intestine 1, 5 and 60 days post induction of the tracing. Modified from Barker et al., 2007.

Posterior studies finally solved the dichotomy of the ISC pool; was the slow-cycling, quiescent +4 population the real ISC, or were the actively CBC LGR5+ cells the ones indispensable for intestinal homeostasis? First, several publications reported that the specific +4 markers *mTert*, *Bmi1*, *Dclk-1*, *Hopx* and *Lrig* were actually co-expressed with *Lgr5* (Itzkovitz et al., 2011, Muñoz et al., 2012). Moreover, by using a novel dimerization-based Cre system, the Winton group were able to lineage trace Histone-2B retaining cells, a widely used marker to distinguish slow-cycling cells. They demonstrated that LRC cells of the intestinal crypt base were lineage-restricted daughters ready to differentiate into enteroendocrine and Paneth cells (Buczacki et al., 2013). Coincident with these results, the Clevers group separated mouse LGR5+ proliferating and quiescent cells by using a Ki67-RFP reporter as a

proliferation marker. They showed that the majority of LGR5-High cells were KI67+, but the LGR5-Low KI67- pool resembled the H2B-slow cycling secretory precursors. Of note, the specific +4 genes were equally expressed amongst the four LGR5/KI67 populations, discarding them as bonafide stem cell markers (Basak et al., 2014). Moreover, Grün *et al* (Grün et al., 2015) showed by single cell profiling that LGR5+ cells constitute a rather homogeneous population.

Taking all these data together, we envision the working model of the intestinal hierarchy as follows: the ISCs are the pool of CBCs intermingled in between Paneth cells in the crypts, and the best marker so far for its identification is *LGR5* gene expression. They are actively proliferating, and, as their daughters move up the crypt-villus axis, they differentiate into the secretory and absorptive progenitors in the transient amplifying compartment, that in turn will acquire the terminally differentiated characteristics of the villus cells. Of note, Paneth cells are the only ones able to migrate downwards upon differentiation, and stay in the crypt providing support to the SC niche. Additionally, our laboratory has identified a subset of LGR5+ cells with slow cycling dynamics. These cells, characterized by *Mex3A* gene expression, retain CBC features and continuously contribute to ISC population (Barriga et al., 2017).

#### *Characterization of the ISC*

Since then, many efforts have been devoted to understanding the properties and division dynamics of the LGR5+ cells in the intestine.

#### Single LGR5 cells reproduce intestinal heterogeneity

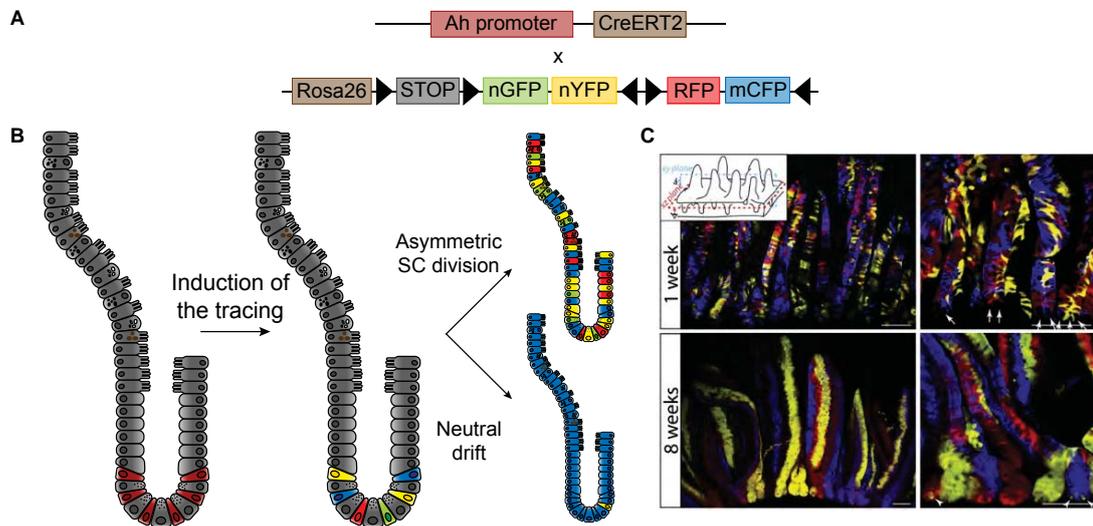
By taking advantage of the organoid technology, which allows the maintenance of *in vitro* 3D cultures that recapitulate tissue heterogeneity (reviewed in Lancaster and Knoblich, 2014), the Clevers group interrogated whether LGR5+ and - cells retained SC capacities. They isolated LGR5-GFP-High and -Low cells from mice intestines and plated them in single cell conditions. 6% of the LGR5-High cells were able to grow 3D organoids that reproduced intestinal crypts, whereas seldom LGR5-Low cells were able to survive (Sato et al., 2009). Later improvement of the system enhanced the single cell survival rate to 60% for the LGR5+ derived cells (Sato et al., 2011b). Moreover, engraftment of *in vitro* grown LGR5-GFP cells isolated from the colonic glands was able to repopulate damaged colonic epithelia giving rise to functional tissue (Yui et al., 2012).

#### ISC follow a pattern of neutral drift dynamics

A key question in ISC biology was whether ISCs self-renewed by asymmetric cell division. Snippert *et al* (Snippert et al., 2010) elegantly solved this issue. Using a multicolor lineage tracing cassette and a general intestinal inducible Cre under the control of the cytochrome P450-promoter-driven Ah-Cre, expressed through the entire intestine, they showed how crypts drifted towards monoclonality, arising each from a single original ISC. From initial mosaic labelling of the whole villus, with several induced clones per crypt, in as short as two weeks homogeneous (same-color) crypts

appeared. By eight weeks, most of them were monoclonal (Figure 1.12). This data fits the stochastic division mode, in which all stem cells have the same proliferation capacity and compete for niche space. Upon division, an ISC can give rise to two cells that remain in the crypt, or one or both daughters can be pushed out and start differentiating, causing the loss of this particular clone. If asymmetric division was the prevailing mechanism, all the initial colors in the crypt would be maintained through generations of SCs. Mathematical modeling (López-García et al., 2010) of division dynamics confirmed that the ISC division rate equaled ISC replacement, postulating a neutral drift pattern of intestinal renewal.

In support of all this data, posterior studies observed that as a result of competition, the CBCs at the central part of the crypt were 3 times more likely than border cells to colonize the crypt (Ritsma et al., 2014).



**Figure 1.12: ISC self-renewal follows a pattern of neutral drift.** **A.** Schematic representation of the Confetti lineage tracing cassette, which randomly recombines upon Cre activation allowing the expression of green, yellow, red or blue fluorescent proteins. **B.** Summary of the experimental hypothesis. The Ah promoter is expressed in all intestinal cells. Upon tamoxifen injection, cells will acquire one of the four confetti colors. If the ISCs follow an asymmetric cell division mode, all the colors present in the crypt at the beginning will be maintained across time. But if they follow a stochastic symmetric division, one ISC will end up occupying the whole crypt, and the villi will thus become monoclonal. **C.** Representative image of the experimental result, showing how all the crypts drift toward monoclonality 8 weeks after the induction of the trace. Modified from Snippert et al., 2010.

### ISC plasticity

Importantly, many studies report plasticity in the stem cell compartment when homeostasis is disrupted. Indeed, in response to specific elimination of LGR5+ cells in transgenic mouse models, the LRC compartment increased their proliferation rates to replenish the damaged ISC pool (Buczacki et al., 2013). Plasticity has been shown to arise from both +4 BMI1+ cells (Tian et al., 2011, Yan et al., 2011), as well as lineage-

committed progenitors such as DLL1+ secretory precursors (van Es and Sato et al., 2012) and ALPI+ enterocytic progenitors (Tetteh et al., 2016).

## 1.4 COLORECTAL CANCER

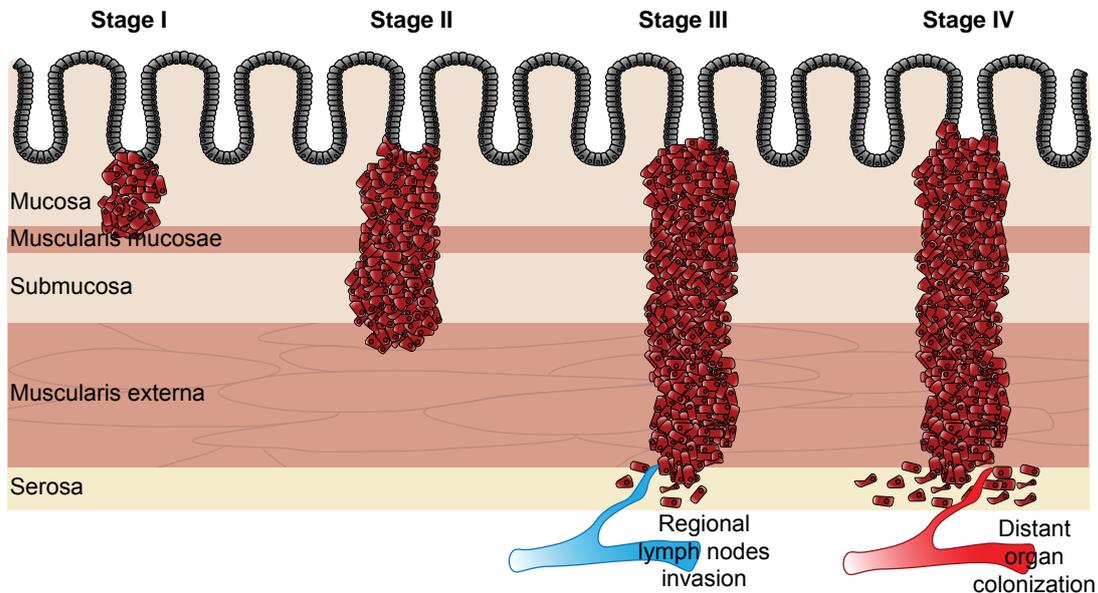
In 2018, 1.8 million new colorectal cancer cases have been diagnosed worldwide. CRC is the second leading cause of death by cancer (9.2%) (Bray et al., 2018), and its increase is tightly related to ageing and western world habits (lack of exercise, increased fat consumption and alcohol abuse) (American Institute for Cancer Research, 2018).

### 1.4.1 CRC staging

The stage at which CRC is diagnosed is critical for its treatment, as the 5-year survival rate is greatly influenced by the stage of the disease. According to the American Joint Committee on Cancer (AJCC: [www.cancerstaging.org](http://www.cancerstaging.org)), CRCs can be classified depending on the status of three elements, in what is called TNM staging:

- T: tumor invasiveness
- N: degree of lymphatic node dissemination
- M: presence of metastatic disease

By looking at these classifiers, four CRC stages have been determined. A higher number correlates with more aggressive disease and is linked to poorer outcome (Figure 1.13). A detailed description is found in Table 1.1:



**Figure 1.13: Colorectal cancer staging.** CRCs are divided in four stages of development depending on the tumor invasive properties. Stage I and II cancers do not present metastatic dissemination yet, whereas stage III is characterized for lymphatic node colonization and stage IV for distant organ metastasis.

Stage	T	N	M	Description
I	1–2	0	0	Small primaries, with mucosa or even muscular invasion
II-A	3	0	0	Tumor invades pericorectal tissues but no dissemination to lymph nodes or distant organs is observed
II-B	4a	0	0	Tumor arrives to the visceral peritoneum
II-C	4b	0	0	Tumor is adhered to other organs, but as the rest of stage II grades, lymph nodes and other organs are clean of disease
III-A	1–2	1	0	Primary tumor is non-invasive but there are already metastasis in regional lymph nodes
III-B	3–4	1	0	Primary tumor is non-invasive but there are already metastasis in regional lymph nodes
III-C	x	2	0	Same features as stage III-B but the number of invaded lymph nodes is 7 or more
IV-A/B	x	x	1–2	Stage IV disease is defined only by the metastasis to distant organs (A= one metastatic site, B=more than one site)

**Table 1.1: Colorectal cancer staging.** Classification of colorectal cancer disease according to tumor invasiveness (T), lymphatic node dissemination (N) and metastatic outgrowth (M) based on AJCC classification. x=any grade

Diagnosed stage I tumors have a 5-year survival rate of 88%, with surgical resection as the preferred treatment. Stage II and III tumors can often be cured by a first round of surgery and adjuvant chemotherapy, but around half of the patients relapse with a more aggressive metastatic disease (mainly in liver and lungs) that decreases the 5-year survival rate to 55-75% in case of stage II disease and to 25-55% for stage III. The worst prognosis is for stage IV diagnosed patients, as no effective treatment for invasive primaries and metastases has yet been developed. The survival rate is lower than 12% for these tumors, highlighting the need to further study CRC biology for the development of new therapies ([www.cancer.org](http://www.cancer.org)).

#### 1.4.2 Types of CRC

Cancer is a genetic disease, caused by the accumulation of so-called driver mutations that enhance anti-apoptotic and proliferative abilities of tumor cells. These mutations can appear randomly, causing sporadic CRC, which accounts for 70-80% of total CRC cases, but they can also be inherited, as is the case for the two best described familial colorectal cancer syndromes:

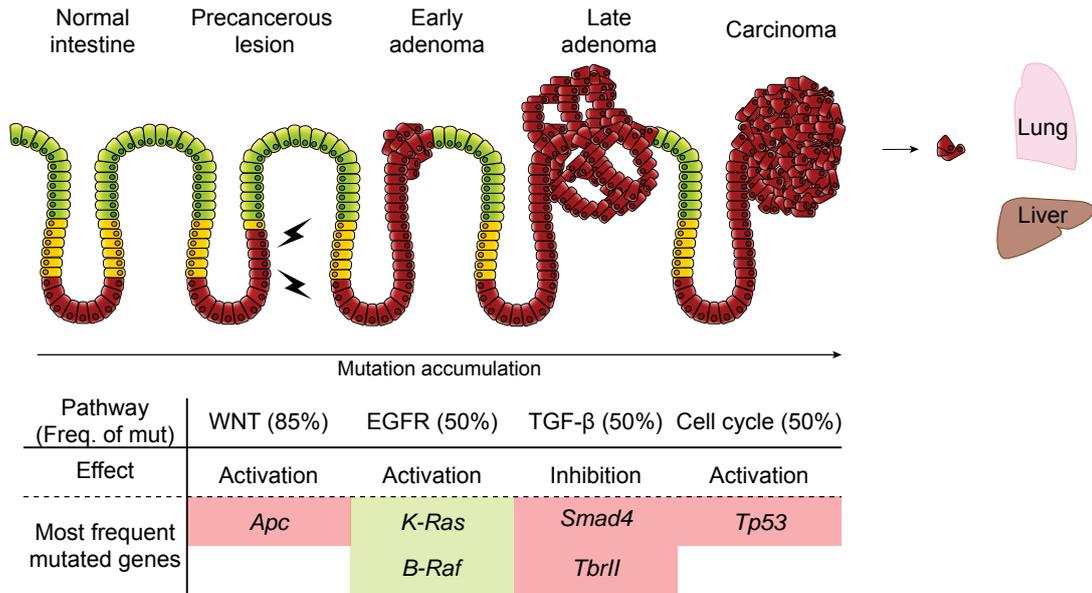
- **Familial Adenomatous Polyposis (FAP):** is an autosomal dominant syndrome caused by inactivating mutations in the tumor suppressor gene APC. Most of the mutations truncate the APC protein, thus leading to enhanced WNT activity and deregulation of the proliferative, apoptotic and migratory capacity of the cells. FAP patients accumulate adenomatous polyps in the colon that inevitably evolve into carcinomas by the age of 30-40 (Galiatsatos and Foulkes, 2006).

- Hereditary Non Polyposis Colorectal Cancer (HNPCC): is caused by mutations in any of the mismatch-repair genes, with MLH1 and MSH2 defects accounting for almost 90% of the cases. DNA repair deficiencies lead to microsatellite instability (MSI), and mutations accumulate in the genome, affecting both tumor suppressor genes and protooncogenes (Lynch and de la Chapelle, 2003). Actually, tumor suppressor and tumor driving genes with repetitive DNA tracts accumulate higher mutation rates in those patients (Duval and Hamelin, 2002). The progression from adenoma to carcinoma is faster in HNPCC than in sporadic or FAP tumors, and patients are usually diagnosed around the age of 45.

#### *CRC progression model*

The study of genetic mutations that precede the spontaneous formation of an advanced colorectal carcinoma has been greatly helped by the knowledge gathered on familiar syndromes. Already in 1990's, Vogelstein proposed an evolutionary model for CRC from benign adenomas to fully blown carcinomas mediated by the accumulation of mutations in key pathways (Fearon and Vogelstein, 1990, Kinzler and Vogelstein, 1996).

The first hit required for the transition from normal epithelia to adenoma is the alteration of the WNT pathway. Mostly mediated by loss of function mutations in the APC gene, the accumulation of nuclear  $\beta$ -catenin imposes a crypt-like proliferative program in the cells (van de Wetering et al., 2002). These aberrant cells display chromosomal instability (CIN) features, and hyper-activate the pro-tumorigenic pathways via MAPK signaling (KRAS, BRAF and PI3KCA mutations). The addition of TP53 mutations, that further enhances genomic instability, and the blockade of the tumor suppressor TGF- $\beta$  pathway, generally by loss of function mutations in the TGF- $\beta$ R2 and the SMAD2,3 or 4, leads to the development of invasive carcinomas (Fearon, 2011) (Figure 1.14). Of note, BRAF mutations are strongly associated with a particular subtype of colon tumors, the sessile serrated adenomas (SSA). They are characterized by a CpG island methylator phenotype (CIMP), which imposes a MLH1 hyper-methylation status on those cells and induces tumorigenesis via MSI mechanisms (Weisenberger et al., 2006, Rad et al., 2013). Importantly, these shared genomic alteration patterns in CRC have been studied in depth through The Cancer Genome Atlas Network (TCGA, 2012).



**Figure 1.14: Genetic progression of colorectal cancer.** Representation of the main driving mutations in ISCs that lead to adenoma and carcinoma formation. The table depicts the most common order in which signaling pathways acquire activating or inhibiting mutations and the most frequently mutated genes in each pathway. Adapted from E.Battle Lab.

#### Modelling CRC progression in mice

Generation of genetic engineered mouse models (GEMMs) have been instrumental to demonstrate Vogelstein's and Fearon's initial theories on the progression of CRC through the acquisition of driver mutations in specific pathways. The first CRC GEMM was found randomly during a genetic screening with N-ethyl-N-nitrosurea (ENU) (Moser et al., 1995). Some mice carried a heterozygous mutation in codon 850 of the *Apc* gene, leading to a truncated protein. Due to the high rate of polyps developed in these mice, they were named *Apc*<sup>Min</sup> (multiple intestinal neoplasia). Following the generation of *Apc*<sup>Min</sup> mice, many other mice strains carrying heterozygous loss of function mutations in the *Apc* gene have been generated. In all of them, polyp formation occurs upon loss of heterozygosity (LOH). Interestingly, different mutations imply a different polyp burden; *Apc*<sup>Min</sup> mice harbor around 30 polyps, *Apc*<sup>Δ716</sup> carry up to 300 intestinal polyps and *Apc*<sup>1638N</sup> only 3-5 (Fodde et al., 1994, Smits et al., 1997).

Complementary to *Apc* mutations, GEMM carrying stabilizing mutations in the β-catenin protein have also been generated (Harada et al., 1999, Romagnolo et al., 1999). The major drawback of all these modified animals is that they only generate benign adenomas that do not progress to advanced disease stages, and even less to metastasis in distant organs. Also, the majority of polyps are found in the small intestine, not in the colon. Combinatorial mutations have somewhat improved the situation, for example *Cdx2* deficiencies induce the polyp formation in the colon (Aoki et al., 2003) and *Smad4* deletions induce aggressive adenocarcinomas (Takaku et al., 1998). Other strategies like inducible *Apc* deletion in *Apc*<sup>15lox/15lox</sup> *Fabp1*<sup>Cre</sup> animals facilitates colon polyp formation via intrarectal tamoxifen injection (Robanus-Mandaag et

al., 2010). Some more metastatic lines were also developed by combining mutations, such as the  $Apc^{CKO/CKO} Kras^{LSL-G12D}$  with a metastatic penetrance of 20% (Hung et al., 2010).

These models prove that the sequential acquisition of mutations is key for sporadic CRC development. In addition to GEMM, colon tumors can also be initiated by inflammation induced by an azoxymethane/dextran sodium sulphate regime (AOM/DSS) (De Robertis et al., 2011).

#### *Modelling CRC progression by organoid culture*

Definitive proof of the mutational landscape required for CRC development comes from the modification of healthy intestinal organoids *in vitro*. The combination of 3D matrixes and a refined cocktail of stem-promoting factors has enabled the derivation primary cultures derived from small of intestine and colon, termed organoids, that *in vitro* reproduce the structure of the gut as if they were “miniguts” (Ootani et al., 2009, Sato et al., 2009, Jung et al., 2009). Sato’s group added, by sequential CRISPR/Cas9 modifications, activating mutations in the oncogenes *KRAS* and *PI3KCA* and loss of function deletions in the tumor suppressor genes *APC*, *SMAD4* and *TP53* to the intestinal “miniguts”. Each additional mutation conferred increased independency from cell culture factors (Table 1.2), and this robustness was translated to *in vivo* growth, complying with Vogelstein’s theory of CRC development. The organoids harboring the five driver mutations, nevertheless, did not present any of other CRC alterations, such as CIN, copy number variations or epigenetic modifications (Matano et al., 2015). In contrast, the Clevers group reported that *APC* and *TP53* mutations are sufficient to induce aneuploidy in human modified organoids (Drost et al., 2015). Additionally, combination of CRISPR/Cas9 mutations in human organoids has enabled the definition of a mutational signature induced by mutations of the MMR pathway, specifically in the *MLH1* and *NTHL1* genes (Drost et al., 2017). Apart from the engineered intestinal organoids, organoids from tumoral tissue have also been developed (Sato et al., 2011b, Mérlos-Suárez et al., 2011, Calon et al., 2012), and the creation of biobanks from patient samples (PDO: patient derived organoids) that faithfully recapitulates the disease upon injection in immunosuppressed animals has supposed a major revolution in the CSC field (van de Wetering et al., 2015). The lack of metastasis upon orthotopic injection in these models suggest that the driver mutations solely are not enough to reproduce the full transition from homeostatic to tumorigenic epithelia (Matano et al., 2015).

	-WNT3A - Rspo	-EGF	-Noggin	+TGF- $\beta$	+nutlin-3	+MEK in- hibitors
WT	Red	Red	Red	Red	Red	Red
A	Green	Red	Red	Red	Red	Red
AK	Green	Green	Red	Red	Red	Red
AKST	Green	Green	Green	Green	Green	Red
AKSTP	Green	Green	Green	Green	Green	Green

**Table 1.2: Mutation accumulation in WNT, EGF, TGF- $\beta$ , TP53 and PI3K pathways confers niche-independent cell growth.** Cell culture media requirements of modified intestinal organoids (green: organoids grow, red: organoids are not able to grow). Mutation accumulation in the different pathways allows cell growth independent of niche factors (i.e without media supplementation) and bypasses antiproliferative drug effects, which reproduces CRC development *in vitro*. Organoid genotype: WT (non mutated), A (APC<sup>-/-</sup>), K (KRAS<sup>G12D</sup>), S (SMAD4<sup>-/-</sup>), T (TP53<sup>-/-</sup>), P (PI3KCA<sup>E545K</sup>). Media elements effect: WNT3A/Rspondin (enhance WNT signaling), EGF (enhances EGF signaling), Noggin (inhibits TGF- $\beta$  pathway), TGF- $\beta$  (activates TGF- $\beta$  pathway), Nutlin-3 (inhibits TP53/MDM2 interaction) and MEK inhibitors (block the MAPK pathway).

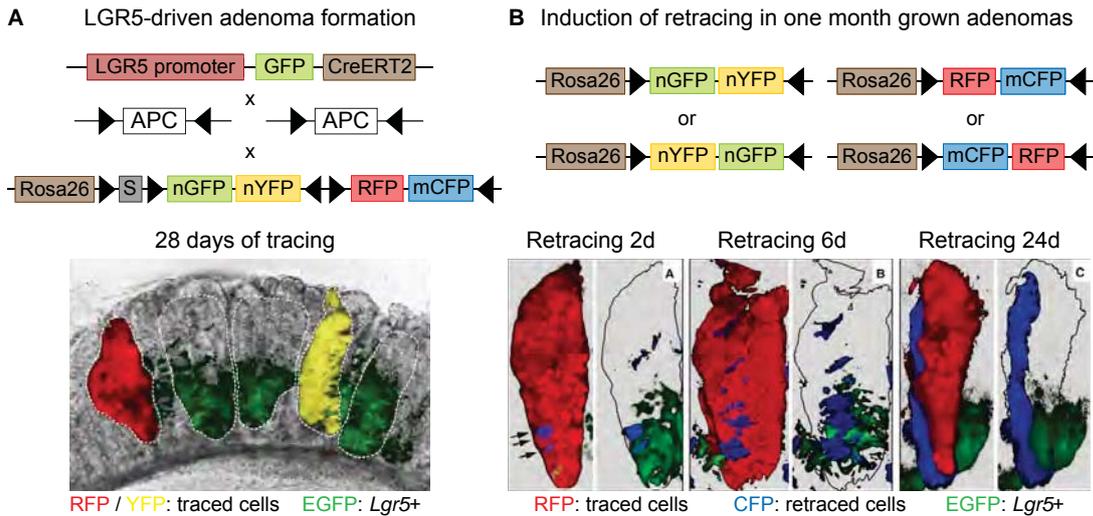
#### 1.4.3 Colon cancer stem cells

The first reports that CRC initiation required the acquisition of a crypt-like phenotype date from 2002 (van de Wetering et al., 2002). They observed that adenoma cells expressed a genetic program similar to that of crypt stem cells. Indeed, later studies confirmed that WNT program upregulation was necessary for adenoma development, as MYC ablation was able to rescue the pro-tumorigenic effect of APC loss in mice (Sansom et al., 2007).

Before the establishment of *Lgr5* as the golden ISC marker (Barker et al., 2007), some researchers already reported the presence of Tumor Initiating Cells (TICs) in colon cancers. By using CD133 (O'Brien et al., 2007, Ricci-Vitani et al., 2007) or CD44 (Dalerba et al., 2007), they were able to separate a population that retained higher tumorigenic potential upon transplantation into mice. Moreover, CD133+ or CD44+ cells were able to differentiate into multiple lineages responding to environmental signals (Vermeulen et al., 2008). The effectiveness of these markers to separate CSC was, nevertheless, challenged by others (Kemper et al., 2010). Definitive proof of the existence of tumor cell populations with distinct initiation potential came from the Clevers group: by inducing an *Apc* deletion specifically in the LGR5+ cells, they generated stable growing adenomas, whereas the same deletion in the transient amplifying and villus compartments gave rise to short lived mutated cells (Barker et al., 2009). Moreover, they showed how despite aberrant WNT activation was present through all the adenoma, only specific subsets of cells retained *Lgr5* expression, suggesting there was indeed cellular heterogeneity within this benign tumors.

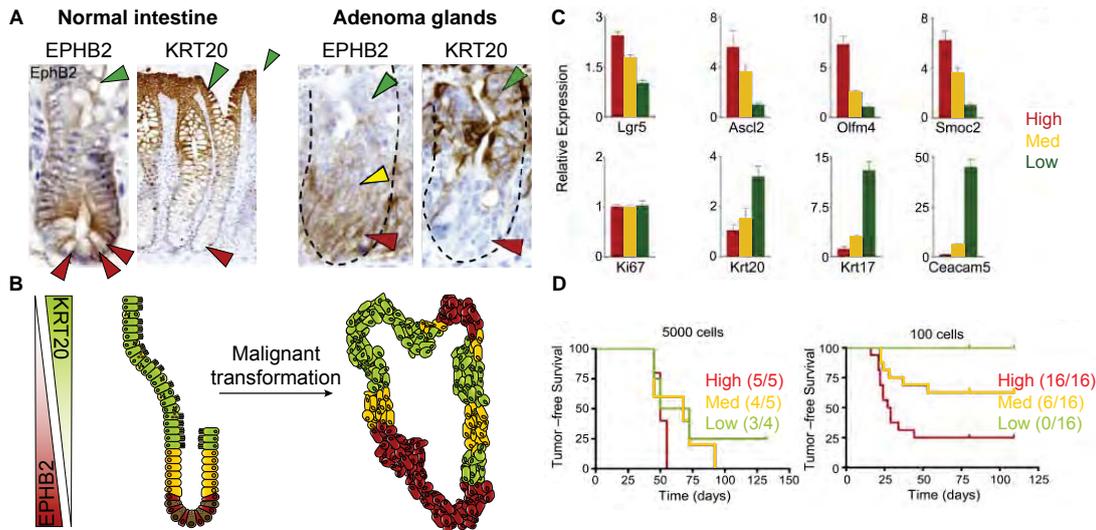
To prove whether this heterogenic population still retained a hierarchical organization, they next combined the multicolor lineage tracing system induced by *Lgr5* (LGR5-CreERT2 R26R-Confetti) with conditional *Apc* deletion in mice (Schepers et

al., 2012). The first tamoxifen pulse induced *Apc* deletion and Confetti recombination to one of the four possible colors in *Lgr5*<sup>+</sup> cells at the same time. Adenomas generated by this system showed glandular-like structures of the same color, evidencing that all the cells arise from the same parental CSC (Figure 1.15A). If one month after adenoma growth a second tamoxifen shot was administered, inducing re-tracing in the *Lgr5*<sup>+</sup> pool, the glands slowly could observe how the glands slowly switched to another color, proving that mouse adenoma growth was fueled by *Lgr5*<sup>+</sup> cells (Figure 1.15B).



**Figure 1.15: Adenomas maintain the intestinal stem cell hierarchy.** **A.** Induction of adenoma formation in mice small intestines by *Apc* deletion in *Lgr5*<sup>+</sup> cells. Cre induction also recombines the R26R-Confetti tracing system. The image depicts monoclonal adenoma glands (delineated by dashed lines) 28 days after induction. **B.** The Confetti cassette can be recombined a second time, provoking a color switch between yellow-green or red-blue. Image depicts the colonization of a gland by a new *Lgr5*-derived clone (blue), proving that *Lgr5*<sup>+</sup> cells fuel tumor growth. Modified from Schepers et al., 2012.

Further studies of the cancer stem cell paradigm in human tumors has been hampered by the lack of reliable antibodies against LGR5 protein. In our group, we overcame this limitation by using a surrogate marker of stemness, the receptor tyrosine kinase EPHB2. EPHB2 mediates cell positioning in the intestine through its interaction with the ephrinB1 ligand (EFNB1) (Batlle et al., 2002). In CRC, downregulation of EPHB2 is required for the loss of tumor cell compartmentalization and progression to aggressive adenocarcinomas (Batlle et al., 2005, Cortina et al., 2007, Solanas et al., 2011). We determined that EPHB2 was a reliable marker to isolate normal ISCs (Jung et al., 2011). By immunohistochemistry, we could also observe a similar pattern between normal crypts and tumor glands, with differentiated (KRT20<sup>+</sup>) and stem-like (EPHB2<sup>+</sup>) compartments (Figure 1.16 A and B). Moreover, using EPHB2 antibodies we separated a cell population from human CRCs that retained a stem cell gene program and displayed higher tumorigenic potential upon reinjection into immunodeficient mice (Figure 1.16 C and D) (Merlos-Suárez et al., 2011).



**Figure 1.16: Colorectal tumor organization mimics that of the normal intestine.** **A.** IHC stainings against EPHB2 and KRT20 in normal and adenomatous glands in human tissue sections. Red arrows point at EPHB2-High/KRT20-Neg cells, green arrows point at EPHB2-Low/KRT20-High cells and yellow arrows point at the intermediate population. **B.** Representation of the stem vs differentiated gradient in the homeostatic crypt, which is maintained during carcinoma progression. **C.** Relative mRNA expression of stem and differentiated genes in the EPHB2-High –Med and –Low sorted populations from human tumors. EPHB2-High cells are enriched in stem cell expression and counter enriched in differentiation markers. **D.** Limiting dilution assay indicates the tumor initiation capacity of EPHB2-High, -Med and –Low sorted cells. In limiting numbers, EPHB2-High are the only cell population that retains tumor formation potential. Modified from Merlos-Suárez et al, 2011.

Altogether, these data demonstrate that despite having the same mutational background, there is heterogeneity within CRC tumor cell populations, and that a certain degree of hierarchy is probably maintained in these tumors. How this difference is possible remains unclear, although emerging evidence point to the stroma as a controller in the differentiation-stemness balance in tumors (Vermeulen et al., 2010).

#### 1.4.4 Tumor microenvironment and the CSC theory

For many years, the research focus has been divided between epithelial cells and tumor microenvironment. Tumor microenvironment comprises several cell types: pericytes, endothelial cells, fibroblasts and immune inflammatory cells, and it greatly influences tumor development (Hanahan and Weinberg, 2011). First evidences of the tumor microenvironment's role in CRC development were found in patients with ulcerative colitis and Crohn's disease; the inflammatory status of the intestine of these patients predisposes to tumor growth (Itzkowitz and Yio, 2004). Indeed, the use of non-steroidal anti-inflammatory drugs protects against CRC development (Chan et al., 2009). Moreover, mouse models for CRC often require inflammation via dextran sodium sulphate (DSS) in order to develop adenomas (Tanaka et al., 2006).

It is thus easy to imagine that, similarly to the healthy stem cell niche, which regulates SC fate and proliferation, there is a cancer stem cell niche dictated by the microenvi-

ronment that regulates CSC behavior in CRC and other tumors. Considering *in vitro* data, there are many reports pointing at the importance of the media factors to determine the CSC phenotype: samples with different mutational background require different supplementation for their growth (Fujii et al., 2016). It is notorious the example of a CRC-derived organoid sensitive to the WNT inhibitor porcupine because it carries a mutation in the *RNF43* negative WNT feedback regulator rather than in *APC* (van de Wetering et al., 2015). Alike the *in vitro* situation, *in vivo* CSCs also depend on these factors to grow, and the tumor microenvironment represents their main source. It has been shown that tumor cells in close proximity to the mesenchyme receive more WNT signaling and retain a stronger stem-like phenotype (Vermeulen et al., 2010). Moreover, a similar effect has been observed regarding Notch pathway; blocking DLL4 ligand in murine vasculature heavily impairs tumor growth (Hoey et al., 2009). A certain degree of BMP response is also maintained in tumor cells, as BMP4 is able to induce CSC differentiation (Lombardo et al., 2011).

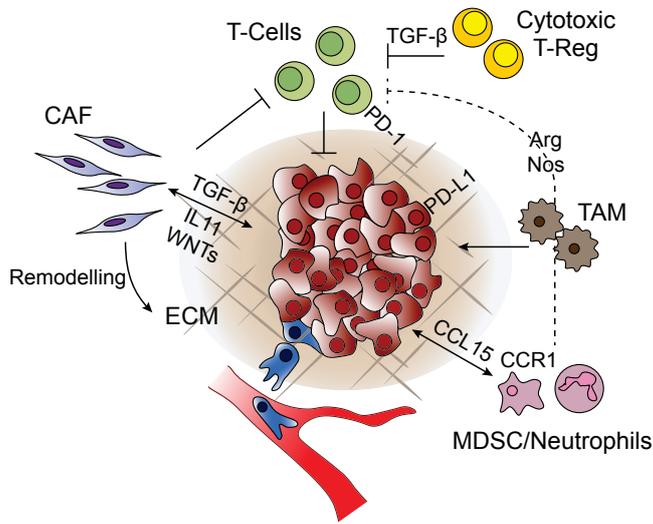
The tumor microenvironment, though, has additional roles than imposing a CSC phenotype. It is well described that a stromal gene program expression in CRC is associated with higher risk of metastasis and poorer prognosis (Calon et al., 2012, Calon et al., 2015). Moreover, the recent CRC classification based on consensus molecular subtypes (CMS) not only groups tumors based on their mutational profile and gene expression, but also takes into account the microenvironment status (Guinney et al., 2015). Importantly, CMS4 tumors are defined as mesenchymal-like and rich in stroma, and patients with CMS4 tumors usually correspond to at stage III and IV and therefore have the lowest relapse-free survival. In the next section, I will summarize the most relevant findings related to stroma-tumor interaction and the metastatic process.

#### 1.4.5 *Metastasis in CRC*

The metastatic process is a rather complex and amazing feat for tumor cells. Epithelial cells must acquire invasive properties, enter the blood vessels, survive the immune system surveillance and finally extravasate and grow in a new organ with a “hostile” environment. These barriers render metastasis as a highly inefficient process at the end, but once they are established, few treatment options exist for CRC patients (Massagué and Obenauf, 2016).

It is straightforward to hypothesize that metastatic cells must derive from cancer stem cells, as they require of stemness potential to recreate the tumor in the distant organs. Research by various groups has focused in elucidating whether CSC markers for the colon would also label metastatic cells, found at the tumor budding edges (Zlobec and Lugli, 2010, van Wyk et al., 2015). Indeed, years ago Brabletz and colleagues already reported increased nuclear  $\beta$ -catenin accumulation at the tumor border (Brabletz et al., 1998), probably mediated by the tumor microenvironment (Brabletz et al., 2001). Despite the efforts, few advances have been made in finding specific markers for colon cancer metastatic cells. The most important contributions in this sense come from the Stassi group, where they define a CD44v6 population with both CSC and metastatic properties (Todaro et al., 2014), and from de Sauvage laboratory, were, by means of LGR5 ablation in knock-in mice, they demonstrated

that LGR5+ cells are required for colorectal metastasis (Melo et al., 2017).



**Figure 1.17: Tumor microenvironment promotes a pro-metastatic phenotype.** Schematic representation of the major players from the TME in tumor progression and dissemination. Tumor cells convert stromal cells into cancer associated fibroblasts via TGF- $\beta$  signaling, which in turn help cancer progression by stimulating proliferation and stemness, remodeling the ECM and inhibiting T-cell response against the tumor. Myeloid lineages and macrophages are also recruited to the tumor, and contribute to its expansion. T-cell cytotoxic effect is blocked by T-Reg activity and PD-L1 expression in the tumor and stroma.

Epithelial tumor cells are primed by the surrounding environment to acquire at least a partial EMT phenotype and migrate, and, in a positive regulation feedback, they also recruit stromal and immune cells to help with the dissemination process (Tauriello et al., 2016) (Figure 1.17). Briefly, the main players from the microenvironment in the acquisition of a metastatic phenotype by epithelial tumor cells are:

- Cancer associated fibroblasts (CAFs): fibroblasts are recruited to the tumor by epithelial cells, and via mediators such as TGF- $\beta$  they undergo a switch towards a tumor-promoting phenotype. They release matrix metalloproteinases that contribute to ECM remodeling, secrete tumor promoting factors such as IL-11 (Calon et al., 2015), and promote stemness via WNT ligand secretion (Kalluri, 2016).
- Endothelial cells: tumor growth is associated with an increase in blood vessel formation. Secretion of angiogenic stimuli like VEGFA and IL-33 is characteristic of tumor cells (Goel and Mercurio, 2014, Zhang et al., 2017), which enhances the creation of an abnormal net of capillaries with defective permeability that also contributes to ease dissemination (Yonenaga et al., 2005).
- Innate immunity: innate immune cells have a dual role in tumor progression. Both macrophages and neutrophils can act either against the tumor cells or favoring their growth and dissemination (Braster et al., 2017, Fridlender et al., 2009). Tumor associated neutrophils are involved in the CCR1-CCL15 recruitment axis in SMAD4 deficient cells: tumor cells secrete CCL15 and attract CCR1+ myeloid cell lineages (neutrophils and myeloid derived suppressor cells (MDSCs)) (Itatani et al., 2013, Inamoto et al., 2016, Yamamoto et al., 2017). Accumulation of myeloid-derived cells has been associated with metastasis outgrowth and poor prognosis in CRC (Wang et al., 2017).
- Adaptive immunity: the development of new CRC models in immunocompetent mice has finally proven the idea that T-cells play a major role in antitu-

mor activity and recurrence prevention (Galon et al., 2006). Work from our lab has shown that a TGF- $\beta$  rich microenvironment promotes T-cell exclusion, mediating an immune-evasion that permits metastasis growth (Tauriello et al., 2018). These findings have opened the door to immunity-based treatments (immunotherapy), where the main effort is to activate the intrinsic immune system of the patient against tumor cells. Indeed combination of TGF- $\beta$  inhibitors and PD-L1 inhibitors induces a strong anti-tumor response, giving new hope to treat advanced CRC patients (Tauriello et al., 2018).

#### *Models of metastatic colorectal cancer*

One of the major drawbacks in the study of colon tumors has been the lack of appropriate models. As explained in **section 1.4.2**, most of the available GEMMs are limited to benign adenomas and some invasive adenocarcinomas that do not progress to metastatic disease, hampering the study of therapies for stage IV human CRC. In the last years, though, the development of new techniques has allowed more accurate reproduction of the disease both *in vitro* and *in vivo*. Recently, our lab generated the first GEMM of metastatic CRC (Tauriello et al., 2018). We crossed mouse strains bearing engineered alleles for four of the most common genetic alterations found in human CRC -Apc<sup>fl/fl</sup>, Kras<sup>LSLG12D</sup>, TGF $\beta$ R2<sup>fl/fl</sup>, Tp53<sup>fl/fl</sup>- and recombined these mutations in intestinal stem cells by means of the Lgr5-creERT2 driver. Quadruple mutant mice developed metastatic intestinal tumors that reproduced several key features of human poor prognosis microsatellite stable CRCs including a stromal rich TGF- $\beta$  -activated TME and T-cell exclusion. We demonstrated that these models are bona-fide surrogates of stromal-rich CMS4 human CRCs (Tauriello et al., 2018). For the first time, they enabled the study the interaction of CRC and the TME in a fully immunocompetent setting. In addition, these tumors were isolated from mice and cultured in organoid format, allowing for further genomic modifications.

#### *Orthotopic transplants*

The organoid culture system, either PDO or CRISPR/Cas9 engineered, enables the long-term maintenance and expansion of heterogeneous CRCs that can be implanted orthotopically in their native environment in immunosuppressed mice. For CRC, the widest used technique is intracaecum injection of *in vitro* grown organoids (Céspedes et al., 2007). Moreover, in our group we also developed a nesting technique, consistent on transplanting small pieces of already grown tumor in the caecum. The most common metastatic sites for orthotopically injected CRC tumors are livers and lungs (Fumagalli et al., 2017, Roper et al., 2017).

### 1.5 GENOME-EDITING TOOLS

As a last chapter of the introduction I cannot fail to mention one of the revolutionary techniques key to the project success, the CRISPR/Cas9 genome-editing tool. During the last decades, the possibilities of customizing the genome of adult cells have refined with the establishment of site-specific endonucleases. The system is based in fusing an unspecific nuclease domain with a DNA-binding specific site. The chimeric

nuclease will induce a double strand break (DSB) at the desired locus, which will be repaired either by non-homologous end joining (NHEJ), introducing mutations, or homologous recombination (HR) if the adequate template is provided. HR repair will introduce any sequence located in between the homology arms, creating a knock-in. The modification of *in vitro* cultured organoids by these means has opened the door to studies previously limited to GEMM, allowing for lineage tracing and clonal analysis in human tumors. The most used genome editing systems are:

#### 1.5.1 Zinc finger nucleases (ZNF)

Cys2-His2 Zinc fingers are one of the most common DNA binding motifs. Each finger consists of 30 amino acids in  $\beta\beta\alpha$  conformation that binds 3-4 base pairs (bp) of DNA. By synthetic generation of three to six zinc finger proteins, DNA sequences of 9 to 18 bp can be specifically recognized in the genome. The combinatorial approach supposes an easy method to target any site in the human genome (Liu et al., 1997).

#### 1.5.2 Transcription activation like elements (TALE)

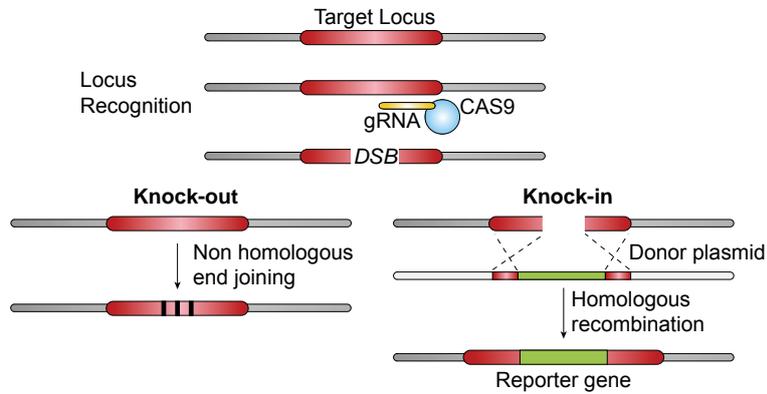
TALE are proteins from the plant invasive bacteria *Xanthomonas*, each domain comprised of 35 amino acids. They recognize single nucleotides within the DNA, thanks to two hypervariable amino acids that determine their specificity. By linking several TALE, any sequence in the genome can be targeted, with higher flexibility than the triplet-based system of the ZFN. TALE domain cloning, nevertheless, is technically complicated due to repetitive sequences (Boch et al., 2009).

#### 1.5.3 CRISPR/Cas9

The bacterial RNA guided system Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) type II has supposed the definitive answer for gene editing. Briefly, bacteria contain foreign DNA segments, termed spacers, in the CRISPR locus. The transduction of spacers generates a CRISPR RNA (crRNA) that binds to a transactivating crRNA (tracrRNA) and together they are able to direct specific DNA cleavage by Cas proteins. Recognition of the site requires a protospacer adjacent motif (PAM) of three nucleotides (NGG) followed by a 20 nucleotide sequence homologous to the crRNA. This system has been reconverted for use in human cells by introducing a plasmid encoding for the guide RNA (crRNA+tracrRNA) and the Cas9 nuclease. Thus, simply by designing the guide sequence, any researcher can easily induce a DSB in a specific genome position (Ran et al., 2013). Another advantage of the CRISPR/Cas9 system is the easiness of multiplexing, i.e targeting several genes at the same time just by using several guide RNA together (Cong et al., 2013) (Figure 1.18).

A great example of the easiness of CRISPR/Cas9 genome editing is the *in situ* modification of mouse colon cells by colonoscopy-mediated delivery of lentiviral vectors containing Cas9 and guide RNA (Roper et al., 2017). Using this system, they were

able to mutate the *Apc* and *Tp53* genes *in vivo*, inducing colon tumor formation.



**Figure 1.18: CRISPR/Cas9 gene editing system.** CRISPR/Cas9 is based on the induction of a double strand break (DSB) in a specific locus within the genome. Cas9 nuclease is translocated to the desired genome position by a specific guide RNA (gRNA) of only 25 bp. The DSB can be either repaired by non-homologous end joining (NHEJ), introducing mutations and thus creating a knock-out, or by homologous recombination (HR) if a template is provided. HR allows the insertion of any fragment of interest in between the homology arms for knock-in generation.

Part II

## OBJECTIVES



## OBJECTIVES

---

This thesis work was aimed to study tumor cell composition and the cell hierarchy that characterize advanced human colorectal cancers. In particular, we have focused on the following objectives:

1. To develop genetic tools that enable the study of the behavior of distinct tumor cell populations within intact tumors.
2. To define the potential of *LGR5* as a marker for human colon cancer stem cells.
3. To analyze the contribution of *LGR5*<sup>+</sup> cells to intestinal cancer growth using lineage tracing strategies on human-derived CRCs.
4. To characterize the heterogeneity of phenotypes and behaviors within the cancer stem cell pool.
5. To analyze the contribution of differentiated-like cells to colorectal cancer growth.
6. To develop 3D imaging systems to visualize and track specific cell populations within tumors.
7. To study the dynamics and the contribution of distinct cell types to tumor growth.
8. To develop mouse orthotopic transplantation models with patient-derived tumors to study the metastatic process.
9. To identify the cell of origin of metastasis.



Part III  
RESULTS



RESULTS

---

The main focus of this project is to investigate the tumor cell hierarchy present in human CRCs. As I detail in the introduction, to a large extent tumor cells and in particular CSCs have been study through transplantation experiments. We developed a new methodology to study distinct tumor cell populations in unperturbed tumors by combining two cutting-edge techniques; CRISPR/Cas9 gene editing tools and patient-derived organoid cultures (PDOs). The PDO culture system has allowed us to indefinitely maintain a collection of human samples growing as stem cells *in vitro* (Sato et al., 2009, Jung et al., 2011). Upon injection into immunodeficient mice, PDOs were capable of regenerating full tumor complexity (Merlos-Suárez et al., 2011, Calon et al., 2012). Moreover, PDO cultures enabled the possibility of creating specific genomic modification using CRISPR/Cas9, and in particular, the insertion of reporter genes and lineage tracing cassettes in selected marker genes.

This thesis results are organized in three chapters. The first chapter represents an extended version of the results published in Cortina, Turon et al. *EMBO Mol Med*, 2017, where we demonstrated the existence of an LGR5+ tumor cell population in human CRCs that behave as a cancer stem-like cells (full article can be found in the Annex section). In the second chapter, we compared the behavior and proliferative dynamics of the LGR5+ population with those of more differentiated-like cells. Finally, I discuss the findings related to the use of *EMP1* as a novel metastatic marker for CRC.

## Author contributions

Dr. Eduard Batlle and Dr. Carme Cortina, as co-directors of the PhD thesis, report that the thesis results are composed of three chapters consisting respectively in 1) an extended version, written by Gemma Turon, of the published paper Cortina, Turon et al., 2017. 2 and 3) unpublished work done by Gemma Turon.

### Chapter I

Results Chapter I of this thesis are an extended version of the publication:

Cortina, C.\*, Turon, G.\*, Stork, D., Hernando-Momblona, X., Sevillano, M., Aguilera, M., Tosi, S., Merlos-Suárez, A., Otto-Attolini, CS., Sancho, E., Batlle, E. (2017). A genome editing approach to study cancer stem cells in human tumors. *EMBO Molecular Medicine*, 9(7), 869–879. <https://doi.org/10.15252/emmm.201707550>

\*these authors contributed equally to this work

EMBO Molecular Medicine had an impact factor of 10.293 in the 2017 edition of the Journal of Citation Reports (Thomson Reuters)

Author contributions of the paper are as follows:

*EB conceived the study and wrote the manuscript. CC and GT designed experimental work, executed experiments, and helped with manuscript writing. DS provided technical support with targeting vector generation and genome editing of organoids. XH-M performed mice work. AM-S and ES helped designing and conceptualizing the study. MS provided crucial help with immunohistochemistry and organoid cultures. CS-OA performed all statistical analyses. MA and ST contributed to image analysis and 3D reconstruction of serial sections.*

As co-first authors CC and GT shared equal work and responsibility for design and execution of all the experiments of the paper.

### Chapter II and Chapter III

Results Chapters II and III of this thesis are unpublished work. Author contributions for these chapters are as follows:

EB conceived the research project. GT designed the experimental work and executed the experiments under the supervision of CC. DS provided technical support with targeting vector generation and genome editing of organoids. XH-M and GT performed mice work together. MS provided technical support with immunohistochemistry. CS-OA provided support with statistical analysis. MA and ST contributed to image analysis and 3D reconstruction of serial sections. ES helped in the interpretation and discussion of the results.

# Chapter 1: Study of LGR5+ cells in human CRCs

## 3.1 CHARACTERIZATION OF LGR5+ CELLS IN HUMAN CRC ORGANOID

### 3.1.1 CRISPR/Cas9 labelling of LGR5 gene

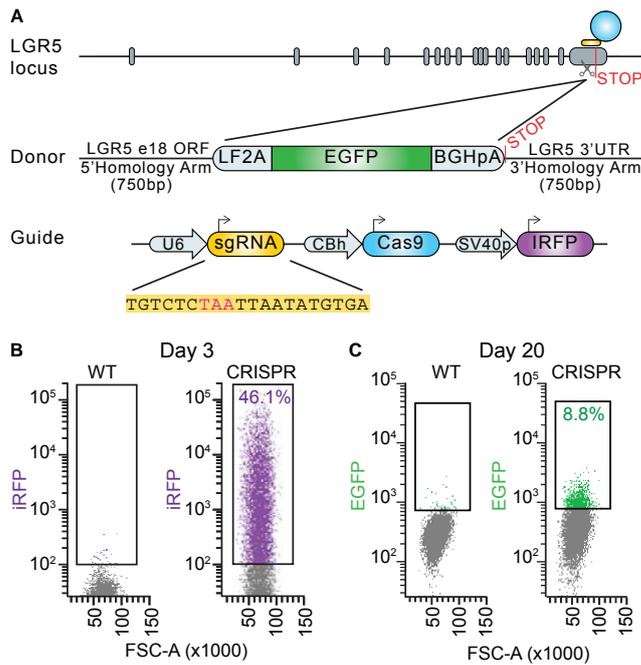
For this study, we selected two CRC patient derived organoids from our collection, PDO6 and PDO7. Both tumors, like the majority of colorectal carcinomas, harbor a combination of canonical CRC mutations in the main driver pathways. PDO6 corresponds to a relatively benign grade 2 CRC and carries no mutations in main components of TGF- $\beta$  and p53 pathways. PDO7, on the contrary, is a highly aggressive carcinoma with mutations in four key pathways for CRC development (WNT, KRAS, TGF- $\beta$  and TP53) (Fearon et al., 2011) (Table 3.1.)

PDO	Stage	WNT pathway	KRAS pathway	PI3K pathway	TP53 /ATM	TGF- $\beta$ pathway
PDO6	II	APC FS (-522/+)	KRAS (A146V/+)	WT	WT	WT
PDO7	IV	APC STOP(R787*)(G13/+)	KRAS (G13/+)	WT	ATM (V182L /N1983S)	SMAD4 (L536R /L536R)

**Table 3.1: CRC driver mutations in the PDOs used in this study.** PDO6 and 7 stage at the time of diagnosis, as well as summary of cancer driver mutations found in the main signaling pathways involved in CRC development in those PDOs. FS: frameshift mutation.

We utilize two vectors, px330-guide-Cas9 containing the gRNA guide and a Cas9 protein sequence, as well as an iRFP marker (modified from the original px330 of Ran et al., 2013), and a donor plasmid containing the reporter construct flanked by homologous recombination arms (Figure 3.1A). gRNA guide was designed to bind and recruit Cas9 nuclease to the 3'UTR of the *LGR5* gene, with the cutting site right before the STOP codon. The reporter included an LF2A-EGFP coding sequence, so that endogenous *LGR5* and inserted *EGFP* would be expressed as a single mRNA molecule and cleaved at the protein level (Szymczak and Vignali, 2005).

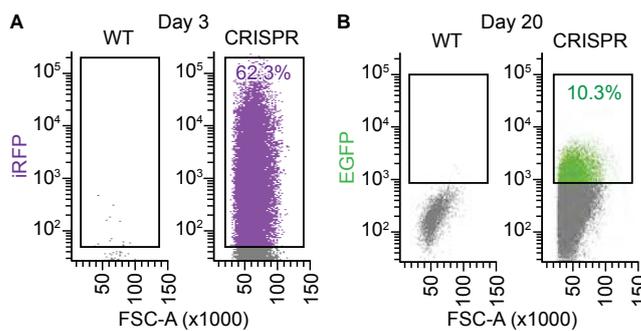
We nucleofected PDO7 with both plasmids at a 3:1 ratio (7+2 $\mu$ g) and FACS (Fluorescent Activated Cell Sorting) separated the cells 3 days after (short term sorting) based on their iRFP positivity (Figure 3.1B). 20 days after the initial sorting, we selected EGFP+ cells by flow cytometry (long term sorting) and plated them in single cell format in 3D conditions (Figure 3.1C). We genotyped single cell derived organoids by PCR, amplifying 3' and 5' ends of the integration site, and also the whole locus to determine whether clones contained correct homo- or heterozygous insertions (Figure A.1A). From here onwards, this PCR-based strategy will be referred as integration PCR: 3'specific, 5'specific and H/H respectively). In addition, we performed



**Figure 3.1: CRISPR/Cas9 design of LGR5-EGFP knock-in in human PDOs.** **A.** Design of LGR5-EGFP donor and CRISPR/Cas9 sgRNA vectors. The blue circle represents the CRISPR/Cas9 protein complex and the yellow box underneath the guide RNA. **B.** Short term FACS-sorting at day 3 post-nucleofection of the cells that have incorporated the guide plasmid. **C.** Flow cytometry profile of CRISPR-engineered cells 20 days post-nucleofection. Adapted from Cortina, Turon et al., 2017.

southern blot analysis of the *EGFP* sequence to ensure that integration occurred exclusively in the *LGR5* locus (Figure A.1B). The efficiency (number of clones obtained with correct integrations) along the process is summarized in Table A.1. We picked two representative homozygous clones, #1 and #2, for further experiments.

We repeated the procedure for PDO6 (Figure 3.2, and Figure A.2), with a similar targeting efficiency (Table A.1). We selected a homozygous clone without off-targets for future experiments (#18).

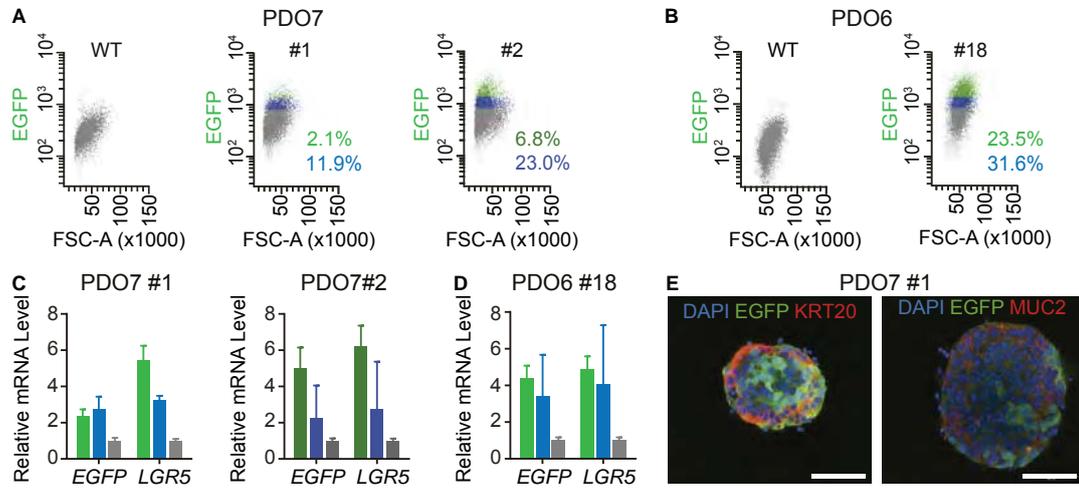


**Figure 3.2: Flow cytometry derivation of PDO6 LGR5-EGFP clones.** **A.** Flow cytometry profile 3 days post-nucleofection of PDO6. iRFP+ cells were isolated for *in vitro* expansion. **B.** Long term purification of the LGR5-EGFP+ population to derive single cell clones. Adapted from Cortina, Turon et al., 2017.

### 3.1.2 *EGFP* knock-in reports *LGR5* levels

We interrogated whether *EGFP* reported *LGR5* in knock-in derived LGR5-EGFP clones. By flow cytometry we separated EGFP-High, -Low and -Negative cell populations and isolated RNA from the sorted fractions (Figure 3.3A and B). RT-qPCR analysis confirmed that *LGR5* mRNA expression paralleled EGFP protein levels (Figure 3.3C and D). Immunofluorescence staining for KRT20 and MUC2 demonstrated that some

cells in the organoids undergo differentiation, despite they were maintained under stem cell conditions. LGR5-EGFP+ cells showed a complementary staining pattern to that of differentiation markers KRT20 and MUC2 (Figure 3.3E).



**Figure 3.3: PDO6 and PDO7 organoids retain heterogeneity in vitro.** **A.** Flow cytometry analysis of PDO7 LGR5-EGFP#1 and #2 *in vitro* grown organoids showing EGFP-High (green) -Low (blue) and -Negative (grey) cells and **B.** of PDO6 LGR5-EGFP#18. **C and D.** EGFP and LGR5 relative mRNA expression level by qrtPCR of the LGR5-High, -Low and -Negative FACS-isolated populations of each clone respectively (PDO7#1 and #2 and PDO6#18). Data is represented as mean $\pm$ s.d. of three technical replicates. **E.** Immunofluorescence of *in vitro* grown organoids of PDO7 LGR5-EGFP#1 stained for EGFP and the differentiation markers KRT20 and MUC2. Scale bars indicate 100  $\mu$ m. Adapted from Cortina, Turon et al., 2017.

### 3.1.3 CRISPR/Cas9 genome editing does not introduce random mutations in PDO7 DNA

Before proceeding to perform *in vivo* experiments, we assessed the introduction of undesired off-target mutations in modified PDOs by the genome editing procedure. For that purpose, we compared parental PDO7 vs LGR5-EGFP#1 derived clone by exome sequencing. This analysis revealed few private mutations present in the clone, yet none of which occurred in known cancer-driving genes (Table 3.2).

CRC driver mutations	<i>de novo</i> missense or non-sense mutations	<i>de novo</i> mutations with predicted high impact (genes)
All present (APC, KRAS, SMAD4, ATM)	64	3 (EIF2AK2, KIAA0101, PPFIBP2)

**Table 3.2: PDO7 WT vs LGR5-EGFP#1 exome sequencing summary.** Analysis of mutations introduced in clone 1 after CRISPR/Cas9 targeting. None of the CRC driver mutations were modified after genome-editing and only three of the 64 newly generated mutations have predicted impact in protein function (detailed in section 6.8).

In addition, we monitored the most probable off-target sites of the LGR5 gRNA predicted by the Zhang laboratory algorithm (<http://tools.genome-engineering.org>). We found no mutations in the predicted sites (Table 3.3).

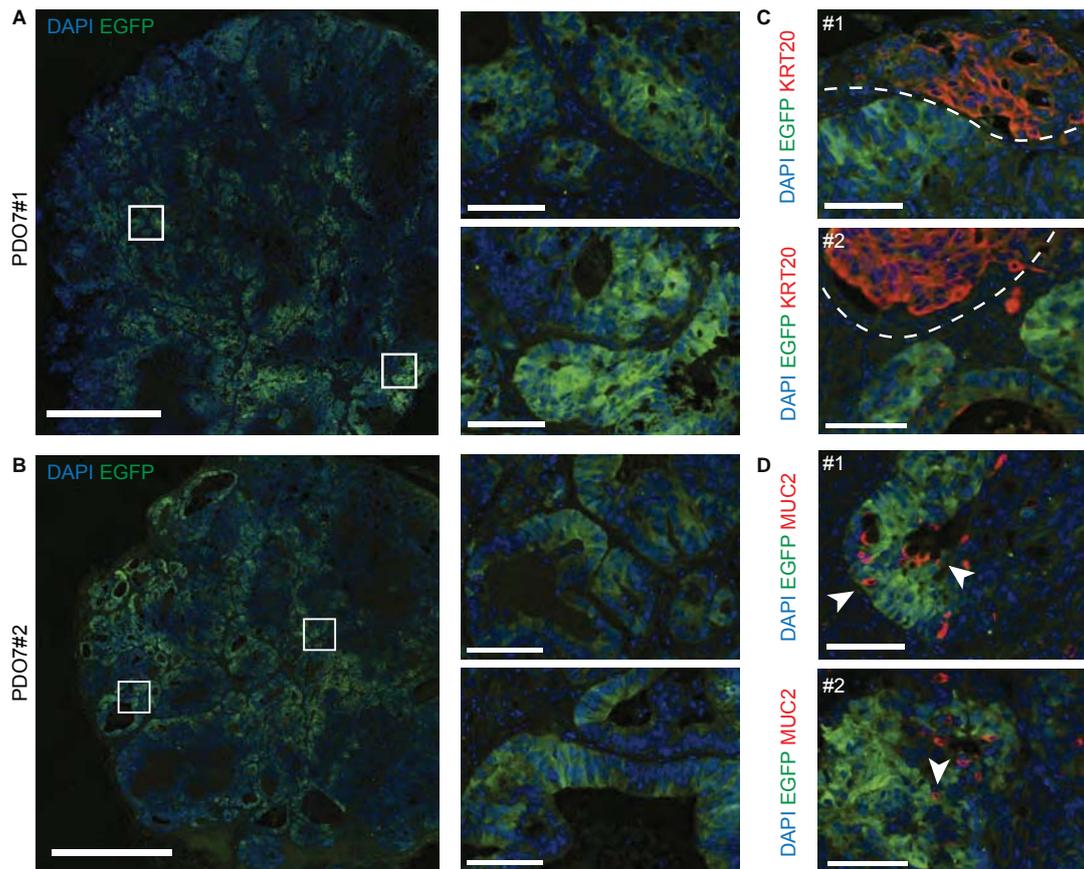
Sequence	Mismatches	Gene	Location	Clone 1
TGTCTCTGAATAGTA GGTGAGGG	4MMs (8:10:13:16)	GPR3 (NM005281)	Exon2: bp521-543 aa141-167	WT
CACCCCTAATTAATA TGTGAGAG	4MMs (1:2:3:5)	SH2D4B (NM207372)	Exon 7: bp 3748-3771 3'UTR	WT
TTTTTCTATTCAATA TGTGAAAG	4MMs (2:4:9:11)	MT1B (NM005947)	Exon 3: bp 325-357 3'UTR	WT
TATCTCTAATGAAAT TGTGAAAG	4MMs (2:11:14:15)	PRND (NM012409)	Exon 2: bp 1118-1140 3'UTR	WT
TATCTATAAATAATT TGTGAAGG	4MMs (2:6:10:15)	USF3 (NM001009899)	Exon 7: bp 13257- 13280 3'UTR	WT

**Table 3.3: Predicted LGR5 CRISPR/Cas9 RNA guide off-target sites.** Alternative sequences recognized by the LGR5-specific gRNA. Number of mismatches between original and alternative sequences are indicated, as well as position of those. Gene and exact location of each sequence is found in the table. None of the predicted hotspots presented mutations in PDO7 LGR5-EGFP#1 (last column).

### 3.1.4 LGR5 marks a stem-like population in human CRC xenografts

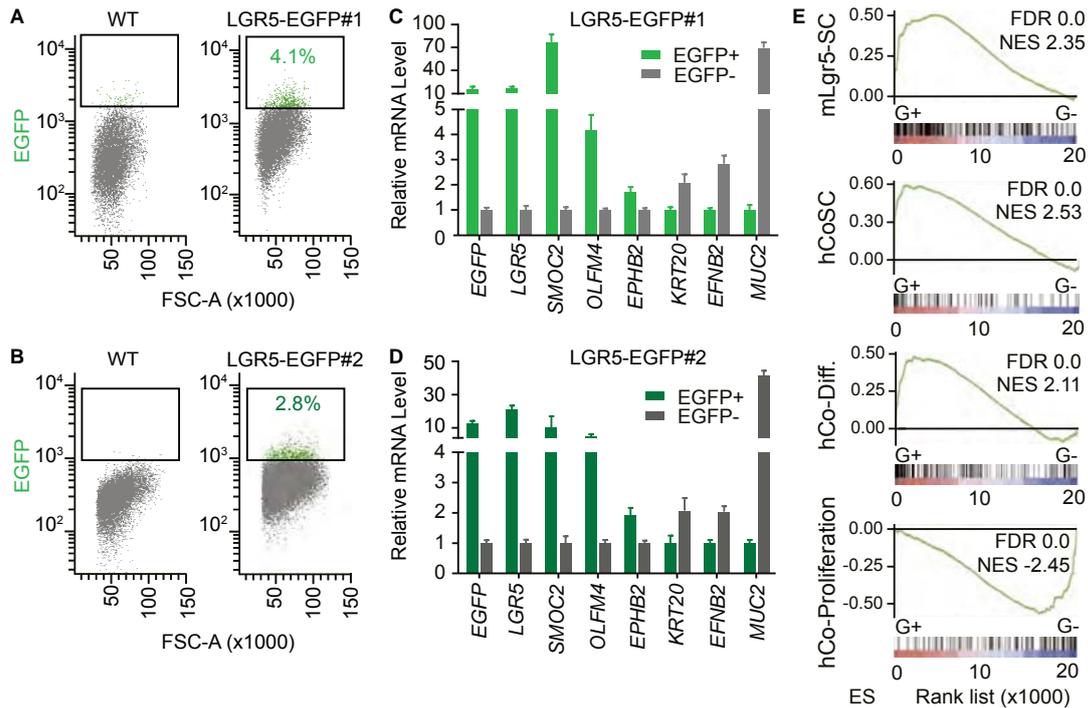
Once we confirmed that the modified clones were correctly targeted, we injected them subcutaneously into immunodeficient Beige/SCID mice (150000 cells per flank) in order to generate patient derived xenografts (PDX). When tumors reached 100-150 mm<sup>3</sup>, mice were sacrificed and tumors analyzed both by immunohistochemistry (IHC) and by disaggregation followed by flow cytometry of alive cells. Histology analysis revealed PDO7 tumors were glandular-like, with LGR5-EGFP+ areas representing 50-60% of the whole tumor (Figure 3.4A and B). Co-staining with differentiation markers corroborated the observations in organoids; there was no overlap between LGR5-EGFP glands and KRT20 or MUC2 expression (Figure 3.4C and D).

In experiments of tumor cell dissociation followed by FACS, we used anti-human EPCAM antibodies to distinguish tumor cells from mouse cells. EGFP+ cells represented 3-4% of the total epithelial compartment (Figure 3.5A and B). According to histological stainings, this 4% of EGFP+ cells corresponded to EGFP-High only, whereas EGFP- cells were actually EGFP-Low and -Negative, possibly because FACS was not sensitive enough to detect all grades of EGFP positivity. We sorted the EGFP+ and EGFP- fractions and performed mRNA extraction followed by gene expression analysis. RT-qPCR revealed a 10-fold enrichment of LGR5 mRNA expression in the



**Figure 3.4: Histology of PDO7 LGR5-EGFP subcutaneous xenografts.** **A.** EGFP immunofluorescence in a paraffin section of LGR5-EGFP#1 xenograft. Magnifications of the glandular like structures are indicated in white rectangles. Scale bars represent 1 mm for whole xenografts and 100  $\mu\text{m}$  for the insets. **B.** Same for in LGR5-EGFP#2 derived tumors. Scale bars represent 1 mm for whole xenografts and 100  $\mu\text{m}$  for the insets. **C.** Dual immunofluorescence staining showing KRT20 and LGR5-EGFP complementary expression domains in #1 and #2, respectively. Dashed lines delimit expression domains in adjacent glands. Scale bars indicate 100  $\mu\text{m}$ . **D.** MUC2 and LGR5-EGFP dual immunofluorescence staining in paraffin sections of #1 and #2. White arrows point at MUC2+ EGFP- cells. Scale bars indicate 100  $\mu\text{m}$ . Adapted from Cortina, Turon et al., 2017.

EGFP+ fraction, as well as of other stem cell markers (*SMOC2*, *OLFM4*, *EPHB2*). We also observed clear counter-enrichment in expression of differentiation marker genes (*KRT20*, *MUC2*, *EFNB2*) (Figure 3.5C and D). We analyzed global gene expression of the two populations using microarrays. Subsequently, we performed gene-set enrichment analysis (GSEA) with signatures of both mouse and human intestinal stem cells (Muñoz et al., 2012, Merlos-Suárez et al., 2011), of the intestinal differentiated cells program (Merlos-Suárez et al., 2011), as well as of crypt proliferative progenitors (Merlos-Suárez et al., 2011) (Figure 3.5C). As expected, EGFP+ cells were highly enriched in the stem cell and proliferation programs and devoid of differentiation marker gene expression.



**Figure 3.5: LGR5-EGFP+ cells are enriched in stem cell genes and devoid of differentiation marker gene expression.** **A and B.** Representative FACS analysis of #1 and #2 disaggregated subcutaneous-derived xenografts. Only EPCAM+/DAPI- cells are plotted. **C and D.** Relative mRNA expression level by qrtPCR of intestinal stem and differentiation genes for the sorted EGFP+ and EGFP- populations. Data is represented as mean±sd of three technical replicates. **E.** GSEA comparing the expression of signatures of mouse LGR5+ cells (mLgr5-SC), human colon stem cells (hCoSCs), differentiated cells (hCo differentiation) or proliferative crypt cells in profiled LGR5-EGFP+ vs LGR5-EGFP- cells from PDO7 LGR5-EGFP#1. Adapted from Cortina, Turon et al., 2017.

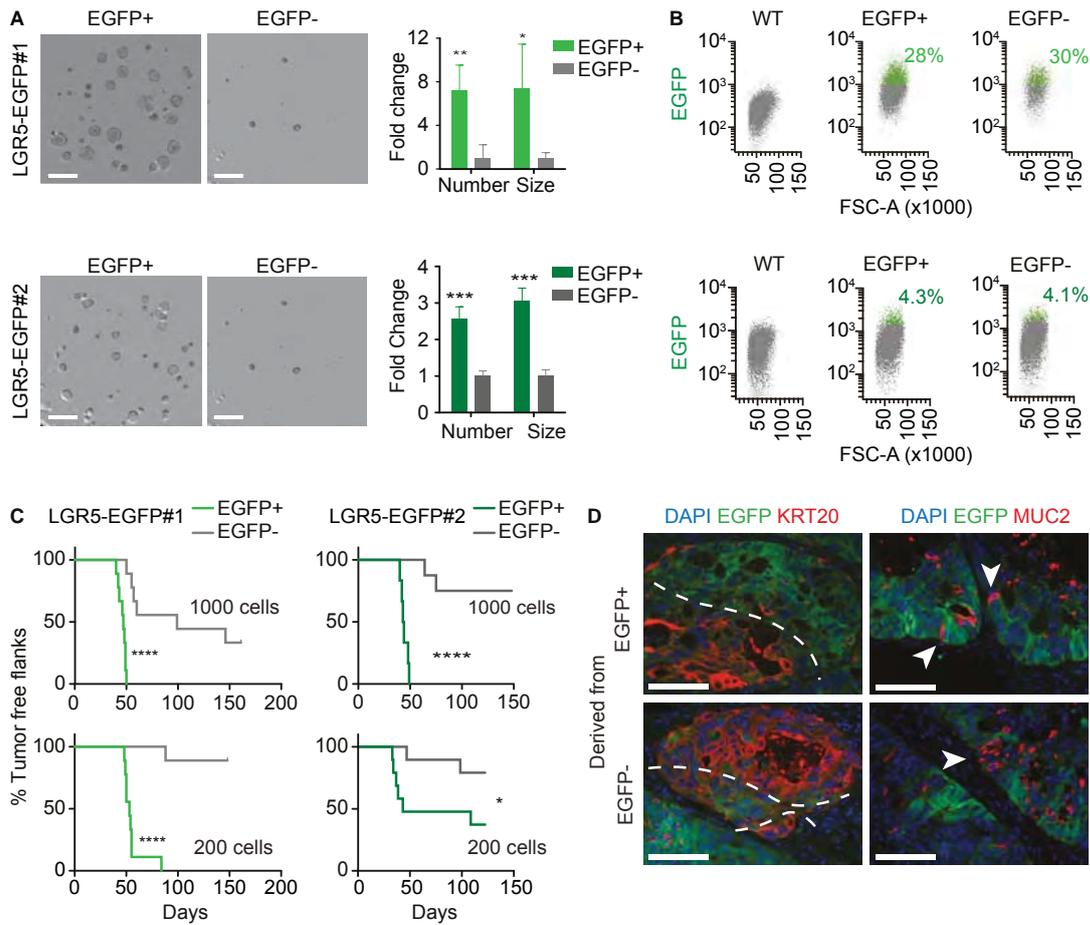
In summary, LGR5-EGFP cells represent a tumor subpopulation identified by expression of stem-like and proliferation genes and counter-enriched in differentiation markers both at the mRNA and protein level. This pattern is reminiscent to that of normal intestinal stem cells.

#### *LGR5+ cells exhibit increased tumor initiating potential*

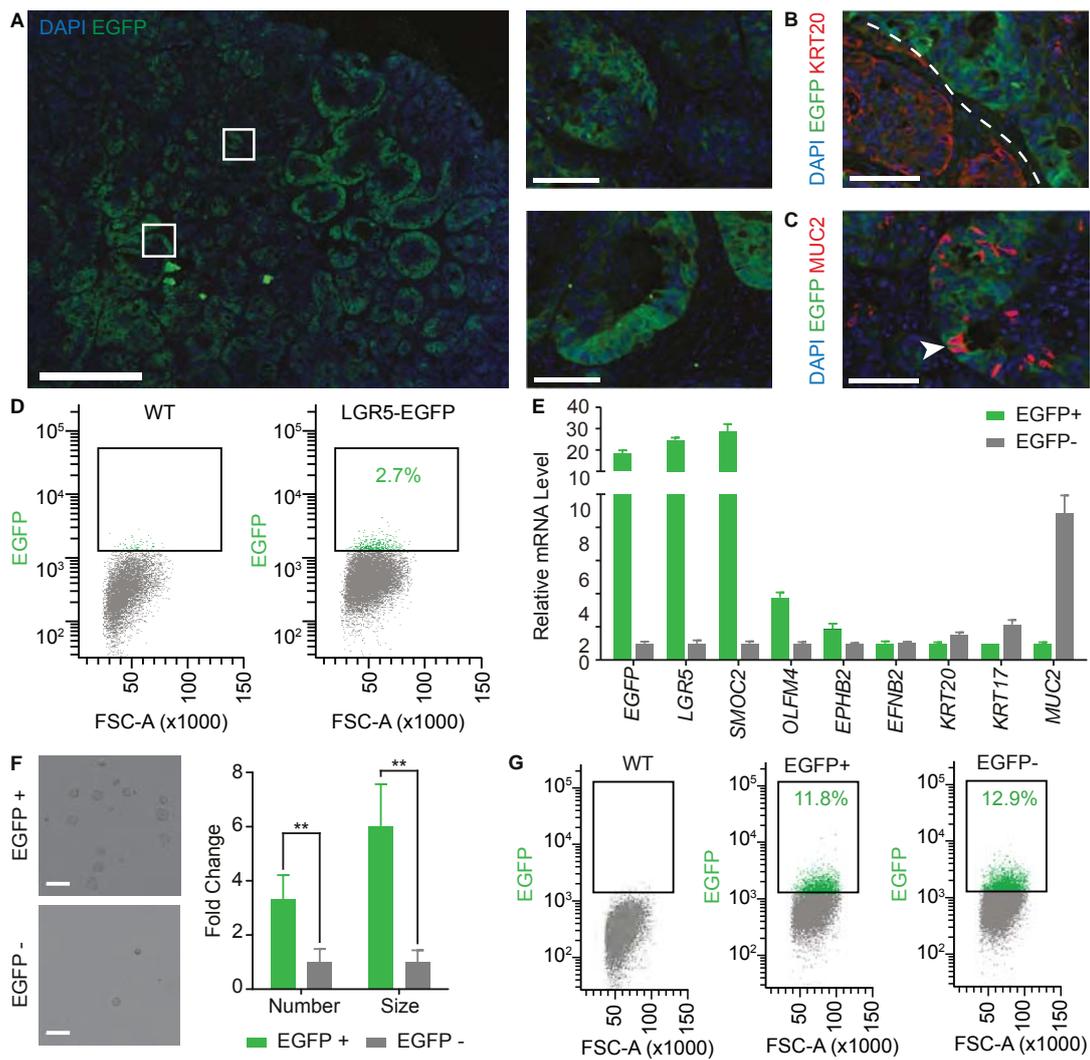
We next isolated the EGFP+ and EGFP- cell populations by FACS and plated them *in vitro* to assess their clonogenic potential. 15 days after plating, EGFP+ purified cells had efficiently formed organoids (50%) whereas negative cells were remarkably less proficient in organoid formation (10-20%) (Figure 3.6A). Of note, grown cultures expanded from both tumor cell populations displayed equivalent EGFP+ and EGFP- cells and therefore returned to the original heterogeneity found *in vitro* (Figure 3.6B). This finding suggests either the occurrence of plasticity, or a residual degree of contamination with EGFP+ cells in the EGFP- fraction. We also reinjected sorted cells into secondary immunodeficient hosts to analyze the tumor initiation potential of EGFP+ and EGFP- population. We injected subcutaneously 200 and 1000 cells per flank of each cell population. In both conditions, we corroborated that the LGR5-EGFP+ population was able to grow tumors with high efficiency whereas LGR5-EGFP- only gen-

erated tumor in a few flanks (Figure 3.6C). Histological analysis of second passage tumors confirmed that grown xenografts derived from both EGFP+ and EGFP- cell fractions recapitulated the morphology and organization of the parental xenografts, with EGFP+ glands devoid of KRT20 and MUC2 (Figure 3.6D). In addition, EGFP intensity and distribution in tumors reproduced that of the parental tumor (Figure A.3), demonstrating that LGR5+ cells can give rise to differentiated tumor cell populations. Similarly to *in vitro* cultures, these data suggest plasticity of tumor cell phenotypes, although again we cannot rule out the LGR5-EGFP- fraction was contaminated with some LGR5+ cells that were undetected by FACS.

PDO6 xenografts had slower growth kinetics than PDO7 and after inoculation in NSG mice required three to four months to generate large (150 mm<sup>3</sup>) tumors. We could nevertheless analyze PDX derived from LGR5-EGFP#18 PDO6 clone. They presented a similar glandular structure to that of PDO7, with a range of 60% EGFP+ cells of different intensities (Figure 3.7A). LGR5-EGFP cells did not express either the differentiation markers KRT20 and MUC2 (Figure 3.7B and C). By flow cytometry, we observed around 3% of EGFP+ cells that express elevated *LGR5* mRNA levels as well as several other ISC marker genes (Figure 3.7D and E). The EGFP-Low/Negative counterparts were enriched in differentiation-like genes (Figure 3.7E). EGFP+ cells purified from xenografts were more proficient in generating colonies than EGFP- cells (Figure 3.7F), but paralleling the results in PDO7 the organoids that grew from the negative fraction had recovered EGFP stem cell levels (Figure 3.7G). Due to its lower tumorigenic potential, we did not manage to assess tumor initiation potential with PDO6.



**Figure 3.6: Stem cell functional assays of the LGR5 population.** **A.** Representative images and quantification of organoid formation assays from cells purified of PDO7 LGR5-EGFP#1 and #2 subcutaneous xenografts (n=4 wells per condition). Data is represented as mean±s.d. Scale bars indicate 1mm. **B.** Representative flow cytometry analysis of 15-days grown organoids from EGFP+ and EGFP- sorted populations. **C.** Tumor initiation capacity (TIC) upon injection of 1000 or 200 cells in secondary immunosuppressed hosts. Graph shows Kaplan-Meier plots (n=9 xenografts). **D.** Dual immunofluorescence staining for KRT20/EGFP and MUC2/EGFP on paraffin sections of xenografts generated by EGFP+ and EGFP- sorted populations respectively. Dashed lines mark stem-like and differentiated-like compartments. White arrows point to secretory cells intermingled in the LGR5- compartment. Scale bars indicate 100µm. *Statistics: differences in organoid formation assay were assessed with Student's t-test and in TIC by log-rank (Mantel-Cox) test: \*P-value < 0.05, \*\*\*P-value < 0.005, \*\*\*\*P-value < 0.0001.* Adapted from Cortina, Turon et al., 2017.



**Figure 3.7: Characterization of LGR5-EGFP+ cells in PDO6.** **A.** Representative images of EGFP expression patterns analyzed by immunofluorescence on a section of PDO6-LGR5-EGFP#18-derived subcutaneous xenograft. White squares indicate position of the insets in the xenograft. Scale bars indicate 1mm for low magnification picture and 100 $\mu$ m for insets. **B.** Dual immunofluorescence demonstrating KRT20 and LGR5 complementary expression domains. Dashed lines mark expression domains of adjacent glands. Scale bars indicate 100 $\mu$ m. **C.** MUC2 and EGFP dual immunofluorescence on paraffin sections of clone #18. White arrows point to LGR5-/MUC2+ cells. Scale bars indicate 100 $\mu$ m. **D.** Flow cytometry profile of EGFP+ and EGFP- fractions from disaggregated xenografts generated by PDO6-LGR5-EGFP#18. Only EPCAM+DAPI- cells are analyzed. **E.** Relative mRNA expression level by qrtPCR of intestinal stem and differentiation genes in EGFP+ vs. EGFP- (EPCAM+) cells isolated from disaggregated xenografts. Data is represented as mean $\pm$ s.d. of three technical replicates. **F.** Representative pictures and quantifications of organoid formation assays generated by EGFP+ and EGFP- cells isolated from xenografts. Scale bars indicate 1mm. (n=4 wells per condition). Differences were assessed with Student's t-test: \*\* p-value<0.01. **G.** Representative flow cytometry analysis of 15 days grown organoids from EGFP+ and EGFP- sorted populations. Adapted from Cortina, Turon et al., 2017.

### 3.1.5 *LGR5+ cell population fuels tumor growth*

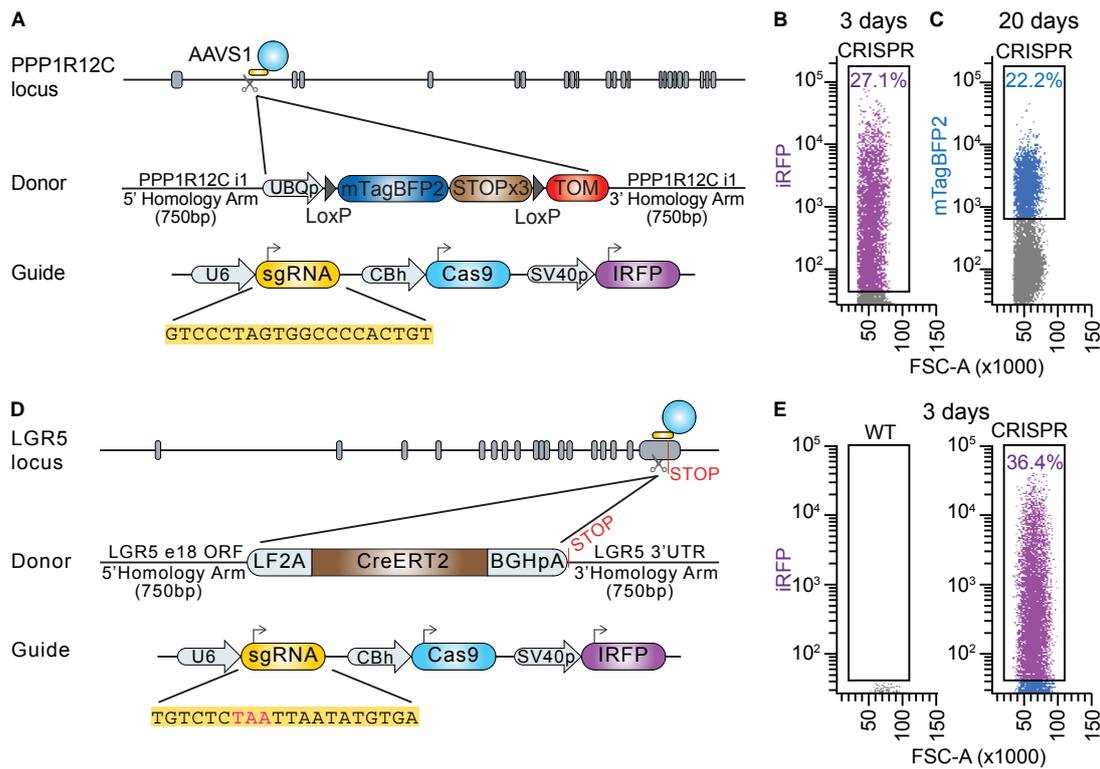
Organoid formation assay and tumor initiation potential assays have been extensively used as surrogates of stemness and are widely used to study the stem cell ability of tumor cell populations. However, these assays require disaggregation of the original tissue and therefore imply alterations in cell microenvironment as well as several other experimental caveats such as the fact that tumor cells remain isolated (Quintana et al., 2008). For a discussion on this topic also read Batlle and Clevers, 2017.

The golden assay to determine the presence of a stem cell-like population in tumors is clonal analysis through lineage tracing, based on labelling the progeny of a certain subpopulation and following its expansion over time.

#### *CRISPR/Cas9 design of the lineage tracing strategy*

We devised a construct to integrate at the genomic safe harbor AAVS1, within intron 1 of the PPP1R12C gene. The integration cassette contained mTagBFP2 (BFP) blue fluorescent protein and a BGH-polyA STOP tract flanked by loxP sites, followed by a TdTOMATO (TOM) reporter. In presence of CreERT2 enzyme, loxP sites would recombine, excising the BFP-STOP cassette and enabling TOM expression, which would be maintained through generations. In a similar fashion as for LGR5-EGFP editing (described in **section 3.1.1**), we designed a guide RNA targeting the desired AAVS1 locus and inserted it in the px330-Cas9-iRFP plasmid (Figure 3.8A). The nucleofection strategy was the same as before, selecting iRFP+ cells at short term and BFP+ cells at long term (Figure 3.8B and C). Clone derivation efficiency is summarized in Table A.1, and genotyping gels and southern blots are shown in Figure A.4. In this case, we chose a clone with a heterozygous integration to perform subsequent gene editing-modification (AAVS1<sup>BFPTOM#20</sup>).

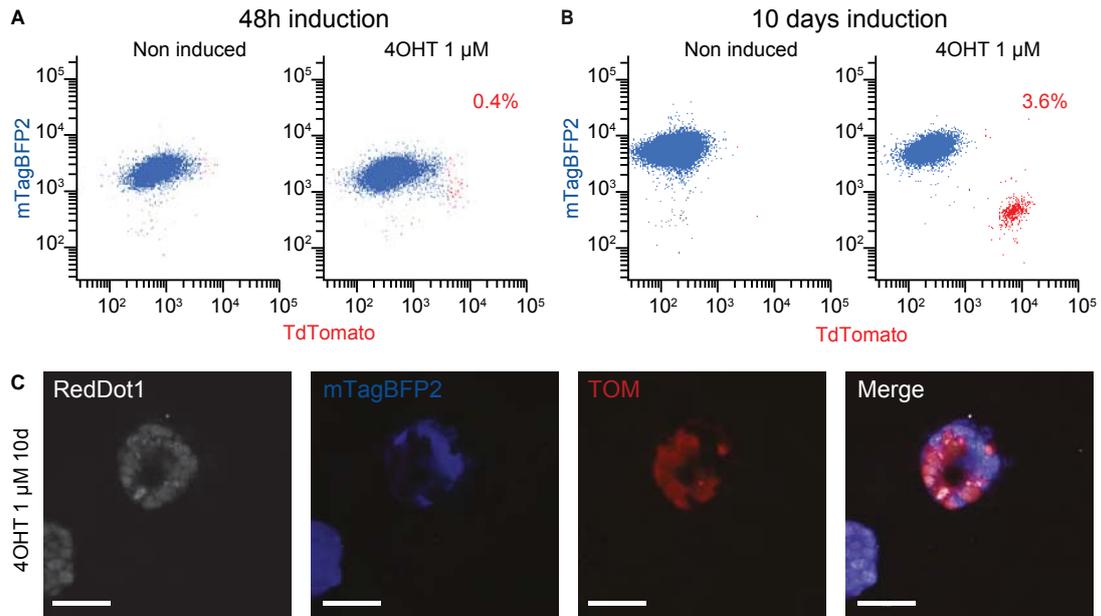
We next modified the LGR5-EGFP donor plasmid described in **section 3.1.1** by switching the EGFP cassette for a CreERT2 gene. We nucleofected AAVS1<sup>BFPTOM#20</sup> cells with the LGR5-guide and Cre-donor plasmid (Figure 3.8D) and selected iRFP+ cells two days after (Figure 3.8E). In these experiments, we directly plated sorted cells to grow single cell derived organoids, as the CreERT2 cassette did not have any fluorescence that enabled long-term selection of targeted cells. The efficiency of such unselected derivation was 16% (Table A.1). We selected clones LGR5-LF2A-CreERT2 (LGR5-Cre) #38 and #48 for further studies (Figure A.5).



**Figure 3.8: Generation of the *LGR5* lineage tracing cassette.** **A.** Scheme of the Lox-BFP-STOP-Lox-TOM AAVS1 knock-in donor and CRISPR/Cas9 gRNA sequences. **B.** FACS purification three days post-nucleofection of the PDO7 iRFP+ cells. **C.** MUC2 and EGFP dual immunofluorescence on paraffin sections of clone #18. White arrows point to *LGR5*<sup>-</sup>/MUC2<sup>+</sup> cells. Scale bars indicate 100 $\mu$ m. **D.** Flow cytometry separation of long-term BFP<sup>+</sup> cells (20 days post-nucleofection). **E.** Short term FACS-sorting of AAVS1<sup>BEPTOM#20</sup> *LGR5*-LF2A-CreERT2 cells for single-cell clone derivation. Adapted from Cortina, Turon et al., 2017.

#### Lineage tracing strategy marks *LGR5*<sup>+</sup> cells

To test the functionality of the lineage tracing system, we first induced CreERT2 activity *in vitro* by culturing modified PDO7 clones with 4-hydroxytamoxifen (4-OHTam) 1 $\mu$ M. A small portion (0.4%) of BFP<sup>+</sup> cells converted to TOM<sup>+</sup> in 48h (Figure 3.9A). This population grew up to 3.6% 10 days post induction (Figure 3.9B and C). The efficiency of CreERT2 recombination was low (1 in every 50 cells *in vitro*), yet this suited well our interest since clonal analysis required tracing induction at single cell level to follow the progeny of individual cells. Moreover, 10 days post induction we did not observe any double BFP<sup>+</sup>TOM<sup>+</sup> cell, supporting a complete recombination of the tracing cassette.

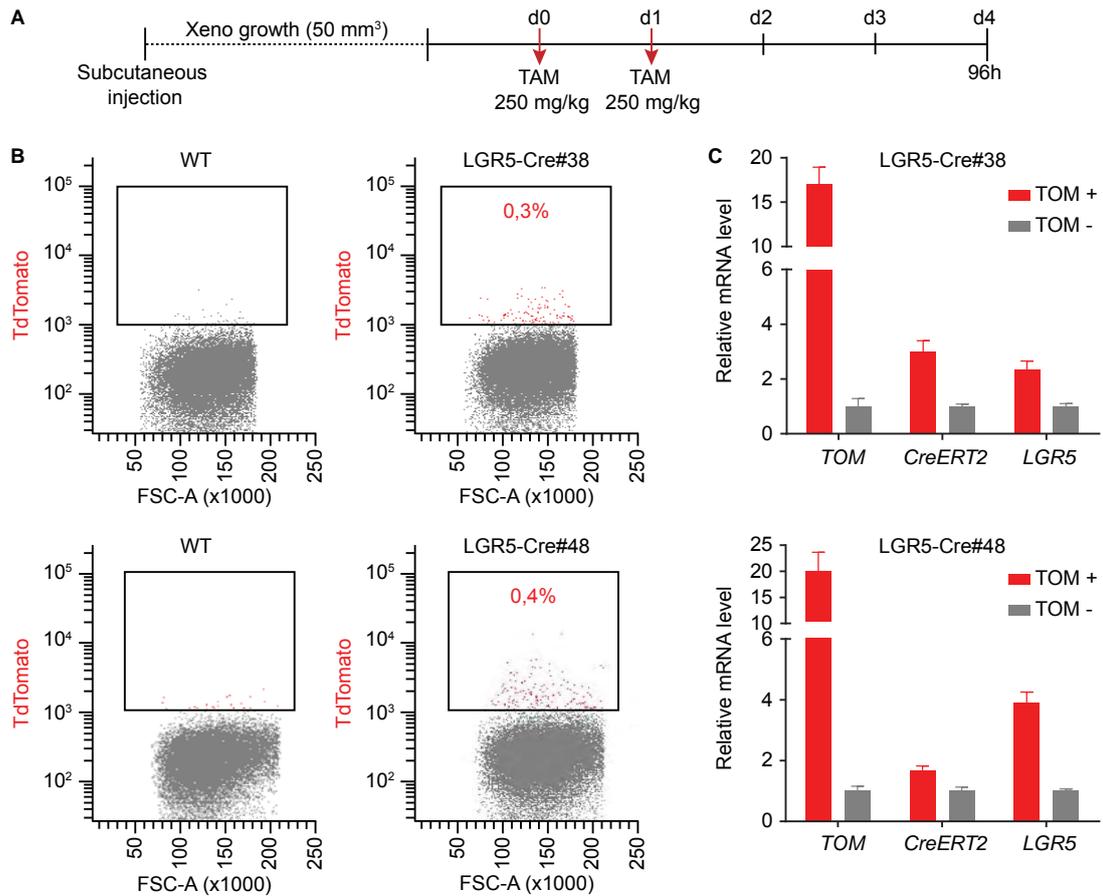


**Figure 3.9: Functioning of the tracing system *in vitro*.** **A.** FACS analysis of BFP and TOM populations in AAVS1<sup>BFP</sup>TOM#20 LGR5-Cre#48 *in vitro* grown organoids 48h after 4-OHT pulse. **B.** FACS profiling of the same clone 8 days later, showing the appearance of the BFP-TOM+ targeted population. **C.** Representative confocal imaging of double knock-in organoids induced with 4-OHTam for 10 days *in vitro*. Scale bars indicate 50 μm. Adapted from Cortina, Turon et al., 2017.

We next performed *in vivo* clonal analysis by injecting subcutaneously in Beige/SCID mice the dual targeted PDO7 (150000 cells per flank). When tumors reached a minimum size of 50 mm<sup>3</sup>, we treated mice with two intraperitoneal shots of tamoxifen (Tam) 250mg/kg, and sacrificed them 96h after the first shot (Figure 3.10A). By FACS of dissociated xenografts, we observed the appearance of a TOM+ cell population in treated tumors (Figure 3.10B), which retained *LGR5* expression as shown by RT-qPCR (Figure 3.10C). This experiment demonstrated for the first time that is possible to perform lineage tracing experiments in human colorectal tumors. Comparing the frequencies of LGR5-EGFP+ cells in PDXs and the induced numbers of TOM+ cell population at 96 hours, we estimated that roughly one in 10-20 LGR5+ cells recombined the tracing cassette in these experiments.

#### *LGR5+ cells give rise to long-lasting clones*

PDO7 subcutaneous xenografts in Beige/SCID mice grow at a fast rate. Therefore, we devised the following tracing strategy to study LGR5 progeny in PDX: after cell inoculation into mice, we waited until tumors reached a size of 50 mm<sup>3</sup> and then treated the animals with two intraperitoneal shots of 250 mg/kg Tam. Tumors were resected from the animals at 96 hours, 14 days and 28 days post induction. After one month, the tumor burden was excessive and animals had to be sacrificed, so pieces of the xenografts (*trochers*) were transplanted under the skin of second hosts, allowing the tumors to grow for one more month (i.e up to 56 days) (Figure 3.11A).

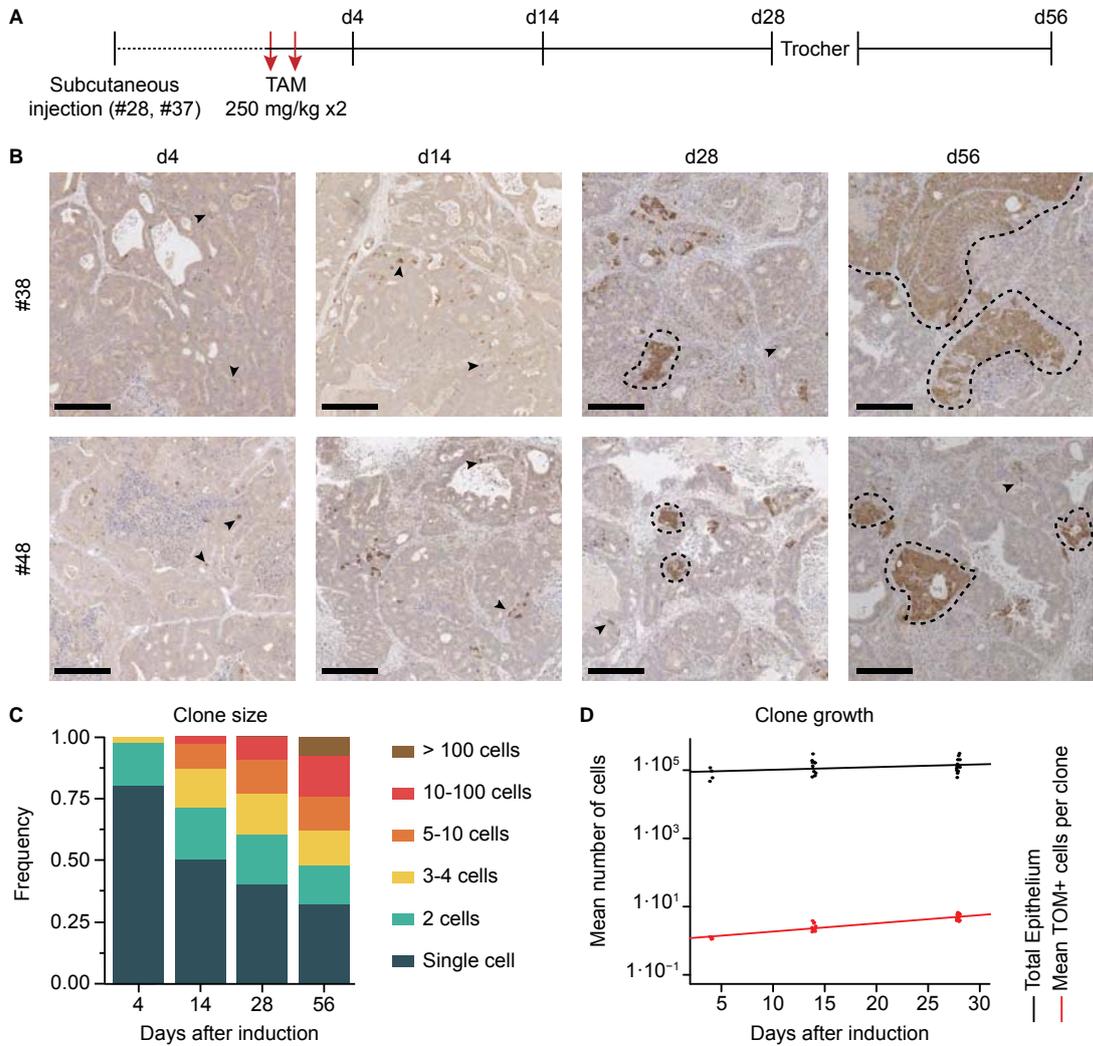


**Figure 3.10: TOM reports LGR5 levels *in vivo*.** **A.** Scheme of the tamoxifen administration dosage in PDO7 subcutaneous xenografts of 50 mm<sup>3</sup>. **B.** Flow cytometry profiling of the TOM-induced population in EPCAM<sup>+</sup>/DAPI<sup>-</sup> cells from disaggregated xenografts 96 hours after Tam induction. Representative examples for both LGR5-Cre#38 and #48 clones. **C.** Relative mRNA expression of sorted TOM<sup>+</sup> and TOM<sup>-</sup> populations at short time point (96h) post induction from the xenografts. Data is represented as mean±s.d. of three technical replicates. Adapted from Cortina, Turon et al., 2017

We analyzed the results by TOM immunohistochemistry in tumor sections at the indicated time points (Figure 3.11B). At 96h (4d) all TOM<sup>+</sup> cells were visualized as single cell entities, meaning that each recombined LGR5 cell will give rise to a unique single-cell derived clone, a condition that is essential to reach valid conclusions in clonal tracing analysis. We quantified vast number of clones in a substantial number of xenografts (detailed numbers of quantifications can be found in **section 6.8.3**). Clones gradually expanded in size with time while the pool of single cell positive cells decreased. Nonetheless, we noticed that a proportion of TOM<sup>+</sup> cells remained as single cells over long periods suggesting that a proportion of the LGR5<sup>+</sup> cancer stem cell pool is in a quiescent state. (Figure 3.11C). Alternatively, this single cell clones may simply represent the borders of large clones that are visualized as individual entities due to the sectioning plane.

By comparing clonal expansion with growth rate of the epithelial compartment (measured by Hematoxylin-Eosin (HE) in histological serial sections) (Figure 3.11D) we

concluded that size of the clones derived from LGR5+ cells scale proportionally to the whole tumor epithelium. Taking into account that many LGR5+ cells were not labelled by TOM, this scaling pattern supports the notion that tumor growth is due to LGR5+ cell growth.



**Figure 3.11: LGR5 progeny expands in vivo in subcutaneous xenografts.** **A.** Schematic representation of the LGR5 lineage tracing experimental setup in subcutaneous xenografts. **B.** Representative IHC using anti-Tomato antibodies on paraffin sections from samples resected at the four indicated time points after tamoxifen treatment. Arrowheads point to single and two cell clones. Dashed lines delimit large clones. Scale bars indicate 250 $\mu$ m. **C.** Clone size frequency per time point according to total cell numbers. Number of clones quantified was 2330 for day 4, 9554 for day 14, 14897 for day 28, and 2476 for day 56. **D.** Correlation of number of epithelial cells per xenograft and number of cells per clone over time (day 4: n = 4; day 14: n = 10; day 28: n = 14 and day 56: n = 24 quantified xenografts, respectively). Adapted from Cortina, Turon et al., 2017.

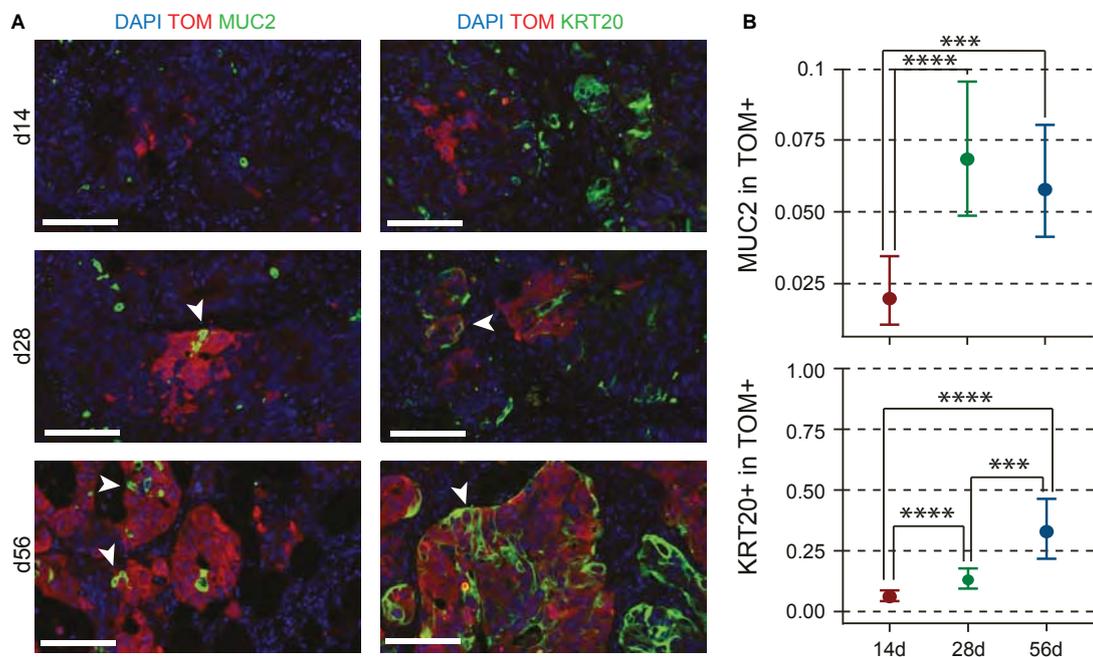
A major limitation in the lineage tracing analysis we have performed is that it is a 2D-based study. Extracting growth kinetics data from a 2D visualization (i.e. histological section) of clones that expand in 3 dimensions in tumor glands with convoluted

patterns imposes a bias in the results and subsequent interpretation. For instance, a small clone visualized in 2D may simply represent a large 3D clone sectioned at the border. To overcome this problem, we attempted to visualize clones in 3D by serial section reconstruction (Movie 1). The reconstruction showed both small (1-4 cells) and large (>4 cells) clones (Figure A.6). This proves that single-cell detected clones are not a result of analysis bias and that some LGR5+ cells remain in single cell status for over one month.

However, this methodology proved to be highly inefficient and costly, and it only permitted reconstruction of small tumor areas. For this reason, we next focused on imaging whole xenografts. We set up a protocol (described in **section 6.9.5**) that allows thick section imaging (up to 400µm in width) with a Selective Plane Illumination Microscopy (SPIM) (Movie 2). This will be a powerful tool for future lineage tracing experiments in the lab. At the time of submitting the thesis, we had not yet implemented this technology in a time-efficient manner.

*Differentiation kinetics are slower in tumors than in healthy epithelia*

Interestingly, we quantified the differentiation rate within LGR5+ derived clones, and observed that differentiation was a progressive process. There were few KRT20+ or MUC2+ cells two weeks post induction but after this time point, the proportion of differentiated cells inside clones increased exponentially (Figure 3.12). This finding implies that the differentiation kinetics was somewhat slower than in homeostatic intestinal tissue, where a cell will migrate upwards toward the villus, differentiate and die in 4-5 days (Van der Flier and Clevers, 2009).

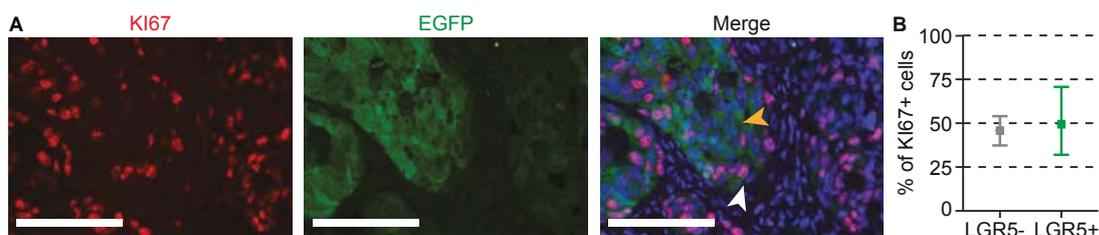


**Figure 3.12: LGR5 progeny gives rise to differentiated cells with time.**

**Figure 3.12: LGR5 progeny gives rise to differentiated cells with time.** **A.** . Dual immunofluorescence stainings showing TOM and differentiation markers MUC2 and KRT20 expression domains at different lineage tracing time points. White arrowheads indicate double-positive cells. Scale bars indicate 100 $\mu$ m. **B.** Quantification of the number of MUC2+ and KRT20+ cells within TOM+ clones at each time point. Data is represented as the 95% confidence intervals of the measurements. Number of clones assessed was 872 (4 days), 372 (day 14), and 69 (day 28) for KRT20 and 387 (day 4), 611 (day 14), and 130 (day 28) for MUC2. *Statistics: P-value was calculated using a generalized linear model with binomial response. \*\*\*P-value < 0.005, \*\*\*\*P-value < 0.0001.* Adapted from Cortina, Turon et al., 2017.

## 3.2 LABELLING PROLIFERATIVE CELLS IN HUMAN CRCS

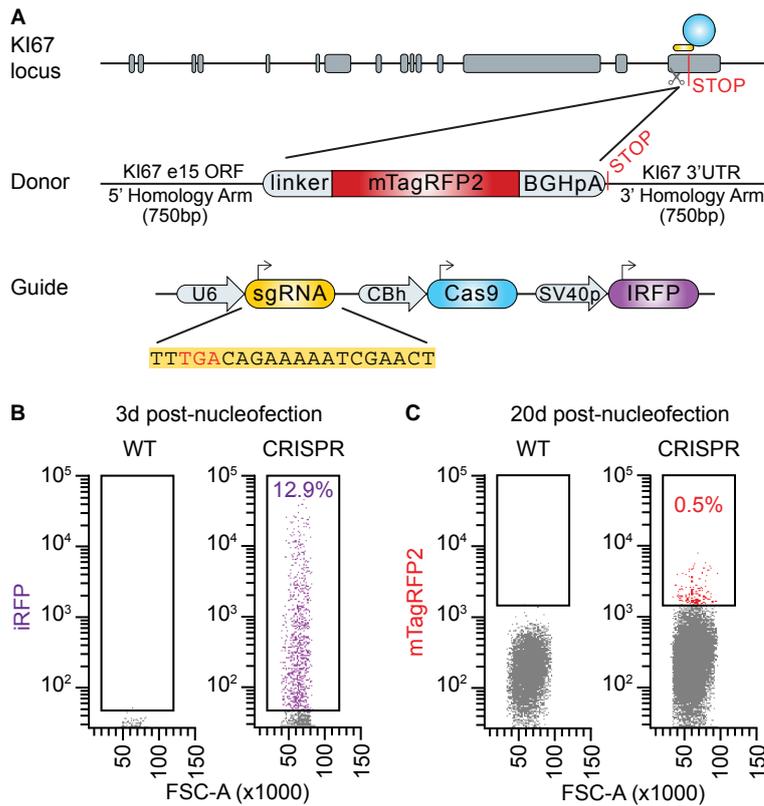
As discussed above, lineage tracing experiments suggest that a subpopulation of LGR5+ cells did not produce progeny over extended periods of time. In addition, KI67 immunostaining in LGR5-EGFP-derived PDX showed that approximately half of the LGR5-EGFP+ cells were KI67+ (Figure 3.13). This suggested the presence of a quiescent LGR5+ cancer stem cell population in human tumors. To investigate this possibility we engineered PDOs that carry a fluorescent label fused to the proliferation marker KI67 as well as the LGR5-EGFP reporter that we described in **section 3.1.1**.



**Figure 3.13: Half of PDO7 LGR5+ cells are KI67+.** **A.** Representative immunofluorescence of PDO7 LGR5-EGFP#1 stained with KI67 antibody. White arrowheads point to double-positive cells; yellow arrowheads point to LGR5+/KI67- cells. Scale bars indicate 100 $\mu$ m. **B.** Quantification of KI67+ cells within the LGR5+ and LGR5- compartments (n = 2749 LGR5+ cells, 1798 LGR5- cells assessed). Data is represented as the 95% confidence intervals of the measurements. *Statistics: P-value was calculated using Student's T-test. Ns: non-significant*

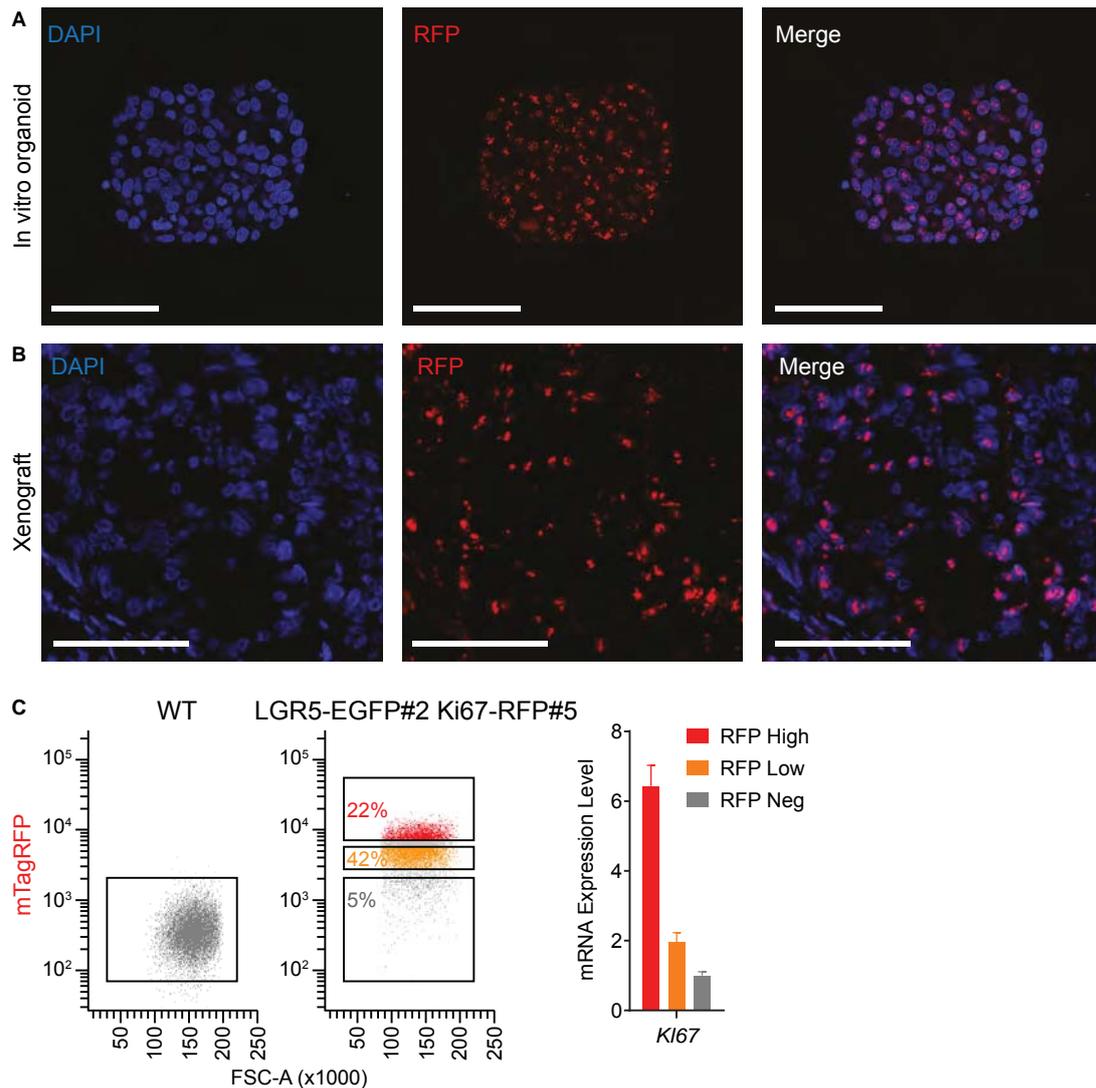
### 3.2.1 Labelling endogenous KI67 protein by CRISPR/Cas9

We designed a KI67-mTagRFP2 (RFP) fusion protein CRISPR/Cas9 construct. In brief, we selected a guide RNA that induced a double strand break before the KI67 stop codon, and designed a donor plasmid with adequate homology arms and a RFP coding sequence in frame with KI67 gene (Figure 3.14A). We followed the same nucleofection and sorting strategy as previously (please see **section 3.1.1**) but using the already targeted PDO7 LGR5-EGFP #1 and #2 clones. Figure Figure 3.14B and C exemplify FACS selection of LGR5-EGFP#2 KI67-RFP+ cells, and integration PCRs are shown in Figure A.7. Total clonal derivation efficiency for both LGR5-EGFP clones is detailed in Table A.1.



**Figure 3.14: PDO7 KI67-RFP fusion protein knock-in generation.** **A.** Design of KI67-TagRFP2 (RFP) donor and CRISPR/Cas9 gRNA vectors. RFP fluorescent protein was inserted as C-terminal fusion with endogenous KI67 protein. **B.** FACS purification at day 3 indicating cells that have incorporated the guide plasmid (iRFP+). **C.** Flow cytometry sorting of RFP+ cells 20 days post-nucleofection. Adapted from Cortina, Turon et al., 2017

Prior to performing any analysis, we confirmed by microscopy the nuclear localization of RFP protein in highly intense nucleolar dots, resembling the subcellular localization of endogenous KI67 protein (Verheijen et al., 1989) (Figure 3.15A). This pattern was also observed in subcutaneous xenografts derived from the same clone (Figure 3.15B). We did not perform southern blot and instead used co-localization as an indication of the absence of expressed off-target integrations. Moreover, RFP-High, Low and Negative populations reported *KI67* mRNA levels *in vitro* (Figure 3.15C). We used this particular clone (LGR5-EGFP#2 KI67-RFP#5) for subsequent experiments.

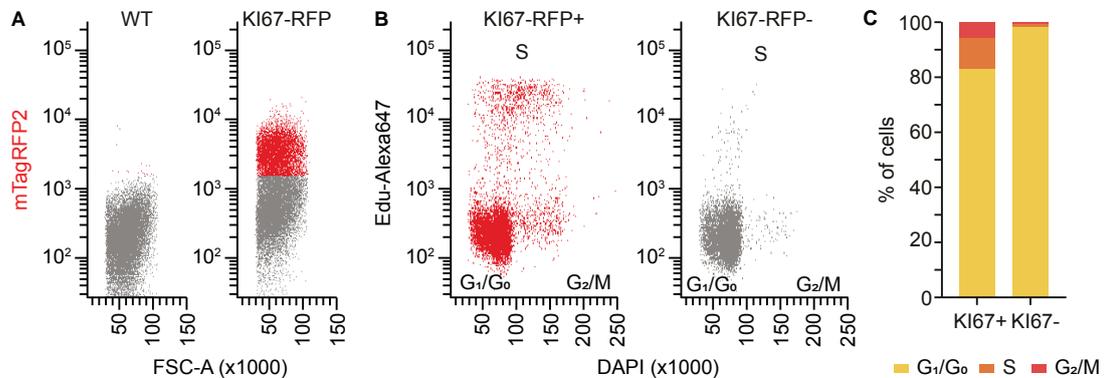


**Figure 3.15: RFP fluorescent marker correlates with KI67 protein.** **A.** Representative confocal images of PDO7-KI67RFP#5 in *in vitro* grown organoids, showing nuclear localization of the RFP reporter. Scale bars indicate 100µm. **B.** Immunofluorescence on paraffin sections of PDO7-KI67RFP#5 subcutaneous xenograft. RFP co-localizes with DAPI nuclear staining. Scale bars indicate 25µm. **C.** Flow cytometry analysis and sorting of 1-week *in vitro* grown #5 according to RFP levels. *KI67* mRNA is expressed mostly in the RFP high fraction as analyzed by RT-qPCR. Data is represented as mean±s.d. of three technical replicates. Adapted from Cortina, Turon et al., 2017

#### *KI67-RFP fusion protein marks actively cycling cells*

To assess whether or not KI67-RFP was a marker for proliferative cells, we injected subcutaneously the modified PDO7 LGR5-EGFP#2 KI67-RFP#5 in Beige/SCID mice and prior to sacrifice animals were treated with the thymidine analog ethynyl deoxyuridine (EdU). EdU is incorporated into the DNA during replication, and is widely used to track proliferating cells in biological systems. Nuclear EdU incorporation can be easily detected by click-it chemistry (Salic and Mitchison, 2008). We FACS-sorted RFP-Positive and -Negative cells from disaggregated xenografts (Figure 3.16A) and we next analyzed the proportion of tumor cells found in each cell

cycle phase by total DAPI content and nuclear EdU incorporation. Flow cytometry quantification showed that indeed RFP+ cell population contained many more cells in the S phase of the cell cycle than the RFP- fraction (Figure 3.16B and C).



**Figure 3.16: KI67-RFP+ cells actively incorporate EdU. A.** Flow cytometry separation of RFP+ and RFP- cell populations of PDO7 LGR5-EGFP#2/KI67-RFP#5 from an EPCAM+/DAPI- disaggregated xenograft. **B.** Cell cycle analysis of the RFP+ and RFP- cells previously sorted. X-axis shows DNA content and Y-axis EdU incorporation (detected by Alexa-647). **C.** Quantification of the frequencies of KI67-RFP+ versus KI67-RFP- cells found in each cell cycle phase. KI67-RFP-Neg fraction is almost devoid of cells in S and G<sub>2</sub>/M phase. Adapted from Cortina, Turon et al., 2017

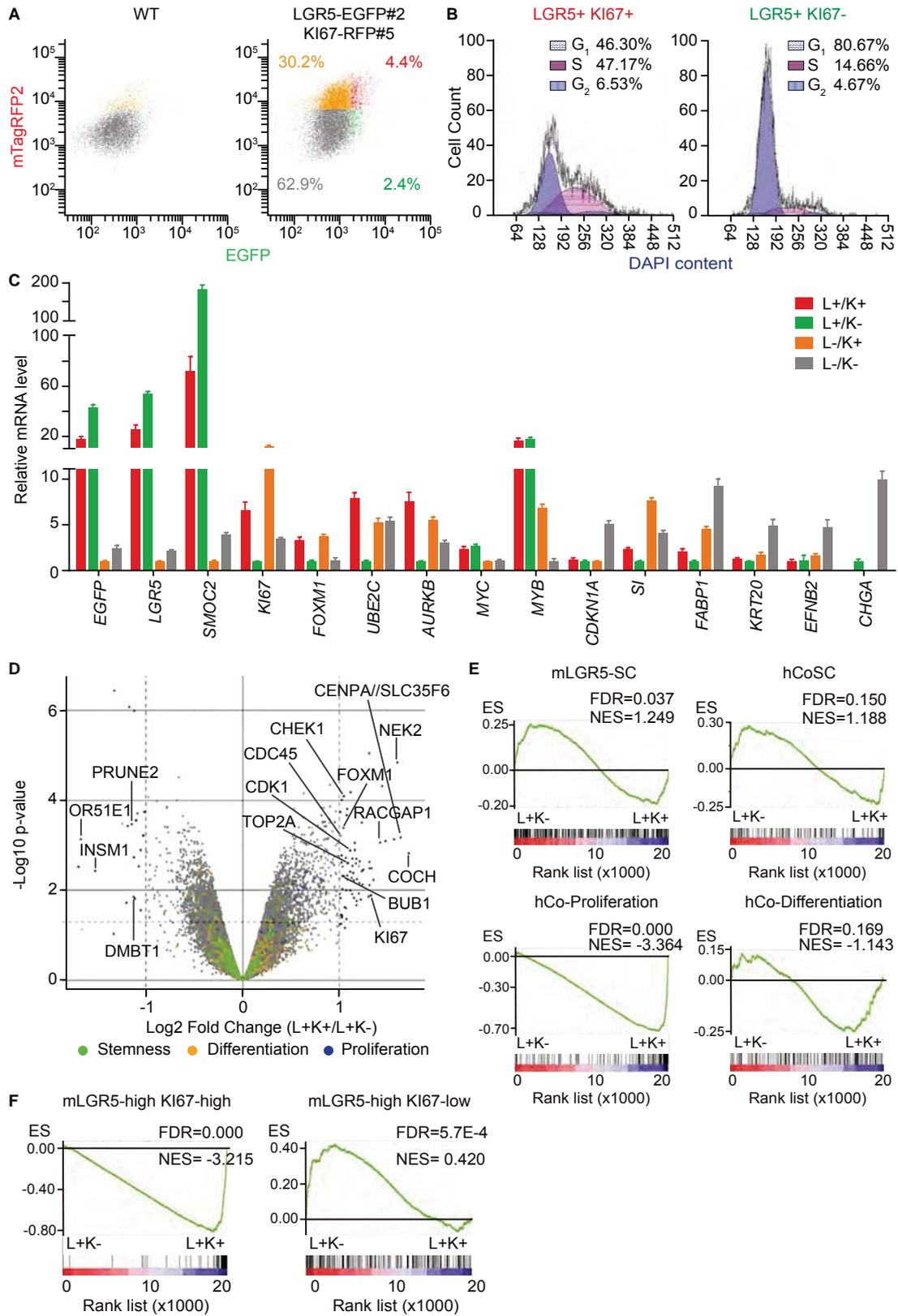
### 3.2.2 One third of the LGR5+ cells are in a slow cycling state

In LGR5/KI67 double labelled PDX, around 30% of LGR5-EGFP+ cells were negative for KI67 (Figure 3.17A). Cell cycle analysis of FACS-purified EGFP+RFP+ and EGFP+RFP- cells indicated that KI67-Negative cells were mostly arrested in G<sub>0</sub>/G<sub>1</sub> phase (Figure 3.17B). We also purified the four populations identified by FACS in tumors (L+K+, L+K-, L-K+, L-K-, where L=LGR5 and K=KI67) and performed gene expression analysis. RT-qPCR expression profile reinforced the cell cycle results, as K+ fractions exhibited high levels of genes involved in proliferation, whereas both L+ populations were clearly stem-like as shown by expression of *LGR5* and *SMOC2* (Figure 3.17C). Of note, L-K- population had higher expression of differentiated genes (*KRT20*, *EFNB2*, *CHGA*) than L-K+ cells, suggesting that the formed could correspond to a mature terminally differentiated population. In contrast, L-K+ cells appeared to be a proliferating transient amplifying population that express markers of adsorptive cells such as *FABP1* and *SI*, among others. Interestingly, L+K- cells retained higher expression of stem cell markers than L+K+ cells, perhaps suggesting that slow cycling stem cells could have enhanced stemness potential than actively cycling LGR5+ stem cells.

By gene expression microarray analysis of L+K+ and L+K- populations we obtained signatures that allowed a more in depth analysis of the expression profile of these two populations. L+K+ fraction was enriched in many proliferation markers, as *UBE2C*, *FOXM1*, *CDC45* and *CDK1* (Figure 3.17D). GSEA analysis using signatures of ISCs, proliferating and differentiated cells confirmed the above observations (Figure 3.17E).

These results point toward the idea that the LGR5+ stem cells in tumors are in two different states -quiescence or proliferation-. In order to validate our findings, we also took advantage of gene expression signatures generated in mouse adenomas for *LGR5* and *KI67* populations (Basak et al., 2014) and found that the signatures obtained perfectly correlated with ours (Figure 3.17F).

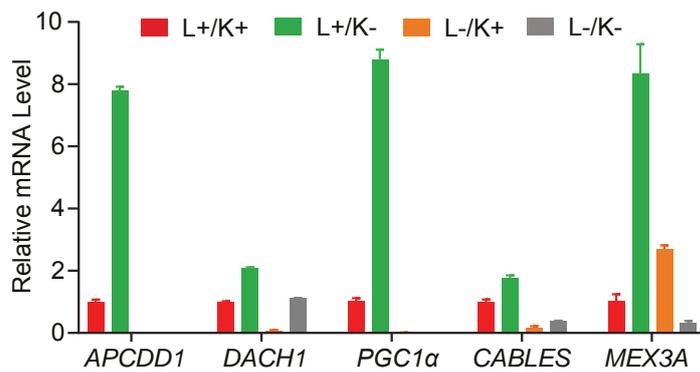
**Figure 3.17: Dual *LGR5* and *KI67* knock-in PDO7 enables separation of quiescent and cycling LGR5+ CRC cells.** **A.** Representative FACS profiles from PDO7 LGR5-EGFP/*KI67*-RFP disaggregated xenografts. Only EPCAM+/DAPI- cells are shown. The four represented populations are: L-K- (gray), L-K+ (orange), L+K+ (red) and L+K- (green). K indicates *KI67*, and L indicates *LGR5*. **B.** Cell cycle analysis LGR5-EGFP+ and *KI67*-RFP+ or *KI67*-RFP- sorted populations. 5363 and 5398 cells were analyzed in each case. **C.** RT-qPCR mRNA expression analysis of proliferation, stem, and differentiation marker genes in the cell populations defined by EGFP/RFP levels. Data is represented as mean±s.d. of three technical replicates. **D.** Volcano plot representing gene expression profile of L+K+ versus L+K- purified populations from #5. Green dots indicate genes belonging to the human colon stem cell signature, orange dots represent genes belonging to the differentiated cell signature, and blue dots depict genes of the crypt proliferative progenitor signature. Well-described genes involved in proliferation are indicated. P-values and fold changes were computed by fitting a linear model with the R package limma. **E.** GSEA comparing LGR5-EGFP+ cells positive or negative for *KI67*-RFP with the signatures used in Fig 3.5. **F.** GSEA with signatures derived from mouse crypt LGR5-High/*KI67*-High or LGR5-High/*KI67*-Low (Basak et al., 2014). Note that the only signatures that are differentially expressed between the two populations correspond to proliferative cells derived either from human crypts or from *KI67*-RFP mice. Adapted from Cortina, Turon et al., 2017



**Figure 3.17: Dual *LGR5* and *KI67* knock-in PDO7 enables separation of quiescent and cycling *LGR5*+ CRC cells.**

### Identification of specific markers for the slow-cycling LGR5+ stem cell pool in human CRCs

By combining two markers (*LGR5* and *KI67*) we demonstrated the existence of a slow-cycling population with stem cell-like gene expression program in human tumors. This was a surprising finding given that in homeostatic conditions ISCs are mostly active and seldom found in quiescent status (Buczacki et al., 2013, Barriga et al., 2017). It is tempting to speculate that the slow-cycling LGR5+ pool in human CRCs might be responsible for chemotherapy resistance and even metastatic dormancy (Suzuki et al., 2006, Kreso et al., 2013). However, few studies have described specific markers for this population (Puig et al., 2018, Buczacki et al., 2018). From the L+K+ and L+K- gene expression profiles generated in the previous experiment (Figure 3.17), we selected a few markers enriched in the L+K- population and validated their expression in the four populations isolated from subcutaneous xenografts by RT-qPCR. Indeed some genes we identify several genes, including *APCDD1* and *MEX3A*, that specifically labelled the quiescent stem cell state in human CRC (Figure 3.18). This finding gave rise to a new research project pursued by Drs. Carme Cortina and Felipe Slebe in our laboratory.



**Figure 3.18: Specific markers of the slow cycling LGR5+ cells.** RT-qPCR analysis of mRNA relative expression levels of putative markers of quiescent stem cells in the mRNA previously generated from L+K+, L+K-, L-K+ and L-K- populations disaggregated from PDO7 LGR5-EGFP KI67-RFP#5 subcutaneous xenografts. Data is represented as mean±s.d. of three technical replicates.

In summary, we have successfully edited the genome of CRC PDOs using reporter and lineage tracing cassettes. This approach enabled the visualization and tracking of a putative CSC population identified by *LGR5* gene expression. LGR5+ cells retained a stem cell program reminiscent of the LGR5+ crypt base columnar cells in healthy epithelia, and exhibited enhanced tumor initiation capacities. Lineage tracing from the *LGR5* locus indicated that its progeny fueled tumor growth. Moreover, tracing experiments also revealed the existence of a quiescent stem cell pool in CRC. We have identified marker genes of this population that will be analyzed in follow-up studies.

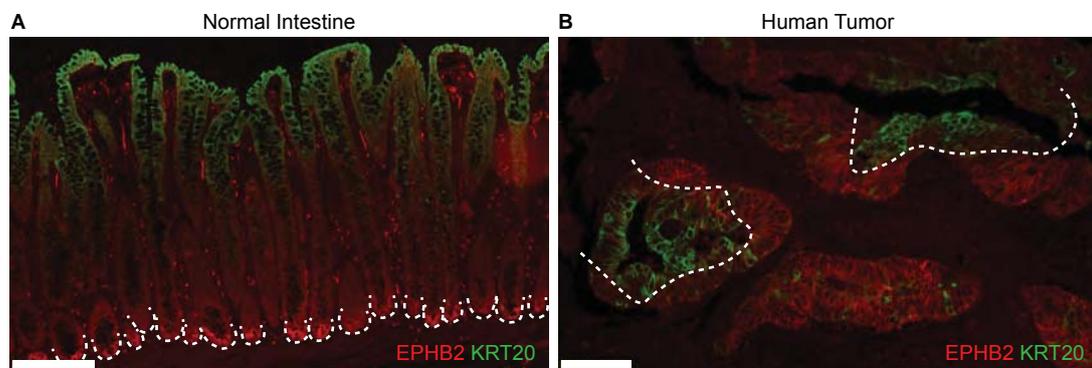
## Chapter 2: Tracing the differentiated compartment of CRC

### 3.3 DEFINING THE DIFFERENTIATED COMPARTMENT OF CRC TUMORS

In the previous chapter we characterized the contribution of LGR5+ cells to tumor growth. The features of this population fit in well with that of a cancer stem cell population (Schepers et al., 2012). To complement this study, we attempted to characterize differentiated-like populations in tumors. At the moment of starting these experiments the accepted view was that differentiated tumor cells were not tumorigenic and did not contribute to tumor growth.

#### 3.3.1 *KRT20 marks differentiated cells in human tumors*

*KRT20* is widely described in the literature as an intestinal pan-differentiation marker (Moll et al., 1990, Dalerba et al., 2011). Moreover, previous data from the lab showed an anti-correlation between stem cell marker gene expression and *KRT20* protein expression, both in normal (Figure 3.19A) and tumor tissues (Figure 3.19B)(Mérlos-Suárez et al., 2011). *KRT20* protein exhibited a complementary expression domain to that of LGR5-EGFP cells, as shown before (Figure 3.4), and *KRT20* mRNA was 2.5-fold downregulated in the LGR5-EGFP+ population (Figure 3.5). For these reasons, we chose *KRT20* as a marker for differentiated tumor cells.

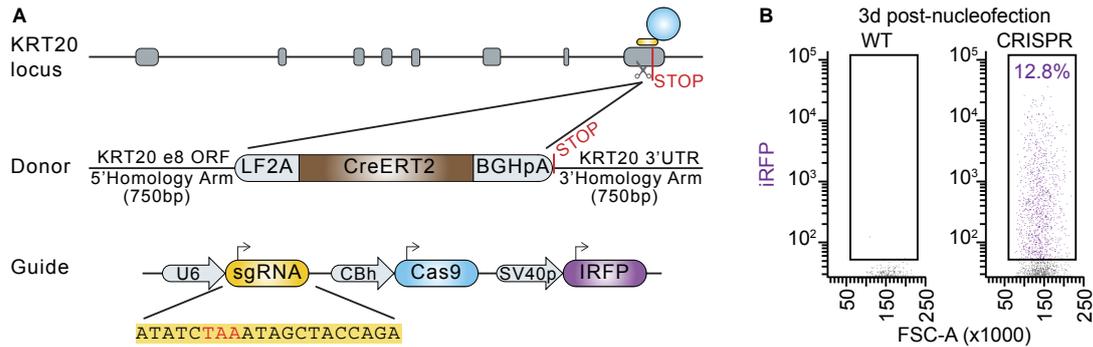


**Figure 3.19: EPHB2 and KRT20 show complementary patterns in healthy intestine and tumor tissue.** **A.** EPHB2 and KRT20 dual immunofluorescence staining in sections from normal human intestine paraffin blocks. Dashed lines indicate crypts, highly positive for EPHB2 and devoid of KRT20. Scale bar indicates 250 $\mu$ m. **B.** Same staining sections from patient tumors. Tumor tissue is composed of glands that resemble healthy crypts, with stem-like (EPHB2) and differentiated-like (KRT20) areas. Dashed lines separate both compartments. Scale bars indicate 100 $\mu$ m.

#### *Generation of PDOs carrying KRT20 lineage tracing cassettes*

We designed the CRISPR/Cas9 *KRT20* targeting vectors by simply modifying the homology arms from the LGR5-LF2A-CreERT2 donor for the sequences correspond-

ing to the *KRT20* gene. We selected a guide that targeted the 3'UTR of *KRT20* (Figure 3.20A). As described for the *LGR5* locus, we nucleofected PDO7 AAVS1<sup>BFPTOM#20</sup> clone and directly plated in single cell format iRFP+ cells at day 3 post nucleofection (Figure 3.20B). Efficiencies of organoid targeting are detailed in Table A.1, and integration PCRs are shown in Figure A.8A. Selected clones were screened by Southern blot for putative off-targets, and we finally chose #7 and #10 for further experiments (Figure A.8B). This gene edition will be referred as KRT20-Cre (KRT20-LF2A-CreERT2).

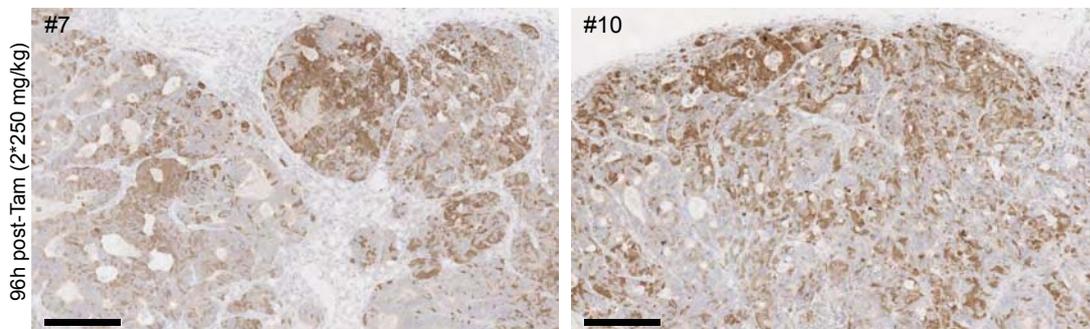


**Figure 3.20: PDO7 AAVS1<sup>BFPTOM#20</sup> KRT20-LF2A-CreERT2 knock-in generation.** **A.** Design of KRT20-LF2A-CreERT2 donor and CRISPR/Cas9 gRNA vectors. **B.** Day 3 short term sorting of the cells that have incorporated the guide plasmid (iRFP+).

Given that we had already validated the lineage tracing strategy *in vitro* using *LGR5* clones (see **section 3.1.5**), we directly performed *in vivo* experiments by injecting KRT20-Cre clones #7 and #10 subcutaneously into Beige/SCID animals.

#### *KRT20-driven CreERT2 is more efficient than LGR5-driven Cre*

We first tested *KRT20* lineage tracing with previously established experimental settings (i.e **section 3.1.5**: 2 intraperitoneal Tam shots of 250 mg/kg), and found increased efficiency of recombination. It was evident that vast areas instead of single cells were labelled 96h after Tam administration (Figure 3.21). This was probably due to the fact that *KRT20* mRNA expression levels are higher than *LGR5* levels (i.e in RT-qPCR analysis, Ct amplification values of 20 (*KRT20*) vs 25 (*LGR5*) in human PDO7 EPCAM+ cells).



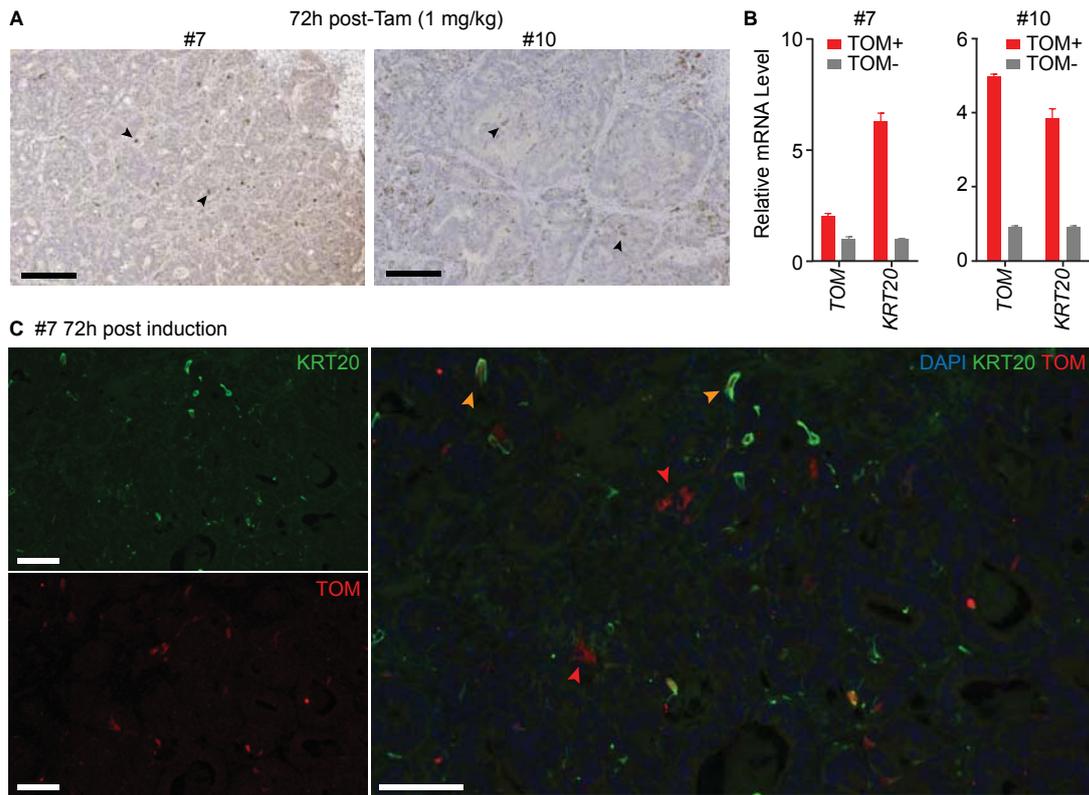
**Figure 3.21: Tamoxifen induces KRT20 lineage tracing from a large number of cells.**

**Figure 3.21: Tamoxifen induces KRT20 lineage tracing from a large number of cells.** TOM IHC on sections from PDX paraffin blocks derived from KRT20-Cre clones 96h post-tamoxifen injection (2\*250 mg/kg). Scale bars indicate 250 $\mu$ m.

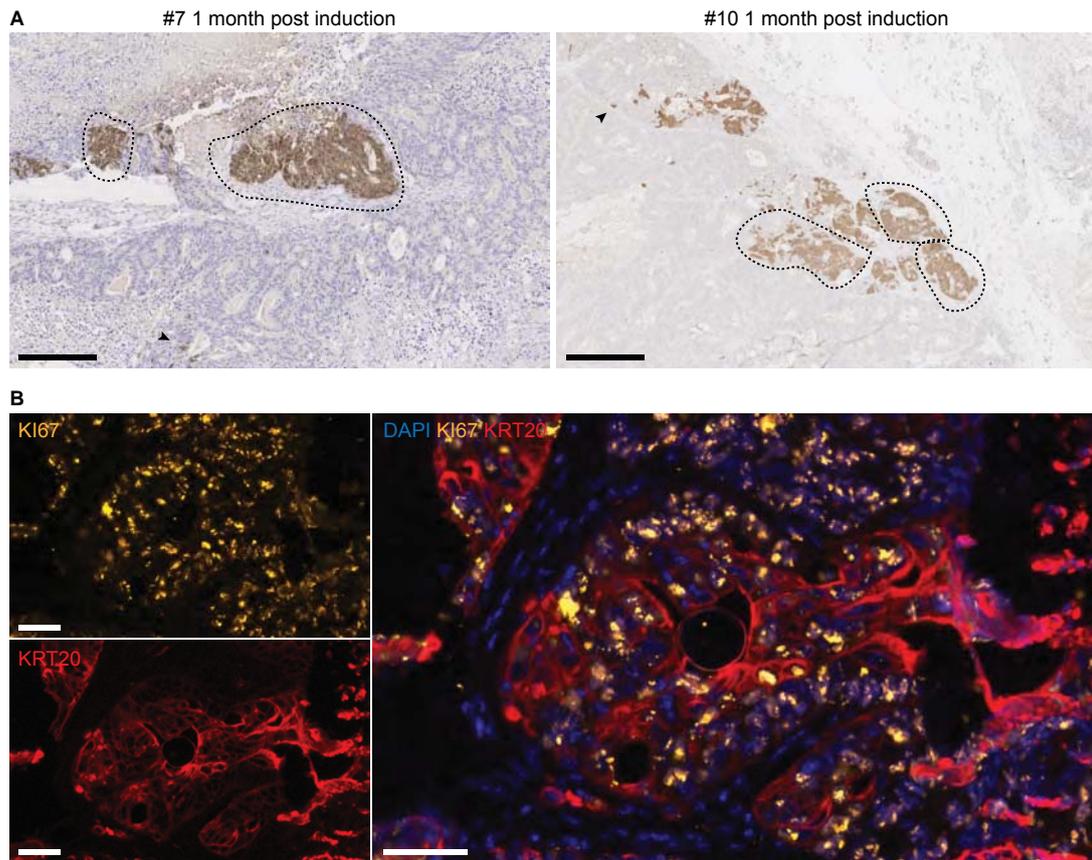
*KRT20-driven progeny persists over time and gives rise to large clones in tumors*

Because of elevated efficiency of recombination of the *KRT20* tracing allele, we lowered administered Tam 250-fold compared to *LGR5* tracing experiment (1 mg/kg), and analyzed clone formation. At 72h post-induction, most TOM+ cells in xenografts were found as single entities (Figure 3.22A), a pattern that was similar to that obtained with *LGR5* tracing system (section 3.1.5). To confirm we were tracing KRT20+ derived cells, we FACS-sorted TOM+ and TOM- populations from dissociated xenografts 72h post induction and analyzed mRNA expression. In both clones, *TOM* levels correlated with *KRT20* mRNA (Figure 3.22B). Of note, we observed that some TOM+ cells did not express KRT20 protein, which suggest that *KRT20* mRNA is expressed as an early event of the differentiation process, before the protein is produced (Figure 3.22C).

Unexpectedly, one month post induction of *KRT20* lineage tracing cassette we could observe large clones (Figure 3.23A), suggesting that at least part of the KRT20+ cells were able to proliferate. To gather more data on the proliferative status of KRT20+ cells in human tumors, we co-stained this protein and KI67 proliferation marker in xenograft sections. Many KRT20+ cells exhibited intense KI67 nuclear staining in PDO7 PDX (Figure 3.23B). These data implies that KRT20 expressing cells do proliferate and contribute to tumor growth.



**Figure 3.22: 1mg/kg of intraperitoneal Tam induces *KRT20* lineage tracing cassette at single cell level.** **A.** TOM immunostaining on sections from paraffin blocks from subcutaneous xenografts derived from *KRT20*-Cre clones 72h post-tamoxifen injection (1mg/kg). Arrowheads point at single TOM+ recombined cells. Scale bars indicate 250 $\mu$ m. **B.** Relative mRNA expression levels of FACS-sorted TOM+ and TOM- populations in subcutaneous xenografts 72h post 1mg/kg Tam injection. Data is represented as mean $\pm$ s.d. of three technical replicates. **C.** Dual immunofluorescence against *KRT20* and TOM proteins in paraffin sections of *KRT20*-Cre#7 subcutaneous xenografts 72h post-Tam injection. Orange arrowheads point at double positive cells and red arrowheads point at TOM+*KRT20*- cells. Many TOM+ are *KRT20*- cells. Scale bar indicates 100 $\mu$ m.



**Figure 3.23: KRT20 progeny in PDO7-derived xenografts grows over time.** A. TOM immunostaining on paraffin blocks from subcutaneous xenografts of KRT20-Cre clones 1 month post-Tam injection (1mg/kg). Dashed lines encircle big clones, arrowheads point at single TOM+ cells. Scale bars indicate 250µm. B. Double immunofluorescence staining for KI67 and KRT20 in subcutaneous xenografts derived from WT PDO7. Scale bars indicate 50µm.

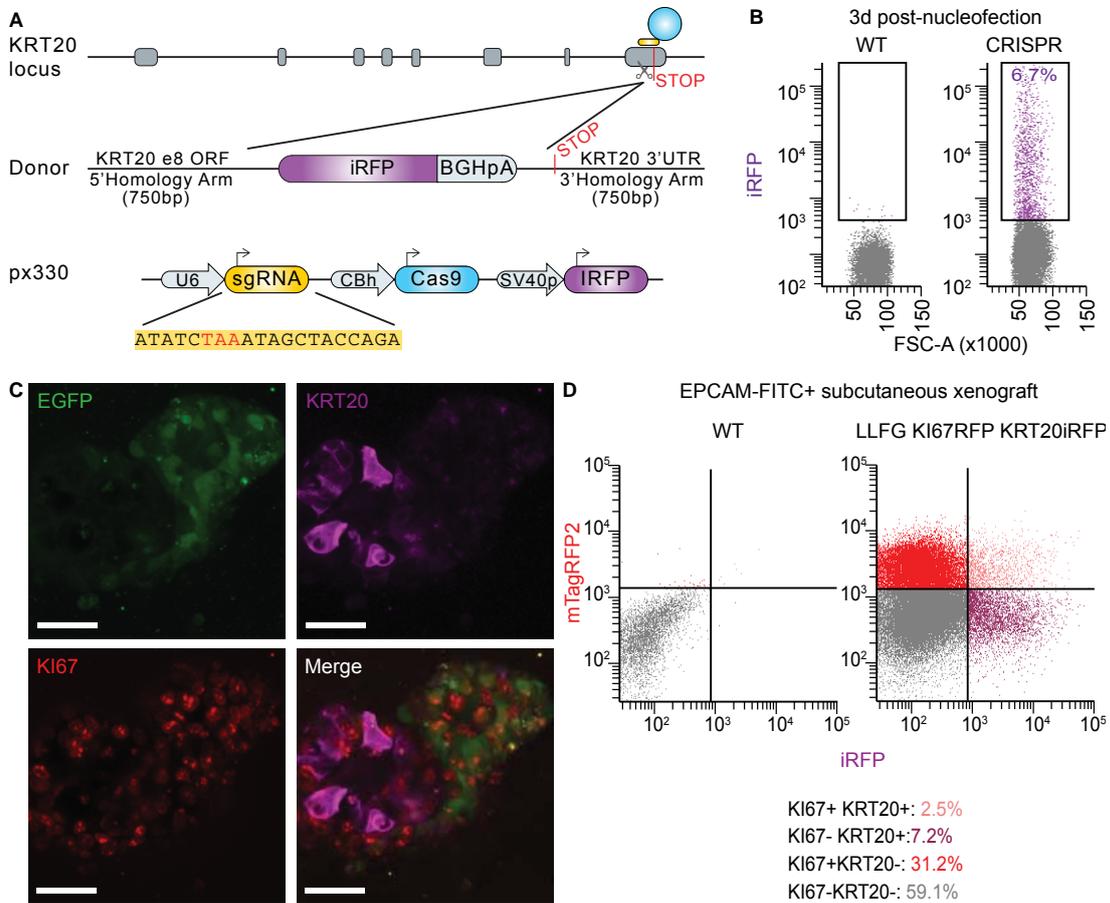
*A large proportion of the KRT20+ cell population proliferates in vivo*

To further study the features of KRT20+ cells, we next sought to create a triple reporter PDO that expressed KRT20-IRFP fusion protein together with LGR5-EGFP and KI67-RFP. These PDOs enabled tracking and purification of stem-like, differentiated-like and proliferative cells *in vivo* in human CRCs.

We nucleofected LGR5-EGFP#2 KI67-RFP#5 with the donor construct (KRT20-IRFP) and KRT20-targeting guide RNA (Figure 3.24A). We sorted IRFP+ cells (i.e. expressed from either the donor or guide plasmid) at short term and plated them in single cell format (Figure 3.24B). Total targeting efficiency was 6.7% (Table A.1). Integration PCR gels are represented in Figure A.9. We found that triple targeted PDO7 already showed *in vitro* cell heterogeneity, with mutually exclusive KRT20/LGR5 patterns (Figure 3.24C).

FACS of disaggregated xenografts derived from triple LGR5/KI67/KRT20 modified PDOs showed that around one third of KRT20+ cells were KI67+ (Figure 3.24D). Of note, virtually all KRT20+ cells in normal intestine or colon from human or mouse

tissues do not proliferate (Jiao et al., 2008). These results confirm that, in CRCs, KRT20 marks a non-stem (LGR5-) population that retains proliferative capacity and whose progeny does not completely exhaust with time.



**Figure 3.24: Generation of triple targeted PDOs.** **A.** Design of KRT20-iRFP fusion protein donor and CRISPR/Cas9 gRNA vectors. **B.** FACS analysis at day 3 post-nucleofection showing iRFP+ cells in the targeted clone. **C.** Confocal imaging of a 7-days *in vitro* grown PDO7 LGR5-EGFP/KI67-RFP/KRT20-iRFP organoid. EGFP and iRFP have complementary expression patterns, whereas KI67-RFP is equally distributed. Scale bars indicate 50 μm. **D.** RFP and iRFP FACS profiling of the EPCAM-FITC+ cells of a disaggregated xenograft derived from triple targeted PDO7 organoids.

### 3.3.2 Characterization of EMP1+ cells

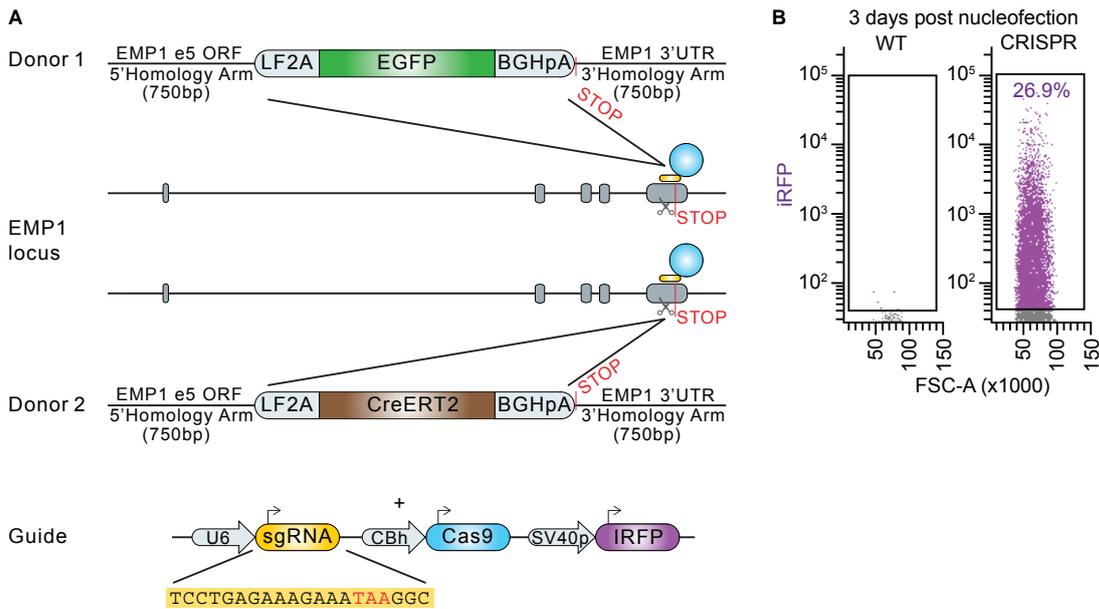
The notion that differentiated tumor cells do not contribute to tumor growth was mostly based on studies that involved dissociation of tumor cells from xenografts and re-inoculation into secondary recipients (Merlos-Suárez et al, 2011, Dieter et al., 2011). The results of the KRT20 lineage tracing experiments described above challenged this view. To further analyze the contribution of differentiated cells to tumor growth, we decided to track differentiated cells using another marker gene. The selected marker fulfilled the following criteria: i) marker gene expression should be enriched in non-stem cell populations, ii) it should represent a pan-differentiation marker rather than

specific of a certain subset of cells (i.e MUC2) iii) it should be amenable for targeting using CRISPR/Cas9 system.

We leverage data previously generated by the lab on the expression profile of normal colon stem cells (Jung et al., 2011), as well as transcriptomic profiles of PDOs to identify genes that comply with the above criteria. This analysis render a single candidate named Epithelial Membrane Protein 1 (*EMP1*). *EMP1*, although poorly described in literature, is consistently downregulated in intestinal stem cell signatures, and, interestingly, is a gene induced when the WNT signaling pathway is switched off in CRC cell lines (unpublished data from the Batlle Lab). Moreover, by RT-qPCR analysis, we detected that *EMP1* mRNA expression was 4-fold enriched within LGR5-EGFP-Negative cells in PDO7 xenografts. Based on these findings, we decided to track the contribution of *EMP1*<sup>+</sup> cells to growth dynamics.

#### *Generation of lineage tracing allele driven by the EMP1 locus*

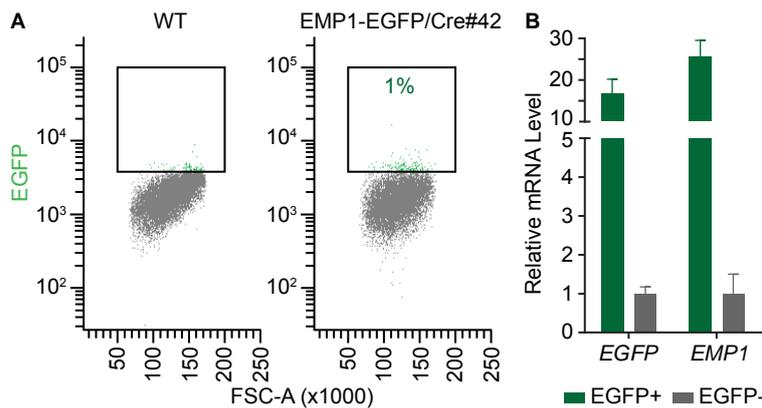
There are no good available antibodies against *EMP1*. This implied that the designed gene editing construct required to insert not only a CreERT2 for tracing purposes but also a fluorescent reporter protein to localize *EMP1*<sup>+</sup> cells in tumors. We tested many combinations of EGFP-CreERT2 fused constructs, with variants of fluorescent protein such as NeonGreen (Shaner et al., 2013) together with several polycistronic linkers (IRES, T2A, LF2A). Yet, they all showed a weak fluorescent signal (data not shown). To overcome this limitation, we nucleofected PDO7 AAVS1<sup>BFPTOM#20</sup> with both an LF2A-CreERT2 and an LF2A-EGFP donor, and subsequently selected clones that had integrated each cassette in one allele of the *EMP1* gene (Figure 3.25A). We grew single cell clones from day 3 IRFP<sup>+</sup> sorted cells (Figure 3.25B), and screened a large number of clones by integration PCR and southern blot to identify correct integrations (Figure A.10). Final efficiency of correctly targeted clones was 2.33% (Table A.1). From here onwards, this gene modification will be named as *EMP1*-EGFP/Cre. We selected clones #28, #37 and #42 among the correctly integrated ones for subsequent *in vivo* experiments.



**Figure 3.25: Dual labelling strategy for the *EMP1* locus.** A. Design of LF2A-CreERT2 and LF2A-EGFP knock-in cassettes and CRISPR/Cas9 gRNA vectors for the *EMP1* gene locus. B. Flow cytometry purification of iRFP+ cells 3 days post-nucleofection.

### *EMP1* marks a differentiated-like population in human tumors

Before proceeding to lineage tracing assays, we studied the *EMP1*-EGFP+ population. First, we analyzed *EMP1*-EGFP/*Cre* clones *in vitro*: flow cytometry analysis revealed that *EMP1*-EGFP+ cells represented only 1% of total population (Figure 3.26A), and gene expression analysis of EGFP+ and EGFP- fractions confirmed that EGFP+ expressed elevated *EMP1* mRNA levels (Figure 3.26B).

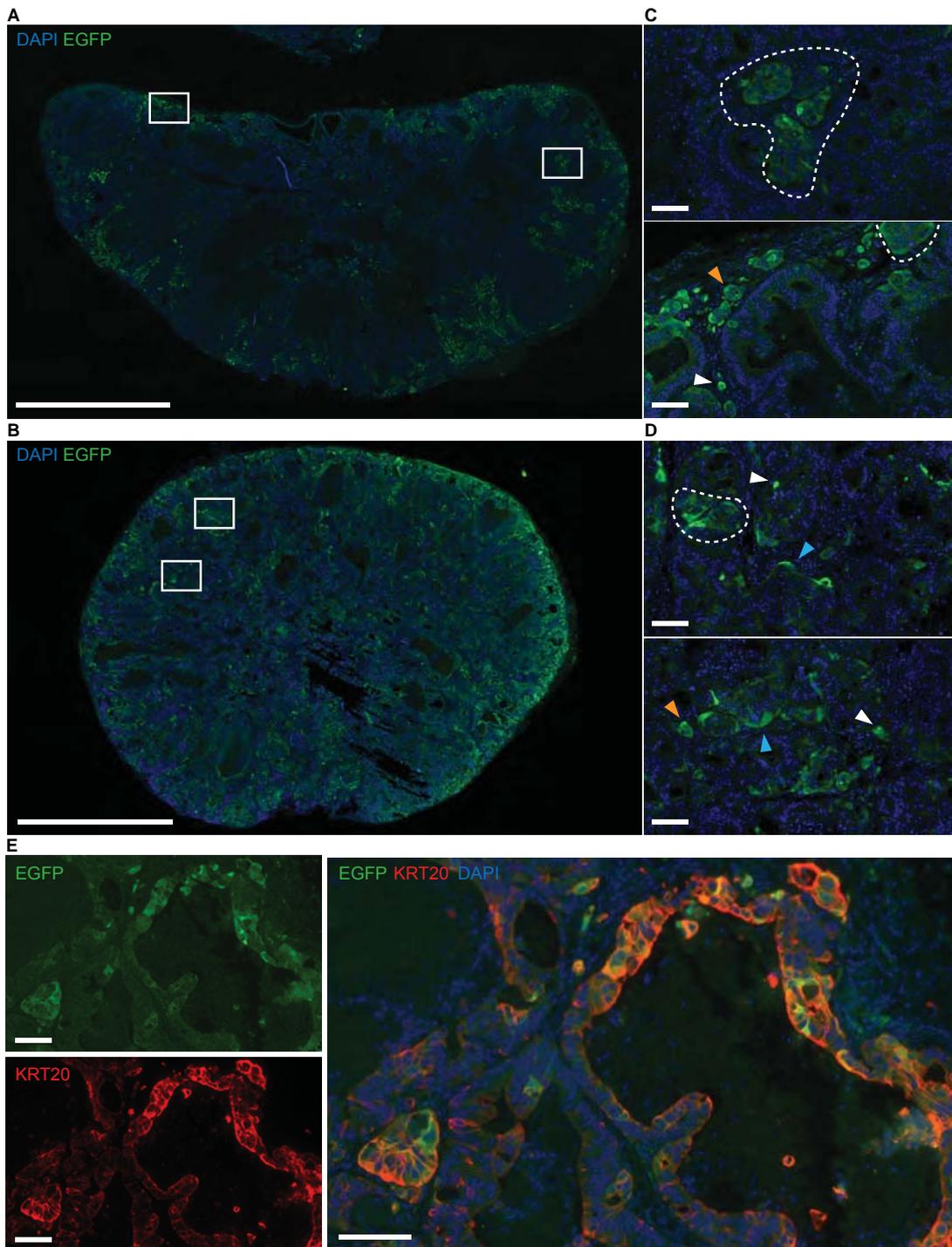


**Figure 3.26: *In vitro* PDO7 grown organoids present *EMP1*-EGFP+ cells.** A. EGFP analysis by flow cytometry of 5-days *in vitro* grown PDO7 *EMP1*-EGFP/*Cre*#42 organoids. B. RT-qPCR analysis of *EMP1* mRNA expression in the EGFP+ vs EGFP- sorted populations. Data is represented as mean  $\pm$  s.d. of three technical replicates.

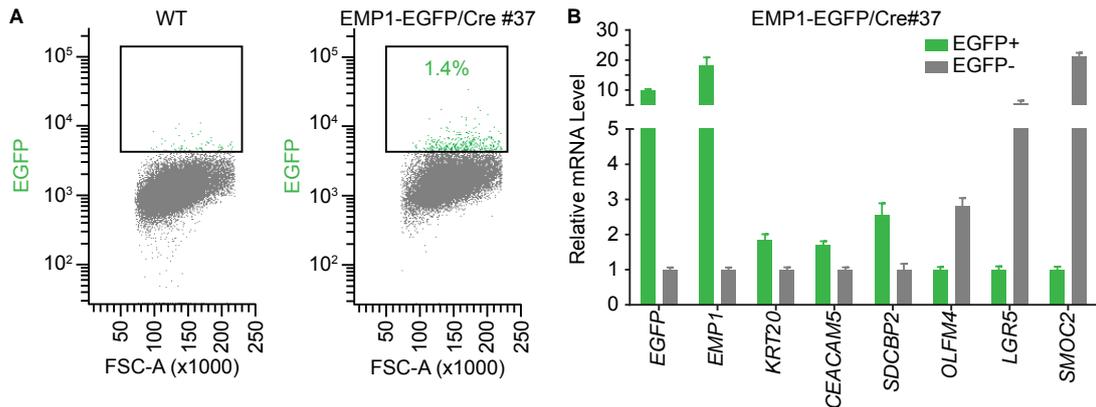
We next generated subcutaneous PDX from these *EMP1* edited clones in Beige/*S-CID* mice. *EMP1*-EGFP+ cells were visualized within tumor glands, accounting for approximately 5 to 10% of total tumor epithelium. We noticed that *EMP1*+ cells accumulated mostly at tumor borders (Figure 3.27A and B). Upon a closer inspection, we identified many *EMP1*+ cells intermingled between tumor stroma rather than in

epithelial glandular structures (Figure 3.27C and D). Moreover, the brightest EMP1-EGFP+ cells displayed morphology consistent with cells undergoing migration and invasion (Figure 3.27C and D). We will further explore this unexpected findings in the next section. In order to confirm that *EMP1* was a marker for a differentiated-like tumor lineage, we immunostained sections of the PDX with KRT20 antibodies and found that indeed most EMP1-EGFP+ cells were KRT20+ (Figure 3.27E).

We also disaggregated xenografts and analyzed them by flow cytometry. EMP1-EGFP+ cells represented around 1% of total EPCAM+ population (Figure 3.28A). As discussed for *LGR5* targeting experiments (see **section 3.1.2**) the discrepancy in the relative abundance of EGFP+ cells detected by FACS versus immunofluorescence on histology samples might be due to the fact that FACS is not sufficiently sensitive to detect EGFP-Low cells compared to the amplification methods used for IHC. Nevertheless, we sorted EGFP-High and EGFP-Negative fractions and confirmed that EMP1+ cells expressed elevated levels of differentiation markers (*KRT20*, *SDCBP2* and *CEACAM5*) but low levels of stem cell genes (*OLFM4*, *LGR5* and *SMOC2*) (Figure 3.28B). The results were largely consistent between clones (Figure A.11).



**Figure 3.27: EMP1-EGFP expression pattern in PDO7 subcutaneous xenografts.** **A and B.** Representative EGFP immunofluorescence in sections from paraffin blocks containing whole EMP1-EGFP subcutaneous xenografts. Scale bar indicates 2.5 mm. **C and D.** Magnifications of tumor areas showing i. EGFP+ glandular-like structures (delimited by dashed lines) ii. Slender EGFP+ cells in the gland borders (blue arrowheads) iii. EGFP+ cells intermingled in the tumor stroma (white arrowheads) iv. Groups of EGFP+ cells migrating together (yellow arrowheads). Scale bars indicate 100 $\mu$ m. **E.** KRT20 and EGFP dual immunofluorescence in EMP1-EGFP derived xenografts. Most of EMP1+ cells express KRT20 protein. Scale bars indicate 100 $\mu$ m

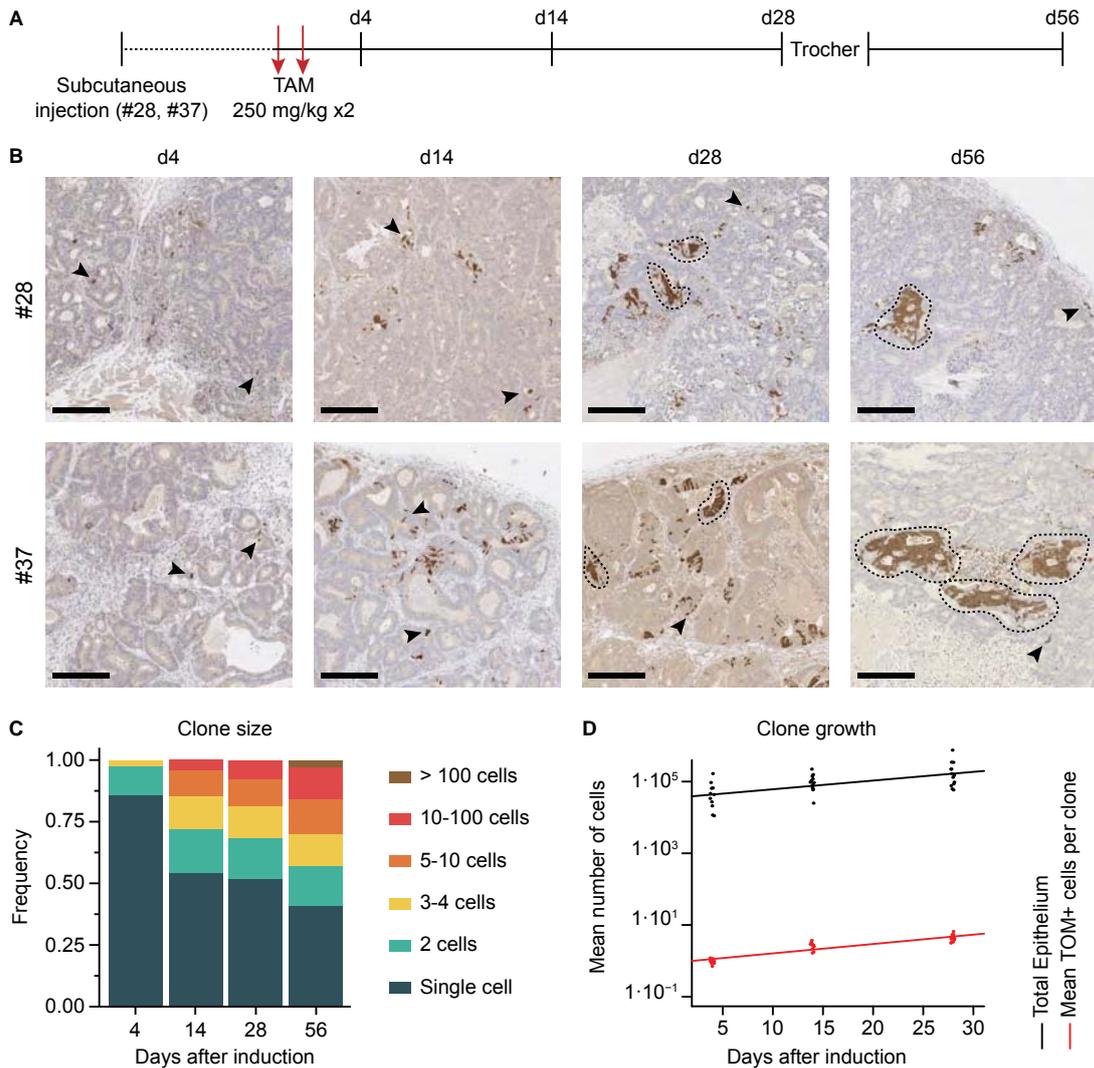


**Figure 3.28: EMP1+ cells are a differentiated-like, non-stem population.** **A.** FACS analysis of EPCAM+ purified cells from disaggregated subcutaneous xenografts generated by injection of PDO7 EMP1-EGFP/Cre#37. **B.** mRNA expression analysis of differentiated and stem marker genes in the EGFP-High and -Negative sorted populations from disaggregated subcutaneous xenografts derived from PDO7 EMP1-EGFP/Cre#37. Data is represented as mean±s.d. of three technical replicates.

*EMP1+ cells contribute to tumor growth with similar output than the LGR5+ and KRT20+ cell populations*

We next performed lineage tracing experiments in PDX grown in Beige/SCID mice with two of the PDO7 EMP1-EGFP/Cre clones. First, we analyzed EMP1-EGFP - Positive and -Negative cells isolated by FACS from disaggregated subcutaneous xenografts. We confirmed that *CreERT2* was specifically expressed within the EGFP+ fraction (Figure A.12). Next, we determined that Tam dosage of 2x250 mg/kg was the most adequate for in vivo lineage tracing of EMP1 progeny in clonal conditions. Indeed, 96h after Tam injection, a TOM+ population of around 0.7% of total EPCAM+ cells was labelled in both clones (Figure A.13A and B). FACS of TOM+ vs TOM- cells followed by RT-qPCR analysis of mRNA expression in EMP1-EGFP/Cre#28 confirmed that at early time-points TOM-labelled cells were enriched in EMP1 expression (Figure A.13C).

Having set the parameters for *EMP1* lineage tracing experiments, we then tracked *EMP1* progeny over time (up to two months of clonal follow-up) as described previously for *LGR5* in **section 3.1.5** (Figure 3.29A). Again, we quantified the size of a significant number of clones by TOM immunohistochemistry in tumor sections. At day 4 post-Tam induction, we observed single TOM+ cells that over time gave rise to clones that increased in size (Figure 3.29B). The clonal expansion kinetics was similar to that of the *LGR5* progeny, with large clones at endpoint, but also a proportion of TOM+ cells that remained as single or as 2-cell clones during the whole experiment (Figure 3.29C). Moreover, EMP1+ cell-derived clones expanded proportionally to total epithelia (Fig 3.32D).



**Figure 3.29: EMP1 progeny expands *in vivo* in subcutaneous xenografts.** **A.** Scheme of EMP1 lineage tracing experimental setup. **B.** Representative IHC using anti-Tomato antibodies on paraffin sections from xenografts at of the four time points after tamoxifen treatment. Arrowheads point to single and two cell clones. Dashed lines delimit large clones. Scale bars indicate 250 $\mu$ m. **C.** Clone size frequency per time point according to number of cells. Number of clones quantified was 4423 for day 4, 20767 for day 14, 17156 for day 28, and 550 for day 56. **D.** Correlation of number of epithelial cells per xenograft and number of cells per clone over time (number of xenografts assessed: n=11, n=13, n=13, n=8 in the time points 4,14, 28 and 56 days respectively).

In summary, we have discovered that differentiated-like tumor cells contribute to tumor growth with a similar kinetics than stem-like LGR5+ cells. This is the first time that clonal analysis has been performed in human advanced CRCs, and these findings go against the established notion that CRCs retain a hierarchical organization similar to the normal intestine (Merlos-Suárez et al., 2011, Schepers et al., 2012). Parallel to our work, a recent publication by the Vermeulen Lab suggests that hierarchy is lost during adenoma to carcinoma transition (Lenos et al., 2018). This new concept is of outmost interest and will help rationalize the design of targeted therapies against specific tumor cell subpopulations.

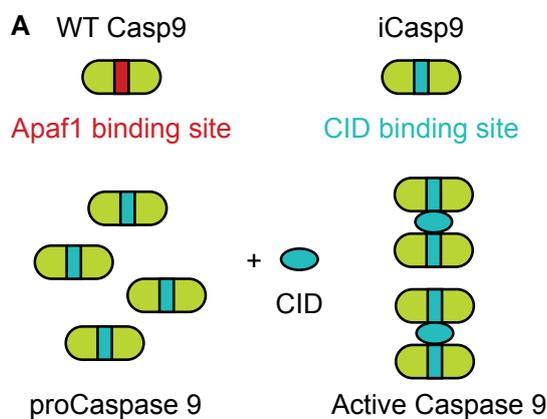
# Chapter 3: *EMP1* is a potential marker for metastatic cells in CRC

## 3.4 *EMP1* DEFINES A SUBSET OF CRC CELLS WITH INVASIVE CHARACTERISTICS

As discussed above, we initially decided to mark the *EMP1* gene to lineage trace differentiated cells in CRC. Yet, visualization of *EMP1*+ cells in tumors revealed that this tumor cell population exhibits a phenotype that resemble that of cells undergoing migration. In particular, *EMP1*+ cell were visualized as individual entities or groups of cells with elongated morphology that often localized within tumor stroma (Christiansen and Rajasekaran, 2006). This third chapter of the thesis focuses on studying the utility of *EMP1* as a novel marker for migratory, possibly metastatic, CRC cells.

### 3.4.1 Tagging *EMP1* with reporter and suicide cassettes using CRISPR/Cas9 technology

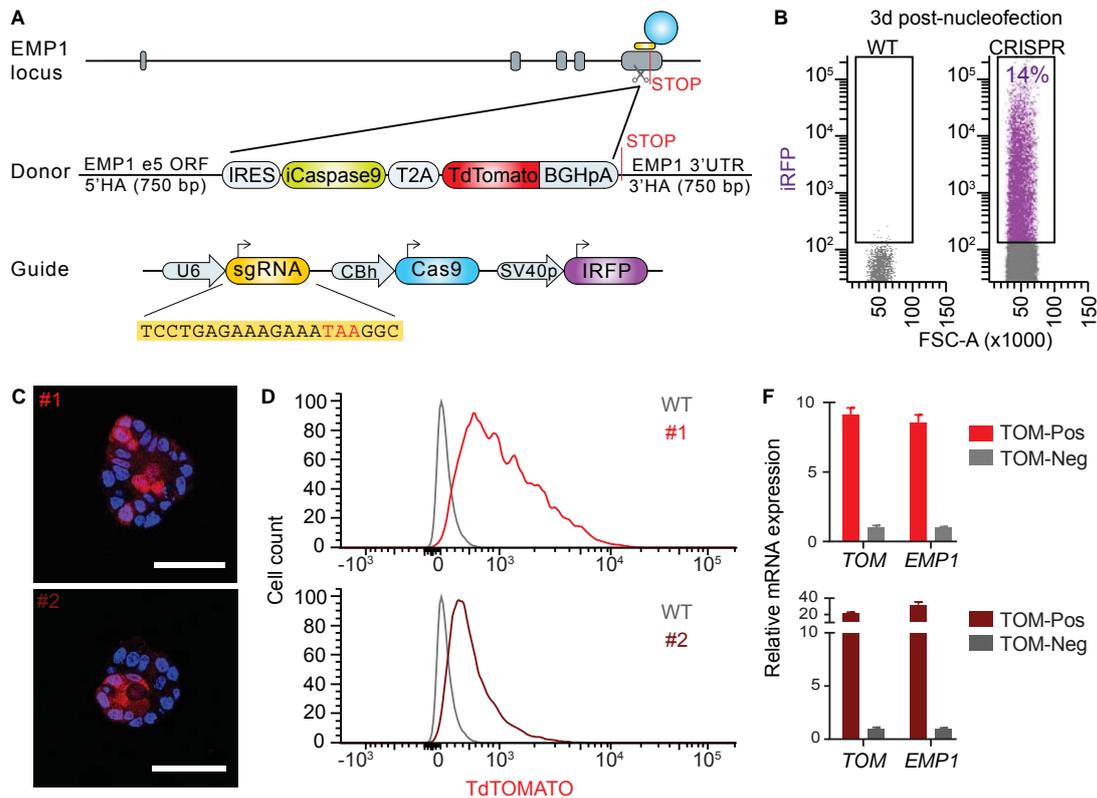
To follow *EMP1*+ cells *in vivo*, a brighter reporter than the previously used EGFP was required. We therefore modified the donor plasmid described in **section 3.3.2** to encode for a TOM protein. This knock-in construct is designed so that the cassette is inserted after the STOP codon and therefore the expression of endogenous *EMP1* locus is unperturbed and cytoplasmic TOM labels cells expressing *EMP1* gene. In addition, this cassette contains an inducible suicide gene that enables experiments of cell ablation (Figure 3.31A). In particular, we inserted an inducible CASPASE9 (iCasp9) in which the APAF-1 binding domain has been substituted by a drug binding domain, so Caspase9 is only able to dimerize and to start the apoptotic cascade in the presence of the Chemical Inducer of Dimerization (CID) (Figure 3.30) (Clackson et al., 1998). Addition of iCasp9 to the construct allows for selective elimination of the *EMP1*+ population upon CID incubation. This cell ablation strategy was previously used to ablate LGR5+ cells in human CRCs using an equivalent strategy to the one described here (Shimokawa et al., 2017).



**Figure 3.30: Inducible Caspase9 starts the apoptotic cascade upon CID-mediated dimerization.** Scheme of the inducible Caspase 9 system. The APAF-1 binding domain for the dimerization and activation of the protein has been substituted by a CID binding domain.

Using the protocol described in previous sections, we nucleofected PDO7 and selected correctly integrated clones (Figure 3.31B). A summary of single cell clone gen-

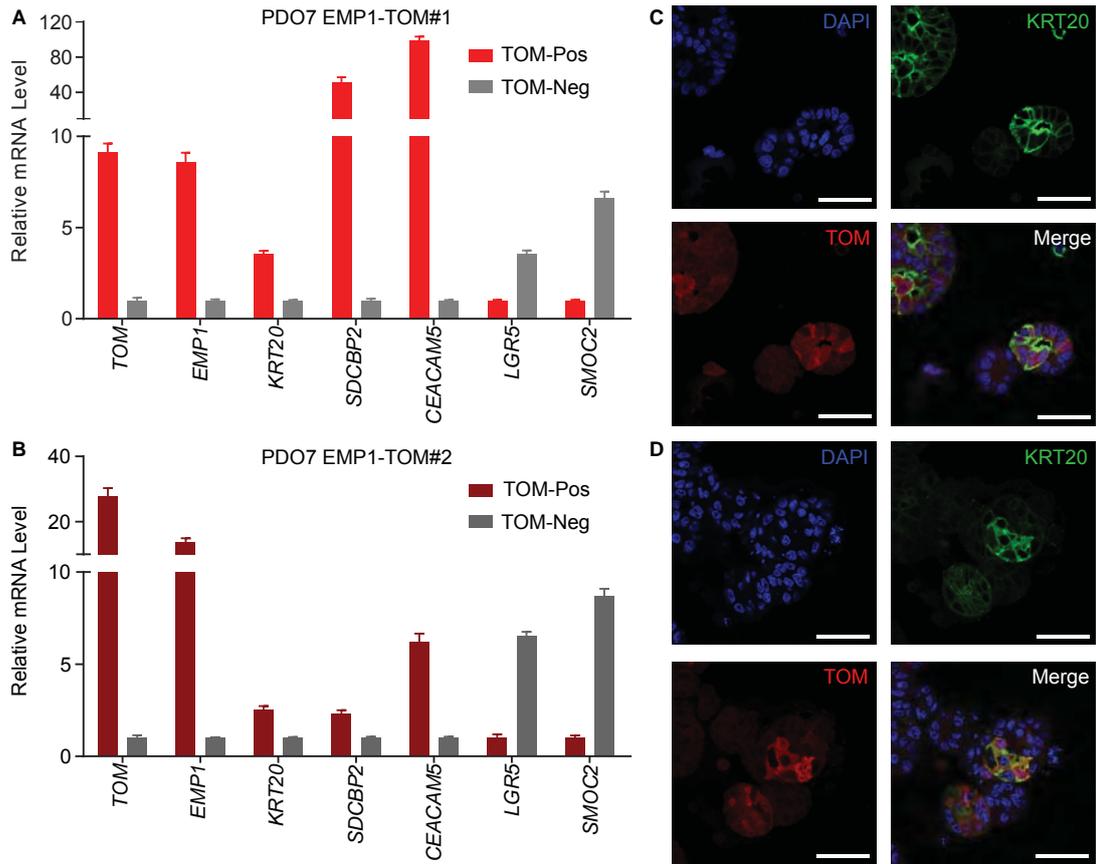
eration results (Figure A.14A) is depicted in Table A.1. We picked #1 (homozygous) and #2 (heterozygous) as representative examples for subsequent experiments. As shown in Figure 3.31C and D, *in vitro* cultured knock-in PDO7 organoids express EMP1-TOM heterogeneously. We also FACS-sorted TOM-High and TOM-Negative cells from organoids established from both clones and observed that, indeed, *TOM* levels correlated with *EMP1* mRNA expression (Fig 3.34F). The edited clones will be named from now on as EMP1-iCT: EMP1-iCaspase9-TOM.



**Figure 3.31: Generation of an inducible suicide cassette under *EMP1* gene control.** **A.** Design of the iCasp9-TOM reporter donor plasmid and CRISPR/Cas9 gRNA vector for the *EMP1* locus. **B.** Short term selection of PDO7 iRFP+ cells 3 days post nucleofection. **C.** Imaging of alive *in vitro* PDO7 EMP1-iCasp9TOM#1 and #2 organoids. Scale bars represent 50 μm. **D.** FACS analysis of TOM levels in *in vitro* grown #1 and #2 organoids. **E.** Relative mRNA expression level of sorted TOM+ and TOM- populations *in vitro* by RT-qPCR. *TOM* expression correlates with *EMP1* mRNA. Data is represented as mean ± s.d. of 3 technical replicates.

### 3.4.2 *EMP1* marks a proliferating differentiated-like population *in vivo*

Taking advantage of EMP1-iCT modified PDO7 clones we characterized *in vitro* the expression pattern of EMP1+ cells. RT-qPCR analysis from TOM+ and TOM- purified cells confirmed that *EMP1* gene was co-expressed with pan-differentiation markers *KRT20*, *SDCPB2* and *CEACAM5* and anti-correlated with stem cell markers *LGR5* and *SMOC2* (Figure 3.32A and B). We also confirmed expression of KRT20 protein in EMP1-TOM+ cells in organoids (Figure 3.32C and D).



**Figure 3.32: EMP1-TOM knock-in marks a differentiated-like population *in vitro*.** **A and B.** mRNA expression level of TOM+ and TOM- populations sorted from 5-days *in vitro* grown PDO7 EMP1-iCT#1 and #2, respectively. Data is represented as mean±s.d. of 3 technical replicates. **C and D.** TOM and KRT20 dual immunofluorescence staining in 5-days grown organoids of PDO7 EMP1-iCT#1 (C) and #2 (D), showing large overlap between both markers. Scale bars indicate 100μm.

We next sought to study EMP1+ cells *in vivo*. We injected EMP1-iCT PDO7 clones subcutaneously into Beige/SCID mice. When tumors reached a size of 150 mm<sup>3</sup>, we sacrificed the mice and disaggregated the PDX for subsequent flow cytometry analysis. We collected the 10% TOM-High, -Low and -Negative fractions (Figure 3.33A). RT-qPCR analysis essentially confirmed *in vitro* data. The EMP1-TOM-High population was enriched in expression of differentiation markers *KRT20* and *SDCBP2*, and counter enriched in stem cell genes *LGR5* and *SMOC2* (Figure 3.33B). Interestingly, *EMP1* expression appeared restricted to the enterocytic lineage, as *MUC2*, a well-defined marker for secretory cells, was highly enriched in TOM-Negative cells. Furthermore, global gene expression analysis using microarray followed by GSEA showed that the intestinal differentiation program was overall enriched in the EMP1-High population, whereas the human intestinal stem cell signature (Merlos-Suárez et al., 2011) was largely enriched in EMP1-Negative cells (Figure 3.33C). In addition, we also compared EMP1-TOM gene expression signature versus LGR5-EGFP+ signature generated in previous experiments from PDO7 (section 3.1.4). This analysis confirmed that *EMP1* and *LGR5* gene expression programs were expressed in a complementary fashion (Figure 3.33D). Moreover, we found that EMP1+ cells were en-

riched in genes of the proliferation program, such as *RAD51*, *BUB1*, *CDCA2*, *CDCA5* and *CCNF* (Figure 3.33E). This finding is in accordance with the results of lineage tracing experiments suggesting that EMP1+ cell proliferate. We validated further this observation by immunostaining on PDXs using antibodies against markers of differentiation (KRT20, MUC2) and proliferation (KI67). EMP1+ cells were largely KRT20+, devoid of MUC2 and many of them presented nuclear KI67 staining (Figure 3.33F). These results were confirmed using xenografts derived from inoculation of clone #2. However, flow cytometry analysis revealed that EMP1-iCT#2 was less bright than EMP1-iCT#1 (Figure A.15A), as we already observed *in vitro*, and therefore FACS-sorting did not offer good resolution to separate cell populations according to EMP1-TOM levels as in clone #1 (Figure A.15B).

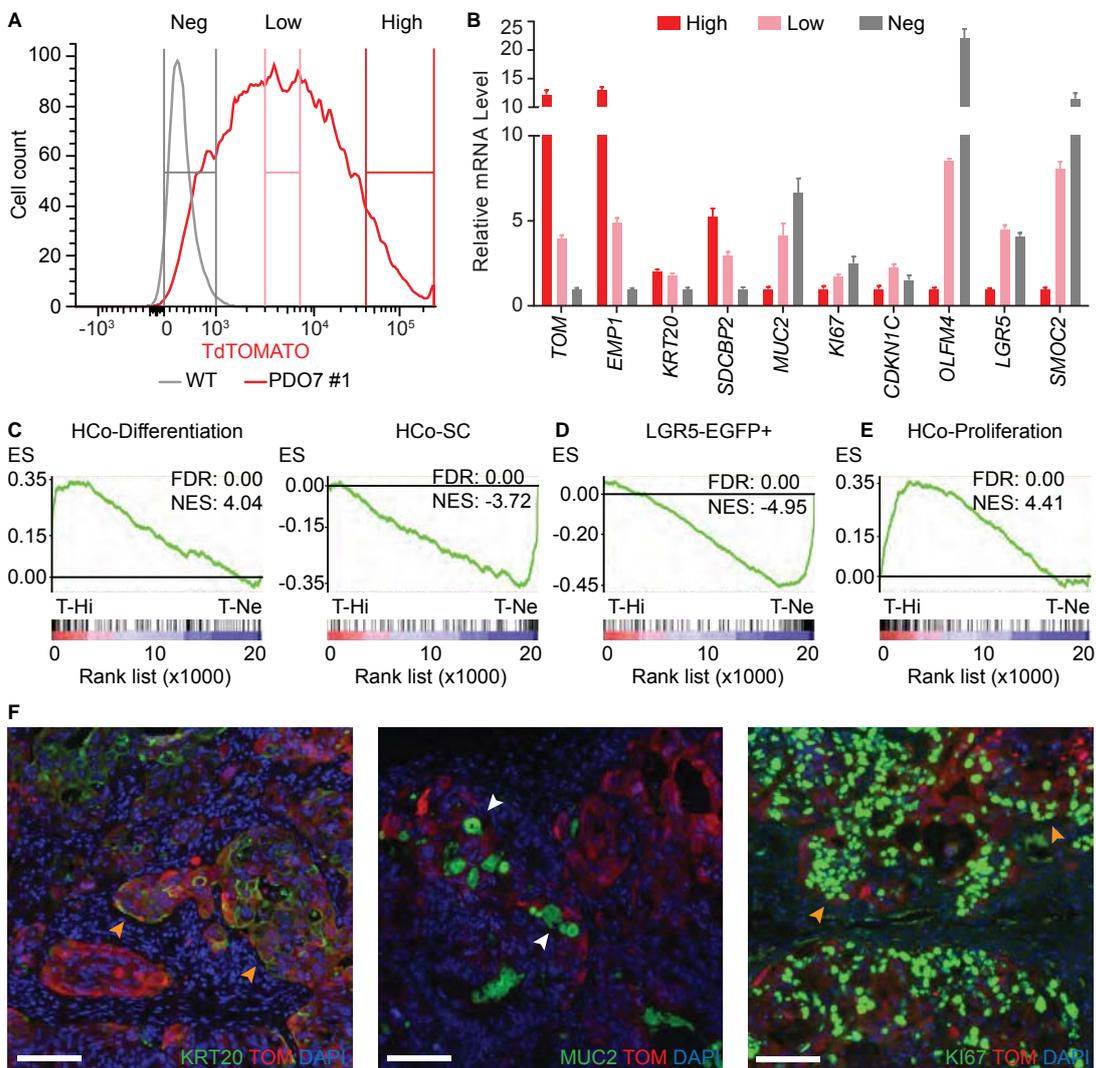
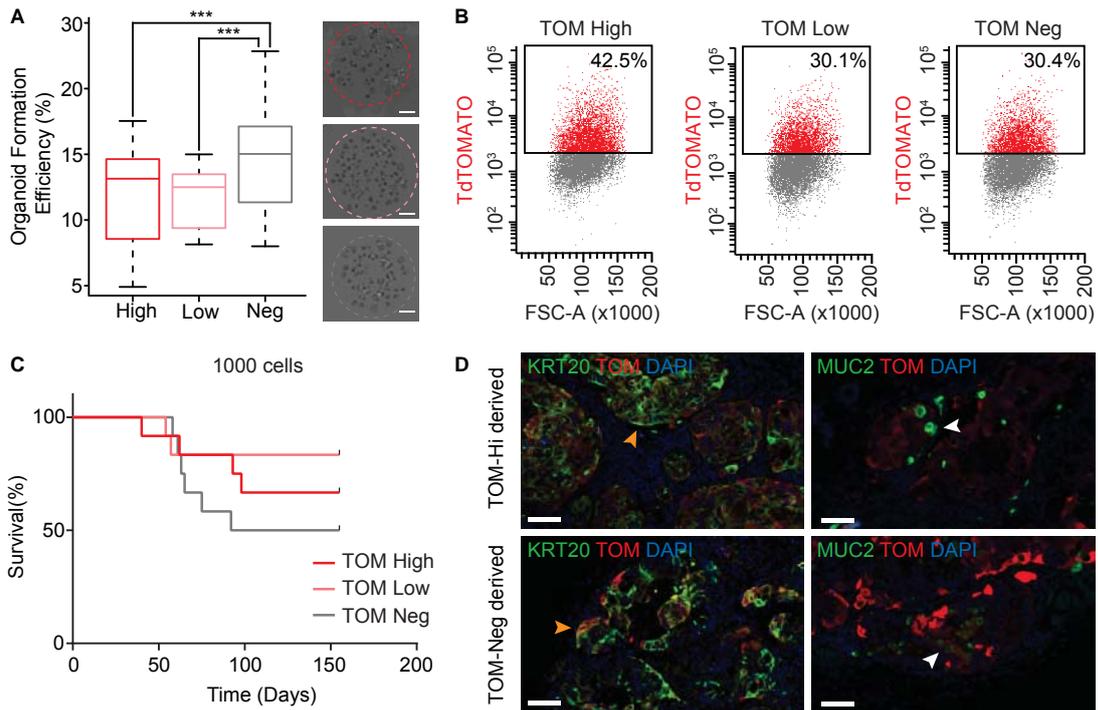


Figure 3.33: Characterization of the EMP1+ population in human xenografts.

**Figure 3.33: Characterization of the EMP1+ population in human xenografts.** **A.** FACS analysis of EPCAM+DAPI- cells from disaggregated EMP1-iCT#1 PDX. **B** RT-qPCR reporting mRNA expression levels of indicated markers in TOM+ vs TOM- sorted populations. Data is represented as mean±s.d. of three technical replicates. **C.** GSEA comparing the expression of signatures of normal differentiated cells (hCo differentiation) and human colon stem cells (hCoSCs) in profiled EMP1-TOM+ vs EMP1-TOM- cells. **D.** GSEA comparing the expression of the previously generated LGR5-EGFP+ signature (in PDO7 CRISPR/Cas9 engineered cells) with the profiled EMP1-TOM+ vs EMP1-TOM- cells. **E.** GSEA comparing a proliferative normal crypt cells signature with profiled EMP-TOM+ vs EMP1-TOM- cells. **F.** TOM/KRT20, TOM/MUC2 and TOM/KI67 dual immunofluorescence stainings on paraffin sections from PDO7 EMP1-iCT#1 xenografts. Yellow arrowheads indicate TOM+KRT20+ or KI67+ cells and white arrowheads indicate TOM-MUC2+ cells. Scale bars indicate 100µm.

We next studied the clonogenic potential of EMP1+ cells. To this end, we FACS-sorted the 10% -High, -Low and -Negative EMP1-TOM cell fractions and plated them in organoid forming conditions. EMP1-Negative cells were slightly more efficient than EMP1-Positive cells in growing organoids (Figure 3.34A). Mean growth efficiency from single cells was around 20% for the EMP1-Negative fraction whereas the EMP1-High cell subpopulation gave rise to clones with an efficiency of 10%. The three sorted populations exhibited *in vitro* plasticity as 15 days after plating they had recovered similar levels of TOM implying regeneration of EMP1-High, -Low and -Neg cells (Figure 3.34B). Cells sorted from EMP1-iCT#2 displayed a very similar behavior (Figure A.15C and D). Finally, we assessed *in vivo* tumor forming capacity of EMP1-High, Low and Negative populations by reinjecting low numbers (1000 cells) of FACS-sorted cells into Beige/SCID mice. These experiments indicated a trend towards EMP1-Negative population being more proficient to initiate tumors than EMP1-Low and -High, yet differences did not reach statistical significance. (Figure 3.34C). In accordance with the clonogenic analysis, we also observed plasticity of cell phenotypes since TIC-grown tumors resembled the parental tumor regardless of whether they were generated by EMP1-High or -Negative cells (Figure 3.34D).

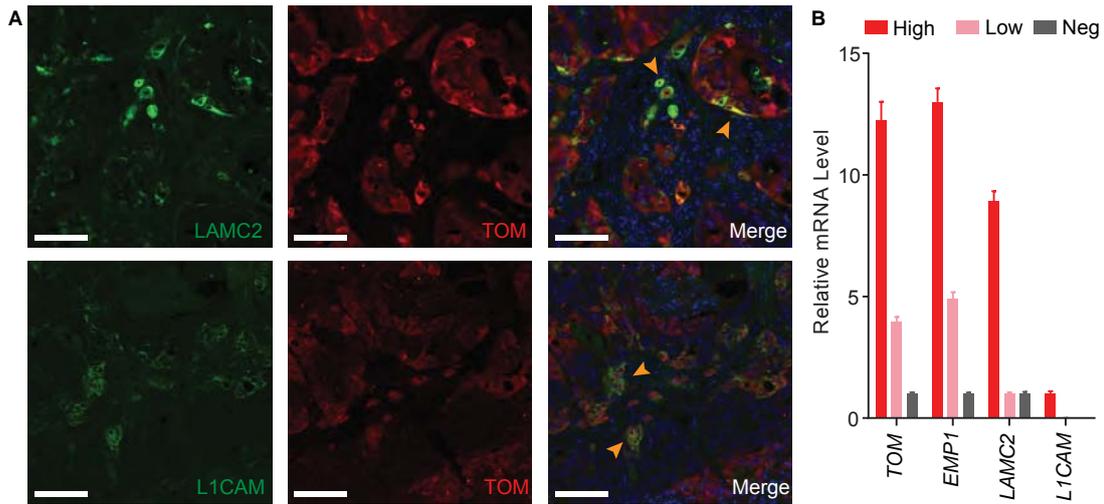
These functional assays fitted the expression profile of the population, being EMP1+ a differentiated-like population with ability to proliferate (as shown by *in vivo* lineage tracing) but less proficient in organoid formation and tumor initiation than EMP1-Negative population (putatively more stem-like).



**Figure 3.34: EMP1<sup>+</sup> population retains growth initiation capacity.** **A.** Organoid formation efficiency of TOM-High, Low and Negative FACS-purified cells from EMP1-iCT#1 derived xenograft, and representative images of each condition. 1500 cells were plated per well, and each experiment assessed the growth of 4 wells per condition. Results are represented as mean $\pm$ s.d. of 3 experiments. Scale bars indicate 1mm. **B** Representative FACS profile of the sorted cells after 15 days of *in vitro* growth. **C.** Kaplan Meier curves showing the tumor initiation capacity of 1000 cells (High, Low or Negative) isolated by flow cytometry upon secondary host inoculation. **D.** Representative TOM/KRT20 and TOM/MUC2 dual immunofluorescence of grown tumors from EMP1-High and EMP1-Negative populations. TOM pattern is similar in both tumor types. Yellow arrowheads point at TOM+KRT20+ areas and white arrowheads signal MUC2+ cells intermingled between the TOM+ areas. Scale bars indicate 100 $\mu$ m. \*\*\*P-value < 0.005.

### 3.4.3 EMP1 expression overlaps with invasive markers

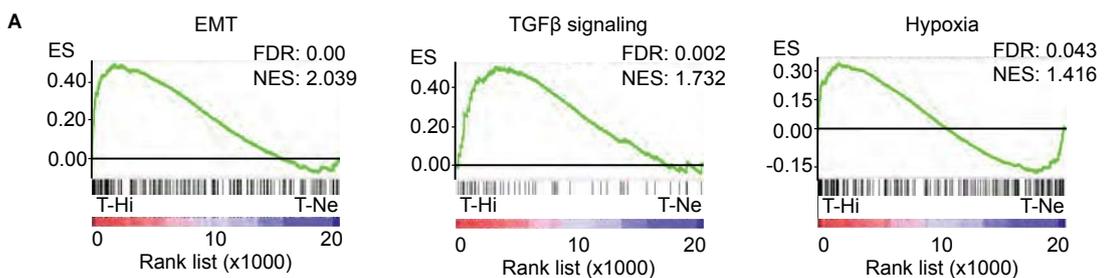
In order to find out whether *EMP1* gene labels migratory cells in CRC, we analyzed its expression domain compared to other well-described invasive markers, LAMC2 and L1CAM (Huang et al., 2017, Er et al., 2018). Immunostaining of xenografts showed that all LAMC2<sup>+</sup> or L1CAM<sup>+</sup> cells were found within the EMP1-TOM<sup>+</sup> population, although there was a large number of EMP1-TOM<sup>+</sup> cells negative for these two markers (Figure 3.36A). RT-qPCR confirmed that *LAMC2* and *L1CAM* genes were up-regulated in *EMP1* TOM-high cells. (Figure 3.36B). We tentatively conclude that *EMP1* marks a population of differentiated, proliferating and migratory cells in CRCs.



**Figure 3.35: EMP1-High cells express invasive markers in PDO7.** **A.** Dual immunofluorescence on section from paraffin blocks of EMP1-iCT#1 xenografts, showing LAMC2+ and L1CAM+ cells within the TOM+ glands (yellow arrowheads). Scale bars indicate 250 $\mu$ m. **B** mRNA expression level by RT-qPCR of TOM-High, Low and Negative FACS-purified fractions. TOM and EMP1 levels correlate with LAMC2 and L1CAM expression. Data is represented as mean $\pm$ s.d. of 3 technical replicates.

*EMP1 signature is enriched in metastasis promoting signaling pathways*

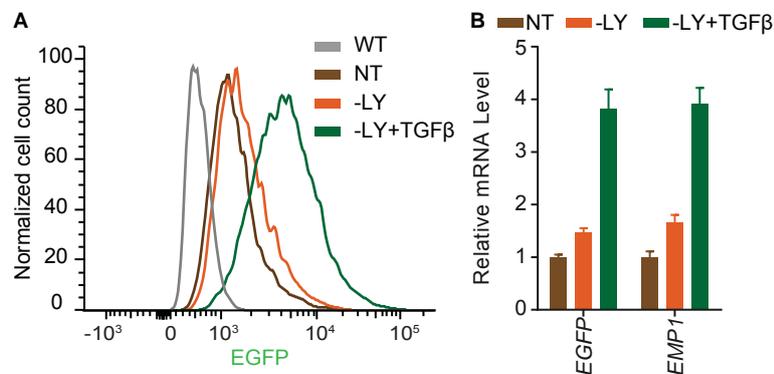
We used the transcriptomic profiles of EMP1-High versus EMP1-Negative CRC cells to assess by GSEA enrichment in particular Gene Ontology (GO) signatures. Complete list of categories are shown in Table A.2. Among others, the most interesting pathways significantly enriched in EMP1-High cells were EMT and TGF- $\beta$  signaling, and, moderately, Hypoxia (Figure 3.36). These pathways are known to play a role in the migration and metastatic process (Calon et al., 2015, Choi et al., 2017, Aiello et al., 2018).



**Figure 3.36: Gene-Ontology pathways upregulated in EMP1-TOM-High cells.** GSEA of the epithelial to mesenchymal transition (EMT), TGF- $\beta$  and Hypoxia signatures from Gene Ontology Pathways against the EMP1-TOM High versus Negative microarray comparison of PDO7 EMP1-iCT#1. EMT and TGF- $\beta$  signaling gene signatures significantly correlate with EMP1-High cells, whereas hypoxia genes are equally distributed between TOM-High and -Neg cells

### *EMP1 gene is upregulated in response to TGF- $\beta$ signaling*

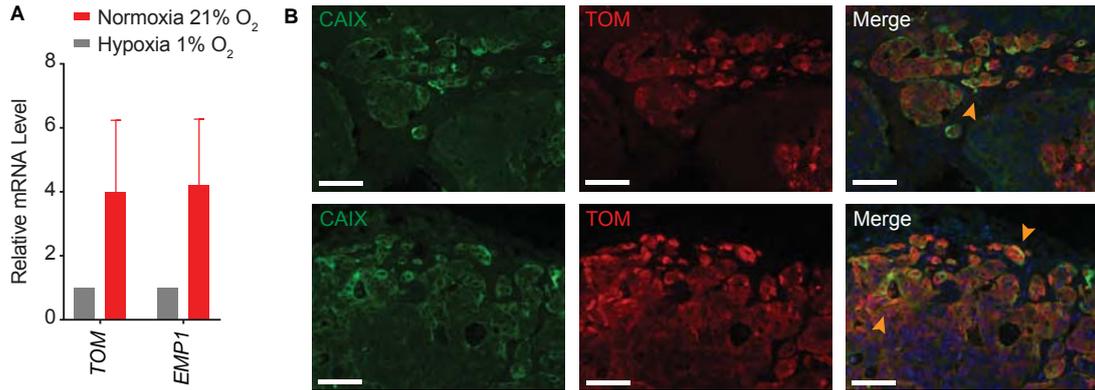
PDO7 harbors a homozygous mutation in SMAD4 and is therefore not responsive to this hormone. Therefore, we used PDO6, which is WT for TGF- $\beta$  pathway, for the following experiment (Table 3.1). We had previously modified PDO6 to incorporate CreERT2 and EGFP knock-ins under the control of EMP1, with a single cell clone-derivation efficiency of 13.95% (Table A.1). We cultured PDO6 EMP1-EGFP#3 for 7 days in complete PDO media (this media contains LY2157299, a TGF $\beta$ R1 small molecule inhibitor as described in **section 6.1.1**), in complete media without LY2157299 and in media minus LY2157299 supplemented with TGF- $\beta$ 1 (5ng/mL). PDO6 organoids grew slower in response to TGF- $\beta$  signaling, suggesting that in this PDO it exerts a cytostatic program, dividing approximately at half of the normal growth rate (data not shown). Both flow cytometry (Figure 3.37A) and relative mRNA expression analysis showed a clear increase in EMP1-EGFP levels upon TGF- $\beta$ 1 stimulation. (Figure 3.37B).



**Figure 3.37: *EMP1* is responsive to TGF- $\beta$  signaling *in vitro* in PDO6.** **A.** FACS analysis of the EGFP positivity in PDO6 EMP1-EGFP in *in vitro* cultured organoids with different media conditions: i. complete PDO media (NT), ii. Media minus the TGF- $\beta$  inhibitor LY2157299 (LY) and iii. Media minus LY and plus TGF- $\beta$ 1. There is a clear increase in the number of EGFP+ cells in response to TGF- $\beta$ 1 addition. **B.** mRNA expression level of the mentioned conditions. Both *EGFP* and *EMP1* mRNAs are upregulated in response to 7 days *in vitro* treatment with TGF- $\beta$ 1 (5ng/ $\mu$ L). Data is represented as mean $\pm$ s.d. of 3 technical replicates.

### *EMP1 gene expression is increased in hypoxia conditions*

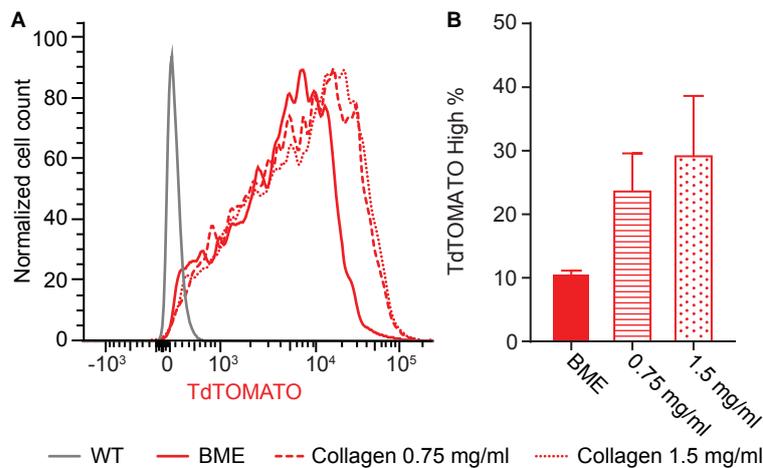
We also tested whether *EMP1* itself was a hypoxia responsive gene, as hypoxia can act as a trigger of the invasive and metastatic behavior (Rankin and Giaccia, 2016). We cultured 2 days-grown PDO7 EMP1-TOM organoids in 1% O<sub>2</sub> for 48 hours and analyzed mRNA expression of both TOM reporter and *EMP1* gene. Indeed, both *TOM* as well as *EMP1* mRNAs were upregulated 4-fold under a hypoxic environment (Figure 3.38A). In addition to these *in vitro* experiments, we observed co-localization of Carbonic Anhydrase IX (CAIX) –a well-described hypoxia-induced gene (Lou et al., 2011) - in *EMP1*-TOM modified PDXs (Figure 3.38B).



**Figure 3.38: PDO7 upregulates *EMP1* in response to hypoxia.** **A.** *TOM* and *EMP1* mRNA expression levels by RT-qPCR after 48h of *in vitro* culture in hypoxia 1% compared to normoxia (21% oxygen). Data is represented as mean±s.d. of 3 experiments. **B.** Representative dual immunofluorescences of PDO7 *EMP1*-iCT#1 xenografts showing that CAIX hypoxic marker correlates with *EMP1*-*TOM* population. Yellow arrowheads mark double positive cells. Scale bars indicate 100µm.

*EMP1+ cells display EMT-like features*

Finally, we hypothesized that, since the gene signature of *EMP1*-High cells is enriched in extracellular matrix components such as laminins (A3, C1, C2), metalloproteinases, extracellular matrix protein 1 and collagen 16A1, *EMP1* expression might be regulated by extracellular matrix elements like collagen. Tumor-modified ECM is enriched in collagen I, and it has been shown that tumor cells acquire mesenchymal migratory traits upon contacting a rigid collagen I matrix (Carey et al., 2017, Peng et al., 2017, Zhang et al., 2018). We cultured our targeted organoids *in vitro* for five days with BME, collagen 0.75 mg/ml or 1.5 mg/ml and subsequently assessed *EMP1*-*TOM* levels by flow cytometry. Indeed, *TOM* levels were upregulated as collagen concentration increased, with more pronounced effects in the *TOM*-High population (Figure 3.39).

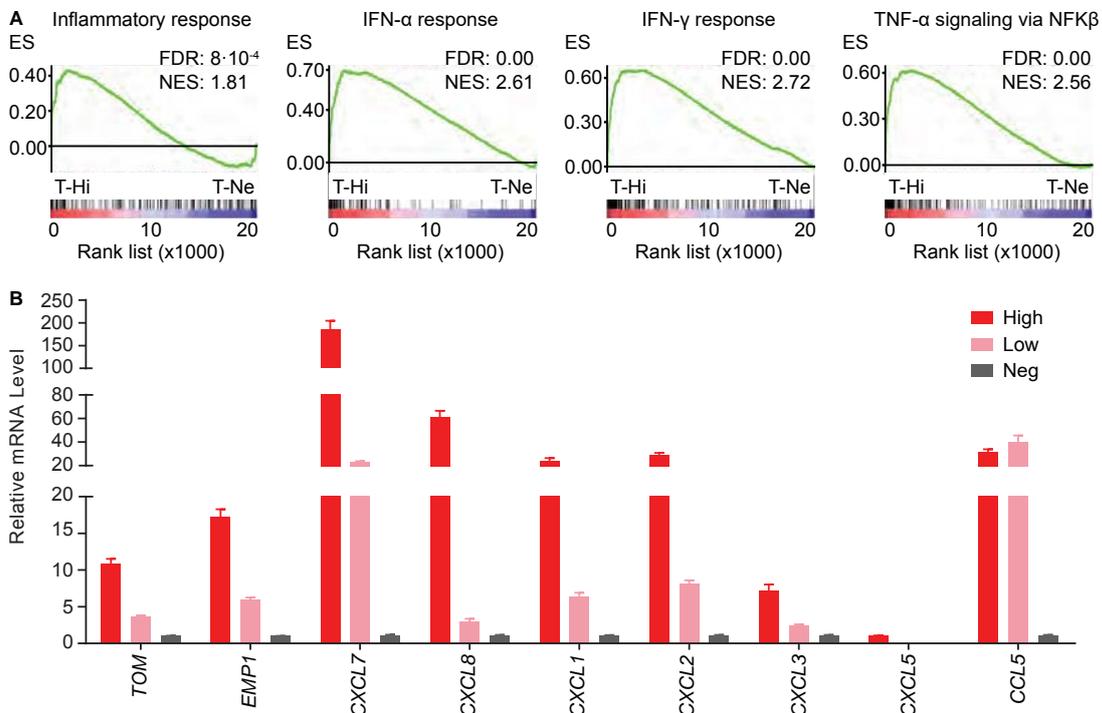


**Figure 3.39: *EMP1* expression increases in response to collagen.**

**Figure 3.39: EMP1 expression increases in response to collagen.** **A.** Flow cytometry analysis of TOM levels in 5-days *in vitro* cultured PDO7 EMP1-iCT#1 organoids with BME, collagen 0.75 mg/ml or 1.5 mg/ml. **B.** of cells in the TOM-High fraction vs exposure to collagen (the TOM -High fraction is defined as the 10 per cent brightest cells in the control condition, BME). EMP1-High population represented around 30 per cent of the total cells in collagen culture 1.5 mg/ml. Data is represented as mean±s.d. of 3 experiments. *Statistics: P-value was calculated using Student t-test. \*P-value<0.05.*

#### 3.4.4 Innate immunity cells are recruited to the tumor by EMP1+ cells

In addition to the mentioned GO terms detailed above, GSEA revealed a very significant enrichment in innate immune-related signatures within the EMP1-High population (Figure 3.40A). Moreover, in EMP1-High vs -Negative gene expression profiling experiments, the top genes defining the EMP1-High population, aside from EMP1 itself, were proteases and chemo-attractants. A list of the 20 most enriched genes is described in Table A.3. We quantified the expression of a range of neutrophil attractant chemokines in EMP1-High -Low and -Negative sorted populations from xenografts by RT-qPCRs (Figure 3.40B). The EMP1-High population expressed not only the chemokines *CXCL1*, 7 and 8 (shown by the microarray analysis), but also *CXCL2*, 3 and 5 as well as *CCL5*. From these results we hypothesize that EMP1+ cells may recruit innate immune cells such as neutrophils and MDSC to the tumor beds, a process that may help metastatic dissemination as has already been shown for other tumor types (Hiratsuka et al., 2006, Wculek et al., 2015, Steele et al., 2016).



**Figure 3.40: EMP1-High cells secrete myeloid attractant chemokines.**

**Figure 3.40: EMP1-High cells secrete myeloid attractant chemokines.** **A.** GSEA of several inflammatory signatures in EMP1-TOM High vs Negative gene expression comparison. **B.** RT-qPCR analysis of indicated genes in TOM-High, -Low and -Negative fractions purified by flow cytometry from EMP1-iCT#1 xenografts. The expression of several chemokines is upregulated in the EMP1-TOM High population. Data is represented as mean $\pm$ s.d. of 3 technical replicates.

#### 3.4.5 Establishment of a conditional EMP1+ cell ablation system

Altogether the accumulated data on EMP1+ cells suggested that this gene could mark a subset of metastatic tumor cells. To functionally dissect the metastatic potential of EMP1+ cells, we performed experiments of cell ablation in human PDOs. To this end, we took advantage of the genetically engineered inducible Caspase9 system integrated under the control of the *EMP1* locus.

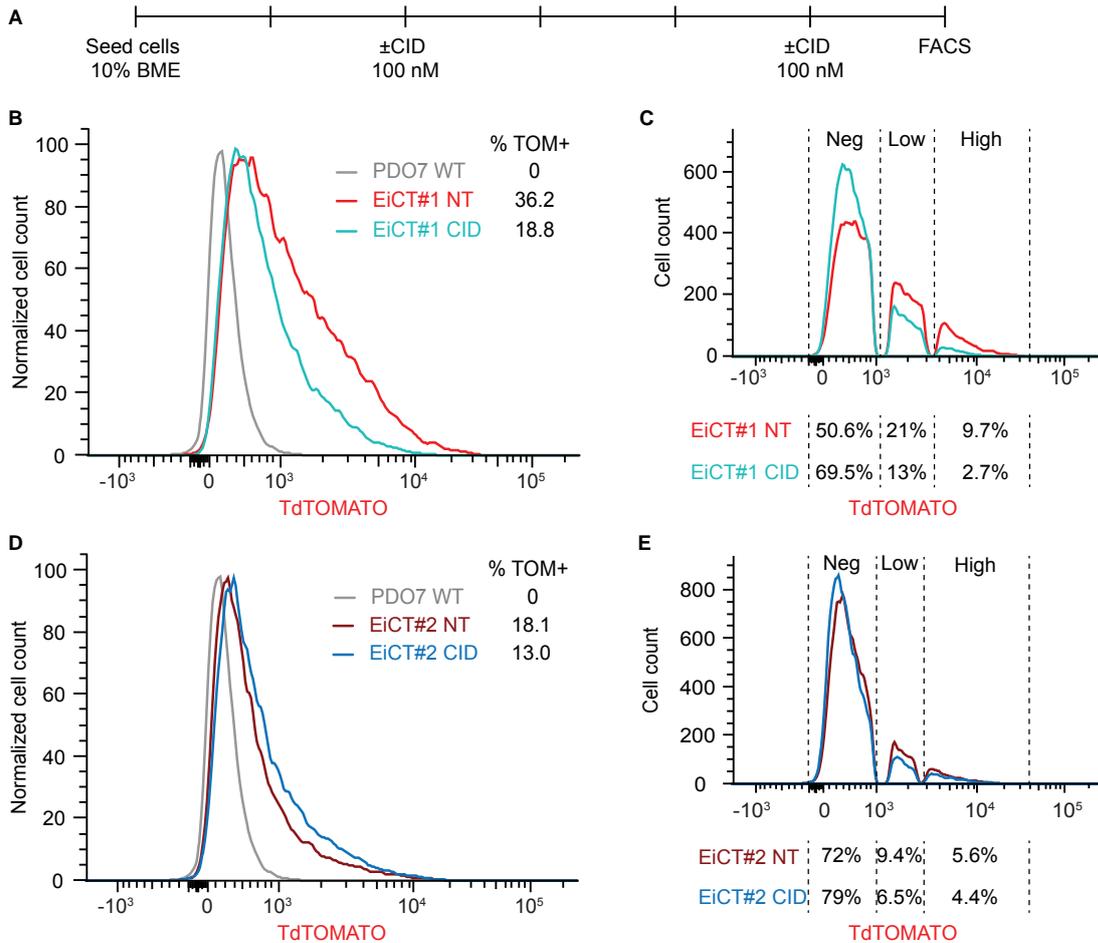
##### *iCasp9 induces EMP1-High cell death in vitro*

We tested the apoptotic potential of *iCasp9* inserted under the control of *EMP1* locus in PDO7. In order to facilitate drug diffusion, we cultured PDO7 EMP1-iCT#1 and #2 in 10% BME suspension for six days. We added CID to the media on days 2 and 5 and analyzed the effect on EMP1+ population by flow cytometry (Figure 3.41A). For homozygous integrated EMP1-iCT#1, FACS data revealed a decrease of 20% on whole TOM+ population (Figure 3.41B), with greater effect in the TOM-High population (Figure 3.41C). The effect was less obvious for heterozygous clone EMP1-iCT#2, with only a decrease of 5% in TOM+ population (Figure 3.41D) and lower affectation of TOM-High cells (Figure 3.41E). This result confirmed that *iCasp9* induced cell death of EMP1-*iCasp9* expressing cells in a dose-dependent manner.

##### *iCasp9 efficiently ablates EMP1-High cells in human xenografts*

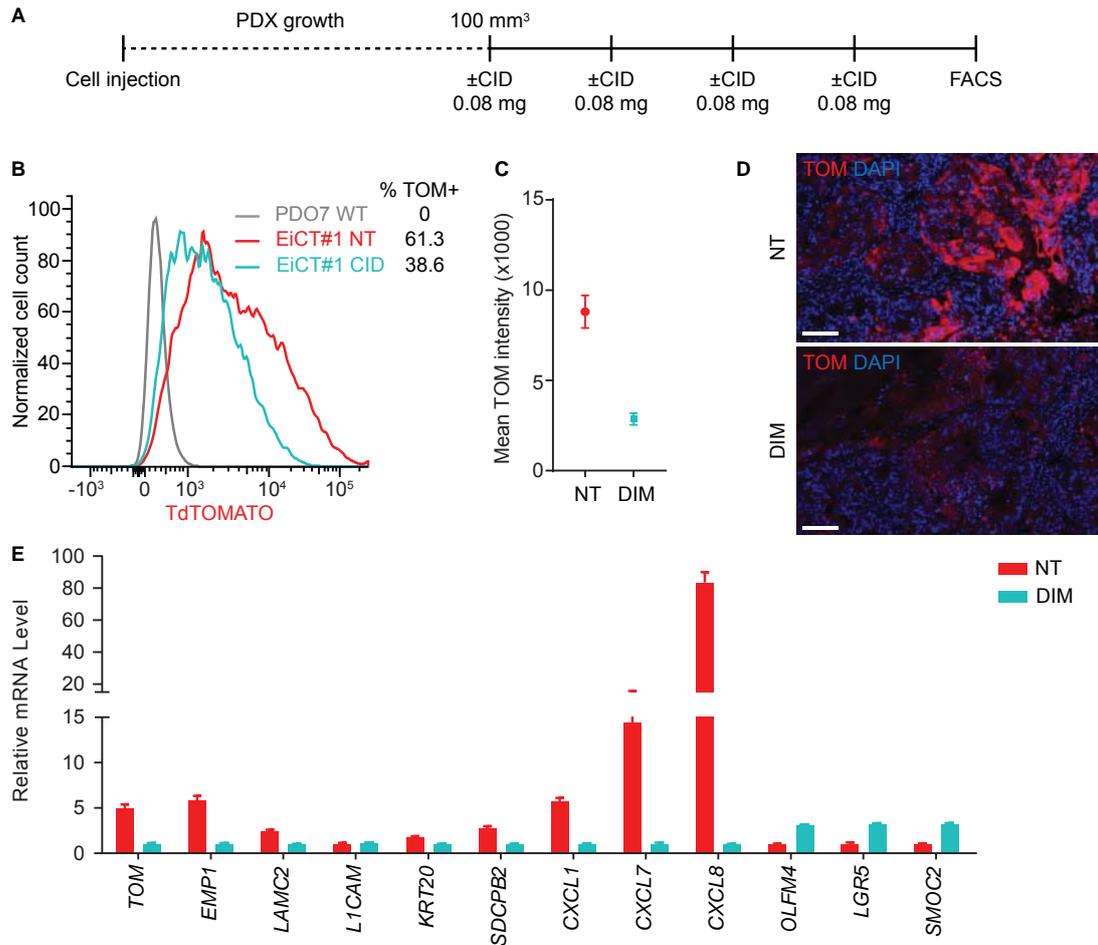
We then injected subcutaneously PDO7 EMP1-iCT#1 and #2 and, once tumors reached a size of 100 mm<sup>3</sup>, we treated the mice for 4 days with CID 0.08mg/kg (Figure 3.42A). We analyzed the effect of *iCasp9* dimerization by flow cytometry in disaggregated xenografts. PDO7 EMP1-iCT#1 responded well to CID-treatment as the 10-20% highest TOM+ cells of total EPCAM+ fraction had disappeared (Figure 3.42B) and total TOM fluorescent intensity in the tumor was decreased by 3-fold (Figure 3.42C). TOM immunofluorescence on xenografts faithfully showed TOM-High cells disappearance (Figure 3.42D). Moreover, mRNA expression analysis of total EPCAM+ population of non-treated vs CID sorted cells showed a reduction in both *TOM* and *EMP1* levels, as well as a decrease in differentiation markers and an increase in stemness genes. Interestingly, chemokine expression was also reduced in CID-treated condition (Figure 3.42E). This results proved that CID treatment was sufficient to eliminate the EMP1-TOM brightest population, which we hypothesized corresponded to the cells undergoing invasion.

As expected, heterozygous EMP1-iCT#2 responded less than clone #1 to the same treatment regime (Figure A.15A). TOM levels were reduced only by 5-10%, with most striking effects on the TOM-High subset (Figure A.15B), and total TOM intensity



**Figure 3.41: CID treatment induces apoptosis of EMP1-High cells in PDO7 *in vitro*.** **A.** Experimental setup for the iCasp9 induction *in vitro*. **B.** Flow cytometry quantification of TOM+ cells in non-treated vs CID-treated organoids of PDO7 EMP1-iCT#1. **C.** Representation of the total number of cells in the Negative, Low and High EMP1 fractions by FACS, and percentage of DAPI- cells that each fraction represents. **D and E.** Same representations for PDO7 EMP1-iCT#2, which was less responsive to CID-induced death *in vitro*.

in the tumor was decreased by 2-fold (Figure A.15C). Again, we observed a mild effect of CID treatment by TOM immunofluorescence on whole xenograft sections (Figure A.15D). In accordance with these results, RT-qPCR analysis of the whole EPCAM+ cell population revealed a moderate decrease in *TOM* and *EMP1* levels. The effect of treatment over differentiation/stemness ratios was neither so pronounced (Figure A.15E).

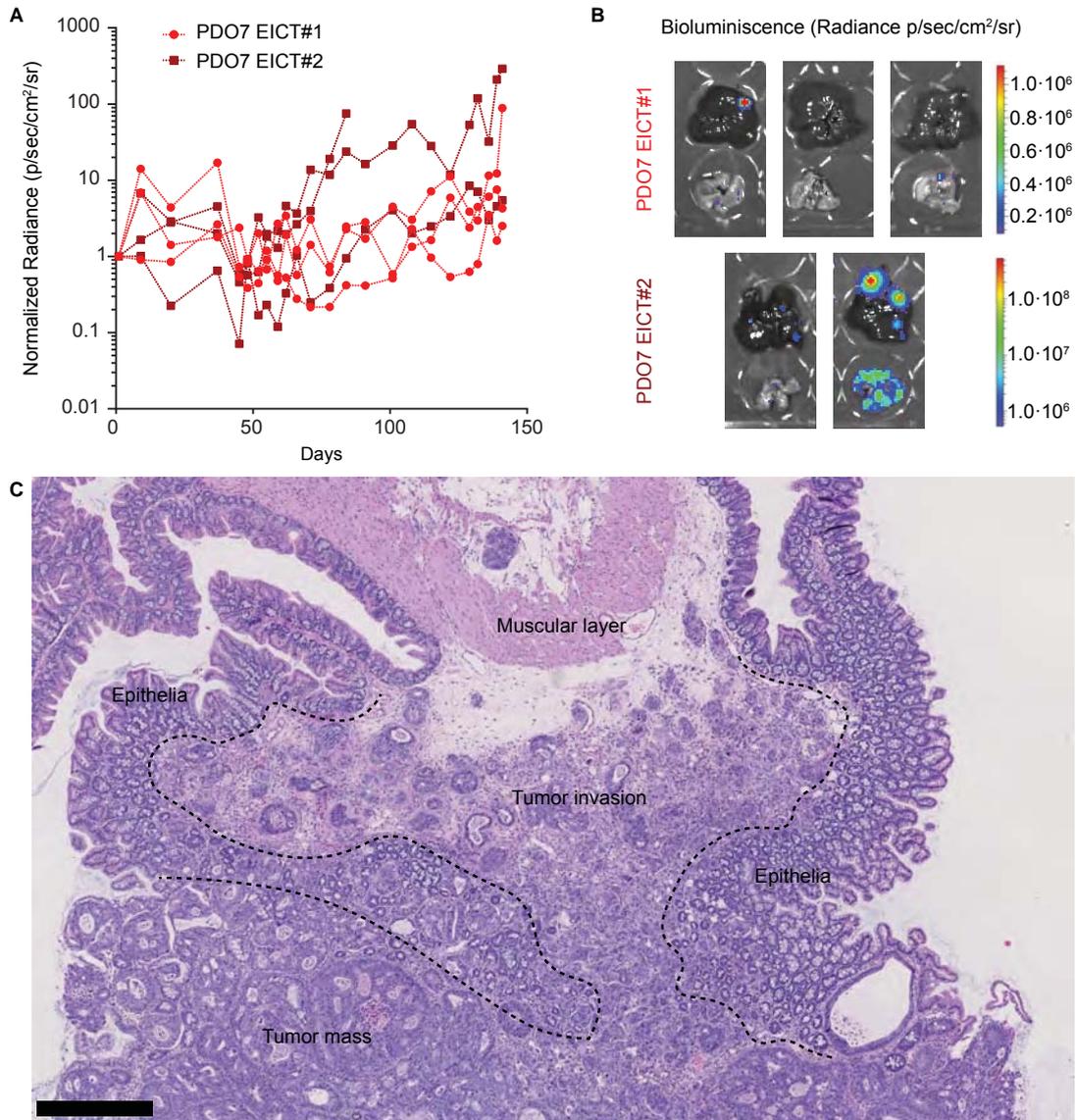


**Figure 3.42: EMP1-High fraction is ablated upon acute CID treatment of the PDX.** **A.** Experimental setup of acute CID treatment in subcutaneous xenografts. **B.** Flow cytometry analysis of TOM levels in DAPI-EPCAM+ cells from non-treated (NT) and dimerized (CID) disaggregated EMP1-iCT#1 xenografts. **C.** Mean TOM fluorescent intensity in DAPI-EPCAM+ cells. Results show mean±s.d. of 3 xenografts per condition. **D.** TOM immunofluorescence staining in paraffin sections of NT and CID xenografts. Note the decrease in total TOM+ cells in the CID condition. Scale bar indicates 100µm. **E.** RT-qPCR mRNA expression analysis for the indicated genes in EPCAM+DAPI- sorted cells from NT vs CID group. *EMP1* gene, as well as the invasive markers (*LAMC2*, *L1CAM* and chemokines) were decreased by treatment. Statistics: *P*-value was calculated using Student's *T*-test. \*\* *P*-value<0.01.

### 3.4.6 Failure of PDO6 and PDO7 to generate metastases precluded analysis of the role of EMP1+ CRC cells in the dissemination of the disease

In sum, by CRISPR/Cas9 knock-in we identified and characterized a subset of CRC cells with apparent invasive characteristics defined by *EMP1* gene expression. In addition, by introducing an inducible suicide cassette under the control of *EMP1* marker gene, we generated a powerful tool to study how the ablation of such EMP1-High population may affect the invasive potential of PDOs. In order to assess whether EMP1+ cells were necessary for the metastatic dissemination, we envisioned the following experiment: injection of PDOs that carry cell ablation cassettes in the caecum of mice followed by chronic CID treatment to eliminate EMP1-High cells in the pri-

primary CRC. As readout of the metastatic potential of non-treated vs CID-treated tumors we will use the number of liver metastatic nodules. As described below, we encountered a number of technical limitations that precluded the successful evaluation of metastatic potential of EMP1+ cells. These included the cell-implantation technique in the caecum, the low dissemination potential of PDOs, and also difficulties in defining the appropriate CID dosage necessary to eliminate EMP1-High cells without eliminating primary tumors. The following sections described the approaches we undertook in order to overcome these drawbacks.

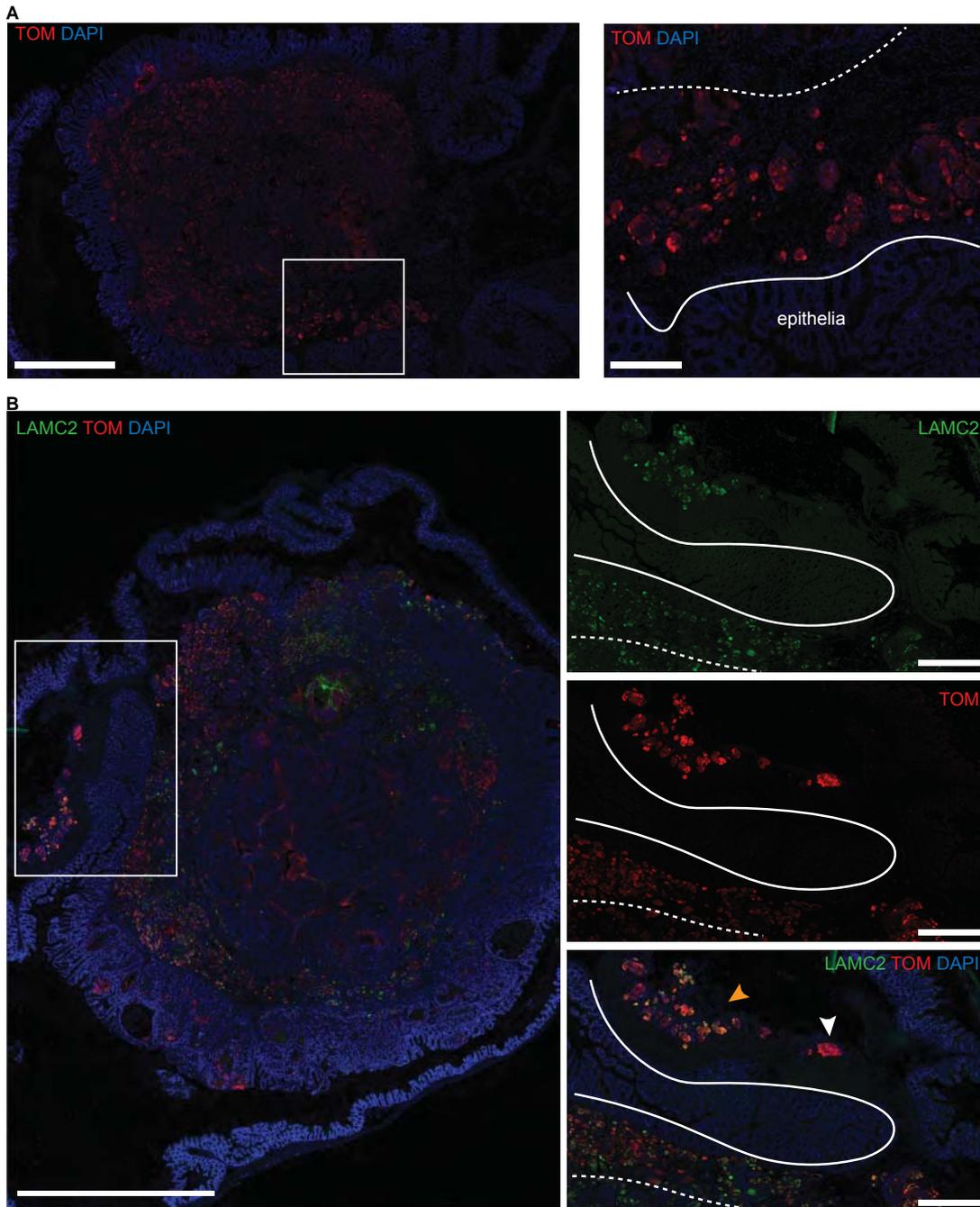


**Figure 3.43: PDO7 generates invasive tumors yet not metastasis upon injection in the caecum of NSG mice.** **A.** Intravital bioluminescence activity over time in mice injected with EMP1-iCT#1 and #2 cells in the caecum. Each dotted line represents an animal. **B.** *Ex-vivo* bioluminescence activity of livers (top) and lungs (bottom) measured after luciferine injection and sacrifice (maximum elapsed time between injection and sacrifice: 3 minutes). **C.** Representative image of a primary tumor invading the healthy epithelia. Dashed lines indicate the border between tumor tissue and mouse intestinal epithelia. Scale bar indicates 100 $\mu$ m.

To assess metastatic potential of EMP1+ cells, we devise a strategy based on depleting EMP1+ cell population from primary CRC. We injected cells from EMP1-iCT clones as small organoids directly in the caecum, between the serosa and the muscular layer (Céspedes et al., 2007). We injected 3 NSG mice of each PDO7 EMP1-iCT clone at cell density of 250000 cells per animal. PDO7 EMP1-iCT clones were transfected with a lentiviral Luciferase (Luc) expressing vector so that their behavior could be tracked *in vivo*. We measured bioluminescence weekly to control primary growth. Luc activity increased slowly until the end of the experiment, and at endpoint (day 141) all animals presented CRCs (Figure 3.43A). We measured *ex-vivo* bioluminescent activity in livers and lungs, but only one animal out of six presented significant metastatic spread (Figure 3.43B). Hematoxylin/eosin staining of liver and lungs tissue sections confirmed that there were no metastases present except in one NSG mouse (not shown).

Despite lack of metastasis, we studied the *EMP1* expression pattern in primary tumor samples. Tumors had a mean size of 70 mm<sup>3</sup>, and had started to invade the normal epithelia, disrupting intestinal crypt architectures (Figure 3.43C). TOM+ cells were present in all tumor areas, yet the brightest ones were found in the invasive fronts (Figure 3.44A). Moreover, LAMC2 pattern largely overlapped with that of EMP1-TOM-High, although some migratory EMP1+ cell clusters did not express LAMC2. Interestingly, tumor buds were highly TOM+ and resembled cells undergoing collective migration (Figure 3.44B). We also stained for EMP1-TOM+ cells in liver and lung sections from the only one tumor that generated metastases in this experiment. Both lung and liver metastatic nodules contained EMP1-TOM+ cells (Figure A.17A and B).

In sum, histological analysis of primary samples corroborated that *EMP1* gene was highly expressed by a subset of migratory-like cells, confirming that it could be used as a marker for this population. Nevertheless, lack of tumor dissemination in this system impeded definitive proof that EMP-High cells were the cell of origin of metastasis.



**Figure 3.44: EMP1+ cells localized at primary tumor invasive fronts and correlated with LAMC2+ cells.** **A.** Representative TOM immunofluorescence in a paraffin section of orthotopically grown PDO7 EMP1-iCT#1. TOM+ cells accumulate at tumor buds, as observed in the magnification. Dashed line indicates the separation between primary mass and invasive front and white line separates tumor and healthy epithelia. Scale bar indicates 1mm and 250 $\mu$ m in the inset. **B.** Example of TOM/LAMC2 dual immunofluorescence in a primary tumor, note that all LAMC2+ cells are also TOM+ (yellow arrowheads). Inset shows a magnification of tumor buds invading the muscular layer, not that all cells are highly positive for TOM and some buds are TOM+LAMC2- (white arrowheads). Dashed line indicates the separation between primary mass and invasive front and solid lines indicates the border between tumor and mouse epithelia. Scale bar indicates 2.5mm for the whole tumor image and 500 $\mu$ m for the inset.

### Efforts to increase metastatic efficiency in EMP1-iCT

Unfortunately, the lack of consistent tumor cell dissemination in this experimental setting impeded the analysis of the role EMP1-High cells in the initiation of metastasis. In order to increase the metastatic potential of PDO7 EMP1-iCT clones, we aimed at establishing organoids from liver and lung metastasis that display increase adaptation to colonize these organs. (Bruns et al., 1999). For this purpose, we injected PDO7 EMP1-iCT#1 in the spleen of NSG mice, a protocol that enable cells to directly enter the blood stream and colonize the liver and lungs. Two months after injection, cells had efficiently formed liver metastases (Figure 3.45A). We disaggregated 6 nodules and seeded them in organoid conditions. Human cells were *in vitro* selected based on puromycin resistance included in the Luc-expressing vector (Figure 3.45B).

We then injected met-derived PDO7 EMP1-iCT#1 organoids orthotopically to evaluate whether they were more metastatic. In particular, we injected three NSG mice with 250000 cells per animal and sacrificed them at day 113 (Figure 3.45C). None of the animals presented metastasis in livers or lungs. One animal developed carcinomatosis in the abdominal musculature through primary adherences, which caused an abnormal increase of bioluminescence (marked by \* in Figure 3.45C). The EMP1-TOM patterns and expression levels of the reporter were similar to those of parental cells derived tumors (Figure 3.45D).

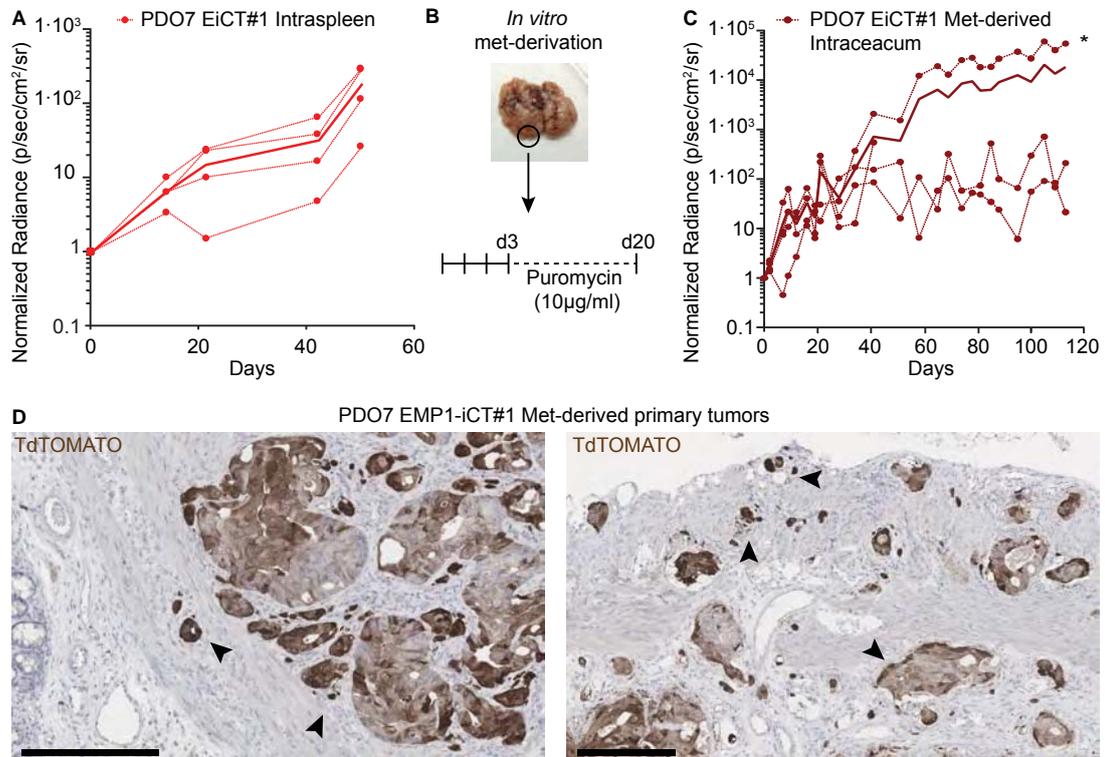
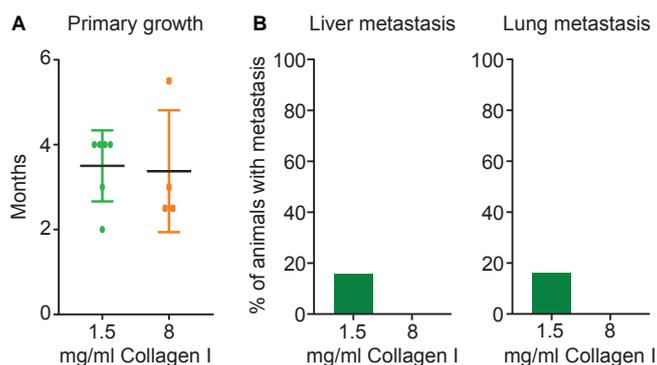


Figure 3.45: Met-derived cells do not present higher invasive potential in PDO7.

**Figure 3.45: Met-derived cells do not present higher invasive potential in PDO7.** **A.** Bioluminescence activity in mice injected via intraspleen with PDO7 EMP1-iCT#1 (500000 cells/animal). Each dotted line represents an animal, and the solid line indicates the mean radiance of the group. **B.** Liver nodules after 60 days of growth and schematic summary of the *in vitro* metastasis-derivation process and culture establishment. **C.** Bioluminescence activity of PDO7 EMP1-iCT#1 Met-derived cells injected intracaecum in NSG mice. \* indicates animal with abdominal carcinomatosis. Solid line represents the mean radiance of the group. **D.** Representative TOM immunostainings of the invasive fronts in primary PDO7 EMP1-iCT#1 Met-derived primary tumors. Arrowheads point at cells invading the muscle. Scale bars indicate 250 $\mu$ m

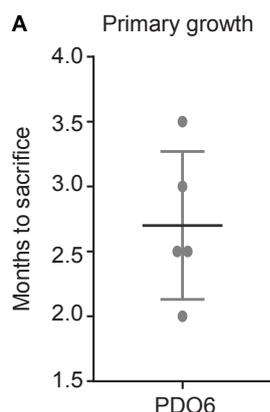
As neither parental nor met-derived PDO7 cells were able to efficiently generate metastases from intracaecum injected cells, we hypothesized that perhaps the choice of orthotopic injection could be impairing growth and spread of human cells in mice. In order to overcome this hurdle, instead of injecting cells we inserted whole pieces of subcutaneous-grown tumors in the caecum of the animals (*nesting*). We first injected PDO7 EMP1-iCT#1 met-derived subcutaneously in Beige/Scid mice, and, when tumors reached a size of 100-150 mm<sup>3</sup>, we collected them and cut the xenografts to stitch 1 mm<sup>3</sup> PDX pieces directly to the caecum of NSG animals. Primary CRCs grew much faster than when injecting cultured organoids directly (60 days compared to 120 days), yet tumors displayed a similar histology (Figure A.18A). Unfortunately, this implantation protocol neither did generate metastasis in livers or lungs (Figure A.18B).

We next tested a recently published protocol of CRC generation from organoids (Fumagalli et al., 2017) that is based on injecting cells in combination with a collagen I matrix. Briefly, we grew organoids for 5 days embedded in BME matrix and one day before injection we re-plated them in collagen I drops (250000 cells per drop). Drops were then inserted below the serosa in the caecum. We compared primary tumor growth and metastasis number in 1.5 and 8 mg/ml collagen drops (Figure 3.46). Only the 1.5 mg/ml condition presented metastases in one animal, an efficiency similar to previous experiments, therefore this setting did not enhance metastatic efficiency of PDO7.



**Figure 3.46: Collagen drop insertion does not increase PDO7 metastatic potential.** **A.** Mean survival of each group post-surgery. Whole collagen drops containing 250000 PDO7 WT cells were injected per animal. Data is represented as mean $\pm$ s.d. (n=6 mice in 1.5mg/ml condition and n=5 mice in 8mg/ml). **B.** Percentage of animals presenting liver or lung metastasis at the time of sacrifice (n=6 mice in 1.5mg/ml condition and n=5 mice in 8mg/ml). Only 1 out of 6 in 1.5mg/ml developed metastases, reproducing the results of standard intraecum cell injection.

Finally, we tested the metastatic capacities of PDO6, another patient-derived organoid that we had edited with cassettes to track EMP1+ cells (see section 3.4.3). We implanted PDO6 tumor pieces in the caecum. Despite primary tumors grew in about three months mice did not presented liver nor lung tumor nodules (Figure 3.47).



**Figure 3.47: Orthotopically implanted PDO6 grows efficiently in primary site.** Survival post-surgery of NSG mice with PDO6 WT nested pieces. Data is represented as mean $\pm$ s.d. n=5 animals.

### 3.5 MOUSE TUMOR ORGANOIDS AS A MODEL SYSTEM

To overcome the limitations of the PDO system to study metastases, we decided to leverage Mouse Tumor Organoid (MTO) biobank recently generated in the lab (Tauriello et al., 2018). As explained in the introduction, these MTOs derived from intestinal tumors arising in GEMMs with compound mutation in the main driver pathways. Some of these MTOs display a high efficiency of metastasis even upon transplantation in immunocompetent syngeneic C57BL/6 mice (Tauriello et al., 2018).

#### *Metastatic derived MTOs are enriched in EMP1 signature*

We first took advantage of the gene expression profile data of *in vitro* cultured MTOs previously generated in the laboratory. In particular we analyzed several quadruple mutant MTOs (L-AKTP, derived from tumors arising in LGR5+ cells and containing mutations in APC, KRAS, TP53 and TGF- $\beta$ R2 genes), and ranked them according

to their enrichment in human *EMP1-High* gene signature that we had previously obtained from EMP1-Tomato PDO7 xenografts. For this analysis we used genes enriched at least 2-fold in EMP1-High vs -Negative cells (Table 3.4). The collection was generated by culturing organoids derived from both spontaneously generated primary tumors (parental) and metastatic tumors. Metastatic tumors were derived from liver nodules that generated after intracaecum or intraspleen injections of parental organoids in C57BL/6 mice. We found that, overall, metastasis-derived MTOs upregulated EMP1 signature compared to their parental organoids, supporting the idea that *EMP1* could be indeed a metastatic marker. We selected the parental MTO93 for further experiments.

MTO	Origin	Parental	Enrichment score
MTO140	Liver metastasis from intracaecum injection	MTO93	0.350368
MTO182	Liver metastasis from intracaecum injection	MTO34	0.349502
MTO49	Liver metastasis from intracaecum injection	MTO34	0.3339777
MTO163	Liver metastasis from intracaecum injection	MTO105	0.328378
MTO48	Liver metastasis from intracaecum injection	MTO34	0.323674
MTO138	Liver metastasis from intracaecum injection	MTO34	0.320654
MTO129	Liver metastasis from intracaecum injection	MTO68	0.299362
MTO34	Parental	—	0.295218
MTO94	Liver metastasis from primary tumor	MTO93	0.293162
MTO113	Liver metastasis from intraspleen injection	MTO34	0.290347
MTO130	Liver metastasis from intraspleen injection	MTO68	0.280313
MTO105	Parental	—	0.272287
MTO95	Liver metastasis from primary tumor	MTO93	0.244984
MTO68	Parental	—	0.206752
MTO93	Parental	—	0.184349

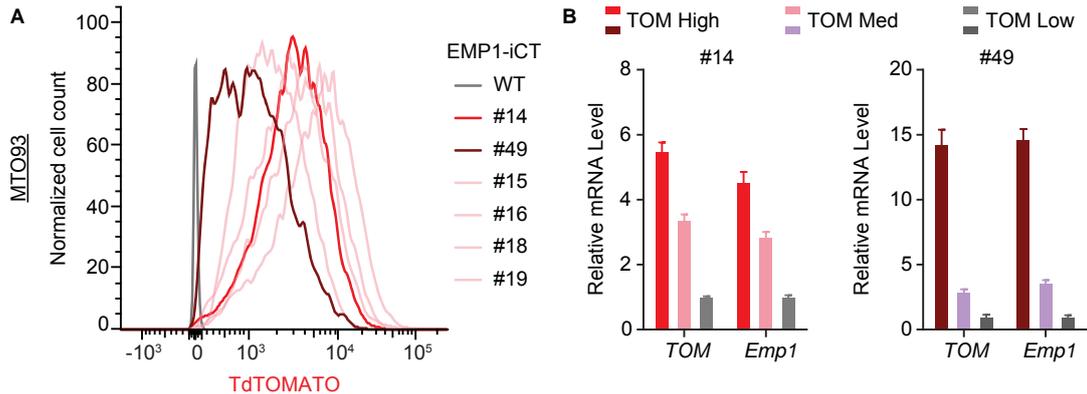
**Table 3.4: Metastasis-derived MTOs are enriched in *EMP1* gene signature.** List of MTOs ranked according to the correlation of their gene expression profile with the EMP1-High signature. The score indicates 0=no enrichment, 1=absolute coincidence. Parental samples are marked in green, and accumulate the lowest scores.

#### *MTOs express high levels of Emp1 in vitro*

We inserted a the iCasp9-TOM ablation cassette in the *Emp1* locus using an approach equivalent to that used for PDOs. Briefly, we swapped the human homology arms of the iCasp-TOM donor for the corresponding *Emp1* mouse sequences. Next we nucleofected MTO93 with donor plasmid and guide-Cas9-iRFP plasmid. FACS-isolation of targeted cells and clonal expansion was performed as previously described; this is, a first round of iRFP+ sorting 3 days post nucleofection and long term sorting based on TOM positivity to derive single cell clones (Figure A.19). Results are summarized in Table A.1 and original integration PCR gels can be found in Figure A.20. Of note,

we did not manage to find any homozygous targeted clone.

We prescreened MTO93 EMP1-iCT derived clones using flow cytometry. This analysis revealed that the clones were largely positive for Emp1-TOM (Figure 3.48A). We picked two representative clones that exhibited different TOM intensities for subsequent analysis. As before, we further confirmed correct integration of donor plasmid by analysis of FACS-purified High, Medium and Low TOM expressing cells. RT-qPCR analysis showed that *TOM* mRNA levels paralleled *Emp1* mRNA expression (Figure 3.48B).



**Figure 3.48: MTOs express high EMP1 levels *in vitro*.** **A.** Flow cytometry analysis of TOM levels in several MTO93 EMP1-iCT clones 5 days *in vitro* grown. Clones selected for further experiments are highlighted in red (#14) and purple (#49) lines. **B.** Relative mRNA expression level by RT-qPCR of FACS-sorted TOM-High, -Medium (Med) and -Low populations. *TOM* expression correlates with *Emp1*. Data is represented as mean±s.d. of 3 technical replicates.

### 3.5.1 EMP1+ cells retain invasive behavior in MTO-derived tumors

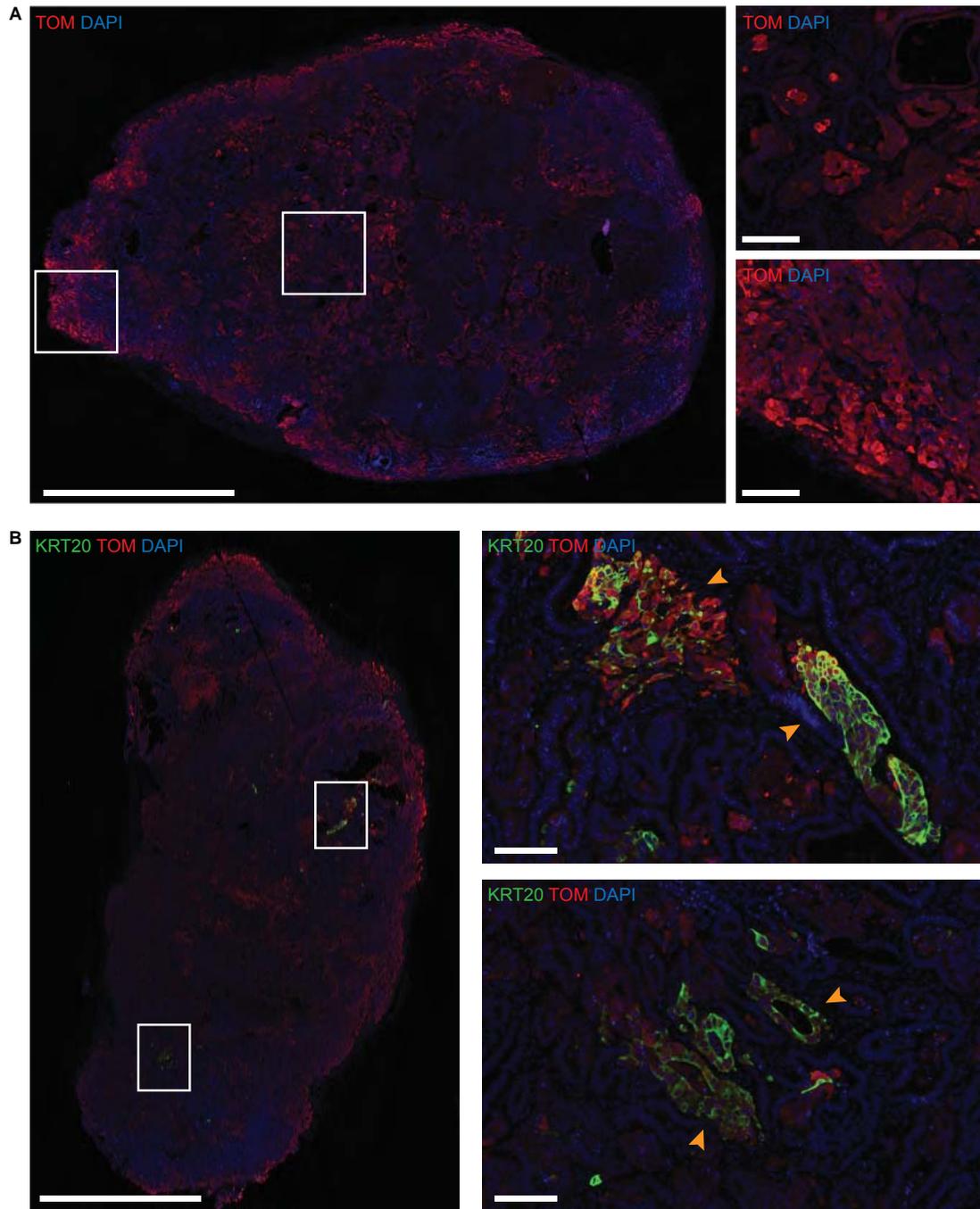
We confirmed that MTO93 generated tumors that reproduced the population heterogeneity present in patient-derived xenografts. Specifically, we were interested in assessing whether *Emp1* was also marking a subpopulation of differentiated-like cells that presented migratory characteristics.

#### *Emp1* marks a differentiated-like population in MTO93-derived xenografts

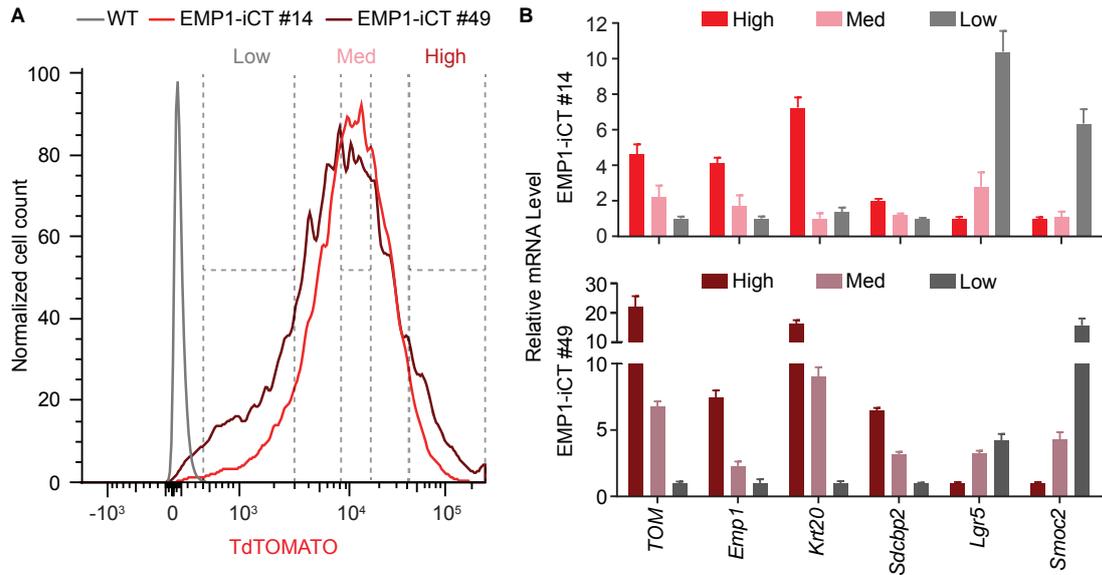
We first injected subcutaneously MTO93 EMP1-iCT cells in C57BL/6 mice and sacrificed the animals when tumors reached 150 mm<sup>3</sup> size. Histological analysis of xenografts showed compact glandular structures surrounded by stromal cells, with 60-70% of the epithelial compartment expressing some degree of TOM reporter. Tumor borders contained many TOM bright cells (Figure 3.49A). These tumors, however, contained fewer KRT20+ than human tumor. Nevertheless KRT20+ areas overlapped with TOM+ regions (Figure 3.49B).

In accordance with immunofluorescence analysis, flow cytometry of MTO93 disaggregated tumors revealed a broad TOM+ spectrum (Figure 3.50A). Isolation and mRNA extraction of TOM-High, -Medium and -Low cell populations confirmed there was

heterogeneity within the TOM population, with enterocytic and pan-differentiation markers (*Krt20*, *Sdcbp2*) mostly expressed by TOM-High cells and stem cell markers (*Lgr5*, *Smoc2*) enriched in the TOM-Neg fraction (Figure 3.50B).



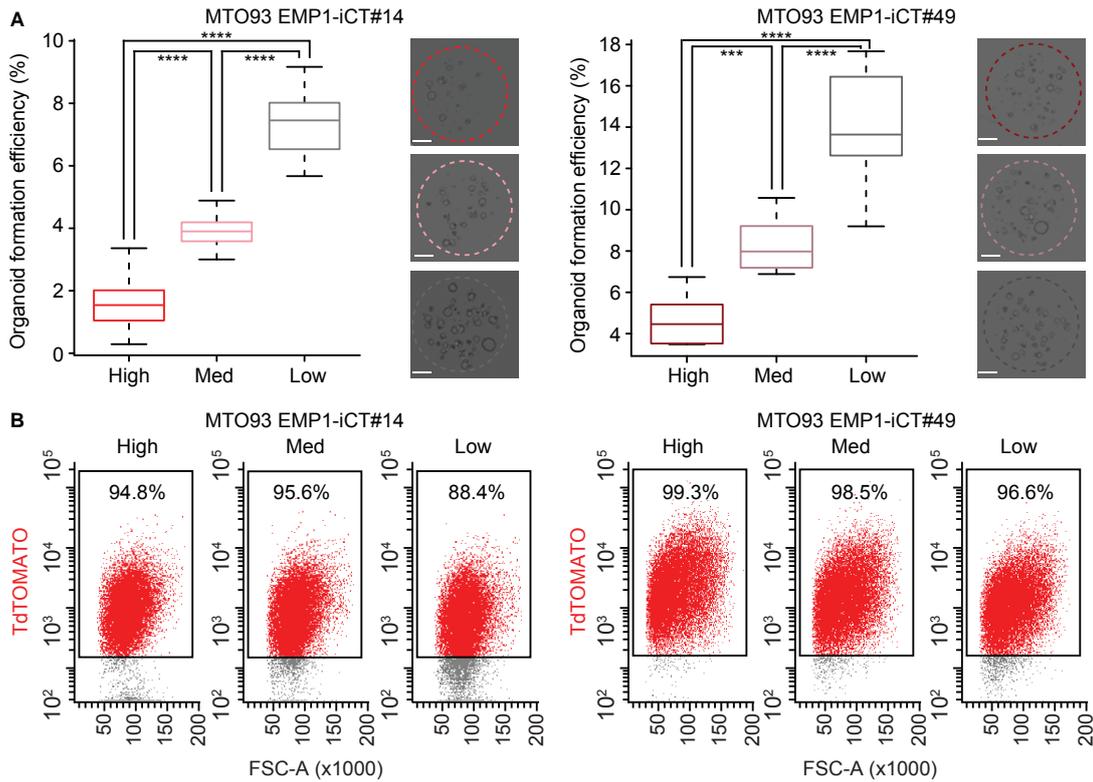
**Figure 3.49: MTO93 forms glandular-like, poorly differentiated tumors.** **A.** Representative EMP1-TOM staining in sections from a xenograft generated by subcutaneous injection of MTO93 EMP1-iCT#14 cells. Glandular structure magnifications are indicated with white squares. Tom brightest cells accumulate at the tumor borders. Scale bars indicate 2.5mm for the whole tumor and 100 $\mu$ m for the insets. **B.** TOM and KRT20 dual immunofluorescence staining on sections from a representative MTO93 EMP1-iCT#49 derived tumor. Very few glands express KRT20 protein, but they correlate with TOM+ cells as indicated in by yellow arrowheads in the insets. Scale bars indicate 2.5mm for the whole tumor and 100 $\mu$ m for the insets.



**Figure 3.50: *Emp1* is expressed by enterocytic-like cells in MTO-derived xenografts. A.** Flow cytometry analysis of TOM levels in EGFP+ cells from disaggregated subcutaneous xenografts of MTO93 EMP1-iCT#14 and #49. Dashed lines indicate the gates for purification of TOM-High, -Med and -Low cells. **B.** RT-qPCR analysis of differentiation and stem cell gene marker expression in the three sorted fractions for both clones.

#### *EMP1-mouse cells display low organoid formation capacity*

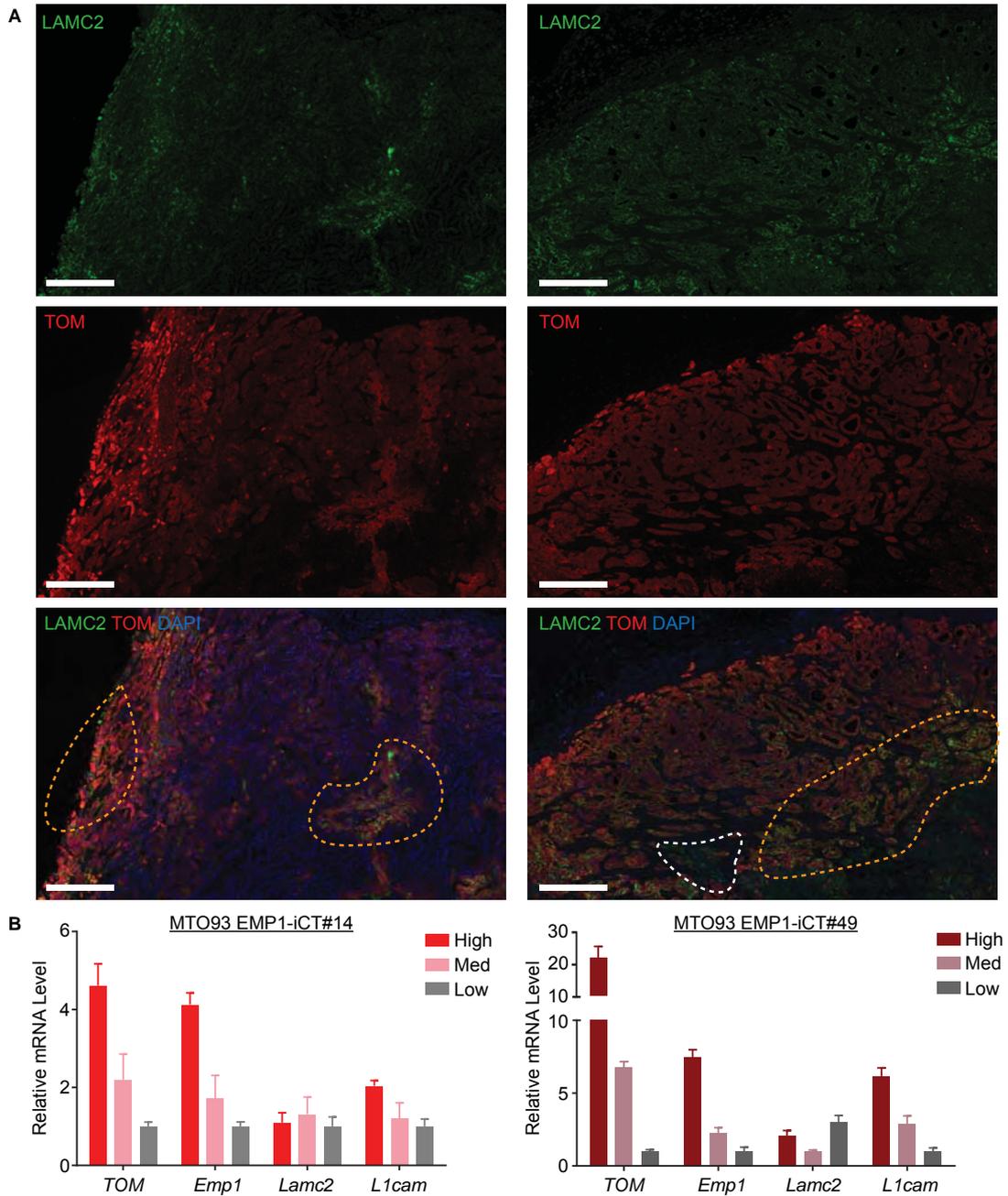
We next assessed organoid forming capacity of TOM-High, -Low and -Negative tumor cells isolated by FACS from disaggregated xenografts. The EMP1-Negative fraction retained the highest organoid formation potential, which is in accordance with the high levels of expression of stem cell genes (Figure 3.51A). We confirmed cell plasticity in the expanded organoids as all three population generated organoids that regained similar levels of TOM expression in only 7 days (Figure 3.51B).



**Figure 3.51: *Emp1* expression determines organoid formation capacity in MTO93. A.** Organoid formation capacity of TOM-High, -Med and -Low FACS-purified cells from disaggregated xenografts for both MTO93 EMP1-iCT#14 and #49, and representative images of each condition. Results are plotted as mean $\pm$ s.d. of 4 experiments (#14) and 2 experiments (#49) of n=4 wells (1500 cells/well) per experiment. Scale bars indicate 1mm. **B.** Representative FACS analysis of TOM levels after 7 days of *in vitro* growth of each subpopulation demonstrating reversion to the parental phenotype.

*EMP1 does not completely correlate with the expression of invasive markers in mouse tumors*

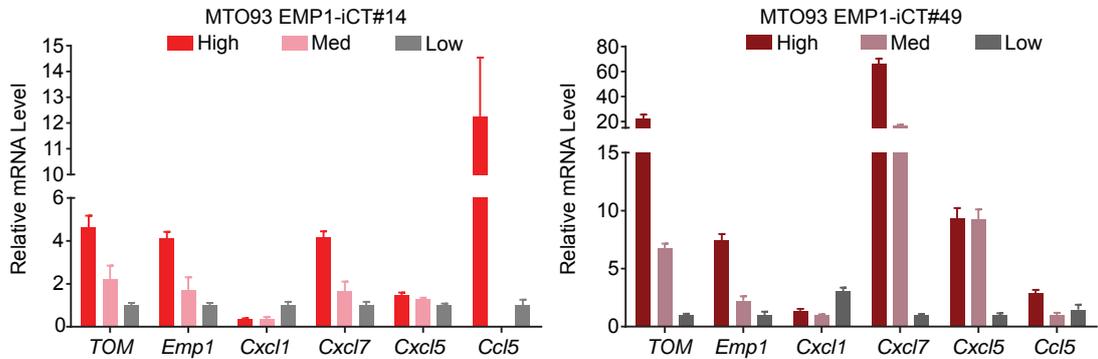
We next used subcutaneous grown tumors to determine the expression of the invasive markers *Lamc2* and *L1cam* in relationship with *Emp1* expression. By immunofluorescence we detected abundant LAMC2+ cells yet, unlike what we observed in humans (see **section 3.4.3**) not every LAMC2+ cells expressed *Emp1* (Figure 3.52A). Correspondingly, *Lamc2* mRNA was not specifically upregulated in any of the three EMP1-TOM sorted populations (Figure 3.52B). L1CAM protein was expressed at very low levels in MTO-derived isografts by IF (data not shown) although its mRNA is marginally enriched in the EMP1-TOM-High cell population (Figure 3.52B).



**Figure 3.52: LAMC2 is broadly expressed in MTO93 xenografts.** **A.** Representative dual immunofluorescence staining of TOM and LAMC2 in MTO93 EMP1-iCT subcutaneous xenografts. LAMC2 is expressed in most cells, and it partially overlaps with TOM+ cells. Yellow dashed lines indicate double positive areas and white dashed lines delimit LAMC2+TOM-areas. Scale bars indicate 250µm. **B.** mRNA expression level of *LAMC2* and *L1CAM* genes in TOM-High, -Med and -Low FACS-purified populations from disaggregated xenografts of MTO93 EMP1-iCT #14 and #49.

### EMP1-High cells are expressing neutrophil attractant chemokines

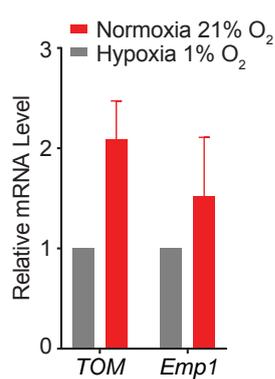
We also studied whether EMP1-high cells in MTO93 expressed elevated chemokine levels similar to their human counterparts. By RT-qPCR we could observe that of the most enriched chemokines in PDO7, *Cxcl7*, *Cdcl5* and *Ccl5* were upregulated in the EMP1-High populations. *CXCL8* does not have a homolog in mouse (Figure 3.53).



**Figure 3.53: EMP1+ cells express neutrophil attractant chemokines in mice.** RT-qPCR analysis reports relative expression levels of several chemokines in MTO93 subcutaneous xenograft-isolated cells according to their *EMP1* levels.

### Effects of hypoxia on *Emp1* expression in MTOs

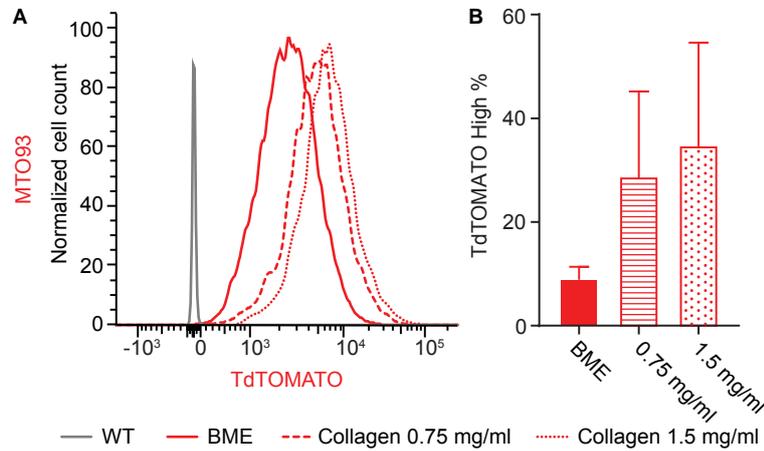
In order to gain more insight into the putative mechanisms by which *Emp1* gene program is induced, we grew MTO93 EMP1-iCT#14 cells 3 days *in vitro* and then kept them for two additional days in normoxia (21% O<sub>2</sub>) or hypoxia (1% O<sub>2</sub>) conditions. mRNA RT-qPCR of *TOM* and *Emp1* levels showed that MTO93 moderately increased *TOM* levels by 2-fold in response to hypoxia (Figure 3.54).



**Figure 3.54: *Emp1* gene expression responds to oxygen availability.** *TOM* and *EMP1* mRNA expression levels by RT-qPCR after 48h of *in vitro* culture in hypoxia compared to normoxia. Data is represented as mean  $\pm$  s.d. of 2 experiments.

### Collagen co-culture induces *Emp1* expression

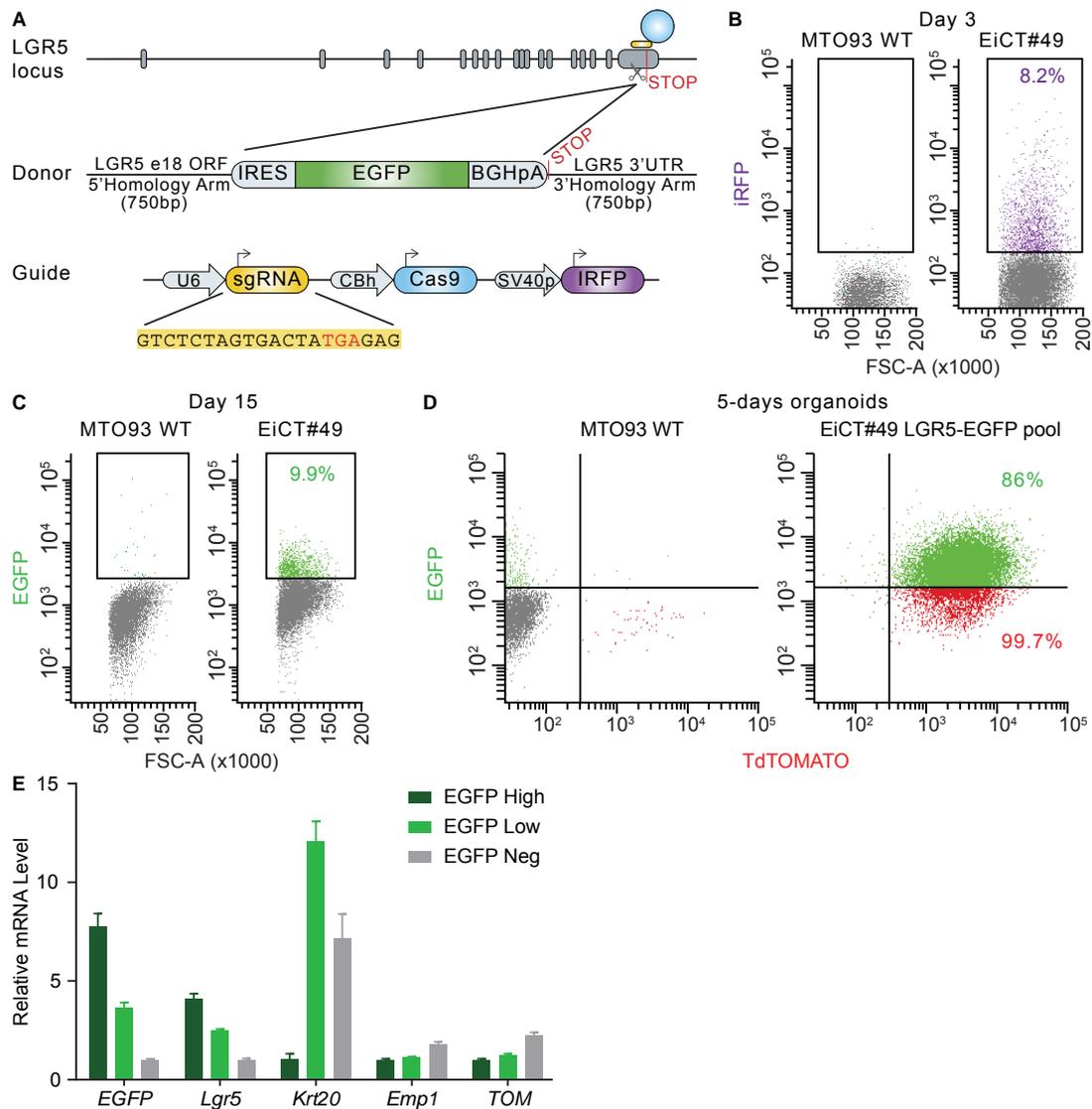
Finally, we tested whether MTO93 was as responsive to extracellular matrix composed of collagen-I as we showed for PDO7 (see section 3.4.3). We cultured the knock-in MTO93 clones *in vitro* for five days with BME, collagen I 0.75 mg/ml or 1.5 mg/ml. TOM levels were upregulated in the collagen I-cultured cells (Figure 3.55).



**Figure 3.55: MTO93 upregulates *Emp1* expression in response to *in vitro* collagen coculture.** **A.** Representative flow cytometry plot of TOM levels in *in vitro* grown MTO93 EMP1-iCT#14 with different 3D matrixes. (BME 100%, collagen 0.75 mg/ml, collagen 1.5 mg/ml). **B.** Quantification of the % of TOM-High cells in each condition (EMP1-High fraction is established as the 10% brightest population in BME100% condition). Data represents mean $\pm$ s.d. of 4 experiments. Statistics: *P*-value was calculated using Student's *t*-test \**p*value<0.05

### 3.5.2 *LGR5* and *EMP1* populations overlap in MTO93

In order to better dissect heterogeneity within the EMP1+ cell population, we inserted an EGFP reporter under the control of *Lgr5* gene in MTO93 EMP1-iCT clones. (Figure 3.56A). Following the same strategy described throughout this thesis, we nucleofected MTO93 EMP1-iCT clones and sorted iRFP+ cells 3 days post nucleofection. 15 days after the short-time sorting we selected EGFP+ cells by FACS and allowed the individual cells to expand monoclonal organoids (Figure 3.56B and C). Only MTO93 EMP1-iCT#49 presented correct *LGR5*-GFP knock-in integration (Figure A.21). Flow cytometry analysis of *in vitro* grown MTO93 EMP1-iCT#49/*LGR5*-EGFP pool revealed most cells were EGFP+ (86%) and TOM+ (99.7%) (Figure 3.56D). Analysis EGFP-High, -Low and -Negative populations confirmed that *EGFP* reported *Lgr5* levels (Figure 3.56E). Due to time constrains we did not derive single-cell clones from the pool.



**Figure 3.56: Tagging *LGR5* gene in MTO93.** **A.** CRISPR/Cas9 design of LGR5-EGFP knock-in for murine cells. **B.** FACS-purification of iRFP<sup>+</sup> cells 3 days post nucleofection. Only MTO93 EMP1-iCT#49 nucleofection is shown. **C.** EGFP<sup>+</sup> cells FACS-separation cells at long-term post-nucleofection. **D.** Representative flow cytometry analysis of 5-days *in vitro* grown MTO93 EMP1-iCT#49 LGR5-EGFP pool organoids. Most of the cells (86%) are double positive. **E.** Relative mRNA quantification of EGFP-High, -Low and -Negative fractions sorted from *in vitro* MTO93 EMP1-iCT#49 LGR5-EGFP organoids. Data represents mean±s.d. of 3 technical replicates.

#### *Emp1* expression allows a better discrimination of the CSCs

We injected MTO93 EMP1-ICT#49 LGR5-EGFP pool subcutaneously in C57BL/6 mice and obtained dual labelled xenografts. These tumors presented a glandular organization with clear separation of the LGR5-EGFP and EMP1-TOM glands overall, yet in some areas we could observe double positive cells (Figure 3.57A). FACS analysis of grown tumors revealed that all cells expressed some level of TOM, and that 50% of the EMP1-TOM<sup>+</sup> population was also EGFP<sup>+</sup> (Figure 3.57B). Despite broad TOM expression, EGFP-LGR5<sup>+</sup> EMP1-TOM-Low cells differentially expressed stem cell markers

(*Lgr5*, *Smoc2*), suggesting they were true CSC. Conversely, *Emp1* gene and differentiation markers *Sdcbp2* and *Krt20*, as well as *Cxcl7*, were heavily downregulated in this population. In addition, TOM+ populations, regardless of their EGFP status (+ or -), expressed similar levels of *Emp1* (Figure 3.57C). Therefore, we can conclude that *Emp1* allowed a better discrimination between stem cell-like cells (EGFP+TOM-) and a fraction of transient amplifying, perhaps more differentiated tumor cells (EGFP+TOM+). A particularly interesting observation is that both EGFP+TOM+ and EGFP-TOM+ cells display upregulation of neutrophil chemoattractants, perhaps suggesting a role for this population in the regulation of innate immunity.

Organoid formation confirmed that the results of this assay are mostly determined by stem cell gene expression, as the EGFP+TOM- population showed the highest capacity to grow organoids, followed by EGFP+TOM+ cells. There were no significant differences between EGFP- TOM+ or TOM- populations (Figure 3.57D).

**Figure 3.57: LGR5 and EMP1 gene expression partially overlap in MTO93 xenografts. A.** Representative immunofluorescences against LGR5-EGFP and EMP1-TOM in MTO93 subcutaneous xenografts. White lines delimit only-TOM or only-EGFP glands and yellow lines delimit double + areas. Scale bars indicate 250 $\mu$ m **B.** Flow cytometry of EPCAM+DAPI- cells from disaggregated xenografts. Note that the TOM-low population has been renamed TOM- for simplicity. Squares indicate the sorting gates. **C.** Relative mRNA expression analysis of indicated marker genes in the four FACS-sorted populations. Data represents mean $\pm$ s.d. of 3 technical replicates. **D.** Organoid formation assay of LGR5+EMP1-, LGR5+EMP1+, LGR5-EMP1+ and LGR5-EMP1- sorted cells. N=2 experiments, 4 wells per experiment (1500 cells/well). *Statistics: P-values were assessed using Student's T-test. P-values: \*\*<0.01 \*\*\*\*<0.0001*

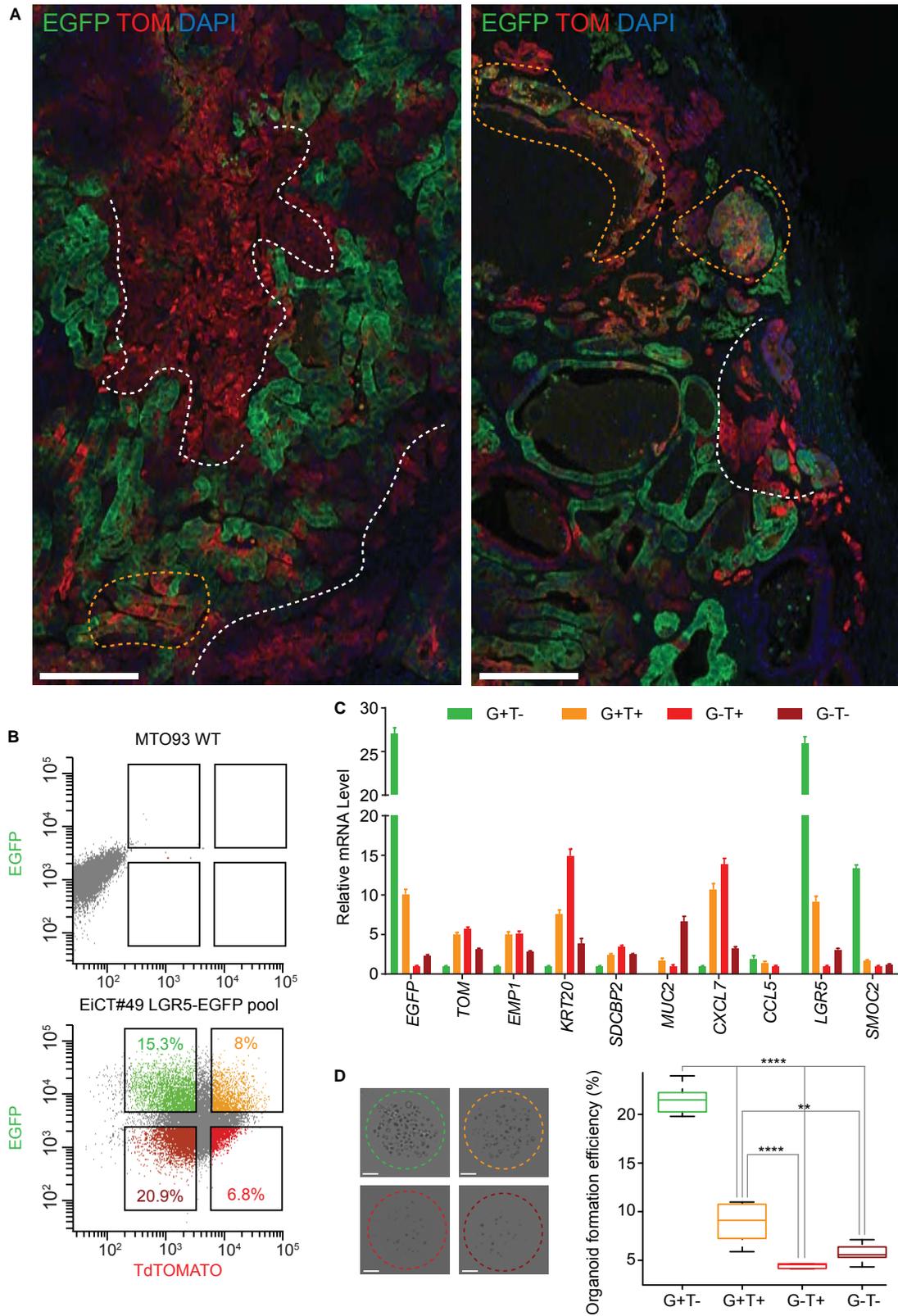
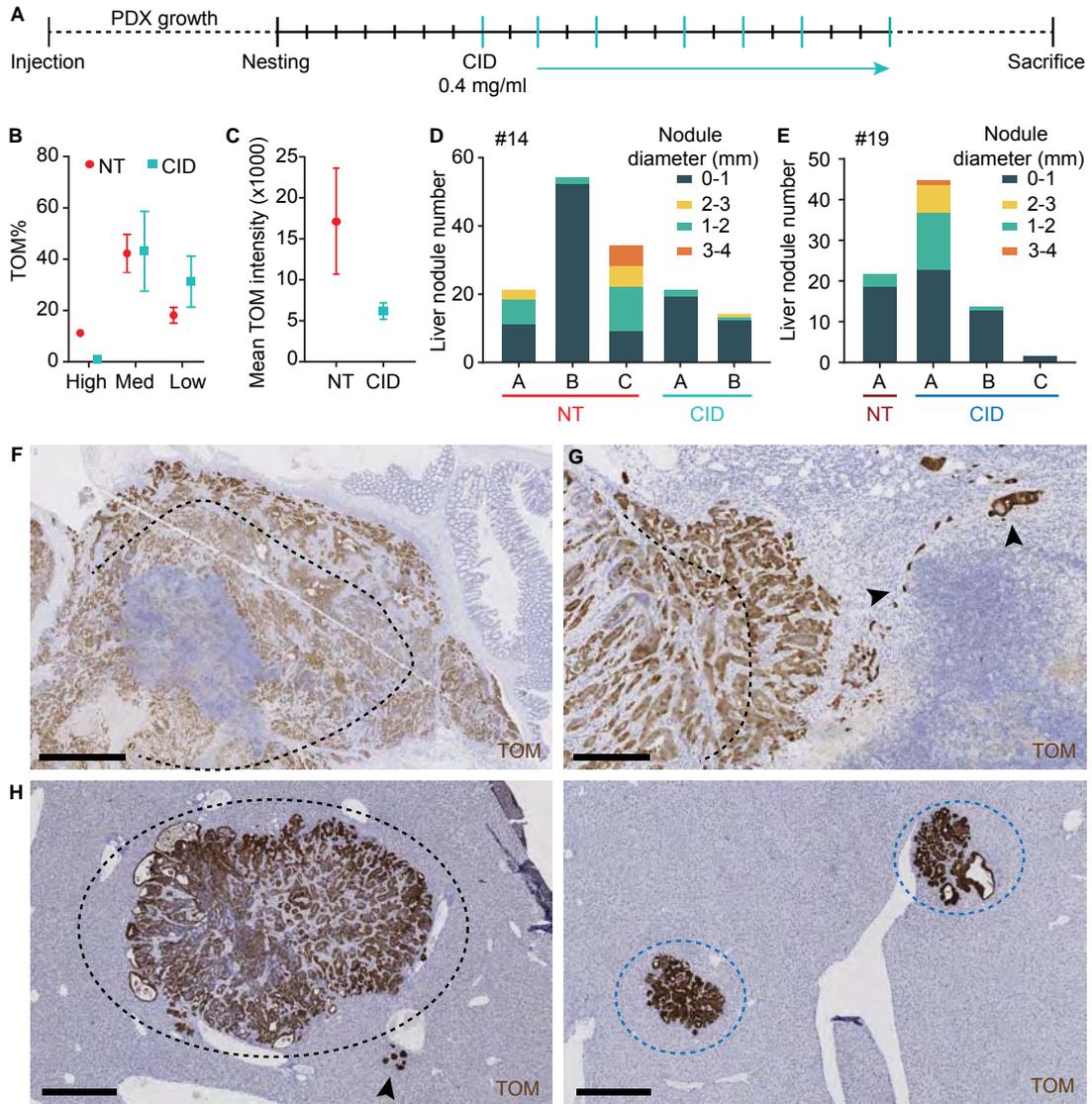


Figure 3.57: *LGR5* and *EMP1* gene expression partially overlap in MTO93 xenografts.

### 3.5.3 Study of EMP1+ cells in MTO-derived CRCs

Our next aim was to specifically eliminate the EMP1-High cells *in vivo* and observe the effect on metastasis formation. We first tested MTO93 primary growth in immunosuppressed Beige/SCID mice, injecting intracaecum 6 mice per clone. Mice were randomized one week after injection in two groups. CID treatment was performed in one group three times a week, 0.04mg/ml (Figure 3.58A).

MTO93 EMP1-ICT#14 and #49 gave rise to large primary CRC in about a month, and tumor disaggregation followed by flow cytometry confirmed the effectiveness of CID treatment to eliminate the EMP1+ population (Figure 3.58B and C). To assess the effect of EMP1+ cell ablation on the invasive potential of the tumors, we counted visible liver nodules in both groups (NT vs CID). There was a tendency to decreased nodule number and size in CID-treated animals but there were no significant differences (Figure 3.58D and E). It is plausible to think that these tumors were too aggressive in nude mice and therefore ablation was enough to prevent cell dissemination or that it occurred before CID treatment. Nevertheless, highest *Emp1* expression corresponded to migratory like cells at tumor margins (Figure 3.58F). We also identified isolated clusters of migrating cells that stained very strongly for TOM (Figure 3.58G). Liver metastases displayed homogenous high levels of EMP1-TOM expression regardless of nodule size (Figure 3.58H).



**Figure 3.58: CID treatment efficiently eliminated EMP1-High cells in MTO93 primary tumors.** **A.** Experimental setup for chronic CID treatment of orthotopic tumors. **B and C.** Liver nodule metastasis number and size per animal in each condition of MTO93 EMP1-iCT#14 and #19 respectively. There is high variability in between groups. **D.** Percentage of cells in each TOM fraction analyzed by flow cytometry of EGFP+DAPI-cells from disaggregated primary tumors. N= 2 tumors in NT group and 4 tumors in CID group. **E.** Mean fluorescent intensity of total EGFP+ population in NT vs CID groups. N=2 and 4 tumors for NT and CID groups respectively. **F.** Representative TOM IHC on paraffin sections from primary tumors. Dashed lines delimit tumor center. Scale bar indicates 250µm. **G.** Immunostaining of TOM cells at tumor buds. Dashed lines delimit tumor center and arrowheads point at bright TOM+ disseminating cells. Scale bar indicates 250µm. **H.** TOM staining of liver nodule metastasis. Black dashed lines encircle big nodules (2-3 mm), blue dashed lines encircle small nodules (0-1 mm) and arrowheads point at single TOM+ disseminated cells.

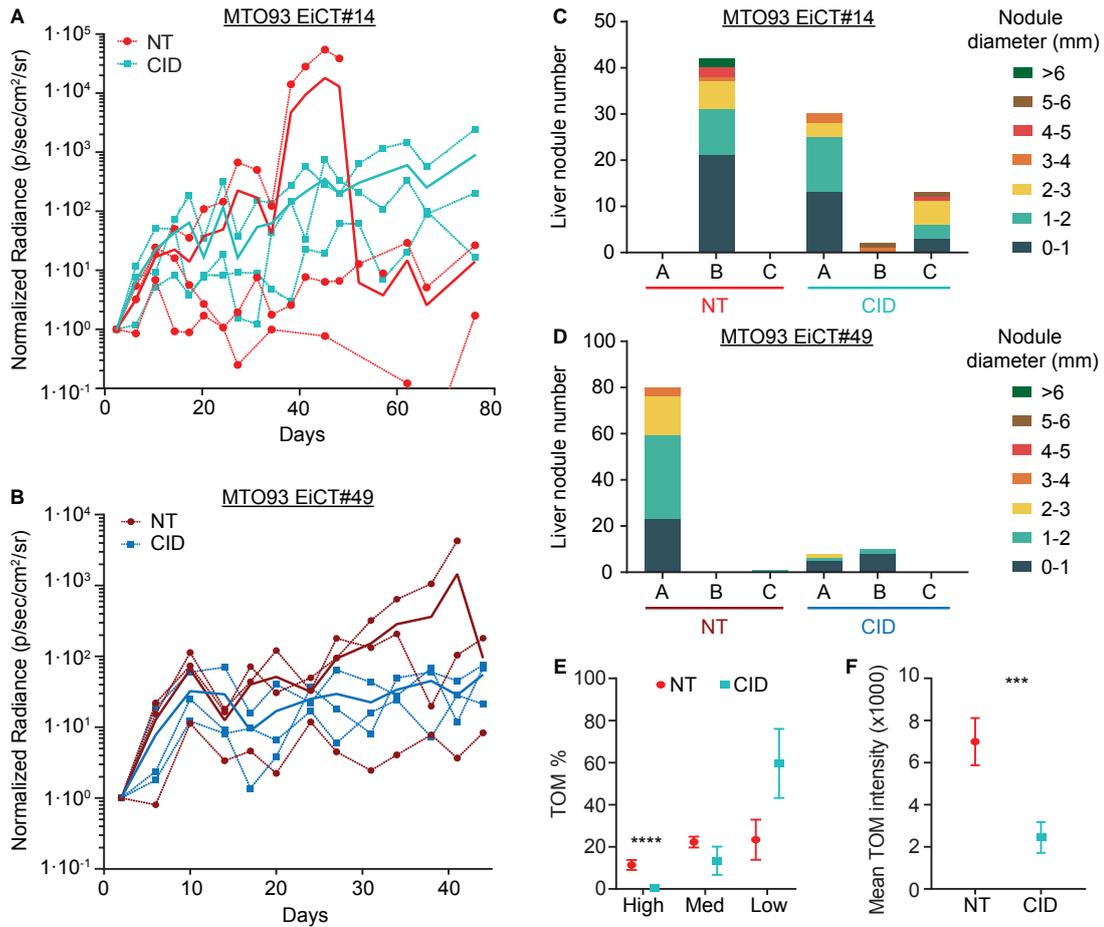
### *MTO93 does not grow homogenously in C57BL6 animals*

One of the major advantages of working with MTOs is that they can develop tumors in immunocompetent mice (a system that better reproduces metastatic dissemination in humans). For this reason, we performed experiments in C57/BL6 mice. We first grew MTO93 EMP1-ICT#14 and #49 cells as subcutaneous tumors in male C57BL/6 mice. When tumors reached a size of around 100 mm<sup>3</sup>, we transplanted 1mm<sup>3</sup> pieces into the caecum of secondary C57BL/6 recipients, using the nesting technique to speed up primary growth rates. We used the same settings as the previous experiment, with 6 injected animals per clone.

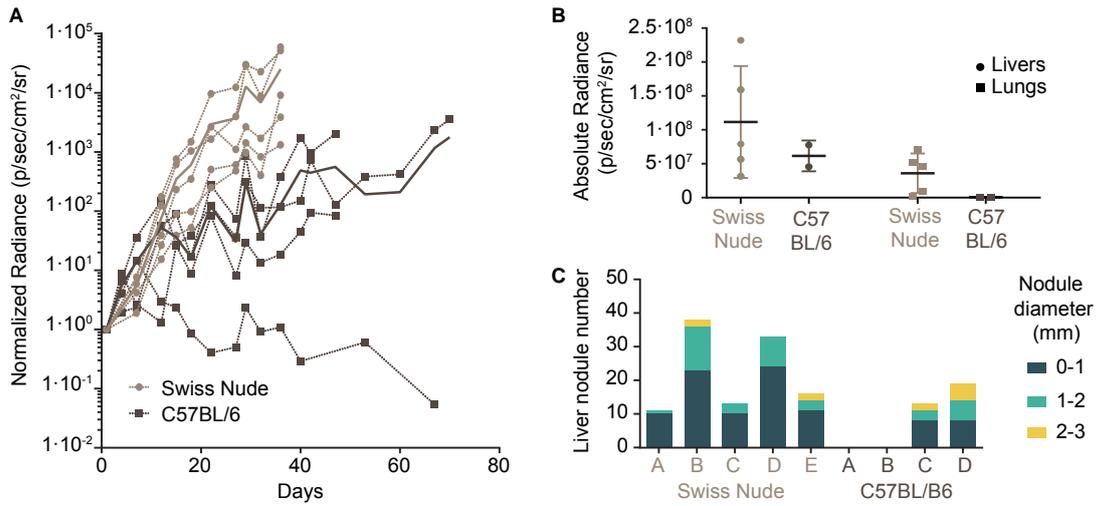
Primary growth was monitored by Luciferase activity two times a week, and CID administration did not fully ablate primary tumor growth as bioluminescence of NT vs CID groups was similar in both clones (Figure 3.59A and B). We ended the experiments at days 76 and 44 respectively due to excessive tumor burden. FACS analysis of disaggregated tumor samples confirmed the effectiveness of the treatment, as the 10% brightest TOM+ cells had completely disappeared in CID-tumors (Figure 3.59C) and total TOM intensity was reduced by 2-3 fold (Figure 3.59D). Unfortunately, liver nodule counting revealed that not every animal developed metastasis (Figure 3.59E and F), which hindered the analysis of the results as we were unable to discriminate whether the differences in liver nodule numbers were due to CID treatment or to biological divergence between animals.

As designing new experimental settings is costly and time-consuming, we first confirmed whether MTO93 was able to produce metastasis consistently in C57BL/6 mice. For this reason, we injected in parallel 4 C57BL/6 mice and 5 SwissNude mice with MTO93 EMP1-iCT#14 (250000 cells/animal) and compared growth in both strains. Bioluminescence escalated faster in Swiss Nude animals, where all developed primary tumors (Figure 3.60A). In C57BL/6 mice we found the same growth variability obtained in the previous experiment suggesting that adaptive immune system blocks metastasis formation in some animals. We assessed the metastatic burden both by ex-vivo bioluminescence (Figure 3.60B) and by liver nodule counting after organ fixation (Figure 3.60C). All Swiss Nude grown tumors generated metastasis 40 days post injection, whereas only two C57BL/6 mice presented liver metastases.

Therefore, MTO93 would only be suitable to study EMP1+ cell dissemination from the primary tumor in immunosuppressed animals, as it is the only experimental setting where it shows consistent primary growth and metastasis development.



**Figure 3.59: MTO93 does not generate metastasis reproducibly in C57BL/6 mice. A and B.** Intravital bioluminescence activity of orthotopically injected MTO93 EMP1-iCT#14-Luc and MTO93 EMP1-iCT#49-Luc respectively. CID treatment started at day 7 post-implantation. Dotted lines represent each animal, solid lines represent the mean of the group. No differences in growth were observed between NT and CID mice. **C.** Distribution of TOM cells in -High, -Med and -Low fractions by FACS analysis of disaggregated primary tumors from NT and CID groups. N=3 tumors NT and 6 tumors CID. P-val \*\*\*\*<0.0001. Differences were assessed by Student T-test in each fraction. **D.** Mean TOM fluorescence intensity quantified by flow cytometry. P-val \*\*\*<0.005. N=3 tumors NT and 6 tumors CID. Differences were assessed by T-student test. **E.** Liver tumor nodule count after animal sacrifice. Animals were named A, B, C in each group. Not all mice develop metastasis, and metastatic burden varied between animals.



**Figure 3.60: MTO93 orthotopic primary tumor growth in Swiss Nude vs C57BL/6.** **A.** Intravital bioluminescence activity of MTO93 EMP1-iCT#14-Luc intracecum injected organoids in Swiss Nude and C57BL/6 animals. **B.** *Ex-vivo* absolute bioluminescence in livers and lungs at experimental endpoint (44 days for Swiss Nude and 70 days for C57BL/6 mice). There is no significant differences between both strains. **C.** Number and size of liver nodules counted after organ fixation.

### 3.5.4 Primary resection allows liver metastasis seeding

Detailed analysis on *Emp1* expression from primary and metastatic tumor sites revealed an additional drawback of our experimental design: metastases expressed high levels of EMP1-TOM (Figure 3.58H). Therefore, results of chronic CID treatment will simply reflect the elimination of the established metastases rather than or in addition to the ablation of migratory cells in primary CRCs. To overcome this problem, we devised a new strategy: organoid injection into the caecum wall followed by resection of whole caecum when a small primary tumors had been already formed. This approach aims at eliminating invasive tumor cells before metastases have developed. Additionally, this experimental setting mimics the situation of a CRC patient undergoing surgery to remove the primary tumor and may represent an excellent model to study disease relapse.

We first determined the minimum time needed for efficient liver seeding from injected primary cells. We inoculated 500000 MTO93 EMP1-ICT#14 cells in the caecum of 10 Beige/SCID animals, and resected one group one week post-injection and another two weeks post-injection (Figure 3.61A). At one week post-injection, animals already presented abdominal carcinomatosis, but primaries could be extirpated. Injected cells had already formed small primary nodules (Figure 3.61B). Of the second group, only one animal survived the surgery, as primary CRC formed attachments to the peritoneal cavity and could not be easily removed. We let the surviving mice live up to one month post-injection, assessing bioluminescence twice a week. Bioluminescent activity did not decrease post-resection due to abdominal carcinomatosis, which we were not able to remove with surgery (Figure 3.61C). At the time of sacrifice mice had develop a significant number of metastases (Figure 3.61D).

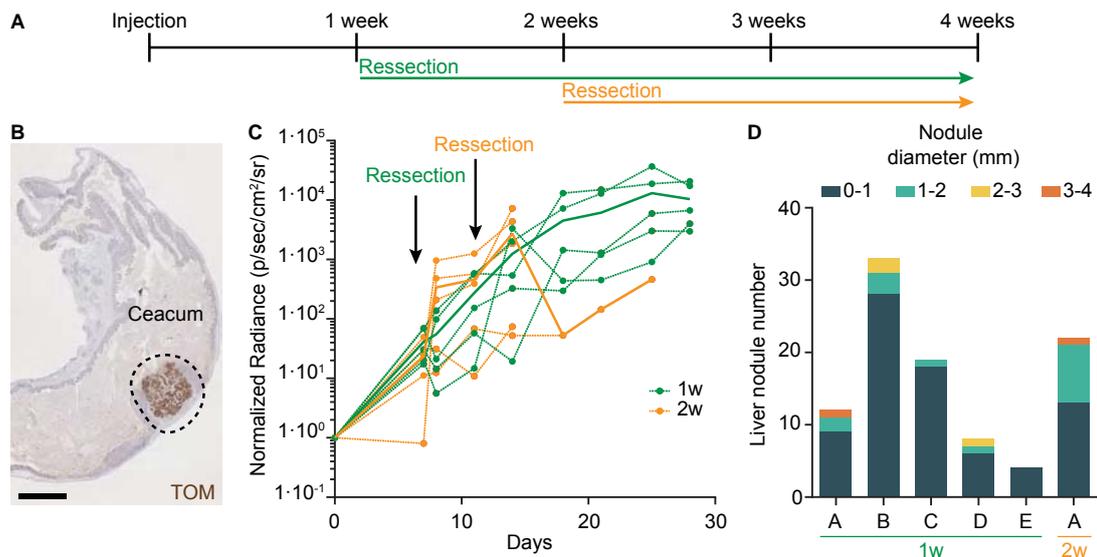
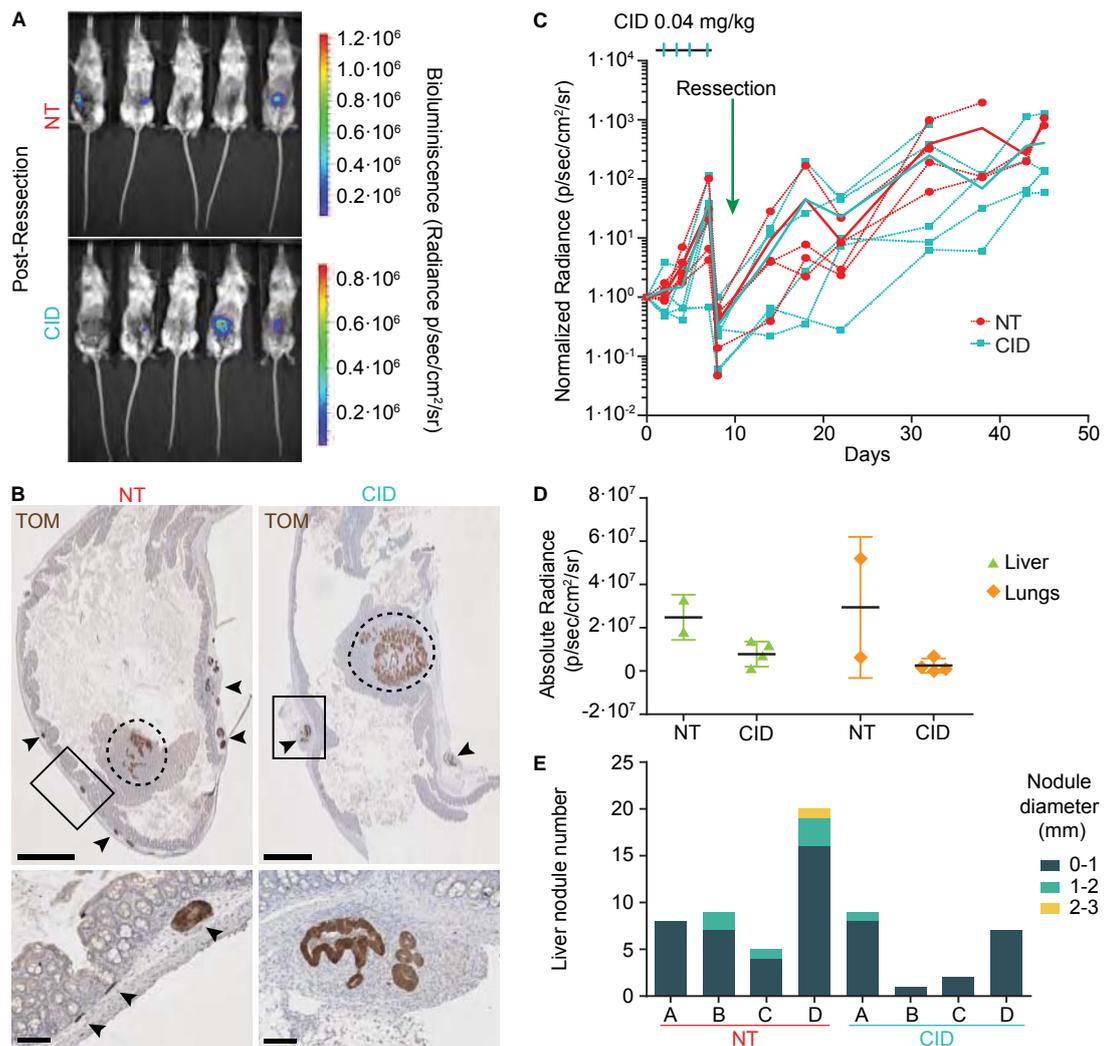


Figure 3.61: Primary tumors resection one week post-injection.

**Figure 3.61: Primary tumors resection one week post-injection.** **A.** Experimental setup of the ceacum resection test. **B.** Representative TOM immunohistochemistry on sections from MTO93 EMP1-iCT#14 orthotopic resected tumors 1 week after surgery. Dashed line delimit the tumor. Scale bar indicates 1mm. **C.** Intravital bioluminescence activity of the Beige/Scid mice bearing MTO93 EMP1-iCT#14-Luc primary tumors. The signal does not decrease after surgery due to abdominal tumor cell dissemination. Solid lines indicate the mean per each group. **D.** Liver nodule metastasis size and number per animal at experimental endpoint.

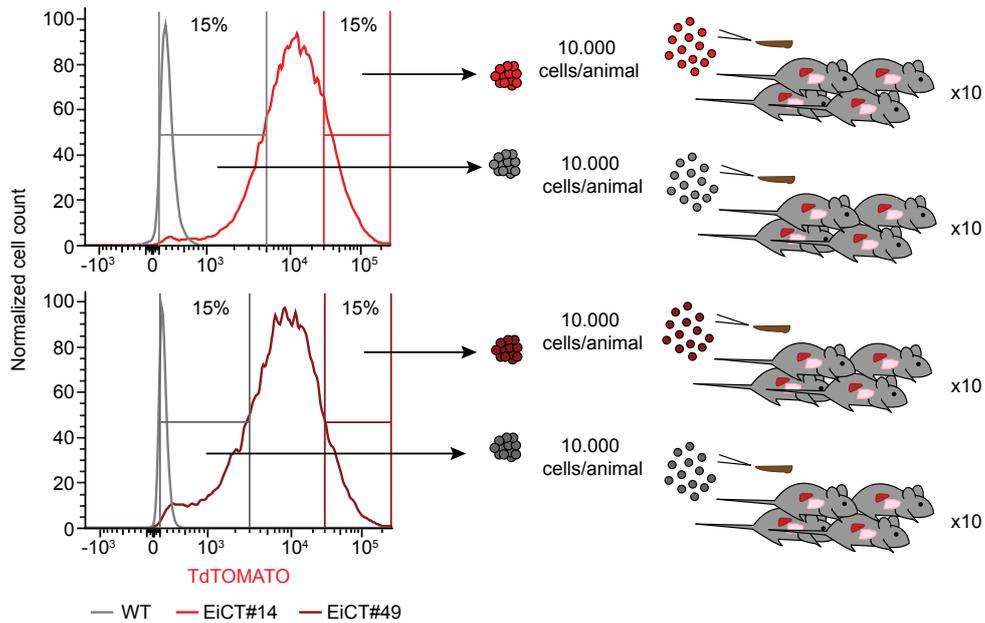
Having determined that the best time point for primary resection in Beige/SCID mice was one week post-injection, we sought to investigate the effect of acute CID treatment during primary growth. We injected 10 Beige/SCID animals with 150000 cells of MTO93 EMP1-ICT#14. Of note, in this case we decreased cell numbers to avoid early abdominal dissemination. Mice were randomized into NT and CID groups. The later receiver 4 shots of CID (0.04 mg) at days 2, 3, 4 and 7 post-injection. Resection of primary tumors was successful, although bioluminescence measurement indicated that mice were not completely clear of tumor cells (Figure 3.62A). Resected caecums presented small primaries, and dissemination through the submucosa was already evident (Figure 3.62B). We followed carcinomatosis and metastasis development for 45 days (Figure 3.62C). Ex-vivo luminescence in livers and lungs was decreased in CID-treated animals (Figure 3.62D), yet liver nodule counting did not render significant differences metastasis burdem (Figure 3.62E). We hypothesize the abdominal carcinomatosis that remained after surgery hampered masked the results as these abdominal tumors possibly kept seeding tumor cells after CID treatment. Anyway, before drawing conclusions, we must improve the experimental conditions to avoid carcinomatosis.



**Figure 3.62: Resection after CID treatment does not reduce metastatic burden.** **A.** Intravital bioluminescence picture of NT and CID-treated mice the day after primary resection. Color bar indicates the radiance intensity. **B.** Representative TOM immunohistochemistry of the resected caecums. Primary tumors are circled with dashed lines, spreading cells are pointed at by arrowheads. Note the decreased spreading in CID group. Scale bars indicate 1 mm and 100 $\mu$ m respectively. **C.** Intravital bioluminescence quantification for the whole experiment duration. The increase is both due to abdominal and liver tumor growth. **D.** Quantification of *ex-vivo* bioluminescence activity in liver and lungs at endpoint. **E.** Liver nodule size and number per animal.

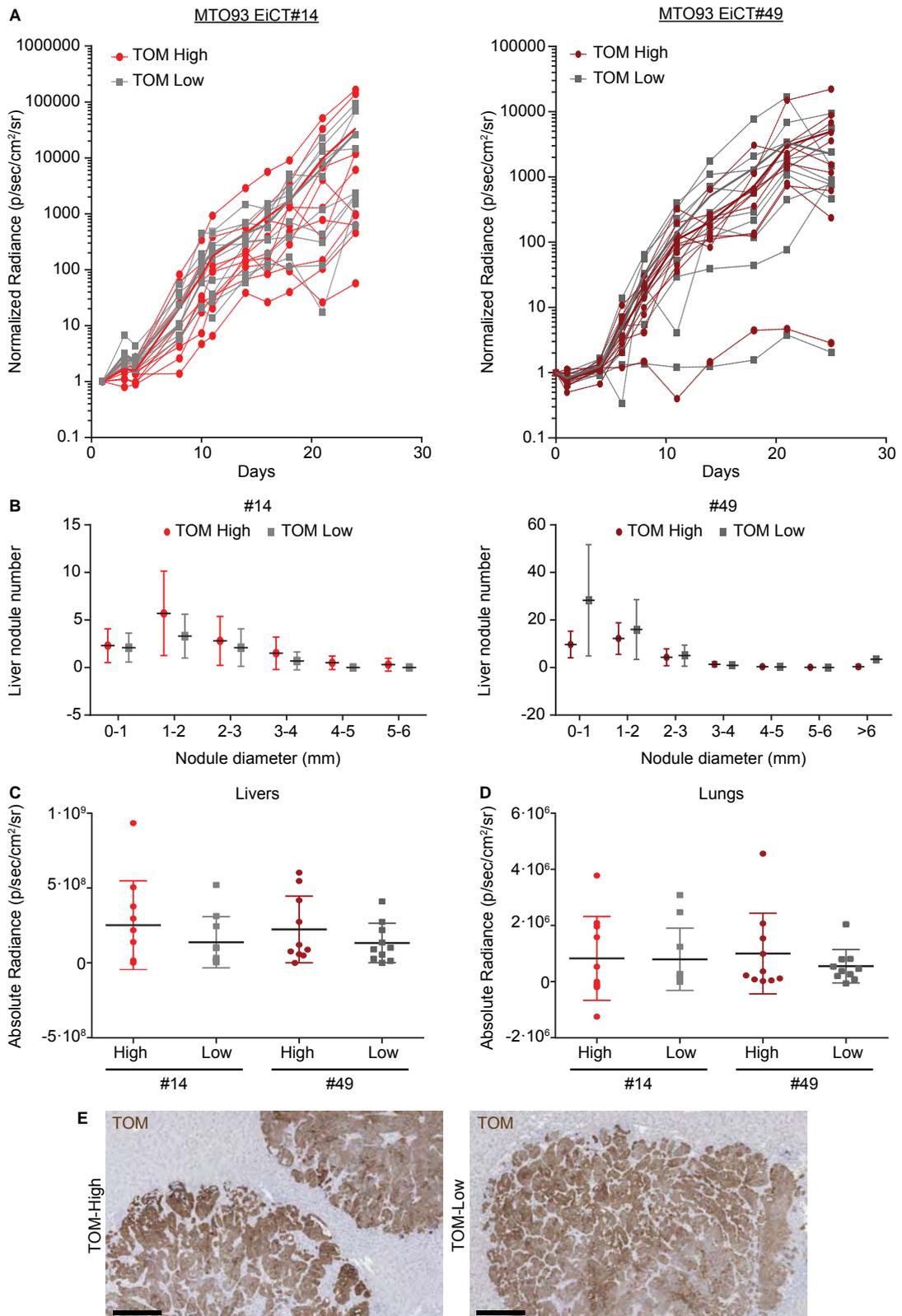
### 3.5.5 EMP1+ cells do not have higher metastatic potential when injected via intraspleen

As orthotopic transplantations of MTOs raised important technical limitations, we switched to a surrogate approach to study metastasis. We assessed whether EMP1+ cells were more capable than the rest to seed the liver upon direct inoculation into the portal circulation (Metastasis Initiation Experiment, MIC). We first generated MTO93 EMP1-iCT#14 and #49 subcutaneous C57/BL6 tumors and FACS-sorted TOM-High and TOM-Low populations. We then reinjected these cell populations through the spleen of C57BL/6 mice (10000 cells per animal) (Figure 3.63).



**Figure 3.63: Metastasis initiation experimental design.** Schematic representation of the metastasis initiation capacity (MIC) experiment). The 15% TOM-highest and -Lowest cells were FACS-sorted per each clone and reinjected via intraspleen into C57BL/6 recipients. N=10 animals per condition

Luciferase activity increased sharply since day 7 post injection (Figure 3.64A), indicating that both the TOM-High and TOM-Low cells had been able to overcome immune surveillance. Growth rate of both populations was similar, and inspection of the livers did not reveal differences in the number or size of metastases (Figure 3.64B). Ex-vivo bioluminescence of liver and lungs corroborated that metastatic seeding was equal in TOM-High vs TOM-Low groups (Figure 3.64C and D). TOM-High and -Low derived tumors presented the same level of TOM staining (Figure 3.64E), indicative of plasticity. Based on this result, we concluded that EMP1+ cells do not have enhanced ability to colonize the liver by themselves when injected directly into the blood stream.



**Figure 3.64: EMP1-High cells do not colonize the liver more efficiently than the rest of the tumor.**

**Figure 3.64: EMP1-High cells do not colonize the liver more efficiently than the rest of the tumor.** **A.** Intravital bioluminescence imaging of intraspleen-injected mice with TOM-High and TOM-Low MTO93 cells of #14 and #49 respectively. **B.** Number and size of metastatic liver nodules at the end of the experiment (30 days). N= 10 animals per condition. **C.** . *Ex-vivo* bioluminescent activity in livers of both groups. N=10 livers per condition. **D.** *Ex-vivo* bioluminescent activity in lungs of both groups. **E.** Representative TOM immunohistochemistry on liver paraffin sections with tumor nodules. Both TOM-High and -Low derived metastasis present the same level of TOM. Scale bars indicate 250µm

Part IV

DISCUSSION



## DISCUSSION

---

Self-renewal in the intestinal epithelium is a well-orchestrated process which makes it a suitable model to study stem cell biology. Intestinal stem cells, located at the bottom of the crypts and marked by *LGR5* gene expression, are a highly proliferative population with the ability to differentiate into all intestinal cell types. The structural organization of the intestine favours a one-way migration of progenitor cells that leave the crypt base and transit towards the top of the villi as they undergo differentiation. Polarized stem cell (WNT, EGF, NOTCH) and differentiation (BMP, TGF- $\beta$ ) signaling pathways determine cell identity along the crypt-villus axis (Clevers, 2013). The intestinal stem cell hierarchy is particularly important for colorectal cancer development, since it is believed that transient amplifying or differentiated cells harboring pro-tumorigenic mutations will eventually be shed into the lumen (Huels and Sansom., 2015). On the contrary, acquisition of mutations in the *LGR5*+ population leads to aberrant cell expansion and adenoma formation (Barker et al., 2009). As cells increase their mutation burden, benign adenomas develop into advanced malignancies with the ability to metastasize. The study of early adenomas has shown that their growth is reminiscent to that of normal epithelia, with a heterogeneous cell composition that relies on *LGR5*+ cells to expand (Scheppers et al., 2012).

Our aim was to study the cellular heterogeneity of human advanced colorectal cancers and to elucidate whether a stem cell hierarchy was maintained in the tumors. At the time of project design, no one had yet performed lineage tracing analysis on human advanced CRCs. Some previous reports of lineage tracing in patient samples, based in CpG methylation patterns (Siegmund et al., 2009) and cytochrome C deficient cells (Humphries et al., 2013) suggested the presence of multiple cancer stem cells in those tumors, yet lack of models for advanced CRC hampered the study of cell composition and tumor expansion in colorectal carcinomas.

We aimed at unraveling the contribution of the putative colon cancer stem cell population, defined by *LGR5* expression, to tumor growth, as well as the behavior and proliferation dynamics of differentiated-like populations. With these studies, we aimed to shed some light on the organization of tumor tissues, a knowledge essential for the development of new therapies. Finally, we were also interested in identifying the cells of origin of metastasis in CRCs. In the following section, I will summarize our findings and relate them to the latest advances in the field.

## Development of new models to study tumor cell heterogeneity

### *Patient derived organoids reproduce human tumor heterogeneity in mice*

One of the major limitations in the study of colorectal tumor cell hierarchy has been the lack of appropriate models. A large number of human 2D CRC cell lines representing the different genetic landscapes found in patients are available (Mouradov et al., 2014), but they do not recapitulate the heterogeneity and organization present in the original tumor. Genetically engineered mouse models (GEMM) failed to generate aggressive adenocarcinomas (Tetteh et al., 2016). The last ten years have seen the establishment of organoid cultures as key to reproduce both healthy and tumor tissues *in vitro* (reviewed in Clevers, 2016), and its utility in maintaining colon tumor patient cells was soon proven by us and many other colleagues (Jung et al., 2011, Sato et al., 2011, Calon et al., 2015, van de Wetering et al., 2015). More recent studies using xenografts derived from subcutaneous injection into mice of human organoids, the patient derived xenografts, have shown that they faithfully reproduce the diversity of human tumors (Linnekamp et al., 2018, Vlachogiannis et al., 2018, Prasetyanti et al., 2019) and are even able to interact with the murine stroma inducing a pro-tumorigenic response in animals (Chao et al., 2017). Taken together, these studies corroborate the suitability of PDOs as a model system to decipher tumor cell heterogeneity.

In order to obtain relevant and translatable results, we selected two PDOs that harbored canonical mutations in the main CRC driver signaling pathways (Table 3.1) (Morin et al., 1997). We have shown that both organoids generate glandular-like, moderately differentiated tumors upon subcutaneous injection into immunodeficient mice. In addition, this histological pattern is maintained upon orthotopic injection in the caecum of the animals. Indeed, PDO6 and 7 are able to generate primary tumors in immunodeficient mice regardless of the implantation method (orthotopic transplants are explained in **section 6.5.2**). However, the metastatic potential of these models is limited, as only few animals displayed liver or lung metastatic nodules even many months after injection. Work by others (Roper et al., 2017, O'Rourke et al., 2017, Fumagalli et al., 2018) shows that some human derived organoids do produce metastases in immunodeficient mice, so we hypothesize that the dissemination ability is intrinsic of each patient derived organoid. PDO7 cells, nevertheless, efficiently generated metastatic niches in the liver upon intraspleen injection, indicating that mutations in key pathways for CRC development provide sufficient niche independency to tumor cells to grow in foreign environments, in agreement with the work of the Clevers and Sato laboratories (Drost et al., 2015, Matano et al., 2015). Moreover, upon orthotopic implantation of quadruple mutant ( $APC^{KO}$ ,  $KRAS^{G12D}$ ,  $TP53^{KO}$ ,  $SMAD4^{KO}$ ) organoids they were able to induce distant metastasis formation (Fumagalli et al., 2017).

To circumvent the problem of lack of liver and lung metastases generated from orthotopic PDO injection, we took advantage of the GEMM-derived organoids generated in the lab (Tauriello et al., 2018). As previously described, quadruple mutant-derived organoids ( $APC^{KO}$ ,  $KRAS^{G12D}$ ,  $TP53^{KO}$ ,  $TGF\beta R2^{KO}$ ) reproduce the CMS4 colorectal

cancer subtype (Guinney et al., 2015), defined by stromal infiltration and high levels of TGF- $\beta$ . CMS4 tumors, thus, display metastatic-like traits and are indeed the ones with worst prognosis (Calon et al., 2015, Isella et al., 2015). These facts validate the use of MTOs for metastatic studies as surrogates of the human disease in mice. Likewise, other groups have recently generated metastatic mouse organoids that recapitulate highly aggressive CRC tumors upon orthotopic injection in immunocompetent animals (Roper et al., 2017, Melo et al., 2017).

#### *CRISPR/Cas9 genome edited organoids retain parental features*

We and others have been the first ones to establish a CRISPR/Cas9-based knock-in strategy to label specific tumor populations in human and mouse CRC derived organoids, a technology so far reserved to animal models (Cortina and Turon et al., 2017, Shimokawa et al., 2017, Melo et al., 2017). Genomic modification is a very powerful tool to simultaneously follow distinct cell states in unperturbed tumors. With our system, we have been able to target more than one gene at a time, generating up to tri-color organoids reporting stemness (LGR5-EGFP), proliferation (KI67-RFP) and differentiation (KRT20-iRFP). A long standing concern of gene editing is the introduction of undesired mutations in the process; CRISPR/Cas9 system has been reported to provoke off-target mutagenesis due to unspecific guide RNA binding (Fu et al., 2013, Pattanayak et al., 2013). By exome sequencing of PDO7 LGR5-EGFP#1, we found very few *de novo* mutations compared to its parental organoid (Table 4.1). Concerned that the analysis method (Mutect2 algorithm) was not properly detecting real *de novo* mutations, as it has an error rate of 6.4 mutations per Mb, we cross checked the analysis with a second algorithm, VarScan2. Both are described to have similar error rates (Kroigard et al., 2016). Indeed, only 17 missense/nonsense mutations coincide between methods, of which only one has a predicted high impact, suggesting that many of the detected changes could actually be false positives (Table 4.1). This analysis confirms that the organoid technology maintains the cells in a genetic stable state, as previously shown by others (Weeber et al., 2015), and that CRISPR/Cas9 genome editing can be safely used in organoids. The final mutation number of our edited PDO is similar to the reported natural evolution of MSS CRC patient derived cells after prolonged *in vitro* culture (Fujii et al., 2016), so Cas9 activity does not introduce significant changes in the genomic landscape of the cells.

We must consider, though, that our knock-in generation involves a long-term FACS selection based on fluorescent marker expression. Marker-based selection of knock-in integrated cells greatly increases positive clone derivation efficiencies (up to 90%), as is the case for LGR5-EGFP or KI67-RFP targeting. Yet, it relies in the fact that *in vitro* cultured organoids from sorted cells are able to revert to the original phenotype regardless of the cell of origin. In our system, we have shown plasticity arising both from stem-like populations (LGR5+) as well as differentiated-like cells (EMP1+) after 15 days of culture. These findings reinforce the idea that our organoid derivation strategy is not biasing subsequent results. In addition, this limitation can be overcome by growing single-cell clones directly after nucleofection without selecting for fluorescent expression, as we have done for *KRT20* and *EMP1* knock-ins, although

Algorithm	CRC driver mutations	<i>de novo</i> missense or non-sense	<i>de novo</i> mutations with predicted high impact (genes)
Mutect2	All present (APC, KRAS, SMAD, ATM)	64	3 (EIF2AK2, KIAA0101, PPFIBP2)
VarScan2	All present (APC, KRAS, SMAD, ATM)	70	6 (CCRN4L, SEMA5A, HOOK3, PRSS3, GLE1, PPFIBP2)
Common	All present (APC, KRAS, SMAD, ATM)	17	1 (PPFIBP2)

**Table 4.1: Exome sequencing results of parental PDO7 vs PDO7 LGR5-EGFP#1 clone using two different analysis softwares.** Summary of *de novo* introduced mutations in PDO7 LGR5-EGFP#1 after CRISPR/Cas9 modification. The analysis was done using two Variant Caller algorithms, Mutect2 and VarScan2. They have an error rate of 6.4 mutations per Mb, suggesting most of the mutations are actually false positives (6.4 mutations\*30 Mb of human exome = 192 false positives). Only 17 mutations appear with both analysis methods, and of the 17 only one has predicted high impact on the protein function.

the efficiency of correctly integrated clones lowers to 1-5%.

The use of monoclonal organoids has advantages and drawbacks. On one side, it restricts the interpretation of the analysis regarding tumor heterogeneity. Intra-tumor heterogeneity (ITH), the existence of clones with different genetic background within the same tumor, is key for cancer development, as well as a major cause of tumor relapse and chemotherapy resistance (reviewed in McGranahan and Swanton, 2017). However, the use of monoclonal organoids ensures that cells of different phenotypes (e.g. stem or differentiated tumor cells) within tumors display identical genotypes and therefore that differences in their behavior are not due to distinct alterations but rather to change in gene programs.

Furthermore, recent studies propose that selective sweeps that change the mutational landscape of the tumor are less frequent than initially thought, due to low fixation probability of new clones in fast cycling cells with spatial constraints (Sottoriva et al., 2015, Baker et al., 2017, Ryser et al., 2018). This suggests that our single-cell derived organoids retain most of the parental tumor composition, and that new intratumor heterogeneity can arise from subsequent evolution of each clone (Giessler et al., 2017). Regardless, to ensure the reliability of our results, we have used several independent genome edited-clones to validate our main observations. In virtually every experiment, we have seen that the different clones behaved equally in terms of proliferation and aggressiveness, and presented similar heterogeneity of stem and differentiated compartments.

In summary, we have combined two powerful techniques, organoid culture and CRISPR/Cas9 technology, to create new tools necessary for the analysis of cell heterogeneity in intact tumors. Beyond the applications and results described in this thesis, this approach opens up the possibility of genetic analysis, something that was before

reserved to developmental processes.

## **Definition of stem and differentiated cell compartments in advanced colorectal tumors**

### *Intestinal stem-like cell numbers increase in tumors*

Up to date, visualization of LGR5+ cells in human tumors had not been possible due to lack of suitable tools. We bypassed this roadblock by inserting fluorescent reporters in the *LGR5* locus of PDOs. In both PDOs analyzed as well as in MTOs, LGR5+ cells represent 50-60% of the total cell number. The glandular-like structures of xenografts are reminiscent of the cell distribution of the healthy epithelia, with extensive differentiated areas, marked by the pan-differentiation *KRT20* marker and counter enriched in *LGR5* expression. The LGR5+ cell proportion suggests an increase in stem cell numbers during tumor progression, a hypothesis validated by *in situ* RNA hybridization of *LGR5* in patient samples (Baker et al., 2015, Martin et al., 2018). It is plausible to think that, like in the healthy epithelia, not all of the LGR5+ cells are actively functioning as stem cells (Kozar et al., 2013, Ritsma et al., 2014). In addition, the distribution of stem and differentiated areas indicates the presence of asymmetric niche factor distribution across the tumor, with some areas retaining crypt-base features that may induce stemness phenotypes (Lotti et al., 2013). Some studies have reported that areas of high WNT signaling occupied by stem cell-like cells are enriched in tumor borders, and the differentiated areas are compressed in the necrotic tumor center (Cernat et al., 2014). This would cause the tumors to expand from the edges and compress the cells at the center (Lamprecht et al., 2017). Our findings contradict these observations, as LGR5+ glands are found across the tumor. It will be very interesting to elucidate the mechanisms that dictate stem cell status in those tumor areas, although we have not been able to tackle that question yet.

### *Differentiated cells are scattered across all tumor*

Besides tracking LGR5+ cells in human and mouse tumors, we have also genetically labelled two other genes, *KRT20* and *EMP1*. *KRT20* is a well-described marker for intestinal differentiated cells (Chan et al., 2009). In agreement, *KRT20* expression pattern complements that of *LGR5*, with glandular regions entirely differentiated and devoid of stem cell markers. *EMP1* is activated upon blockade of the WNT signaling pathway (unpublished observations from Battle lab) and it is consistently downregulated in stem cell signatures. In accordance, *EMP1*+ cells expression profile overlaps with differentiation genes such as *KRT20*, *CEACAM5* and *SDCBP2*, and is counter-enriched in the stem cell markers *LGR5*, *SMOC2* and *OLFM4*. Therefore, we have used it as a differentiation marker to validate the results obtained with *KRT20* lineage-tracing alleles. Unexpectedly, despite the fact that *EMP1* is expressed across the tumor, highly positive cells are grouped in the xenografts boundaries. This cellular distribution is also confirmed in orthotopically-grown tumors. Indeed, the CMS4-like MTO93 expresses high levels of this marker, perhaps suggesting that its expression is enriched in metastatic tumors. We hypothesize that *EMP1* is a marker gene of both differentiated and invasive cells. We further detail the utility of *EMP1* gene as invasive marker

in the next sections, but we will first consider its implications in the tumor hierarchy as a differentiated-like cell marker. This pattern suggests that perhaps differentiated cells (or at least WNT-OFF) cells are the ones undergoing migration and perhaps operate as the cell of origin of metastasis. This hypothesis, however, goes against recent finding on LGR5+ cell as the cells of origin of metastasis (Melo et al., 2017). This question will need to be addressed by comparing the effects of EMP1+ versus LGR5+ cell ablation in metastatic tumors.

## **Identification of determinants of tumor initiating potential**

### *Tumor initiation depends on stem cell program expression*

It has been long hypothesized that tumor growth could only be initiated and sustained by cancer stem cells (Barker et al., 2009). Indeed, CRC-SC separated by *CD44*, *CD133* or *EPHB2* markers (O'Brien et al., 2007, Dalerba et al., 2007, Ricci-Vitani et al., 2007., Merlos-Suárez et al., 2011) were more proficient in tumor initiation upon injection into secondary mice hosts than their negative counterparts. In line with these preliminary experiments, our purified LGR5-EGFP cells showed higher organoid formation capacity *in vitro* and tumor initiation potential *in vivo*. Moreover, by using *EMP1* as a surrogate of differentiation we have confirmed that non-LGR5 cells have less organoid formation capacity and retain only a limited tumor initiation potential. Organoid formation assay using LGR5-EGFP/*EMP1*-TOM double-labelled MTO93 confirmed that LGR5+ cells were the only ones retaining clonogenic potential, and that the subset of LGR5+*EMP1*+ cells had the characteristics of a transient-amplifying-like population. Our findings correlated with work done by colleagues using CRISPR/Cas9 engineered mouse tumor organoids (Melo et al., 2017).

### *Plasticity between CSC and differentiated cell compartments*

An important question in the CSC field is whether plasticity arises from non-stem cells. In homeostatic setting, it has already been proven that hierarchies are not as strict as initially thought, as damage in hair, skin, kidney or liver, for example, is rescued by de-differentiation of progenitors and post-mitotic cells to their respective SC lineages (Ito et al., 2005, Kusaba et al., 2014, Tarlow et al., 2014, Hoeck et al., 2017). Likewise, in the normal intestine differentiated-like populations are able to replenish the LGR5+ pool upon homeostasis perturbation (Tian et al., 2011, Takeda et al., 2011, Powell et al., 2012, van Es et al., 2012, Tetteh et al., 2016), so it is expected that tumor cells also present a certain degree of plasticity. Our experiments corroborate this theory, as tumors arising from LGR5- or *EMP1*+ populations perfectly recapitulate the parental heterogeneous cell composition. In line with these findings, plasticity of CRC tumors at all development stages has also been recently observed using *ASCL2* as stem cell reporter (Oost et al., 2018).

A fundamental issue, though, is whether cells must revert to a stem-cell state before giving rise to new tumors. In the healthy intestine, indeed, abolition of both LGR5+ and +4 cells impedes intestinal regeneration (Metcalf et al., 2014), indicating a requirement for the cells to transition through stem-like states during plasticity-

mediated tissue regeneration. To formally address this issue, we will need to specifically eliminate the LGR5+ population in our tumors and observe whether they are still able to grow. While we were working on this aspect, two studies published in Nature tackle this issue. In Shimokawa et al., 2017 paper, they ablated LGR5+ cells by inducible Caspase9 in human-CRC patient derived xenografts. In this setting, KRT20 progeny increased proliferation and acquires CSC phenotype. Importantly, they proved that plasticity of differentiated cells depended on reversion to an LGR5+ cell status, as continuous LGR5+ cell ablation impeded tumor regeneration. On the contrary, Melo et al., 2017 proposed that primary tumor growth was maintained by LGR5- cells without need of reversion, as continuous LGR5+ cell elimination did not cause tumor shrinkage. Interestingly, they suggested this effect might be mediated by the microenvironment, as metastatic nodules did depend upon LGR5-driven growth. Therefore, these data are contradictory but may suggest that the degree of plasticity of LGR5- cells depends on the tumor type or model system analyzed.

### **Intestinal stem cell hierarchy is lost during malignant transformation**

#### *Tumor initiation capacity does not reflect cell behavior in intact tumors*

The xenograft transplantation assays were long considered as the gold standard to assess cancer stem cell potential as they proved the ability of a single cell to recreate the whole original disease in a host. The field of CSCs has been largely biased by the use of this experimental approach. More recently, the improvement of genetic labelling techniques has allowed the visualization of cells in their native environment, a factor essential in solid tumors as their fate depends on specific niche factors and cell-to-cell interactions (Morrison and Spradling, 2008). Indeed, many studies show how cells with multipotent capacity in transplantation assays actually have a more limited ability to differentiate when analyzed in intact tissues (reviewed in Batlle and Clevers, 2017). Results using the lineage tracing techniques have also challenged the Hematopoietic-SC model, providing evidence that a transient progenitor population maintains hematopoiesis in the adult mice (Busch et al., 2015). Therefore, it is becoming increasingly evident discordancies in the assessment of stem cell potential by lineage tracing or by transplantation approaches. The results obtained in this thesis are another example of such discrepancy.

#### *Differentiated and stem-like derived clones persist over time in tumors*

Our initial hypothesis was that advanced CRC tumors displayed a cell hierarchy reminiscent to that of the normal intestine, based on previous findings in mouse adenomas (Schepers et al., 2012). To prove this, we labelled the progeny of LGR5+, KRT20+ and EMP1+ cells respectively with an inducible CreERT2 system. Due to low levels of transgene expression in *LGR5* and *EMP1* knock-ins, we estimate we only labelled around 5-10% of each subpopulation. *KRT20* gene, on the contrary, is highly expressed in differentiated-like tumor cells, rendering a much more efficient recombination. By Tam dosage titration, we managed to perform lineage tracing at the clonal level (tracking single cells) of the *LGR5*, *KRT20* and *EMP1* lineages. A major concern of this type of experiments is also the leakiness of the Cre enzyme (He et al., 2017),

but we never found TOM clones in untreated animals (data not shown).

If our hypothesis that intestinal tumors rely on a stem cell hierarchy to expand over long periods of time was true, clones derived from LGR5+ cells would expand exponentially and colonize the whole tumor, as in the healthy intestine (Barker et al., 2007). On the other hand, clones derived from differentiated cells, marked by *KRT20* or *EMP1* expression, would eventually disappear. By following the progeny of LGR5+, *KRT20*+ or *EMP1*+ cells in patient derived xenografts for as long as two months, we have observed that the three cell lineages contribute similarly to tumor growth. The clones derived from the three populations persisted and expanded for the whole length of the experiment. The system is limited due to the fast PDX growth rate, yet two months clone expansion is sufficient to prove that differentiated cell lineages do not disappear with time. An important consideration to interpret these data is that the *KRT20* gene might be early transcribed in differentiating cells, days before it is translated. This is supported by the finding that some TOM+ cells 72h post-induction of *KRT20* lineage tracing did not express yet *KRT20* protein. Therefore, the *KRT20* allele lineage tracing may label early progenitors rather than fully differentiated cells. Similar caveats could be raised by the use of *EMP1* gene.

Nevertheless, our findings are further supported by the work of Vermeulen and colleagues (Lenos et al., 2018). They chose a marker-free lineage tracing approach, labelling random single cells in the PDX, and clones arising from these cells proliferate independently of their origin. They also report that tumors grow preferentially in the outer regions, something we have not observed. This could be explained because they analyze xenografts of up to 1000 mm<sup>3</sup>, whereas our tumors harbor a maximum size of 300 mm<sup>3</sup>. The center of such big tumors becomes highly necrotic, and is therefore expected that cell proliferation can only happen in the border. On the contrary, by using a similar *LGR5* and *KRT20* lineage tracing system, Shimokawa et al., 2017 report that *LGR5*-derived clones grow over time whereas *KRT20* progeny is lost with time, with the exception of a few of clones. These controversial results can be explained in one hand by the fact that we use different PDOs for the study, and in the other hand because accurate 3D analysis of clonal proliferation in our xenografts is required to confirm if *LGR5*+ cells really have an advantage over the rest of the tumor to proliferate. Of note, we have resolved the technical limitations (xenograft clearing, antibody penetration and microscope resolution) for 3D imaging during project development, so future studies will be able to answer these questions.

#### *Differentiation kinetics in tumors*

Apart from assessing the proliferation dynamics of CSC-derived cells, we have quantified their differentiation potential. Interestingly, *KRT20*+ and *MUC2*+ cells within the *LGR5*-derived clones appear in substantial numbers only one month after tracing, in contrast with the normal epithelium where cells differentiate and die in a range of 5-7 days (Potten and Loeffle, 1987). The slow differentiation rate might be due to tissue organization; in normal intestine, stem cell signals (WNT, EGF, NOTCH) accumulate at the crypt bottom and are counteracted by a differentiation-inducer BMP/TGF- $\beta$  gradient higher up in the villi (Clevers, 2013). In tumors, stem cell promoting signals

are probably less constrained since tissue architecture is disrupted (Vermeulen et al., 2010). This fact is supported by the increased number of LGR5+ stem cells found in PDO6 and 7 tumors (50-60%) (Cortina and Turon et al., 2017). Indeed, by single-cell profiling, Li et al., 2017 report that 93% of the cells in CRC tumors retain stem or transient amplifying expression programs, whereas only 10% of the total intestinal cells are LGR5+ in homeostatic conditions (Barker et al., 2007).

### *Quiescent cancer stem cells*

Lineage tracing revealed that almost half of the LGR5-derived clones remained in single to two-cell status weeks after induction of the tracing. 3D reconstructions proved that many LGR5+ cells remained in single cell status, and that this result was not a bias of the 2D analysis. This finding suggests the presence of a putative quiescent LGR5+ cell population, which is somewhat surprising, as about 90% of the LGR5+ cells in healthy intestine are proliferating (Barker et al., 2007, Schepers et al., 2012, Basak et al., 2014). Cell cycle status has been assessed by KI67 protein levels, as it is a protein strictly associated with cell cycle progression through G<sub>1</sub>/M/G<sub>2</sub> phases (Scholzen and Gerdes, 2000, Miller et al., 2018). and lost during quiescence. Double labelling of LGR5 and KI67 in PDO7 enabled cell cycle analysis of LGR5+KI67+ vs LGR5+KI67- populations, corroborating that a large proportion of LGR5+KI67- cells are arrested in G<sub>1</sub>/G<sub>0</sub> cell cycle phase. Moreover, these cells display an overall down-regulation of the proliferation program of crypt progenitors (Jung et al., 2011). In contrast, LGR5-KI67+ population shows traits of absorptive differentiation (*FABP1* and *SI* expression) and LGR5-KI67- cells appear terminally differentiated (*KRT20+*), reminiscent of the homeostatic differentiation pathway (Dalerba et al., 2011). Moreover, the LGR5+KI67- gene signature overlaps with the mouse intestinal LGR5+KI67- transcriptomic profile defined in Basak et al., 2014.

Quiescent stem cells have been identified in normal intestine, although their identity remains controversial (Buckzaki et al., 2013, Srinivasan et al., 2016, Richmond et al., 2018). In addition, our group defined a subset of slow cycling LGR5+ cells marked by the expression of *Mex3A* in healthy mouse epithelium (Barriga et al., 2017). RT-qPCR mRNA analysis showed that LGR5+KI67- tumor cells are highly expressing *MEX3A* among other markers (Figure 3.18), indicative that perhaps the LGR5+KI67- subset of tumor cells are reminiscent of a quiescent CSC population. The presence of slow cycling LGR5+ cells that mediate tumor regrowth has also been discovered in other tumors (Sánchez-Dánes et al., 2018), underscoring the importance of better characterizing this subpopulation. The LGR5-EGFP/KI67-RFP double knock-in in PDOs provides a valuable tool to deepen in the study of quiescent cancer stem cells.

In summary, we hypothesize that cellular hierarchy is lost during the transition from early lesions to advanced colorectal carcinomas, and that phenotypic heterogeneity in tumor cell composition is most likely dictated by the microenvironment or other extrinsic signals, in agreement with the work of others (Prasetyanti et al., 2013, Lenos et al., 2018). Our findings complete earlier studies based on mice adenomas that report the presence of CSC in colorectal tumors (Schepers et al., 2012, Kozar et al., 2013, Nakanishi et al., 2013). Adenomas still retain some degree of tissue organization, with

crypt-niche defined areas, and their dependency on stem cell hierarchy decreases as the tumors evolve. Indeed, a similar evolution and loss of hierarchy has been shown for squamous cell carcinomas (Driessens et al., 2012).

Moreover, we propose that the stem cell program is required for tumor initiation in transplantation experiments (Barker et al., 2009, Schwitalla et al., 2013), but does not determine an advantage for growth in intact tumors, as different subpopulations fuel tumor expansion in CRCs. We cannot rule out, nevertheless, that some CRC tumors do retain a hardwired hierarchical organization, as observed by Shimokawa et al., 2017. In addition, we report the presence of a quiescent cancer stem cell population. Finally, we also report plasticity between tumor subpopulations, coincident with increasing evidence provided by others (Shimokawa et al., 2017, Melo et al., 2017).

### **Proliferation in CRC tumors depends on protein synthesis capacity**

I would like to conclude this part of the discussion by linking our findings to the work of the lab by Morral and Stanisavljevic et al, 2019 (*Nature, under revision*). They observed that tumor cells in contact with the stroma retain most of the ribosomal RNA (rRNA) expression, and are thus actively producing proteins, whereas tumor cells further away from the stromal niche shut down protein synthesis. Interestingly, this phenomenon is independent of *LGR5* marker expression, and puts forward a model where cell proliferation is dependent on biosynthetic capacity rather than stemness potential. The data perfectly correlates with our findings that tumor subpopulations fuel tumor growth regardless of their stem or differentiation status. Moreover, they define a terminally differentiated population that has lost all protein synthesis capacity. This represents a bias in all lineage tracing experiments, as this subset of differentiated cells will never be able to express the marker necessary for their tracing since they no longer synthesize protein. It also explains why the mRNA of the pan-differentiation marker *KRT20* is expressed in early differentiating cells, as the protein has to be produced while synthesis machinery is still switched on. *KRT20* protein long half-life ensures it is retained in the cell during terminal differentiation (Morral and Stanisavljevic, *under revision*).

### **EMP1+ cells as the cell of origin of metastasis**

In chapters I and II of the results section of this thesis we focused on studying the relevance of the stem cell hierarchy in advanced CRC tumors. As we have discussed, overall tumor organization is reminiscent of that of the homeostatic crypt, including the presence of stem-like and differentiated-like cells organized in defined compartments. However, the expression of one or another genetic program does not determine the potential of cells to proliferate and persist for long time in the tumors. Following the same rationale, metastatic cells might not necessarily express a CSC gene program, a concept that is widespread yet lacks robust experimental proofs (Oskarsson et al., 2014). In chapter III, we have tried to characterize the cell of origin of metastasis using *EMP1* as a marker gene for this population. We focused our efforts on the study of the metastatic process in CRC because metastases are the main cause of death in

these patients (Misiakos et al., 2011, Riihimaki et al., 2016).

The metastatic capacity of a cell is determined by several factors. First, the acquisition of mutations in key pathways (WNT, EGF/MAPK, TP53 and TGF- $\beta$ ) endorses CRC cells with niche-independent growth. However, metastasis is a stringent process and not all CRC cells are able to survive in distant organs despite harboring such mutations. It has been proposed that only a subset of long-term self-renewing cells retained the ability to form metastasis (Dieter et al., 2011). Importantly, several studies prove that primary and metastatic disease fundamentally share the same somatic mutations (Jones et al., 2008, Baldus et al., 2010, Brannon et al., 2014). This result indicates that metastatic capacity cannot depend on the mutational status solely, but rather must be dictated by microenvironmental cues that predispose or endow cells at the periphery of the tumors with the capacity to migrate and form metastasis. Many studies show how the interplay between tumor microenvironment (TME) and cancer epithelial cells is key for the acquisition of metastatic traits (reviewed in Tauriello et al., 2016).

In spite of the intensive research in the field, it is not clear which cells acquire the ability to migrate, since most experiments to test metastatic capacities have been performed by isolating tumor cells and transplanting them into recipient hosts. The ability to identify these cells from the tumor bulk would provide means for the development of targeted therapies against metastasis. For instance, Mangues and colleagues were able to reduce tumor cell intravasation by specifically targeting CXCR4+ cells, a marker for migratory cells (Céspedes et al., 2018). Based on these ideas, we aimed to define a robust marker for the cells that acquire invasive properties in CRC.

#### *EMP1 is highly expressed at the tumor borders*

By addition of a TOM fluorescent reporter under *EMP1* control, we visualized *EMP1*+ cells evenly distributed across tumor glands. The observation, nonetheless, that the *EMP1*-High cells accumulated in tumor buds, both at the invasive borders and intermingled in stromal areas of subcutaneous xenografts and orthotopic tumors, prompted us to further investigate the migratory capacities of these cells. Tumor buds at the invasion fronts (i.e. the spreading of single or small groups of cells at the tumor edges) are associated with metastatic disease in CRC (Prall et al., 2005, Zlobec et al., 2014, van Wyk et al., 2015). CRC metastases, mostly localized at the liver, display a similar organization to the primary tumor (Merlos-Suárez et al., 2011). In consonance, we have also observed a similar distribution of *EMP1*-TOM+ cells in the primary and metastatic nodules

Of note, the expression of *EMP1* is very broad in CRC, with TOM labelling around 60% of tumor cells in PDO7 cells and most cells in MTO93 derived xenografts (>90%). This effect is likely enhanced due to the long half-life of TdTOMATO protein (Muzumdar et al., 2007) (i.e. it might be that descendants of *EMP1* retain certain TOM level although they do not longer express *EMP1*). The lack of good antibodies against *EMP1* protein hampers the assessment of the endogenous expression of *EMP1* in tumor samples. Nevertheless, cells at CRC invasion fronts express elevated levels of

this gene. Based on this observation, we established a threshold for *EMP1* expression of 10-15% highest TOM+ cells to define a subpopulation with putative invasive and migratory abilities. Indeed, these TOM-Bright cells express elevated *LAMC2* and *L1CAM* levels, two markers of migratory cells. *LAMC2* encodes for the  $\gamma 2$  subunit of Laminin5, an epithelium attachment molecule widely associated with tumor budding and invasive properties (Lenander et al., 2001, Shinto et al., 2005, Huang et al., 2017, Galatenko et al., 2018). *L1CAM* is a neuronal cell adhesion molecule that facilitates motility and blood vessel entry (Gavert et al., 2005, Valiente et al., 2014, Er et al., 2018). Recent work by the Massagué lab suggests a role for *L1CAM* in metastasis initiation in multiple tumor types (Er et al., 2018). It is worth pointing out that co-staining of those markers with EMP1-TOM in PDX tumors revealed that some EMP1-High buds were lacking both markers. We speculate that *EMP1* is a broader maker of cell migration than *LAMC2* and *L1CAM*

#### *Invasive capacity and stemness might be uncoupled*

The fact that we propose a differentiated-like population as the cell of origin of metastasis contradicts the general belief that metastatic cells are part of the CSC pool (Brabletz et al., 2012, Baker et al., 2015, Dame et al., 2018). In support of this hypothesis, direct liver seeding via intraspleen-injection showed that EMP1-High cells formed metastatic nodules with the same efficiency as the rest of tumor populations, despite being poorly tumorigenic in organoid formation and tumor initiation assays. This implies tumor initiation and metastatic growth do not depend upon the same mechanisms, and that distant organ colonization may not require of a CSC phenotype. Actually, during the development of the project a publication by De Sauvage laboratory proposed LGR5+ cells are indispensable for the metastatic process (Melo et al., 2017). Using quadruple mutant MTOs harboring an LGR5-GFP-DTR (diphtheria toxin receptor) knock-in, they prove that selective LGR5+ cell ablation impedes liver metastasis growth. Interestingly, though, they also report that diphtheria toxin treatment does not affect tumor invasiveness, as budding can still be observed in the primary sites.

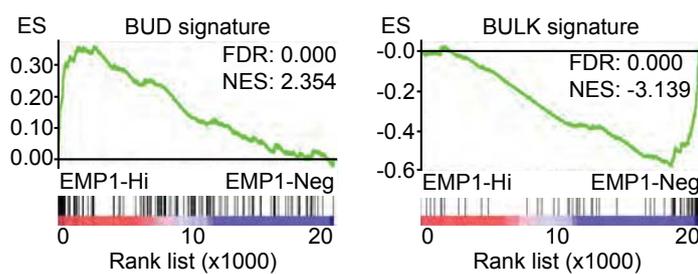
We hypothesize that the EMP1-High cells are a differentiated-like population endorsed with the capacity to invade the adjacent tissues and putatively extravasate and colonize distant organs. Once micro-metastases are established, these cells are able to revert their phenotype and recapitulate the original tumor composition (Merlos-Suárez et al., 2011) regardless of the cell of origin, as we have observed in metastases-initiation experiments (**section 3.5**). Moreover, these data fit in well with the phenotypic plasticity of CRC cells during organoid formation and tumor initiation assays. It would be highly relevant to understand if the EMP1-High cells immediately revert to LGR5+ upon arriving to the metastatic site or whether the metastases are able to grow independently of CSC marker expression. The experiments by Melo and colleagues on LGR5+ cell ablation points to a need of stem cell reversion for micro metastasis outgrow. This does not exclude our thesis that the cells that detach from the primary tumor and initiate the metastatic cascade are actually non-CSC like. Furthermore, other have reported low levels of stem cell markers in the invasive borders: earlier studies already showed that less than 5% of the tumor buds expressed *CD44*

or *CD133* (Hostettler et al., 2010) and *LGR5* was present in around 15% of them only (Kleist et al., 2011)

Thanks to the EMP1-TOM reporter, we have accumulated a number of further evidence to support the idea that EMP1-High cells are part of the metastatic CRC population.

#### *EMP1 gene signature correlates with invasive hallmarks*

Metastasis-derived MTOs are enriched in EMP1-High gene signature expression compared to their parental organoids. Our studies also revealed that the gene expression of EMP1-High vs Negative is enriched in genes express in the tumor buds whereas the reverse comparison is enriched in genes present in the tumor bulk (De Smedt et al., 2017). i.e. The Bud gene signature is highly upregulated in the EMP1-High fraction whereas the Bulk signature is expressed by the EMP1-Neg cells (Fig 4.3). Moreover, EMP1-High cells upregulate genes involved in EMT. EMT has long been associated to tumor invasion, as cells need to gain certain mesenchymal traits to migrate (Nieto et al., 2016). We propose that EMP1-High cells display a partial EMT that favors cell migration, either as single cells or in collective migratory fashion, which is proposed to improve organ colonization (Aceto et al., 2014, Cheung et al., 2016, Gkoutela et al., 2019). Stroma surrounding the tumor facilitates migration by creating collagen I fibers that serve as migratory paths for the invading cells (Provenzano et al., 2006, Conklin et al., 2011). Indeed, *in vitro* grown organoids express higher levels of EMP1-TOM when cultured with collagen I, indicating that *EMP1* is upregulated during the invasion process by changes in the extracellular matrix or its stiffness. We also found that among the most highly expressed genes in the EMP1-High cells metalloproteinases (*MMP7*) and proteases (*KLK10*, *CTSE*), which are involved in matrix remodeling and metastasis progression (Adachi et al., 1999, Zlobec et al., 2010, Kobayashi et al., 2015, Bufu et al., 2018)



**Figure 4.1: EMP1-High cell signature is enriched in tumor bud gene expression.** Gene set enrichment analysis of EMP1-High vs Negative gene expression array of PDO7 in specific gene signatures for the tumor buds (BUD) and tumor center (BULK). EMP1-High cells correlate with budding gene expression. The gene signatures were defined in De Smedt et al., 2017.

In addition, we have observed that *EMP1* expression responds to low oxygen concentrations. It is well described that hypoxic tumor microenvironments favor cell dissemination via *HIF1 $\alpha$*  mediated effects (reviewed in Rankin and Giaccia, 2016). In addition, *EMP1* level also increased in response to TGF- $\beta$  activation in PDO6. TGF- $\beta$  is largely produced by CAFs and its levels are high at invasion fronts. These results

suggest a putative mechanism for *EMP1* gene program activation in the tumor edge: cells at the tumor border are in close contact with stromal cells and extracellular matrix components that induce an EMT-like phenotype. A hypoxic environment reinforces this phenotype, and TGF- $\beta$  responsive tumors respond to signals secreted by fibroblasts to enhance their migratory properties.

#### *Linking EMP1-High cells to the recruitment of pro-metastatic myeloid cells*

A relevant finding is the identification of a chemokine expression program specific of EMP1-High cells. In PDO7, EMP1-High cells express cytokines involved in myeloid cell recruitment; *CXCL1*, *CXCL7*, *CXCL8* and *CCL5*. MTO93 partially reproduced this effect by expression of *Cxcl7* and *Ccl5* chemokines. Myeloid lineages, comprising neutrophils and myeloid derived suppressor cells (MDSCs) play important roles during the metastatic dissemination in multiple tumor types through matrix remodeling, adaptive immunity suppression, EMT promotion and pre-metastatic niche formation (reviewed in Leach et al., 2017). We are not the first ones to report an enrichment in chemokine secretion at the CRC tumor invasive fronts (Kobayashi et al., 2015). This finding proposes a novel metastatic mechanism of EMP1+ CRC tumor cells, and putative blockade of CXCR1/2 receptors might greatly improve treatment options, as shown for breast cancer (Acharyya et al., 2012, Wculek et al., 2015).

#### *EMP1 protein itself might play a role in colonization*

So far, we have focused in using *EMP1* as a marker for a cell population, without analyzing its function. *EMP1* gene encodes for a cell adhesion glycoprotein of the PMP22 family, and its functions are poorly described (Taylor et al., 1995). Its role in cancer and metastasis is controversial: some reports associate *EMP1* expression with chemotherapy resistance and poorer survival in leukemia and lung cancers (Jain et al., 2005, Aries et al., 2014), whereas others suggest a role for *EMP1* in suppressing proliferation and invasion in breast and CRC tumors (Sun et al., 2014a, b and c).

A very recent publication, though, demonstrates that *EMP1* expression is upregulated in prostate cancer cells upon stroma-tumor cell interaction, coincident with the increased *EMP1* levels we observe at CRC invasive fronts. Moreover, they report that *EMP1*-overexpressing tumor cells display higher mobility *in vitro* and metastasis formation capacity (Ahmat Amin et al., 2018). These studies propose that *EMP1* mediates this pro-metastatic effect via Rac1 activation. Rac1 is a small GTPase involved in cell adhesion and migration processes in a variety of systems (reviewed in Raftopoulou and Hall, 2004). These findings suggest a mechanism by which EMP1 protein would be implicated in enhancing cell migration in the tumor leading edge. To validate this hypothesis we are currently developing *EMP1* knockouts in our PDO and MTO collection

#### *Ablation of EMP1-High cells is essential to prove their role in metastasis*

Taking all the results together, we propose *EMP1* as a broad marker comprising both differentiated non-invasive and invasive cells, being EMP1-High the ones located at

invasion fronts. To demonstrate the role of EMP1-High cells in metastasis we engineered PDO7 and MTO93 to carry an EMP1-inducible Caspase9-TOM knock-in, and proved the efficiency of iCasp9 in eliminating the 10% Highest EMP1-population in all organoids, both *in vitro* and *in vivo*. The 10% highest fraction is the one we have described as invasive-like and neutrophil attractant, since the EMP1-Med and -Low are devoid of such metastatic-like expression programs.

Despite our efforts, we have not been able to obtain a consistent experimental setting to study metastasis from primary orthotopic tumors using with our organoid models. This limitation hampered the functional analysis of EMP1-High cell in metastasis. We continue to work on these experiments and hope to set up a suitable system in the near future.

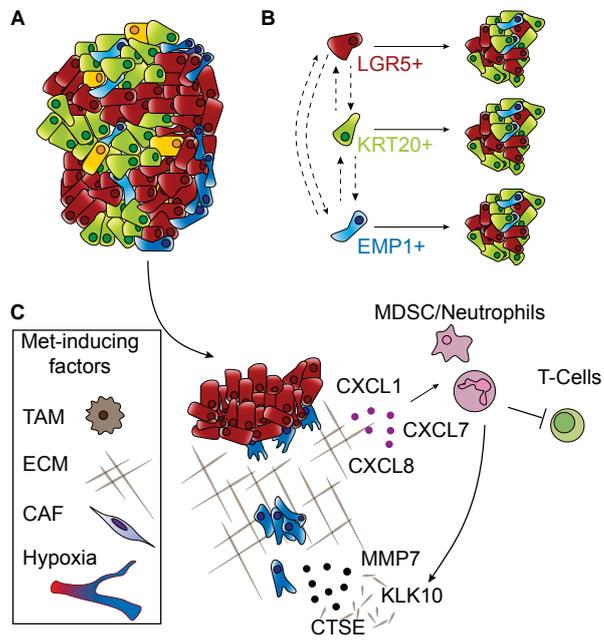
One of the more promising approaches is the development of models for CRC metastasis that recreate extirpation of the primary CRC and subsequent disease-relapse similar to the process observed in patients. This new model of CRC relapse may enable the observation of metastasis development independently of primary growth, extending animal survival rates as well and allowing longer development of metastases observation. We believe this strategy will be of great help not only to our project, but also for future studies on treatments in the adjuvant setting.

In addition to these experiments, our interests are also focused in isolating tumor-circulating cells disseminated from primary transplanted tumors. This would provide valuable data about the nature of those cells, as few studies have been able to characterize them. In Grillet et al., 2017, they suggest CTCs from CRC patients harbor CSC features, although they did not take into account plasticity of the cells once they are seeded in culture in stem cell promoting media. Our aim is to identify whether the disseminated cells are EMP1+ or, on the contrary, harbor a stem-like phenotype. A plausible hypothesis, taking into account the work of Melo et al., 2017, would be that EMP1+ cells are the ones migrating from the primary tumor but, once they are in CTC status or upon reaching the metastatic site, they revert to a stem-like state.

### **Working model and therapeutic implications**

Our findings (Figure 4.2) have implications for the development of future therapies. The actual standard of care for patients is based on 5-fluorouracil chemotherapy combined with oxaliplatin or irinotecan, which eliminates the bulk of proliferative tumor but moits quiescent or slow-cycling cells. Development of targeted therapies against key pathways, such as EGF, renders good initial responses but are soon overcome by acquired resistances (Hata et al, 2016., Russo et al., 2018). Many efforts have been focused during the last years in eliminating the CSC, either by LGR5-specific antibodies (Junttila et al., 2015) or by inhibition of CSC specific proteins, such as BMI1 (Kreso et al., 2014). These treatments, nevertheless, have considerable side-effects as normal ISC also rely on such markers. Moreover, loss of stem cell hierarchy and high levels of plasticity in advanced CRCs limit the utility of such drugs. New focus on targeting the tumor microenvironment to elicit an anti-tumoral response in stromal

and immune cells holds more promise of success, especially in blocking metastatic spread (Tauriello et al., 2018).



**Figure 4.2: Working model.** **A.** CRC tumors are composed of a mixture of differentiated-like (green) and stem-like (red) cells. **B.** All tumor subpopulations contribute to tumor expansion, as we have seen by *LGR5*, *KRT20* and *EMP1* lineage tracing. **C.** In response to metastasis inducing factors, some cells at the border of the tumor acquire migratory capabilities, recruiting MDSC and neutrophils to the tumor site and disrupting the ECM. We have identified these invasive cells by *EMP1* gene expression.

Part V

## CONCLUSIONS



CONCLUSIONS

---

The main conclusions from this thesis are:

1. The combination of CRISPR/Cas9 gene editing and patient-derived organoid cultures enables the study of tumor cell populations by lineage tracing and cell ablation experiments in human CRCs within an intact environment.
2. Advanced human colorectal tumors retain cellular heterogeneity, with stem-like and differentiated-like compartments reminiscent of the healthy intestinal epithelium.
3. Expression of the *LGR5* gene defines a subpopulation of cells in human CRCs that retains the expression of a stem cell program and that displays elevated tumorigenic potential according to tumor initiation assays.
4. *LGR5*<sup>+</sup> CRC cells generate tumor cells that undergo progressive differentiation with a slow kinetics.
5. Around 50% of the *LGR5*<sup>+</sup> cancer stem cell pool is composed by cells that are not actively cycling. These cells remain quiescent for long periods of time. CRCs contain a differentiated-like population marked by *KRT20* expression, which include both absorptive-like and secretory-like cells.
6. Differentiated tumor cells exhibit less tumorigenic potential than *LGR5*<sup>+</sup> cells as shown by tumor initiation assays.
7. Approximately one third of the differentiated-like cells in human tumors are in a proliferative state.
8. Tumor expansion is achieved thanks to the proliferation of both cancer stem cells and differentiated-like pools of cells.
9. There is extensive phenotypic plasticity. Stem cell-like cell generate differentiated tumor cells but also differentiated tumor cells produce stem cell-like cells in vitro and in vivo. This results support the existence of shallow non-hardwired hierarchy in human CRCs.
10. *EMP1* labels invasive cells in human and mouse CRCs.
11. *EMP1*<sup>+</sup> CRC cells express markers of differentiation, are comprised within the *LGR5* low/negative tumor cell population and express elevated levels of chemoattractants for myeloid cells.



Part VI

MATERIALS AND METHODS



## MATERIALS AND METHODS

---

The methods described herein are an adaptation from Cortina, Turon et al., 2017.

### 6.1 ORGANOID CULTURES

#### 6.1.1 *Patient Derived Organoids*

Tumor sample used to expand PDO6 was obtained from an individual treated at Hospital de la Santa Creu i Sant Pau, under informed consent and approval of the Tumor Bank Committees according to Spanish ethical regulations. The study followed the guidelines of the Declaration of Helsinki, and patient identity for pathological specimens remained anonymous in the context of this study. Tumor cells were grown as organoids embedded in BME2 (Basement Membrane Extract 2, AMSbio) using a modification of the media described by the Clevers lab (van de Wetering et al., 2015) (Advanced DMEM/F12, 10 mM HEPES, 1x Glutamax; 1x B-27 without retinoic acid, 20 ng/ml bFGF (basic fibroblast growth factor); 50 ng/ml EGF (epidermal growth factor), 1  $\mu$ M LY2157299 and 10  $\mu$ M Y27632, recombinant R-SPONDIN1 (1  $\mu$ g/ml) and recombinant Noggin (100 ng/ml). PDO7, a kind gift of G. Stassi (University of Palermo), was obtained from the dissociation of whole CRCs in suspension as described elsewhere (Lombardo, Scopelliti et al., 2011). Upon arrival to our laboratory, they were cultured with the medium described above. PDOs were passaged once a week by Trypsin-EDTA disaggregation 20 min at 37°C, and fragmented organoids were seeded again in BME2 drops. All cells were tested weekly for mycoplasma contamination with negative results.

For *in vitro* lineage tracing studies, 5-days grown organoids embedded in 100% BME2 were treated for 24h with 4-hydroxytamoxifen (4-OHT, Sigma H6278). For *in vitro* EMP1+ cell ablation, CRISPR/Cas9 modified organoids were grown in 10% BME2 suspension, and treated at days 2 and 5 with CID 100nM (AP20187, ApexBio).

#### *Collagen cultures*

For collagen cultures, cells were resuspended in Collagen I 1.5 mg/ml or 8 mg/ml (Corning, 45354249) and seeded in 25 $\mu$ l drops. After one hour of incubation at 37°C, PDO media was added to the plates.

#### 6.1.2 *Mouse Tumor Organoids*

MTO93 was derived by Dr. Tauriello as described in Tauriello et al., 2018. The organoids were maintained in MTO culture medium (Advanced DMEM/F12, 10 mM HEPES, 1X Glutamax; 1X B-27 without retinoic acid, 50 ng/ml EGF (epidermal growth factor), 1  $\mu$ M LY2157299 and recombinant Noggin (100 ng/ml). MTOs were passaged once

a week by Trypsin-EDTA disaggregation 20 min at 37°C, and fragmented organoids were seeded again in BME2 drops. All cells were tested weekly for mycoplasma contamination with negative results. *In vitro* EMP1+ cell ablation of targeted MTOs was done with same settings as PDOs.

### *Collagen cultures*

For collagen cultures, cells were resuspended in Collagen I 1.5 mg/ml or 8 mg/ml (Corning, 45354249) and seeded in 25µl drops. After one hour of incubation at 37°C, MTO media was added to the plates.

## 6.2 CRISPR/CAS9 KNOCK-IN DESIGN

### 6.2.1 Donor plasmid construction

750 bp (*LGR5*, *KI67*, *AAVS1*, *KRT20* and *EMP1* constructs) of 5' homology arm (HA) and 3' HA were amplified from PDO7 gDNA and MTO93 gDNA respectively or synthesized by gene synthesis (Thermo Fisher) and cloned in pShuttle or pDONR vectors. LF2A-EGFP-BGHpA, linker-tagRFP2, LF2A-CreERT2-BGHpA, Lox-tagBFP2-3xpA-Lox-tdTomato-BGHpA, linker-IRFP insertion cassettes were generated by gene synthesis (Thermo Fisher) and cloned in the 5'HA-3'HA previously engineered pShuttle or pDONR vectors. Inducible Caspase9-TdTOMATO insertion cassette was kindly provided by Dr. Sato (Shimokawa et al., 2017).

### 6.2.2 Guide RNA design

Guide RNAs were designed using the <http://crispr.mit.edu> web tool. To select for the most suitable sgRNAs, we applied the following criteria: i. localization of the sgRNA as near as possible to the desired site of insertion to maximize homologous recombination efficiency, ii. Cas9-mediated double strand break downstream of STOP codon to prevent NHEJ-induced indels in the ORF, iii. guides selected to anneal at the intersection between the 5' homology arm and 3' homology arm so that the donor plasmid is protected from Cas9 cut, iv. minimum off-target score according to <http://crispr.mit.edu> and maximum Doench activity score (Doench et al., 2014). sgRNAs used to modify PDOs were as follows: *LGR5* (TGCTCTAATTAATATGTGA), *KI67* (TTTGACAGAAAATCGAACT), *AAVS1* (GTCCCTAGTGGC-CCCACTGT), *KRT20* (ATATCTAAATAGCTACCAGA), *EMP1* (TCCTGAGAAAGAAA TAAGGC). gRNA used to modify MTOs were as follows: *LGR5* (GTCTCTAGTGAC-TATGAGAG), *EMP1* (GAAATAAGCCGAATACGCTCA).

### 6.2.3 px330-iRFP Cas9 plasmid construction

Px330 Cas9 plasmid from Feng Zhang's laboratory was obtained from Addgene (ref. 42230) and was modified by the introduction of a SV40promoter-iRFP expression cassette downstream of Cas9 by FseI - EcoRI. In addition the BbsI site of iRFP was

silenced by site-directed mutagenesis. gRNAs were cloned in px330-iRFP as described in <http://www.genome-engineering.org/crispr>

#### 6.2.4 CRISPR/Cas9 knock-in generation in PDOs and MTOs

##### *Nucleofection*

For PDO and MTO nucleofection, 2 million of single-cell trypsinized PDO cells were nucleofected with 7 µg of donor plasmid and 2 µg of px330-iRFP Cas9 corresponding plasmids using Lonza nucleofector kit V (VVCA-1003) and program A-32 in an Amaxa-II nucleofector following manufacturer protocol. For PDO nucleofection with two donors simultaneously (LF2A-EGFP/LF2A-CreERT2), we nucleofected 2 million cells with 6 µg of the Cre donor, 2 µg of the EGFP donor and 1.5 µg of the guide plasmid.

##### *FACS strategy and generation of single-cell derived organoids*

Nucleofected cells were embedded in BME2 drops and cultured for 2-3 days in PDO or MTO medium. We then isolated cells that were iRFP+ (i.e. had incorporated the targeting vector) by FACS (Red B channel) and cultured them in 3D for about 18-20 days. For the AAVS1 lineage tracing cassette, iRFP+/tagBFP2+ double positive cells were sorted. The LGR5-LF2A-CreERT2, KRT20-LF2A-CreERT, KRT20-iRFP, EMP1-LF2A-EGFP/CreERT2 and EMP1-iCasp9-TOM (PDO7) knock-ins were directly plated in single cell format after iRFP+ purification. For the rest of the integrations we let the cells 18-20 days in culture, and we then observed the emergence of a cell population that expressed the marker gene of interest suggesting that some cells had integrated reporter vector. We selected the cell population positive for expression of EGFP (LGR5-LF2A-EGFP, BlueB channel), mtagRFP2 (KI67-RFP, GreenD channel) mtagBFP2 (AAVS1-BFP-TOM, VioF channel) and TdTOMATO (EMP1-iCasp9-TOM, GreenD channel, only MTOs) respectively, which were at the same time negative for iRFP. Cells were seeded in a 96-well format to derive single-cell clones. Wells with more than one cell per drop were discarded.

##### *Specific genotyping PCRs*

Single-cell derived clones were lysed in buffer consisting of 10 mM Tris, 1 mM EDTA, 1% Tween 20 and 0.4 mg/ml proteinase K for 1h at 55°C. The lysate was directly used in the specific integration PCR. For the 5' specific integration PCR a forward primer upstream of the 5' homology arm and a reverse primer at the beginning of the inserted cassette were used. Similarly, for the 3' specific integration PCR a forward primer at the end of the inserted cassette and a reverse primer downstream of the 3' homology arm were used. The PCR conditions were as follows: DNA Polymerase (BioTools 10012-4103) 95°C 2min x38 (95°C 30s – 55°C 30s – 72°C 1:30min) 72°C 5min - hold 16°C. Used primer sequences are shown in Table 6.1.

##### *Southern blot*

Clones that were correctly targeted, based on 5' and 3' specific integration PCRs, were further checked for off-target cassette insertions by southern blot. Genomic DNA was extracted using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma

Locus-insertion	PCR	Primer
LGR5-LF2A-EGFP	5'specific	F: GTTTTTTGTCCATTTTTGCTTCAG R: ACCACCCCGGTGAACAGC
	3'specific	F: TCGGCATGGACGAGCTGTACAAG R: AGTGGAATTCCCTTCTGAGCTTTG
LGR5-LF2A-CreERT2	5'specific	F: GTTTTTTGTCCATTTTTGCTTCAG R: CGAACATCTTCAGGTTCTGCGGG
	3'specific	F: TTCCTTGCAAAAGTATTACATCAC R: AGTGGAATTCCCTTCTGAGCTTTG
KI67-tagRFP2	5'specific	F: CCGTCGAGTCTTACAATAAAAACG R: TCAATTAAGTTTGTGCCCCAG
	3'specific	F: ATGGTGTCTAAGGGCGAAGAG R: TACAGAGTACTGGTGTCACTTCCTG
AAVS1-BFPTOM	5'specific	F: GGACCACTTTGAGCTCTACTGG R: GGGCCATTTACCGTAAGTTAIGTA
	3'specific	F: GCATTGTCTGAGTAGGTGTCATT R: ATGAGATGGTGGACGAGGAAGG
KRT20-LF2A-CreERT2	5'specific	F: GCACCCTGGAAGAGAGAGGTAAG R: CGAACATCTTCAGGTTCTGCGGG
	3'specific	F: TTCCTTGCAAAAGTATTACATCAC R: CCGTAGAGACAGGAAGTAGATTAG
KRT20-iRFP	5'specific	F: GCACCCTGGAAGAGAGAGGTAAG R: CTCCGAGTCGAACACATCGGCCCGC
	3'specific	F: CATCACCGAACGCCGTACTACG R: CCGTAGAGACAGGAAGTAGATTAG
EMP1-LF2A-EGFP	5'specific	F: CTTCTGTGTCATTGCCCTCCTGG R: CTTGAAGAAGATGGTGCCTCCTGG
	3'specific	F: CAGCAGAACACCCCATCGGCGACG R: GAGGAAGAGTAATAGGATTGTGC
EMP1-LF2A-CreERT2	5'specific	F: CTTCTGTGTCATTGCCCTCCTGG R: AAGCATGTTTAGCTGGCCCAAATGT
	3'specific	F: CAAGAACGTGGTGGCCCTCTATGAC R: GAGGAAGAGTAATAGGATTGTGC
EMP1-iCasp9-TOM (PDO)	5'specific	F: CTTCTGTGTCATTGCCCTCCTGG R: AAGACGGCAATATGGTGGAAAAT
	3'specific	F: GCATTGTCTGAGTAGGTGTCATT R: GAGGAAGAGTAATAGGATTGTGC
EMP1-iCasp9-TOM (MTO)	5'specific	F: CTAATGCTCCTGGTTCTGTCTTG R: AAGACGGCAATATGGTGGAAAAT
	3'specific	F: GCATTGTCTGAGTAGGTGTCATT R: AGTGTGATGAAAGCTGTTGAGAG

**Table 6.1: Primers used for the specific integration PCRs.** Primer sequences for assessing the correct integration of each knock-in. The PCRs are designed to amplify a fragment that includes the homology arms (before and after them respectively) and part of the knock-in.

Locus-insertion	Primer	R.E
LGR5-LF2A-EGFP	F: CCACCATGGTGAGCAAGGGCGAGG R: TTACTTGACAGCTCGTCCATGCC	HindIII
LGR5-LF2A-CreERT2	F: ATGGACATGTTTCAGGGATCGCCAGG R: GATTACGTATATCCTGGCAGCGATC	PvuII
AAVS1-BFPTOM	F: GGGCATGGCACC GG CAGCACC R: CCTACTTGACAGCTCGTCCATGCC	PvuII
KRT20-LF2A-CreERT2	F: ATGGACATGTTTCAGGGATCGCCAGG R: GATTACGTATATCCTGGCAGCGATCC	MfeI
EMP1-LF2A-EGFP	F: CCACCATGGTGAGCAAGGGCGAGG R: TTACTTGACAGCTCGTCCATGCC	NdeI
EMP1-LF2A-CreERT2	F: ATGGACATGTTTCAGGGATCGCCAGG R: GATTACGTATATCCTGGCAGCGATC	NdeI

**Table 6.2: Southern blot probes.** Primers used for southern blot probe generation and restriction enzymes used to digest each construct. R.E, restriction enzyme

G1N70-1KT). 10 µg of genomic DNA were digested overnight with the appropriate restriction enzyme (Table 6.2) and separated on a 0.8% agarose gel. DNA fragments were transferred by capillarity to a Hybond-N+ membrane (GE Healthcare RPN303B) overnight. Probes were generated by PCR (protocol as described in the previous section, primers in Table 6.2) and radioactively labelled with  $\alpha$ -[32P]dCTP using the MegaPrime labelling kit (GE Healthcare RPN1604). Hybridization with the probe was carried out overnight at 60°C. Probes were detected using a Phosphorimager plate.

### 6.3 LENTIVIRAL INFECTIONS

Lentivirus containing Luciferase-Puromycin or EGFP-Luciferase constructs were generated in HEK293T cells transfected with packaging vectors. Virus producers HEK293T cells were cultured in DMEM 10% FBS medium and media was collected at days 1,2 and 3 post infection. Virus were extracted from the supernatant using the LentiX-Concentrator reagent (Clontech PT4421-2). For PDO and MTO infection, organoids were seeded in BME2 layers and treated with three rounds of infection (viral load 5X) in presence of polybrene (8µg/ml). Selection was based on Puromycin resistance or EGFP fluorescence by FACS.

### 6.4 MOUSE STUDIES

#### 6.4.1 Subcutaneous xenografts

Beige/Scid (CB17-PrkdcscidLystbg/J) female mice were purchased at PRRB Animal Facility, and C57BL/6 (C57BL/6Jrj) male mice were purchased at Janvier Labs. They were used at 6 weeks of age at the beginning of the experiments, which took place during a maximum of 21 weeks until the animals showed signs of distress. Animals

were kept in Barcelona Science Park Animal Facility (SEA-PCB). All mouse experiments were approved by the Animal Care and Use Committee of Barcelona Science Park (CEEA-PCB) and the Catalan government. 150000 cells (PDO7, MTO93) or 2 million cells (PDO6) in a format of 3 to 5-days grown organoids (MTOs) or 5 to 7-days (PDOs) were injected subcutaneously into C57BL/6 male mice and Beige/Scid female mice respectively in 50% BME2-HBSS with a maximum of 4 xenografts per animal. Tumor volume was measured with manual calipers and using the formula (length x width x height)/2. Mice were sacrificed when tumors reached a maximum of 300 mm<sup>3</sup>, the animal displayed ulceration in one of the xenografts or showed symptoms of distress. Xenografts were resected and disaggregated as previously described in Merlos-Suárez et al., 2011. Human epithelial cells from disaggregated xenografts were stained with hEPCAM-PeCy7 1/150 (eBioScience 25-9326-42) or hEPCAM-APC-Vio770 1/100 (Miltenyl Biotec 130-101-161). Mouse epithelial cells from disaggregated xenografts were stained with mEPCAM-APCCy7 1/150 (BioLegend 118217). Immune and endothelial cells were discarded using anti-mCD16CD32 blocking antibody 1/200 (Tonbo Biosciences 70-0161-U500) and anti mCD31-BV421 1:200 (rat, BD biosciences, 562939cloneMEC13.3) / anti-mCD45RB-BV421 1:200 (rat, BD biosciences, 562849clone16A). DAPI 1µg/ml was added to distinguish alive/dead cells. The cell suspension was analyzed with a BD Aria Fusion FACS or Aria FACS according to the following criteria: i. debris was discarded by FSC-A/ SSC-A gating ii. Aggregates by FSC-A/FSC-W gating iii. Alive cells by DAPI negative signal (VioF ultraviolet or violet channel 360 V) iv. Human EPCAM positive cells were selected by Green A (414 V) or Red A channels (500 V) and mouse tumor cells were selected based on mouse EPCAM positivity (Red A channel 500V) or constitutive EGFP expression (Blue B channel, 450V). Cells that were positive for the criteria mentioned above were analyzed for EGFP (Blue B channel, 500 V), TdTomato, tagRFP2 (Green D channel, 500 V) or iRFP (Red B channel, 600V).

Mice bearing 50 mm<sup>3</sup> xenografts were treated with two consecutive doses of intraperitoneal tamoxifen 250 mg/kg (Sigma, T5648) for LGR5, KRT20 and EMP1 lineage tracing assays and with one dose at 1 mg/kg for KRT20 specific lineage tracing.

For acute cell ablation experiments, mice bearing 100-150 mm<sup>3</sup> subcutaneous tumors were treated four consecutive days with CID 0.08 mg/kg (for PDOs) or 0.04 mg/kg (for MTOs) (AP20187, ApexBio).

#### *Tumor initiation assays*

Viable (DAPI-) human EPCAM+ single cells from disaggregated xenografts were sorted according to their EGFP positivity and subsequently transplanted into new recipient mice at limiting dilutions. We injected 200 or 1000 cells per flank in 100 µl of BME2-HBSS 50% - 50% (n=12). Tumor volume was measured twice a week. When a xenograft reached 300 mm<sup>3</sup>, it was resected from the animal. The experiment finished when all xenografts were grown or when the animals were 21 weeks old. Differences were assessed with Log-Rank (Mantel Cox) test.

#### *Organoid formation assays*

Human EPCAM + alive single cells from disaggregated xenografts or EGFP+ alive single cells from MTO xenografts were isolated by FACS according to their LGR5-

EGFP or EMP1-TOM positivity as described before and seeded *in vitro* in 25  $\mu$ l BME2 drops containing 1000 cells/drop (n=4) per condition for LGR5-EGFP experiments and 1500 cells/drop per condition in EMP1-TOM experiments. Then, plates were scanned with a ScanR inverted microscope at day 1 post-seeding to quantify the exact number of cells seeded per drop and at experimental endpoints (day 14 post-seeding for the PDOs and day 7 post-seeding for the MTOs). Full drops were scanned taking overlapping pictures at 4x magnification and at 8 different z-stacks with a separation of 200  $\mu$ m among them. Z-stacks of each field of view were projected in a single image and the full drop was digitally reconstructed by stitching the different image projections using an Image J custom-made macro developed for this purpose. Total number and mean size of cells (i.e. any object with a diameter larger than 5  $\mu$ m) or organoids (diameter larger than 400  $\mu$ m) were counted. Differences were assessed with Student's t-test.

#### *DAPI content analysis*

A minimum of 50000 cells for each population of interest were sorted with a BD Aria Fusion FACS as described and fixed with EdU-cell cycle kit fixative for 15 min (Life Technologies, C10424). DAPI (1 $\mu$ g/ml) was incubated for 1 hour and cells were subsequently analyzed with a BD Aria FACS gated from EPCAM-positive. Cell cycle stages were determined using WinCycle v4.0.

#### *Cell cycle analysis*

Mice bearing subcutaneous xenografts were injected with EdU 80 mg/kg 3 h before their sacrifice. Xenografts were resected, disaggregated and stained with human-EPCAM-FITC (Ab112067, dilution 1/75) and DAPI at 1 $\mu$ g/ml as described above. 100000 DAPI negative cells were sorted, fixed with EdU-cell cycle kit fixative for 15 min and subsequently stained with click-iT Alexa-647 reagent and DAPI following manufacturer's protocol (Life Technologies, C10424). Cells were subsequently sorted in a FACS Aria according to their Edu-Alexa647 signal (red A channel 500 V) and their DNA content (vioF channel 340 V) gated from EPCAM-positive cells which were tagRFP2-Positive or tagRFP2-Negative.

#### 6.4.2 *Orthotopic transplantations*

NSG (NOD.Cg-Prkdcscid Il2rgtm1wjl/SZJ) and SwissNude (CrI:NU(Ico)-Fox1nu) female mice were purchased at Charles River for PDO6 and PDO7 orthotopic transplantations, and C57BL/6 male mice were purchased for MTO93 primary tumor injections. According to Animal Care Standards, all mice received analgesia (Buprenorphine 3  $\mu$ g) post-injection. In all cases, primary tumor growth was assessed by abdominal palpation and by luciferase activity. Intravital bioluminescence was measured twice a week with an IVIS-Spectrum (Perkin Elmer) imager. Luciferase activity was analyzed 1 minute after retro orbital injection of 50  $\mu$ l D-Luciferin at 15 mg/ml (Resem BV). Values were normalized per mouse on day 0, and plotted using GraphPad (v7). Liver nodule metastasis was scored macroscopically and results were analyzed with GraphPad (v7).

For long term cell ablation experiments in primary tumors, mice were treated since day 7 after surgery with three shots a week of CID 0.04 mg/kg (AP20187, ApexBio).

#### *Intracaecum cell injection*

500000 cells (PDO7, MTO93) in a format of 3 to 5-days grown organoids (MTOs) or 5 to 7-days (PDOs) were injected intracaecum in between the serosa and muscular layer in 30% BME2-HBSS (15  $\mu$ l) using a 30G needle, modified from (Céspedes et al., 2007).

#### *Nesting technique*

100-150 mm<sup>3</sup> subcutaneous xenografts were fractionated in 1 mm<sup>3</sup> pieces for caecum implantation. Tumor pieces were stitched to tip of the caecum, which was then folded over them to prevent carcinomatosis (Tauriello et al., 2018).

#### *Collagen drop transplantation*

3 to 5 days grown PDO7 organoids were recovered and seeded in collagen I drops (1.5 or 8 mg/ml), with the equivalent of 250000 cells per drop. After overnight growth in collagen I, the drops were transplanted into the caecum of mice following the protocol described in Fumagalli et al., 2017.

#### *Primary tumor resection*

MTO93 3-days grown organoids were recovered from cell culture and the equivalent to 500000 or 150000 cells were injected intracaecum (as explained above). One or two weeks post injection the whole caecum was surgically resected to eliminate primary tumor growth.

#### 6.4.3 *Intraspleen injections*

NSG (NOD.Cg-Prkdcid Il2rgtm1wjl/SZJ) female mice were purchased at Charles River for PDO7 intraspleen transplantations. According to Animal Care Standards, all mice received analgesia (Buprenorphine 3  $\mu$ g) post-injection. *In vitro* grown PDO7 EMP1-iCT cells were resuspended in HBSS for injection, using 500000 cells in 70 $\mu$ l HBSS per mouse. Intrasplenic injections were performed with a 30G needle as previously described (Warren et al., 1995).

#### *Metastasis initiation experiment*

MTO93 cells were FACS-sorted as described before and resuspended in HBSS for injection, using 10000 cells in 70 $\mu$ l per mouse. Intrasplenic injections were performed with a 30G needle as previously described (Warren et al., 1995).

### 6.5 QUANTITATIVE REAL-TIME PCR

1000 cells from either EPCAM-stained disaggregated xenografts or from *in vitro* cultured PDOs or MTOs fulfilling the fluorescence criteria stated for each experiment

were sorted directly in 45  $\mu$ l of picoprofile lysis buffer. RNA and cDNA was extracted and amplified as previously described in González-Roca et al., 2010. Concentration of the obtained cDNA was measured with a Nanodrop. 5 ng of cDNA were used per each qrtPCR well. RT-qPCRs were performed with TaqMan assays and TaqMan Universal PCR Master Mix (Applied Biosystems 4369016) or SYBR-Green PCR Master Mix (Applied Biosystems 4309155) in triplicates following manufacturer's instructions. For human mRNAs, the following TaqMan assays (ThermoFischer) were used: APCDD1 (Hs00537787m1), AURKB (Hs00177782m1), CABLES (Hs01106667m1), CDKN1A (Hs00355782m1), CDKN1C (Hs00175938m1), CEACAM5 (Hs00944025m1), CHGA (Hs00900370m1), CreERT2 (custom), CXCL1 (Hs00236937m1), CXCL2 (Hs00601975m1), CXCL7 (Hs00234077m1), CXCL8 (Hs00174103m1), DACH1 (Hs00362088m1), EFNB2 (Hs00187950m1), EGFP (Mr04097229mr), EMP1 (Hs00608055m1), EPHB2 (Hs00362096m1), FOXM1 (Hs00153543m1), KRT17 (Hs00356958m1), KRT20 (Hs00300643m1), LAMC2 (Hs01043711m1), L1CAM (Hs01109748m1), LGR5 (Hs00173664m1), MKI67 (Hs01032443m1), MUC2 (Hs03005094m1), MYB (Hs00193527m1), MYC (Hs00905030m1), OLFM4 (Hs00197437m1), PGC1 $\alpha$  (Hs01016719m1), SDCBP2 (Hs00210404m1), SMOC2 (Hs0159663m1), TdTomato (custom), UBE2C (Hs00964100g1). All qrtPCRs were normalized to B2M (Hs99999907m1) and PPIA (Hs99999904m1) expression.

For mouse mRNA, the following TaqMan assays (ThermoFischer) were used: CCL5 (Mm01302427m1), CXCL1 (Mm004207460m1), CXCL7 (Mm00470163m1), EMP1 (Mm00515678m1), KRT20 (Mm01306857m1), LAMC2 (Mm00500494m1), L1CAM (Mm00493049m1), LGR5 (Mm0043889m1), MKI67 (Mm01278671m1), MUC2 (Mm01276696m1), SDCBP2 (Mm01245859m1), SMOC2 (Mm00491553m1). All qrtPCRs were normalized to ACTB (Mm00725412s1) and PPIA (Mm002342430g1) expression.

And the following primers for SYBR Green detection were used:  
hCCL5 (F: 5'GACACCACACCCTGCTGCT3' /R: 5'TACTCCTTGATGTGGGCACG3'),  
hCXCL3 (F: 5'CGCCCAAACCGAAGTCATAG3' /R: 5'GCTCCCCTTGTTTCAGTATCT3'),  
hCXCL5 (F: 5'AGCTGCGTTGCGTTTGT3' /R: 5'TGGCGAACACTTGCAGATT3'),  
mCXCL5 (F: 5'GTTCCATCTCGCCATTCATG3' /R: 5'GCGGCTATGACTGAGGAAGG3'),  
hFABP1 (F: 5'AAGACAGTGGTTCAGTTGGAAG3' / R: 5'TGAGTTCGGTCACAGACTTGAT3')  
hSI (F: 5'TCCAGCTACTACTCGTGTGAC3' / R: 5'CCCTCTGTTGGGAATTGTTCTG3').

## 6.6 MICROARRAY ANALYSIS

Affymetrix array data were normalized using RMA background correction and summarization (Irizarry et al., 2003) as implemented in the "affyPLM" package (Bolstad, 2004) from R. Annotations for the Human PrimeView array version na32 were downloaded from Affymetrix (<https://www.affymetrix.com/analysis/index.affx>). Technical metrics were computed for each sample as described in (Eklund and Szallasi, 2008). All samples passed quality controls. A linear model was fitted to find differentially expressed genes between conditions of interest with technical metrics (pm iqr, pm median, RNA degradation) and biological replicates as covariates. The "lmFit" function from the "limma" package (Ritchie et al., 2015) was used for fitting the model. Microarray data for PDO7-LGR5-EGFP High vs. Negative cells is available

at GSE92960. Microarray data for PDO7-LGR5-EGFP/KI67-tagRFP2 is available at GSE92961. Microarray data for PDO7-EMP1 High vs Negative is not available online. Gene set enrichment analysis (GSEA) as implemented in (Subramanian, 2005) was performed on ranked lists with all genes in the array. For each gene the fold change of the most variable probe was used as representative. We also ran GSEA on previously published gene sets (Merlos-Suarez et al., 2011, Muñoz et al., 2012). We also downloaded the expression matrix from GEO accession GSE52813 (Basak et al., 2014) and performed differential expression analysis between groups “Lgr5-HiKi67-Hi” and “Lgr5-HiKi67-Low”. We defined signatures by filtering genes with a p-value lower than 0.05 and fold change larger (smaller) than 1.5 (-1.5).

## 6.7 WHOLE EXOME SEQUENCING

The library preparation for capturing of selected DNA regions (Agilent Human All Exon 50Mb v5, Agilent) was performed according to Agilent’s SureSelect protocol v1.8 for Illumina paired-end sequencing. In brief, 3.0 µg of genomic DNA was sheared on a Covaris E210 instrument (Covaris). The fragment size (150-300 bp) and the quantity were confirmed with an Agilent 2100 Bioanalyzer 7500 chip. Agilent paired-end adaptors were ligated to fragmented DNA followed by PCR amplification (6 cycles) with SureSelect Primer and SureSelect Pre-capture Reverse PCR primers and hybridized for 24 h at 65 °C. The hybridization eluate was PCR amplified (16 cycles) in order to add the index tags using SureSelectXT Indexes for Illumina. The final library quality, control of the size and the concentration were determined on an Agilent 2100 Bioanalyzer 7500 chip.

Each library was sequenced on an Illumina HiSeq 2000 instrument in a fraction of a sequencing lane, HiSeq PE Cluster Kit v3, following the manufacturer’s protocol, with a paired end run of 2x101 bp, with coverage of at least 96% of the target region covered by at least 10 sequencing reads (C10 >96%). Image analysis, base calling and quality scoring of the run were processed using the manufacturer’s software Real Time Analysis (RTA 1.13.48, HCS 1.5.15.1) and followed by generation of FASTQ sequence files by CASAVA. Samples were aligned to the human genome version hg19 with bwa (Li and Durbin, 2009) using the mem algorithm. Somatic mutations were detected using Mutect2 with default parameters (McKenna et al., 2010) and compared to those obtained by VarScan2 (Koboldt et al., 2012). To predict the effects of single nucleotide changes, we used the SnpEff program (Cingolani et al., 2012). Sequences have been deposited at ENA (PRJEB18738). The control sample was previously submitted to ENA with accession number PRJEB7932 (Organoid P8) (Calon et al., 2015).

## 6.8 IMMUNOSTAINING AND QUANTIFICATIONS

### 6.8.1 *Immunostaining and confocal imaging of in vitro cultured PDOs*

For LGR5-EGFP PDO7 modified organoids, cells were seeded and grown for 2 weeks without trypsinization and then harvested using Matrisperse Cell Recovery Solution (Corning, 354253). Full organoids were seeded at 100000 cells per well in microscopy chamber slides in thin layers of BME (20 µl per chamber). For alive imaging,

organoids were 30 min incubated at 37 °C with Hoechst 33342 1:1000 (Molecular Probes, R37605) or RedDot1 1:200 (Biotium, 40060). For immunocytostainings, samples were fixed with 4% paraformaldehyde for 15 min and blocked with 20 mM glycine for 20 min. Permeabilization was achieved with 0.5% Triton X-100 for 30 min at room temperature, then a second blocking step with 1% BSA for 1 h at room temperature was performed.

For EMP1-TOM PDO7 modified organoids, cells were seeded and grown for 7 days without trypsinization and fixed o/n in formalin 37%. The fixative was washed for 2h in PBS and then the drops were transplanted to a histology microcassette for paraffin embedding. 4 µm sections were stained according to standard procedures. Antigen retrieval was carried out with boiling Tris-EDTA buffer for 20 min, then samples were blocked with Peroxidase-Blocking Solution (Dako: S202386) for 10 min at room temperature.

Primary antibodies were incubated overnight at 4 °C and after that, secondary antibody was added for 1 h at room temperature in the dark. DAPI (1:5000) was added after secondary antibody staining for 15 min at room temperature. Washes were performed in between each step with PBS. Primary antibodies against KRT20 (mouse, Dako, M7018), MUC2 (mouse, AbCam, Ab118964) GFP (goat, Abcam, ab6673) and TOM (goat, Sicgen, Ab8181-200) were used at 1:100 dilution. Secondary antibodies donkey anti-goat and anti-mouse conjugated to Alexa-488 (DαM,DαG) and Alexa-647 (DαM, DαR, DαG) (Life Technologies, A21202, A11055, A31571, A31573, A21447 respectively) were used at 1:200 dilution. Images were taken with LEICA SP5 confocal microscope.

### 6.8.2 Immunostaining on paraffin sections

Whole xenografts or primary tumors were cut in halves and fixed in formalin o/n. The fixative was removed and tissue was washed for 2 h in PBS before paraffin embedding. Immunostainings were performed on 4 µm tissue sections according to standard procedures. Briefly, for immunohistochemistry (IHC) antigen retrieval was carried out with boiling Tris-EDTA buffer for 20 min, then samples were blocked with Peroxidase-Blocking Solution (Dako: S202386) for 10 min at room temperature and incubated with primary antibody.

For the lineage tracing quantifications, primary antibody against TdTomato (rabbit, Rockland, 600-401-379) was used at 1:200 (overnight incubation). Bridging was performed with a Goat anti-Rabbit antibody 1:200 (Jackson Immunologic, 111-005-003) for 1 h at room temperature. Secondary antibody Anti-goat-HRP (Immunoresearch, DPVG55HRP) 1:200 was added for 30 min at room temperature. Slides were developed using DAB (Dako, K346811) and counterstained with hematoxylin prior to mounting with DPX (Panreac, 255254.1608). Washes were performed in between steps with wash buffer (Dako, K800721). Serial sections were used for hematoxylin-eosin staining according to standard procedures.

For EMP1-TOM detection via immunohistochemistry, a similar procedure but using anti-TdTOMATO (Sicgen Ab8181-200) without bridge was performed (1:100 for PDO samples, 1:200 for MTO samples). Secondary antibodies and developing reagents were the same.

For immunofluorescence the applied protocol was the same as for IHC with the fol-

lowing primary antibodies: TdTomato or RFP (rabbit, Rockland, 600-401-379 or goat, Sicgen Ab8181-200) 1:200, KRT20 (mouse, Dako, M7018), KI67 (mouse, Novocastra, 6004992), LAMC2 (rabbit, Sigma, HPA024638), L1CAM (rabbit, Sigma, HPA005830) MUC2 (mouse, AbCam, Ab118964) and secondary antibodies: donkey anti-goat conjugated to Alexa-488 (Life Technologies A21447) and donkey anti-mouse conjugated to Alexa-568 (Life Technologies A10037). Bridging for rabbit-TdTomato antibody was performed with Goat anti-Rabbit antibody 1:200 (Jackson Immunologic, 111-005-003) for 1 h at room temperature. DAPI was added at 1:2500 after secondary antibody incubation and slides were mounted with Fluorescent mounting media (Dako 53023).

### 6.8.3 Quantification of lineage tracing clone number and area

All images for lineage tracing quantification were acquired using a NanoZoomer-2.0 HT C9600series scanner (Hamamatsu, Photonics, France) with the 20X objective and coupled to a mercury lamp unit L11600-05 and using NDP.scan2.5 software U10074-03 (Hamamatsu, Photonics, France). All brightfield images were visualized with a gamma correction set at 1.8 with the NDP.view 2U123888-01 software (Hamamatsu). The total epithelial area was estimated from hematoxylin-eosin stained sections of the xenografts by manually training a machine learning pixel classifier (Trainable Weka Segmentation, Fiji ImageJ distribution (Witten et al., 1999)). Local colorimetric, textural and structural features were included in the feature set used by the classifier. To reduce the level of details and speed up the operation the images were downscaled by a factor 8 prior to classification.

TdTomato IHC scans were downsampled by a factor 4 and analyzed by manually training a machine learning pixel classifier (Ilastik software (Sommer et al., 2011)). This time, only intensity based local features were included and the clones were identified as TdTomato+ 2D connected particles in the resulting classification masks.

Clones at 4 days were counted manually rather than with Ilastik as the relatively weak levels of staining imposed severe limitations to image analysis. We also noticed that secondary tumors generated after passaging using a trocher displayed non-labeled cells intermingled within clones. This issue complicated clones identification as they often appeared fragmented, which biased their consecutive count and area estimation. To limit this problem we applied a post-processing step aiming at clustering neighbor 2D connected particles in the classification masks. It was implemented as a custom ImageJ macro (ADM, IRB Barcelona) leveraging morphological closing by a disk of user defined size (set to 8 pixels for all conditions). In order to estimate levels of background noise in area quantification only clones from day 4 were used. The logarithms of the areas were clustered using the Mclust function from the "mclust" package (Fraley et al., 2012) in the R statistical environment (R Development., 2008), finding 2 clusters through the Bayesian Information Criterion. The cluster with lower mean was considered to be background. Clones with area lower than the 25 percentile of the remaining cells were filtered out from the dataset. Since most clones from day 4 consist of a single cell, the median of these areas was used as cell size for each experiment. Clone sizes were defined as consisting of 1 cell, 2, 3 or 4, 5 to 10, 11 to 100 and more than 100 cells. The results of the analysis are summarized in Table 6.3. KI67+ on LGR5-EGFP tumors was performed on a total number of: n=2749 Lgr5+

Time (days)	Number of tumors	Number of clones assessed for KRT20+	Number of clones assessed for MUC2+
14	3	872	387
28	8	372	611
56	8	69	130

**Table 6.3: Quantification of TOM+KRT20+ or MUC2+ cells in LGR5 lineage tracing.** Number of tumors and Tdtomato+ clones assessed by their KRT20+ / MUC2+ status in each timepoint.

cells, 1798 Lgr5- cells.

Means and confidence intervals for the fraction of cells containing the differentiation markers were computed with a generalized linear model with mixed effects and binomial family. The technical variables xenograft and clone were taken as random effects and the function glmer from the R package "lme4" (Bates et al, 2015) was used to fit the model, compute p-values and estimate fixed effects. The confint function was used to compute confidence intervals for these estimates.

#### 6.8.4 Quantification of LGR5-EGFP+ cells within the xenografts

EGFP immunofluorescence intensity was quantified on Nanozoomer scanned slides using the Histogram function of the Fiji software (21). Background noise was set using the mean of several hand-picked background areas within each quantified xeno. Total quantified numbers were: PDO7#1: 4 xenos, PDO7#1 EGFP+ TIC derived: 4 xenos, PDO7#1 EGFP- TIC derived: 3 xenos.

#### 6.8.5 Visualization of 1 month grown LGR5-derived clones

##### *Serial sectioning and 3D reconstruction*

26 paraffin embedded serial sections (3  $\mu\text{m}$  thick) of 1 month induced PDO7#1 LGR5-CreERT2 subcutaneous xenograft were stained for Tdtomato as described above and scanned using NanoZoomer 2.0HT (Hamamatsu) digital scanner at 20X magnification. Regions of interest were cropped at 5x magnification and transformed to .tiff images with the NDPtoolkit software (Hamamatsu). Therefore images were aligned using the Register Virtual Stack Slices plugging of Fiji software (Schindelin et al., 2012). Then hyperstack images were cropped, converted to 8-bit grayscale and inverted. A manually trained machine learning pixel classifier (Ilastik software, (Sommer et al., 2011) was used to create a 3D binary mask and the clones were identified as 3D connected particles. This mask was post processed to discard spurious particles by considering that a valid clone must span at least 2 consecutive sections and a minimum volume of 400  $\mu\text{m}^3$ . The filtered binary mask was overlaid over a volume rendering of the original stack to visualize the clones in context (Imaris 3D software).

*Thick sections and SPIM microscopy*

A PDO7 LGR5-Cre#1 derived xenograft 1 month induced was fixed in PFA 4% for two hours and embedded in 4% low melting point agarose after PBS washes. Embedded tissue was cut in 200  $\mu$ m sections with a vibratom. Sections were then incubated for 5 days with antibody against TdTOMATO 1:200 (rabbit, Rockland, 600-401-379) and 5 more days with a Goat anti-Rabbit bridge 1:200 (Jackson Immunologic, 111-005-003). Secondary antibody D $\alpha$ G 647 1:200 (Life Technologies, A21447) was finally incubated for 5 days and DAPI 1:500 was added for 5h. 24h PBS washes were performed in between steps. After staining, samples were cleared for 2 days in RapiClear solution (RC147001) and imaged using an Elyra PS1-LSM880 SPIM microscope. Movies were created using IMARIS software v9.

Part VII  
BIBLIOGRAPHY



## BIBLIOGRAPHY

- Aceto, N., Bardia, A., Miyamoto, D. T., Donaldson, M. C., Wittner, B. S., Spencer, J. A., ... Maheswaran, S. (2014). Circulating Tumor Cell Clusters are Oligoclonal Precursors of Breast Cancer Metastasis. *Cell*, 158(5), 1110–1122.
- Acharyya, S., Oskarsson, T., Vanharanta, S., Malladi, S., Kim, J., Morris, P. G., ... Massagué, J. (2012). A CXCL1 paracrine network links cancer chemoresistance and metastasis. *Cell*, 150(1), 165–178.
- Adachi, Y., Yamamoto, H., Itoh, F., Hinoda, Y., Okada, Y., Imai, K. (1999). Contribution of matrilysin (MMP-7) to the metastatic pathway of human colorectal cancers. *Gut*, 45(2), 252–258.
- Ahmat Amin, M., Shimizu, A., Zankov, D., Sato, A., Kurita, S., Ito, M., ... Ogita, H. (2018). Epithelial membrane protein 1 promotes tumor metastasis by enhancing cell migration via copine-III and Rac1. *Oncogene*, 37, 5416–5434.
- Aiello, N. M., Maddipati, R., Norgard, R. J., Balli, D., Li, J., Yuan, S., ... Stanger, B. Z. (2018). EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Developmental Cell*, 45(6), 681–695.
- Al-Hajj, M., Wicha, M., Benito-Hernandez, A., Morrison, S., Clarke, M. (2003). Prospective identification of tumorigenic breast cancer cells. *PNAS*, 100(7), 3983–3988.
- Aoki, K., Tamai, Y., Horiike, S., Oshima, M., Taketo, M. M. (2003). Colonic polyposis caused by mTOR-mediated chromosomal instability in *Apc+/ $\Delta$ 716 Cdx2 +/-* compound mutant mice. *Nature Genetics*, 35(4), 323–330.
- Ariès, I. M., Jerchel, I. S., van den Dungen, R. E. S. R., van den Berk, L. C. J., Boer, J. M., Horstmann, M. A., ... den Boer, M. L. (2014). EMP1, a novel poor prognostic factor in pediatric leukemia regulates prednisolone resistance, cell proliferation, migration and adhesion. *Leukemia*, 28(9), 1828–1837.
- Baker, A.-M., Graham, T. A., Elia, G., Wright, N. A., Rodriguez-Justo, M. (2015). Characterization of LGR5 stem cells in colorectal adenomas and carcinomas. *Scientific Reports*, 5, 8654.
- Baker, A.-M., Huang, W., Wang, X.-M. M., Jansen, M., Ma, X.-J., Kim, J., ... Graham, T. A. (2017). Robust RNA-based in situ mutation detection delineates colorectal cancer subclonal evolution. *Nature Communications*, 8, 1998.
- Baldus, S. E., Schaefer, K.-L., Engers, R., Hartleb, D., Stoecklein, N. H., Gabbert, H. E. (2010). Prevalence and Heterogeneity of KRAS, BRAF, and PIK3CA Mutations in Primary Colorectal Adenocarcinomas and Their Corresponding Metastases. *Clinical Cancer Research*, 16(3), 790–799.
- Barker, N., Ridgway, R. A., van Es, J. H., van de Wetering, M., Begthel, H., van den Born, M., ... Clevers, H. (2009). Crypt stem cells as the cells-of-origin of intestinal

cancer. *Nature*, 457(7229), 608–611.

Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., ... Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature*, 449(7165), 1003–1007.

Barriga, F. M., Montagni, E., Mana, M., Mendez-Lago, M., Hernando-Momblona, X., Sevillano, M., ... Batlle, E. (2017). *Mex3a* Marks a Slowly Dividing Subpopulation of *Lgr5*+ Intestinal Stem Cells. *Cell Stem Cell*, 20, 1–16.

Basak, O., van de Born, M., Korving, J., Beumer, J., van der Elst, S., van Es, J. H., Clevers, H. (2014). Mapping early fate determination in *Lgr5*+ crypt stem cells using a novel *Ki67*-RFP allele. *The EMBO Journal*, 33(18), 2057–2068.

Batlle, E., Bacani, J., Begthel, H., Jonkeer, S., Gregorieff, A., van de Born, M., ... Clevers, H. (2005). EphB receptor activity suppresses colorectal cancer progression. *Nature*, 435(7045), 1126–1130.

Batlle, E., Clevers, H. (2017). Cancer stem cells revisited. *Nature Medicine*, 23(10), 1124–1134.

Batlle, E., Henderson, J. T., Begthel, H., van den Born, M. M. W., Sancho, E., Huls, G., ... Clevers, H. (2002). Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell*, 111(2), 251–63.

Benveniste, P., Frelin, C., Janmohamed, S., Barbara, M., Herrington, R., Hyam, D., Iscove, N. N. (2010). Intermediate-Term Hematopoietic Stem Cells with Extended but Time-Limited Reconstitution Potential. *Cell Stem Cell*, 6(1), 48–58.

Bjerknes, M., Cheng, H. (1981). The stem-cell zone of the small intestinal epithelium. II. Evidence from paneth cells in the newborn mouse. *American Journal of Anatomy*, 160(1), 65-75.

Bjerknes, M., Cheng, H. (1981). The stem-cell zone of the small intestinal epithelium. III. Evidence from columnar, enteroendocrine, and mucous cells in the adult mouse. *American Journal of Anatomy*, 160(1), 77–91.

Bjerknes, M., Cheng, H. (1981). The stem-cell zone of the small intestinal epithelium. V. Evidence for controls over orientation of boundaries between the stem-cell zone, proliferative zone, and the maturation zone. *American Journal of Anatomy*, 160(1), 105-112.

Bjerknes, M., Cheng, H. (1981). Methods for the Isolation of the Intact Epithelium From the Mouse Intestine. *The Anatomical Record*, 199, 656-674.

Blanpain, C., Fuchs, E. (2009). Epidermal homeostasis: a balancing act of stem cells in the skin. *Nature Reviews Molecular Cell Biology*, 10(3), 207–217.

Blanpain, C., Fuchs, E. (2014). Stem cell plasticity. Plasticity of epithelial stem cells in tissue regeneration. *Science*, 344(6189), 1242281.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., ... Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959), 1509–1512.

- Bolstad, B. M. (2004). Low level Analysis of High density Oligonucleotide Array Data: Background, Normalization and Summarization. University of California, Berkeley.
- Bonnet, D., Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine*, 3(7), 730–737.
- Borst, P. (2012). Cancer drug pan-resistance: pumps, cancer stem cells, quiescence, epithelial to mesenchymal transition, blocked cell death pathways, persists or what? *Open Biology*, 2(5), 120066.
- Brabletz, T., Jung, A., Hermann, K., Günther, K., Hohenberger, W., Kirchner, T. (1998). Nuclear overexpression of the oncoprotein beta-catenin in colorectal cancer is localized predominantly at the invasion front. *Pathology, Research and Practice*, 194(10), 701–704.
- Brabletz, T., Jung, A., Reu, S., Porzner, M., Hlubek, F., Kunz-Schughart, L. A., ... Weinberg, R. A. (2001). Variable-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *PNAS*, 98(18), 10356–10361.
- Brabletz, T., Jung, A., Spaderna, S., Hlubek, F., Kirchner, T. (2005). Migrating cancer stem cells - an integrated concept of malignant tumour progression. *Nature Reviews Cancer*, 5(9), 744–749.
- Brannon, A. R., Vakiani, E., Sylvester, B. E., Scott, S. N., McDermott, G., Shah, R. H., ... Berger, M. F. (2014). Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biology*, 15, 454.
- Braster, R., Bögels, M., Beelen, R. H. J., Van Egmond, M. (2017). The delicate balance of macrophages in colorectal cancer; their role in tumour development and therapeutic potential. *Immunobiology*, 222, 21–30.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CANCER J CLIN* (Vol. 68). American Cancer Society.
- Breault, D. T., Min, I. M., Carlone, D. L., Farilla, L. G., Ambruzs, D. M., Henderson, D. E., ... Hole, N. (2008). Generation of mTert-GFP mice as a model to identify and study tissue progenitor cells. *PNAS*, 105(30), 10420–10425.
- Bruns, C. J., Harbison, M. T., Kuniyasu, H., Eue, I., Fidler, I. J. (1999). In vivo selection and characterization of metastatic variants from human pancreatic adenocarcinoma by using orthotopic implantation in nude mice. *Neoplasia*, 1(1), 50–62.
- Buczacki, S. J. a, Popova, S., Biggs, E., Koukorava, C., Buzzelli, J., Vermeulen, L., ... Winton, D. J. (2018). Itraconazole targets cell cycle heterogeneity in colorectal cancer. *The Journal of Experimental Medicine*, 215(7), 1891–1912.
- Buczacki, S. J. a, Zecchini, H. I., Nicholson, A. M., Russell, R., Vermeulen, L., Kemp, R., Winton, D. J. (2013). Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature*, 495(7439), 65–69.

- Bufu, T., Di, X., Yilin, Z., Gege, L., Xi, C., Ling, W. (2018). Celastrol inhibits colorectal cancer cell proliferation and migration through suppression of MMP3 and MMP7 by the PI3K/AKT signaling pathway. *Anti-Cancer Drugs*, 29(6), 530–538.
- Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., ... Rodewald, H.-R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, 518(7540), 542–546.
- Cadigan, K. M. (2008). Wnt-beta-catenin signaling. *Current Biology*, 18(20), 943–947.
- Calon, A., Espinet, E., Palomo-Ponce, S., Tauriello, D. V. F., Iglesias, M., Céspedes, M. V., ... Batlle, E. (2012). Dependency of colorectal cancer on a TGF- $\beta$ -driven program in stromal cells for metastasis initiation. *Cancer Cell*, 22(5), 571–84.
- Calon, A., Lonardo, E., Berenguer-llergo, A., Espinet, E., Hernando-momblona, X., Iglesias, M., ... Batlle, E. (2015). Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nature Genetics*, 47(4), 320–329.
- Carey, S. P., Martin, K. E., Reinhart-King, C. A. (2017). Three-dimensional collagen matrix induces a mechanosensitive invasive epithelial phenotype. *Scientific Reports*, 7, 42088.
- Celià-Terrassa, T., Meca-Cortés, O., Mateo, F., Martínez de Paz, A., Rubio, N., Arnal-Estapé, A., ... Thomson, T. M. (2012). Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *The Journal of Clinical Investigation*, 122(5), 1849–68.
- Cernat, L., Blaj, C., Jackstadt, R., Brandl, L., Engel, J., Hermeking, H., ... Horst, D. (2014). Colorectal Cancers Mimic Structural Organization of Normal Colonic Crypts. *PLoS ONE*, 9(8), e104284.
- Céspedes, M. V., Espina, C., García-Cabezas, M. A., Trias, M., Boluda, A., Gómez del Pulgar, M. T., ... Mangués, R. (2007). Orthotopic microinjection of human colon cancer cells in nude mice induces tumor foci in all clinically relevant metastatic sites. *The American Journal of Pathology*, 170(3), 1077–1085.
- Céspedes, M. V., Unzueta, U., Aviñó, A., Gallardo, A., Álamo, P., Sala, R., ... Mangués, R. (2018). Selective depletion of metastatic stem cells as therapy for human colorectal cancer. *EMBO Molecular Medicine*, 10, e8772.
- Chan, A. T., Ogino, S., Fuchs, C. S. (2009). Aspirin Use and Survival After Diagnosis of Colorectal Cancer. *JAMA*, 302(6), 649–659.
- Chan, C. W. M., Wong, N. A., Liu, Y., Bicknell, D., Turley, H., Hollins, L., ... Bodmer, W. F. (2009). Gastrointestinal differentiation marker Cytokeratin 20 is regulated by homeobox gene CDX1. *PNAS*, 106(6), 1936–1941.
- Chao, C., Widen, S. G., Wood, T. G., Zatarain, J. R., Johnson, P., Gajjar, A., ... Hellmich, M. R. (2017). Patient-derived Xenografts from Colorectal Carcinoma: A Temporal and Hierarchical Study of Murine Stromal Cell Replacement. *Anticancer Research*, 37(7), 3405–3412.
- Cheng, H., Leblond, C. P. (1974). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine I. Columnar cell. *American Journal*

of Anatomy, 141(4), 461–479.

Cheshier, S. H., Morrison, S. J., Liao, X., Weissman, I. L. (1999). In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *PNAS*, 96(6), 3120–5.

Cheung, K. J., Padmanaban, V., Silvestri, V., Schipper, K., Cohen, J. D., Fairchild, A. N., ... Contributed, J. S. B. (2016). Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *PNAS*, 113(7), E854–863.

Choi, B.-J., Park, S.-A., Lee, S.-Y., Cha, Y. N., Surh, Y.-J. (2017). Hypoxia induces epithelial-mesenchymal transition in colorectal cancer cells through ubiquitin-specific protease 47-mediated stabilization of Snail: A potential role of Sox9. *Scientific Reports*, 7(1), 1–15.

Christiansen, J., Rajasekaran, A. (2006). Reassessing Epithelial to Mesenchymal Transition as a Prerequisite for Carcinoma Invasion and Metastasis. *Cancer Research*, 66(17), 8319–8326.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92.

Clackson, T., Yang, W., Rozamus, L. W., Hatada, M., Amara, J. F., Rollins, C. T., ... Holt, D. A. (1998). Redesigning an FKBP-ligand interface to generate chemical dimers with novel specificity. *PNAS*, 95(18), 10437–10442.

Clevers, H. (2013). The intestinal crypt, a prototype stem cell compartment. *Cell*, 154(2), 274–84.

Clevers, H. (2015). What is an adult stem cell? *Stem Cells*, 350(6266), 1319–1320.

Clevers, H. (2016). Modeling Development and Disease with Organoids. *Cell*, 165(7), 1586–1597.

Clevers, H., Batlle, E. (2013). SnapShot: The Intestinal Crypt. *Cell*, 152, 1198.

Conboy, I. M., Rando, T. A. (2002). The regulation of Notch signaling controls satellite cell activation and cell fate determination in postnatal myogenesis. *Developmental Cell*, 3(3), 397–409.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), 819–823.

Conklin, M. W., Eickhoff, J. C., Riching, K. M., Pehlke, C. A., Eliceiri, K. W., Provenzano, P. P., ... Keely, P. J. (2011). Aligned Collagen Is a Prognostic Signature for Survival in Human Breast Carcinoma. *The American Journal of Pathology*, 178(3), 1221–1232.

Cortina, C., Palomo-Ponce, S., Iglesias, M., Fernández-Masip, J. L., Vivancos, A., Whissell, G., ... Batlle, E. (2007). EphB–ephrin-B interactions suppress colorectal cancer progression by compartmentalizing tumor cells. *Nature Genetics*, 39(11), 1376–1383.

- Cortina, C., Turon, G., Stork, D., Hernando-Momblona, X., Sevillano, M., Aguilera, M., ... Batlle, E. (2017). A genome editing approach to study cancer stem cells in human tumors. *EMBO Molecular Medicine*, 9(7), 869–879.
- Cotsarelis, G., Sun, T. T., Lavker, R. M. (1990). Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell*, 61(7), 1329–37.
- Crosnier, C., Vargesson, N., Gschmeissner, S., Ariza-McNaughton, L., Morrison, A., Lewis, J., ... Nüsslein-Volhard, C. (2005). Delta-Notch signalling controls commitment to a secretory fate in the zebrafish intestine. *Development*, 132(5), 1093–1104.
- Dalerba, P., Dylla, S. J., Park, I.-K., Liu, R., Wang, X., Cho, R. W., ... Clarke, M. F. (2007). Phenotypic characterization of human colorectal cancer stem cells. *PNAS*, 104(24), 10158–63.
- Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P. S., Rothenberg, M. E., Leyrat, A. A., ... Quake, S. R. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nature Biotechnology*, 29(12), 1120–1127.
- Dame, M. K., Attili, D., McClintock, S. D., Dedhia, P. H., Ouillet, P., Hardt, O., ... Colacino, J. A. (2018). Identification, isolation and characterization of human LGR5-positive colon adenoma cells. *Development*, 145, dev153049.
- de Lau, W., Kujala, P., Schneeberger, K., Middendorp, S., Li, V. S. W., Barker, N., ... Clevers, H. (2012). Peyer's patch M cells derived from Lgr5(+) stem cells require SpiB and are induced by RankL in cultured "miniguts." *Molecular and Cellular Biology*, 32(18), 3639–3647.
- De Robertis, M., Massi, E., Poeta, M. L., Carotti, S., Morini, S., Cecchetelli, L., ... Fazio, V. M. (2011). The AOM/DSS murine model for the study of colon carcinogenesis: From pathways to diagnosis and therapy studies. *Journal of Carcinogenesis*, 10, 9.
- De Smedt, L., Palmans, S., Andel, D., Govaere, O., Boeckx, B., Smeets, D., ... Sagaert, X. (2017). Expression profiling of budding cells in colorectal cancer reveals an EMT-like phenotype and molecular subtype switching. *British Journal of Cancer*, 116(1), 58–65.
- Dieter, S. M., Ball, C. R., Hoffmann, C. M., Nowrouzi, A., Herbst, F., Zavidij, O., ... Glimm, H. (2011). Distinct types of tumor-initiating cells form human colon cancer tumors and metastases. *Cell Stem Cell*, 9(4), 357–365.
- Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., ... Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology*, 32(12), 1262–1267.
- Driessens, G., Beck, B., Caauwe, A., Simons, B. D., Blanpain, C. (2012). Defining the mode of tumour growth by clonal analysis. *Nature*, 488, 527–530.
- Drost, J., van Boxtel, R., Blokzijl, F., Mizutani, T., Sasaki, N., Sasselli, V., ... Clevers, H. (2017). Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science*, 358(6360), 234–238.

- Drost, J., van Jaarsveld, R. H., Ponsioen, B., Zimmerlin, C., van Boxtel, R., Buijs, A., ... Clevers, H. (2015). Sequential cancer mutations in cultured human intestinal stem cells. *Nature*, 521(7550), 43–47.
- Duval, A., Hamelin, R. (2002). Mutations at Coding Repeat Sequences in Mismatch Repair-deficient Human Cancers: Toward a New Concept of Target Genes for Instability.
- Eklund, A. C., Szallasi, Z. (2008). Correction of technical bias in clinical microarray data improves concordance with known biological information. *Genome Biology*, 9(2), R26.
- Er, E. E., Valiente, M., Ganesh, K., Zou, Y., Agrawal, S., Hu, J., ... Massagué, J. (2018). Pericyte-like spreading by disseminated cancer cells activates YAP and MRTF for metastatic colonization. *Nature Cell Biology*, 20(8), 966–978.
- Farin, H. F., Van Es, J. H., Clevers, H. (2012). Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. *Gastroenterology*, 143(6), 1518–1529.
- Fearon, E., (2011). Molecular Genetics of Colorectal Cancer. *Annu. Rev. Pathol. Mech. Dis*, 6, 479–507.
- Fearon, E., Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, 61(5), 759–767.
- Feil, R., Rgen Wagner, J., Metzger, D., Chambon, P. (1997). Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains. *Biochemical and Biophysical Research Communications*, 237, 752–757.
- Ferraro, F., Celso, C. Lo, Scadden, D. (2010). Adult stem cells and their niches. *Advances in Experimental Medicine and Biology*, 695, 155–68.
- Fischer, K. R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S. T. C., ... Gao, D. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature*, 527(7579), 472–476.
- Flavahan, W. A., Wu, Q., Hitomi, M., Rahim, N., Kim, Y., Sloan, A. E., ... Hjelmeland, A. B. (2013). Brain tumor initiating cells adapt to restricted nutrition through preferential glucose uptake. *Nature Neuroscience*, 16(10), 1373–82.
- Flentjar, N., Chu, P., Ng, A. Y. -N., Johnstone, C. N., Heath, J. K., Ernst, M., ... Pritchard, M. A. (2007). TGF- $\beta$ RII Rescues Development of Small Intestinal Epithelial Cells in Elf3-Deficient Mice. *Gastroenterology*, 132(4), 1410–1419.
- Fodde, R., Eedermann, W., Yang, K., Van Leeuwen, C., Carlson, C., Renault, B., ... Kucherlapati, R. (1994). A targeted chain-termination mutation in the mouse Apc gene results in multiple intestinal tumors. *PNAS*, 91, 8969–8973.
- Fraley, C., Raftery, A. E., Murphy, T. B. (2012). mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Retrieved from <http://cran.r-project.org/web/packages/mclust/index.html>.

- Franken, N. A. P., Rodermond, H. M., Stap, J., Haveman, J., van Bree, C. (2006). Clonogenic assay of cells in vitro. *Nature Protocols*, 1(5), 2315–2319.
- Fre, S., Huyghe, M., Mourikis, P., Robine, S., Louvard, D., Artavanis-Tsakonas, S. (2005). Notch signals control the fate of immature progenitor cells in the intestine. *Nature*, 435(7044), 964–968.
- Fridlender, Z. G., Sun, J., Kim, S., Kapoor, V., Cheng, G., Ling, L., ... Albelda, S. M. (2009). Polarization of Tumor-Associated Neutrophil (TAN) Phenotype by TGF- $\beta$ : N1 versus N2 TAN. *Cancer Cell*, 16(3), 183–194.
- Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., Sander, J. D. (2013). High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, 31(9), 822–826.
- Fuchs, E., Chen, T. (2013). A matter of life and death: self-renewal in stem cells. *EMBO Reports*, 14(1), 39–48.
- Fujii, M., Shimokawa, M., Date, S., Takano, A., Matano, M., Nanki, K., ... Sato, T. (2016). A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. *Cell Stem Cell*, 18(6), 827–838.
- Fumagalli, A., Drost, J., Suijkerbuijk, S. J. E., van Boxtel, R., de Ligt, J., Offerhaus, G. J., ... van Rheeën, J. (2017). Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proceedings of the National Academy of Sciences of the United States of America*, 114(12), E2357–E2364.
- Fumagalli, A., Suijkerbuijk, S. J. E., Begthel, H., Beerling, E., Oost, K. C., Snippert, H. J., ... Drost, J. (2018). A surgical orthotopic organoid transplantation approach in mice to visualize and study colorectal cancer progression. *Nature Protocols*, 13(2), 235–247.
- Galatenko, V. V., Maltseva, D. V., Galatenko, A. V., Rodin, S., Tonevitsky, A. G. (2018). Cumulative prognostic power of laminin genes in colorectal cancer. *BMC Medical Genomics*, 11(Suppl 1), 9.
- Galiatsatos, P., Foulkes, W. D. (2006). Familial Adenomatous Polyposis. *The American Journal of Gastroenterology*, 101(2), 385–398.
- Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagè, C., ... Pagè, F. (2006). Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science*, 313, 1960–1964.
- Gavert, N., Conacci-Sorrell, M., Gast, D., Schneider, A., Altevogt, P., Brabletz, T., Ben-Ze'ev, A. (2005). L1, a novel target of beta-catenin signaling, transforms cells and is expressed at the invasive front of colon cancers. *The Journal of Cell Biology*, 168(4), 633–642.
- Gerbe, F., Sidot, E., Smyth, D. J., Ohmoto, M., Matsumoto, I., Dardalhon, V., ... Jay, P. (2016). Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. *Nature*, 529(7585), 226–230.
- Gerbe, F., Van Es, J. H., Makrini, L., Brulin, B., Mellitzer, G., Robine, S., ... Jay, P. (2011). Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secre-

tory cell type in the intestinal epithelium. *Journal of Cell Biology*, 192(5), 767–780.

Giessler, K. M., Kleinheinz, K., Huebschmann, D., Balasubramanian, G. P., Dubash, T. D., Dieter, S. M., ... Glimm, H. (2017). Genetic subclone architecture of tumor clone-initiating cells in colorectal cancer. *The Journal of Experimental Medicine*, 214(7), 2073–2088.

Gkountela, S., Castro-Giner, F., Szczerba, B. M., Vetter, M., Landin, J., Scherrer, R., ... Aceto, N. (2019). Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding. *Cell*, 176, 98–112.

Goel, H. L., Mercurio, A. M. (2013). VEGF targets the tumour cell. *Nature Reviews Cancer*, 13(12), 871–882.

Gomez, C., Chua, W., Miremedi, A., Quist, S., Headon, D. J., Watt, F. M. (2013). The interfollicular epidermis of adult mouse tail comprises two distinct cell lineages that are differentially regulated by Wnt, Edaradd, and Lrig1. *Stem Cell Reports*, 1(1), 19–27.

Gregorieff, A., Stange, D. E., Kujala, P., Begthel, H., Van Den Born, M., Korving, J., ... Clevers, H. (2009). The Ets-Domain Transcription Factor Spdef Promotes Maturation of Goblet and Paneth Cells in the Intestinal Epithelium. *Gastroen*, 137(4), 1333–1345.

Grillet, F., Bayet, E., Villeronce, O., Zappia, L., Lagerqvist, E. L., Lunke, S., ... Pannequin, J. (2017). Circulating tumour cells from patients with colorectal cancer have cancer stem cell hallmarks in ex vivo culture. *Gut*, 66(10), 1802–1810.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., ... van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*, 525(7568), 251–255.

Guinney, J., Dienstmann, R., Wang, X., De Reyniès, A., Schlicker, A., Soneson, C., ... Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, 21(11), 1350–1356.

Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuperwasser, C., Lander, E. S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146(4), 633–644.

Hanahan, D., Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, 100(1), 57–70.

Hanahan, D., Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144, 646–674.

Harada, N., Tamai, Y., Ishikawa, T.-O., Sauer, B., Takaku, K., Oshima, M., Taketo, M. M. (1999). Intestinal polyposis in mice with a dominant stable mutation of the  $\beta$ -catenin gene. *The EMBO Journal* (Vol. 18).

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

Hata, A. N., Niederst, M. J., Archibald, H. L., Gomez-Caraballo, M., Siddiqui, F. M., Mulvey, H. E., ... Engelman, J. A. (2016). Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nature*

Medicine, 22(3), 262–269.

He, L., Li, Y., Li, Y., Pu, W., Huang, X., Tian, X., . . . Zhou, B. (2017). Enhancing the precision of genetic lineage tracing using dual recombinases. *Nature Medicine*, 23(12), 1488–1498.

Hiratsuka, S., Watanabe, A., Aburatani, H., Maru, Y. (2006). Tumour-mediated up-regulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. *Nature Cell Biology*, 8(12), 1369–1375.

Hoeck, J. D., Biehs, B., Kurtova, A. V., Kljavin, N. M., de Sousa e Melo, F., Alicke, B., . . . de Sauvage, F. J. (2017). Stem cell plasticity enables hair regeneration following Lgr5+ cell loss. *Nature Cell Biology*, 19(6), 666–676.

Hoey, T., Yen, W.-C., Axelrod, F., Basi, J., Donigian, L., Dylla, S., . . . Gurney, A. (2009). DLL4 Blockade Inhibits Tumor Growth and Reduces Tumor-Initiating Cell Frequency. *Cell Stem Cell*, 5(2), 168–177.

Holohan, C., Van Schaeybroeck, S., Longley, D. B., Johnston, P. G. (2013). Cancer drug resistance: an evolving paradigm. *Nature Reviews Cancer*, 13(10), 714–726.

Hostettler, Isabel, Zlobec, I., Terracciano, L., Lugli Isabel Hostettler, A., Lugli, A. (2010). ABCG5-positivity in tumor buds is an indicator of poor prognosis in node-negative colorectal cancer patients. *World Journal of Gastroenterol*, 16(6), 732–739.

Huang, D., Du, C., Ji, D., Xi, J., Gu, J. (2017). Overexpression of LAMC2 predicts poor prognosis in colorectal cancer patients and promotes cancer cell proliferation, migration, and invasion. *Tumor Biology*, 39(6), 1–7.

Huels, D. J., Sansom, O. J. (2015). Stem vs non-stem cell origin of colorectal cancer. *British Journal of Cancer*, 113(1), 1–5.

Humphries, A., Cereser, B., Gay, L. J., Miller, D. S. J., Das, B., Gutteridge, A., . . . Graham, T. A. (2013). Lineage tracing reveals multipotent stem cells maintain human adenomas and the pattern of clonal expansion in tumor evolution. *PNAS*, 110(27), E2490–E2499.

Hung, K. E., Maricevich, M. A., Richard, L. G., Chen, W. Y., Richardson, M. P., Kunin, A., . . . Kucherlapati, R. (2010). Development of a mouse model for sporadic and metastatic colon tumors and its use in assessing drug treatment. *PNAS*, 107(4), 1565–70.

Illa-Bochaca, I., Fernandez-Gonzalez, R., Shelton, D. N., Welm, B. E., Ortiz-de-Solorzano, C., Barcellos-Hoff, M. H. (2010). Limiting-Dilution Transplantation Assays in Mammary Stem Cell Studies. In *Protocols for Adult Stem Cells* (pp. 29–47).

Inamoto, S., Itatani, Y., Yamamoto, T., Minamiguchi, S., Hirai, H., Iwamoto, M., . . . Kawada, K. (2016). Loss of SMAD4 Promotes Colorectal Cancer Progression by Accumulation of Myeloid-Derived Suppressor Cells through the CCL15-CCR1 Chemokine Axis. *Clinical Cancer Research*, 22(2), 492–501.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., Speed, T. P. (2003). Exploration, normalization, and summaries of high density

oligonucleotide array probe level data. *Biostatistics*, 4(2), 249–264.

Isella, C., Terrasi, A., Bellomo, S. E., Petti, C., Galatola, G., Muratore, A., . . . Medico, E. (2015). Stromal contribution to the colorectal cancer transcriptome. *Nature Genetics*, 47(4), 312–319. 4

Itatani, Y., Kawada, K., Fujishita, T., Kakizaki, F., Hirai, H., Matsumoto, T., . . . Taketo, M. M. (2013). Loss of SMAD4 from colorectal cancer cells promotes CCL15 expression to recruit CCR1+ myeloid cells and facilitate liver metastasis. *Gastroenterology*, 145(5), 1064–1075.e11.

Ito, M., Liu, Y., Yang, Z., Nguyen, J., Liang, F., Morris, R. J., Cotsarelis, G. (2005). Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nature Medicine*, 11(12), 1351–1354.

Itzkovitz, S., Lyubimova, A., Blat, I., Maynard, M., Van Es, J., Lees, J., . . . Van Oudenaarden, A. (2012). Single molecule transcript counting of stem cell markers in the mouse intestine. *Nature Cell Biology*, 14(1), 106–114.

Itzkowitz, S. H., Yio, X. (2004). Inflammation and Cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 287(1), G7–G17.

Jaenisch, R., Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell*, 132(4), 567–82.

Jain, A., Tindell, C. A., Laux, I., Hunter, J. B., Curran, J., Galkin, A., . . . Agus, D. B. (2005). Epithelial membrane protein-1 is a biomarker of gefitinib resistance. *PNAS*, 102(33), 11858–11863.

Jenny, M., Uhl, C., Roche, C., Duluc, I., Guillermin, V., Guillemot, F., . . . Gradwohl, G. (2002). Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. *The EMBO Journal*, 21(23), 6338–6347.

Jensen, J., Pedersen, E. E., Galante, P., Hald, J., Heller, R. S., Ishibashi, M., . . . Madsen, O. D. (2000). Control of endodermal endocrine development by Hes-1. *Nature Genetics*, 24(1), 36–44.

Jiao, Y.-F., Nakamura, S.-I., Sugai, T., Yamada, N., Habano, W. (2008). Serrated Adenoma of the Colorectum Undergoes a Proliferation versus Differentiation Process: New Conceptual Interpretation of Morphogenesis. *Clinical Translational Research Oncology*, 74, 127–134.

Jones, S., Chen, W.-D., Parmigiani, G., Diehl, F., Beerwinkler, N., Antal, T., . . . Markowitz, S. D. (2008). Comparative lesion sequencing provides insights into tumor evolution. *PNAS*, 105(11), 4283–4288.

Jung, P., Sato, T., Merlos-Suárez, A., Barriga, F. M., Iglesias, M., Rossell, D., . . . Batlle, E. (2011). Isolation and in vitro expansion of human colonic stem cells. *Nature Medicine*, 17(10), 1225–1227.

Junttila, M. R., Mao, W., Wang, X., Wang, B.-E., Pham, T., Flygare, J., . . . Polson, A. G. (2015). Targeting LGR5+ cells with an antibody-drug conjugate for the treatment of

- colon cancer. *Science Translational Medicine*, 7(314), 314ra186.
- Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nature Reviews Cancer*, 16(9), 582–598.
- Kayahara, T., Sawada, M., Takaishi, S., Fukui, H., Seno, H., Fukuzawa, H., ... Chiba, T. (2003). Candidate markers for stem and early progenitor cells, Musashi-1 and Hes1, are expressed in crypt base columnar cells of mouse small intestine. *FEBS Letters*, 535, 131–135.
- Kedrin, D., Gala, M. K. (2015). Genetics of the Serrated Pathway to Colorectal Cancer. *Clinical and Translational Gastroenterology*, 6(e84).
- Kemper, K., Grandela, C., Medema, J. P. (2010). Molecular identification and targeting of colorectal cancer stem cells. *Oncotarget*, 1(6), 387–395.
- Kinzler, K. W., Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell*, 87(2), 159–170.
- Kleist, B., Xu, L., Li, G., Kersten, C. (2011). Expression of the adult intestinal stem cell marker Lgr5 in the metastatic cascade of colorectal cancer. *International Journal of Clinical and Experimental Pathology*, 4(4), 327–335.
- Kobayashi, T., Masaki, T., Nozaki, E., Sugiyama, M., Nagashima, F., Furuse, J., ... Ohkura, Y. (2015). Microarray Analysis of Gene Expression at the Tumor Front of Colon Cancer. *Anticancer Research*, 35(12), 6577–81.
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., ... Wilson, R. K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, 22(3), 568–576.
- Koch, S. (2017). Extrinsic control of Wnt signaling in the intestine. *Differentiation*, 97, 1–8.
- Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P. J., Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genetics*, 19(4), 379–383.
- Kozar, S., Morrissey, E., Nicholson, A. M., van der Heijden, M., Zecchini, H. I., Kemp, R., ... Winton, D. J. (2013). Continuous Clonal Labeling Reveals Small Numbers of Functional Stem Cells in Intestinal Crypts and Adenomas. *Cell Stem Cell*, 13(5), 626–633.
- Kreso, A., O'Brien, C. a, van Galen, P., Gan, O. I., Notta, F., Brown, A. M. K., ... Dick, J. E. (2013). Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science*, 339(6119), 543–548.
- Kreso, A., van Galen, P., Pedley, N. M., Lima-Fernandes, E., Frelin, C., Davis, T., ... O'Brien, C. a. (2014). Self-renewal as a therapeutic target in human colorectal cancer. *Nature Medicine*, 20(1), 29–36.
- Krieger, T., Simons, B. D. (2015). Dynamic stem cell heterogeneity. *Development*, 142, 1396–1406.

- Krøigård, A. B., Thomassen, M., Laenkholm, A.-V., Kruse, T. A., Larsen, M. J. (2016). Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. *Plos One*, 11(3), e0151664.
- Kusaba, T., Lalli, M., Kramann, R., Kobayashi, A., Humphreys, B. D. (2014). Differentiated kidney epithelial cells repair injured proximal tubule. *PNAS*, 111(4), 1527–1532.
- Lamprecht, S., Schmidt, E. M., Blaj, C., Hermeking, H., Jung, A., Kirchner, T., Horst, D. (2017). Multicolor lineage tracing reveals clonal architecture and dynamics in colon cancer. *Nature Communications*, 8, 1406.
- Lancaster, M., Knoblich, J. (2014). Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science*, 345(6194), 283.
- Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., ... Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*, 367(6464), 645–648.
- Leach, J., Morton, J. P., Sansom, O. J. (2017). Neutrophils: Homing in on the myeloid mechanisms of metastasis. *Molecular Immunology*.
- Leblond, C. P., Messier, B. (1958). Renewal of chief cells and goblet cells in the small intestine as shown by radioautography after injection of thymidine-H3 into mice. *The Anatomical Record*, 132(3), 247–259.
- Leblond, C. P., Stevens, C. E. (1948). The constant renewal of the intestinal epithelium in the albino rat. *The Anatomical Record*, 100(3), 357–377.
- Lenander, C., Habermann, J. K., Öst, Å., Nilsson, B., Schimmelpenning, H., Tryggvason, K., Auer, G. (2001). Laminin-5  $\gamma$ 2 chain expression correlates with unfavorable prognosis in colon carcinomas. *Analytical Cellular Pathology*, 22, 201–209.
- Lenos, K. J., Miedema, D. M., Lodestijn, S. C., Nijman, L. E., van den Bosch, T., Romero Ros, X., ... Vermeulen, L. (2018). Stem cell functionality is microenvironmentally defined during tumour expansion and therapy response in colon cancer. *Nature Cell Biology*, 20, 1193–1202.
- Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760.
- Linnekamp, J. F., Hooff, S. R. van, Prasetyanti, P. R., Kandimalla, R., Buikhuizen, J. Y., Fessler, E., ... Medema, J. P. (2018). Consensus molecular subtypes of colorectal cancer are recapitulated in in vitro and in vivo models. *Cell Death and Differentiation*, 25(3), 616–633.
- Liu, Q., Segal, D. J., Ghiara, J. B., Barbas, C. F., III. (1997). Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *PNAS*, 94(11), 5525–5530.
- Lombardo, Y., Scopelliti, A., Cammareri, P., Todaro, M., Iovino, F., Ricci-Vitiani, L., ... Stassi, G. (2011). Bone Morphogenetic Protein 4 Induces Differentiation of Colorectal Cancer Stem Cells and Increases Their Response to Chemotherapy in Mice. *Gastroenterology*, 140(1), 297–309.

- Lopez-Garcia, C., Klein, A. M., Simons, B. D., Winton, D. J. (2010). Intestinal Stem Cell Replacement Follows a Pattern of Neutral Drift. *Science*, 330(6005), 822–825.
- Lorenz, E., Uphoff, D., Reid, T. R., Shelton, E. (1951). Modification of Irradiation Injury in Mice and Guinea Pigs by Bone Marrow Injections. *JNCI*, 12(1), 197–201.
- Lotti, F., Jarrar, A. M., Pai, R. K., Hitomi, M., Lathia, J., Mace, A., ... Rich, J. N. (2013). Chemotherapy activates cancer-associated fibroblasts to maintain colorectal cancer-initiating cells by IL-17A. *The Journal of Experimental Medicine*, 210(13), 2851–2872.
- Lou, Y., McDonald, P. C., Oloumi, A., Chia, S., Ostlund, C., Ahmadi, A., ... Dedhar, S. (2011). Targeting tumor hypoxia: suppression of breast tumor growth and metastasis by novel carbonic anhydrase IX inhibitors. *Cancer Research*, 71(9), 3364–3376.
- Lynch, H. T., de la Chapelle, A. (2003). Hereditary Colorectal Cancer. *New England Journal of Medicine*, 348(10), 919–932.
- Mani, S. A., Guo, W., Liao, M.-J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., ... Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, 133(4), 704–15.
- Martin, M. L., Zeng, Z., Adileh, M., Jacobo, A., Li, C., Vakiani, E., ... Paty, P. B. (2018). Logarithmic expansion of LGR5+ cells in human colorectal cancer. *Cellular Signalling*, 42, 97–105.
- Massagué, J., Obenauf, A. C. (2016). Metastatic colonization by circulating tumour cells. *Nature*, 529(7586), 298–306.
- Matano, M., Date, S., Shimokawa, M., Takano, A., Fujii, M., Ohta, Y., ... Sato, T. (2015). Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nature Medicine*, 21(3), 256–262.
- May, R., Riehl, T. E., Hunt, C., Sureban, S. M., Anant, S., Houchen, C. W. (2008). Identification of a Novel Putative Gastrointestinal Stem Cell and Adenoma Stem Cell Marker, Doublecortin and CaM Kinase-Like-1, Following Radiation Injury and in Adenomatous Polyposis Coli/Multiple Intestinal Neoplasia Mice. *Stem Cells*, 26(3), 630–637.
- Mcdermott, S. P., Eppert, K., Lechman, E., Doedens, M., Dick, J. E. (2010). Comparison of human cord blood engraftment between immunocompromised mouse strains. *Blood*, 116, 193–200.
- McGranahan, N., Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*, 168(4), 613–628.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.
- Melo, F. de S. e, Kurtova, A. V., Harnoss, J. M., Kljavin, N., Hoeck, J. D., Hung, J., ... de Sauvage, F. J. (2017). A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer. *Nature*, 543(7647), 676–680.
- Merlos-Suárez, A., Barriga, F. M., Jung, P., Iglesias, M., Céspedes, M. V., Rossell, D., ... Pals, S. T. (2011). The Intestinal Stem Cell Signature Identifies Colorectal Cancer

- Stem Cells and Predicts Disease Relapse. *Cell Stem Cell*, 8(5), 511–524.
- Merrell, A. J., Stanger, B. Z. (2016). Adult cell plasticity in vivo: de-differentiation and transdifferentiation are back in style. *Nature Reviews Molecular Cell Biology*, 17(7), 413–425.
- Metcalfe, C., Kljavin, N. M., Ybarra, R., de Sauvage, F. J. (2014). Lgr5+ stem cells are indispensable for radiation-induced intestinal regeneration. *Cell Stem Cell*, 14(2), 149–159.
- Miller, I., Min, M., Yang, C., Tian, C., Gookin, S., Carter, D., Correspondence, S. L. S. (2018). Ki67 is a Graded Rather than a Binary Marker of Proliferation versus Quiescence. *Cell Reports*, 24, 1105–1112.
- Misiakos, E. P., Karidis, N. P., Kouraklis, G. (2011). Current treatment for colorectal liver metastases. *World Journal of Gastroenterology*, 17(36), 4067–4075.
- Moll, R., Schiller, D. L., Franke, W. W. (1990). Identification of protein IT of the intestinal cytoskeleton as a novel type I cytokeratin with unusual properties and expression patterns. *The Journal of Cell Biology*, 111(2), 567–580.
- Montgomery, R. K., Carlone, D. L., Richmond, C. A., Farilla, L., Kranendonk, M. E. G., Henderson, D. E., . . . Breault, D. T. (2011). Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *PNAS*, 108(1), 179–184.
- Morel, A.P., Lièvre, M., Thomas, C., Hinkal, G., Ansieau, S., Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PloS One*, 3(8), e2888.
- Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., Kinzler, K. W. (1997). Activation of  $\beta$ -catenin-Tcf signaling in colon cancer by mutations in  $\beta$ -catenin or APC. *Science*, 275(5307), 1787–1790.
- Morrison, S. J., Kimble, J. (2006). Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature*, 441(7097), 1068–74.
- Morrison, S. J., Spradling, A. C. (2008). Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell*, 132(4), 598–611.
- Moser, A., Luongo, C., Gould, K., McNeley, M., Shoemaker, A., Dove, W., (1995). ApcMin: A mouse model for intestinal and mammary tumorigenesis. *European Journal of Cancer*, 31(7–8), 1061–1064.
- Moser, A. R., Pitot, H. C., Dove, W. F. (1990). A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science*, 247(4940), 322–324. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2296722>
- Mouradov, D., Sloggett, C., Jorissen, R. N., Love, C. G., Li, S., Burgess, A. W., . . . Sieber, O. M. (2014). Molecular and Cellular Pathobiology Colorectal Cancer Cell Lines Are Representative Models of the Main Molecular Subtypes of Primary Cancer. *Molecular and Cellular Pathobiology*, 74(12), 1–10.
- Munoz, J., Stange, D. E., Schepers, A. G., van de Wetering, M., Koo, B.-K., Itzkovitz, S., . . . Clevers, H. (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent “+4” cell markers. *The EMBO Journal*, 31(14), 3079–3091.

- Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L., Luo, L. (2007). A global double-fluorescent Cre reporter mouse. *Genesis*, 45(9), 593–605.
- Nakanishi, Y., Seno, H., Fukuoka, A., Ueo, T., Yamaga, Y., Maruno, T., . . . Chiba, T. (2013). *Dclk1* distinguishes between tumor and normal stem cells in the intestine. *Nature Genetics*, 45(1), 98–103.
- Nieto, M. A., Huang, R. Y.-J., Jackson, R. A., Thiery, J. P. (2016). EMT: 2016. *Cell*, 166(1), 21–45.
- O'Brien, C. A., Pollett, A., Gallinger, S., Dick, J. E. (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature*, 445(7123), 106–110.
- O'Rourke, K. P., Loizou, E., Livshits, G., Schatoff, E. M., Baslan, T., Manchado, E., . . . Lowe, S. W. (2017). Transplantation of engineered organoids enables rapid generation of metastatic mouse models of colorectal cancer. *Nature Biotechnology*, 35(6), 577–582.
- Oost, K. C., van Voorthuijsen, L., Fumagalli, A., Lindeboom, R. G. H., Sprangers, J., Omerzu, M., . . . Snippert, H. J. G. (2018). Specific Labeling of Stem Cell Activity in Human Colorectal Organoids Using an ASCL2-Responsive Minigene. *Cell Reports*, 22(6), 1600–1614.
- Ootani, A., Li, X., Sangiorgi, E., Ho, Q. T., Ueno, H., Toda, S., . . . Kuo, C. J. (2009). Sustained in vitro intestinal epithelial culture within a Wnt-dependent stem cell niche. *Nature Medicine*, 15(6), 701–706.
- Oskarsson, T., Batlle, E., Massagué, J. (2014). Metastatic Stem Cells: Sources, Niches, and Vital Pathways. *Cell Stem Cell*, 14(3), 306–321.
- Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, 31(9), 839–843.
- Pinto, D., Gregorieff, A., Begthel, H., Clevers, H. (2003). Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes and Development*, 17(14), 1709–1713.
- Potten, C. S., Booth, C., Tudor, G. L., Booth, D., Brady, G., Hurley, P., . . . Okano, H. (2003). Identification of a putative intestinal stem cell and early lineage marker; *musashi-1*. *Differentiation*, 71, 28–41.
- Potten, C. S., Kovacs, L., Hamilton, E. (1974). Continuous labelling studies on mouse skin and intestine. *Cell Proliferation*, 7(3), 271–283.
- Potten, C. S., Loeffler, M. (1987). A comprehensive model of the crypts of the small intestine of the mouse provides insight into the mechanisms of cell migration and the proliferation hierarchy. *Journal of Theoretical Biology*, 127(4), 381–391.
- Potten, C. S., Owen, G., Booth, D. (2002). Intestinal stem cells protect their genome by selective segregation of template DNA strands. *Journal of Cell Science*, 115(Pt 11), 2381–2388.

- Powell, A. E., Wang, Y., Li, Y., Poulin, E. J., Means, A. L., Washington, M. K., ... Coffey, R. J. (2012). The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. *Cell*, 149(1), 146–158.
- Prall, F., Nizze, H., Barten, M. (2005). Tumour budding as prognostic factor in stage I/II colorectal carcinoma. *Histopathology*, 47, 17–24.
- Prasetyanti, P. R., van Hooff, S. R., van Herwaarden, T., de Vries, N., Kalløe, K., Rodermond, H., ... Medema, J. P. (2019). Capturing colorectal cancer inter-tumor heterogeneity in patient-derived xenograft (PDX) models. *International Journal of Cancer*, 144(2), 366–371.
- Prasetyanti, P., Zimmerlin, C., Bots, M., Vermeulen, L., De Sousa E Melo, F., Medema, J. (2013). Regulation of stem cell self-renewal and differentiation by Wnt and Notch are conserved throughout the adenoma-carcinoma sequence in the colon. *Molecular Cancer*, 12, 126.
- Provenzano, P. P., Eliceiri, K. W., Campbell, J. M., Inman, D. R., White, J. G., Keely, P. J. (2006). Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Medicine*, 4(1), 38.
- Puig, I., Tenbaum, S. P., Chicote, I., Arqués, O., Martínez-Quintanilla, J., Cuesta-Borrás, E., ... Palmer, H. G. (2018). TET2 controls chemoresistant slow-cycling cancer cell survival and tumor recurrence. *The Journal of Clinical Investigation*, 128(9), 3887–3905.
- Quintana, E., Shackleton, M., Sabel, M. S., Fullen, D. R., Johnson, T. M., Morrison, S. J. (2008). Efficient tumour formation by single human melanoma cells. *Nature*, 456(7222), 593–598.
- Rad, R., Cadiñ, J., Rad, L., Varela, I., Strong, A., Kriegl, L., ... Bradley, A. (2013). A Genetic Progression Model of Braf V600E-Induced Intestinal Tumorigenesis Reveals Targets for Therapeutic Intervention. *Cancer Cell*, 24, 15–29.
- Raftopoulou, M., Hall, A. (2004). Cell migration: Rho GTPases lead the way. *Developmental Biology*, 265(1), 23–32.
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), 2281–2308.
- Rankin, E. B., Giaccia, A. J. (2016). Hypoxic control of metastasis. *Science*, 352(6282), 175–180.
- Rao, J. N., Wang, J.-Y. (2010). *Intestinal Architecture and Development. Regulation of Gastrointestinal Mucosal Growth*. Morgan and Claypool Life Sciences.
- Ricci-Vitiani, L., Lombardi, D. G., Pilozzi, E., Biffoni, M., Todaro, M., Peschle, C., De Maria, R. (2007). Identification and expansion of human colon-cancer-initiating cells. *Nature*, 445(7123), 111–115.
- Richmond, C. A., Rickner, H., Shah, M. S., Ediger, T., Deary, L., Zhou, F., ... Breault, D. T. (2018). Stem Cell Reports Signaling Is Required for Reserve Intestinal Stem Cell Activation during Intestinal Regeneration Following Acute Inflammation. *Stem Cell*

Reports, 10, 17–26.

Riihimäki, M., Hemminki, A., Sundquist, J., Hemminki, K. (2016). Patterns of metastasis in colon and rectal cancer. *Scientific Reports*, 6, 29765.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47–e47.

Ritsma, L., Ellenbroek, S. I. J., Zomer, A., Snippert, H. J., De Sauvage, F. J., Simons, B. D., ... Van Rheenen, J. (2014). Intestinal crypt homeostasis revealed at single stem cell level by in vivo live-imaging. *Nature*, 507(7492), 362–365.

Robanus-Maandag, E. C., Koelink, P. J., Breukel, C., Salvatori, D. C. F., Jagmohan-Changur, S. C., Bosch, C. A. J., ... Smits, R. (2010). A new conditional Apc-mutant mouse model for colorectal cancer. *Carcinogenesis*, 31(5), 946–952.

Rodríguez-Colman, M. J., Schewe, M., Meerlo, M., Stigter, E., Gerrits, J., Pras-Raves, M., ... Burgering, B. M. T. (2017). Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature*, 543(7645), 424–427.

Romagnolo, B., Berrebi, D., Saadi-Keddoucci, S., Porteu, A., Pichard, A. L., Peuchmaur, M., ... Perret, C. (1999). Intestinal dysplasia and adenoma in transgenic mice after overexpression of an activated beta-catenin. *Cancer Research*, 59(16), 3875–3879.

Rompolas, P., Deschene, E. R., Zito, G., Gonzalez, D. G., Saotome, I., Haberman, A. M., Greco, V. (2012). Live imaging of stem cell and progeny behaviour in physiological hair-follicle regeneration. *Nature*, 487(7408), 496–9.

Roper, J., Tammela, T., Cetinbas, N. M., Akkad, A., Roghanian, A., Rickelt, S., ... Yilmaz, Ö. H. (2017). In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis. *Nature Biotechnology*, 35(6), 569–576.

Russo, M., Lamba, S., Lorenzato, A., Sogari, A., Corti, G., Rospo, G., ... Bardelli, A. (2018). Reliance upon ancestral mutations is maintained in colorectal cancers that heterogeneously evolve during targeted therapies. *Nature Communications*, 9, 2287.

Ryser, M. D., Yu, M., Grady, W., Siegmund, K., Shibata, D. (2018). Epigenetic Heterogeneity in Human Colorectal Tumors Reveals Preferential Conservation And Evidence of Immune Surveillance. *Scientific Reports*, 8, 17292.

Sada, A., Jacob, F., Leung, E., Wang, S., White, B. S., Shalloway, D., Tumbar, T. (2016). Defining the cellular lineage hierarchy in the interfollicular epidermis of adult skin. *Nature Cell Biology*, 18(6), 619–31.

Salic, A., Mitchison, T. J. (2008). A chemical method for fast and sensitive detection of DNA synthesis in vivo. *PNAS*, 105(7), 2415–2420.

Sánchez-Danés, A. slow-cycling L. tumour population mediates basal cell carcinoma relapse after therapy, Larsimont, J.-C., Liagre, M., Muñoz-Couselo, E., Lapouge, G., Brisebarre, A., ... Blanpain, C. (2018). A slow-cycling LGR5 tumour population mediates basal cell carcinoma relapse after therapy. *Nature*, 562(7727), 434–438.

- Sancho, E., Batlle, E., Clevers, H. (2004). Signaling pathways in intestinal development and cancer. *Annu. Rev. Cell Dev. Biol.*, 20, 695–723.
- Sancho, P., Barneda, D., Heeschen, C. (2016). Hallmarks of cancer stem cell metabolism. *British Journal of Cancer*, 114(12), 1305–12.
- Sangiorgi, E., Capecchi, M. R. (2008). *Bmi1* is expressed in vivo in intestinal stem cells. *Nature Genetics*, 40(7), 915–920.
- Sansom, O. J., Meniel, V. S., Muncan, V., Phesse, T. J., Wilkins, J. A., Reed, K. R., ... Clarke, A. R. (2007). *Myc* deletion rescues *Apc* deficiency in the small intestine. *Nature*, 446(7136), 676–679.
- Sasaki, N., Sachs, N., Wiebrands, K., Ellenbroek, S. I. J., Fumagalli, A., Lyubimova, A., ... Analyzed Data, A. V. O. (2016). *Reg4* + deep crypt secretory cells function as epithelial niche for *Lgr5* + stem cells in colon. *PNAS*, 113(37), 5399–5407.
- Sato, T., Stange, D. E., Ferrante, M., Vries, R. G. J., Van Es, J. H., Van Den Brink, S., ... Clevers, H. (2011). Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium. *YGAST*, 141, 1762–1772.
- Sato, T., van Es, J. H., Snippert, H. J., Stange, D. E., Vries, R. G., van den Born, M., ... Clevers, H. (2011). Paneth cells constitute the niche for *Lgr5* stem cells in intestinal crypts. *Nature*, 469(7330), 415–418.
- Sato, T., Vries, R. G., Snippert, H. J., Van De Wetering, M., Barker, N., Stange, D. E., ... Clevers, H. (2009). Single *Lgr5* stem cells build crypt–villus structures in vitro without a mesenchymal niche. *Nature*, 459, 262–265.
- Sauer, B., Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *PNAS*, 85, 5166–5170.
- Schepers, A. G., Snippert, H. J., Stange, D. E., Van Den Born, M., Van Es, J. H., Van De Wetering, M., Clevers, H. (2012). Lineage Tracing Reveals *Lgr5* + Stem Cell Activity in Mouse Intestinal Adenomas. *Science*, 337, 730–735.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676–682.
- Scholzen, T., Gerdes, J. (2000). The Ki-67 protein: From the known and the unknown. *Journal of Cellular Physiology*, 182(3), 311–322.
- Schonhoff, S. E., Giel-Moloney, M., Leiter, A. B. (2004). Minireview: Development and Differentiation of Gut Endocrine Cells. *Endocrinology*, 145(6), 2639–2644.
- Schwitalla, S., Fingerle, A. A., Cammareri, P., Nebelsiek, T., Gö Ktuna, S. I., Ziegler, P. K., ... Greten, F. R. (2013). Intestinal Tumorigenesis Initiated by Dedifferentiation and Acquisition of Stem-Cell-like Properties. *Cell*, 152, 25–38.
- Shaner, N. C., Lambert, G. G., Chammas, A., Ni, Y., Cranfill, P. J., Baird, M. A., ... Wang, J. (2013). A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nature Methods*, 10(5), 407–409.

- Shimokawa, M., Ohta, Y., Nishikori, S., Matano, M., Takano, A., Fujii, M., ... Sato, T. (2017). Visualization and targeting of LGR5+ human colon cancer stem cells. *Nature*, 545(7653), 187–192.
- Shinto, E., Tsuda, H., Ueno, H., Hashiguchi, Y., Hase, K., Tamai, S., ... Matsubara, O. (2005). Prognostic implication of laminin-5 gamma 2 chain expression in the invasive front of colorectal cancers, disclosed by area-specific four-point tissue microarrays. *Laboratory Investigation*, 85, 257–266.
- Siegmund, K. D., Marjoram, P., Woo, Y.-J., Tavaré, S., Shibata, D. (2009). Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *PNAS*, 106(12), 4828–4833.
- Singh, S. K., Hawkins, C., Clarke, I. D., Squire, J. A., Bayani, J., Hide, T., ... Dirks, P. B. (2004). Identification of human brain tumour initiating cells. *Nature*, 432(7015), 396–401.
- Smits, R., Kartheuser, A., Jagmohan-Changur, S., Leblanc, V., Breukel, C., de Vries, A., ... Fodde, R. (1997). Loss of Apc and the entire chromosome 18 but absence of mutations at the Ras and Tp53 genes in intestinal tumors from Apc1638N, a mouse model for Apc-driven carcinogenesis. *Carcinogenesis*, 18(2), 321–7.
- Snippert, H. J., van der Flier, L. G., Sato, T., van Es, J. H., van den Born, M., Kroon-Veenboer, C., ... Clevers, H. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell*, 143, 134–144.
- Solanas, G., Cortina, C., Sevillano, M., Batlle, E. (2011). Cleavage of E-cadherin by ADAM10 mediates epithelial cell sorting downstream of EphB signalling. *Nature Cell Biology*, 13(9), 1100–1107.
- Sommer, C., Straehle, C., Köthe, U., Hamprecht, F. A. (2011). *ilastik: Interactive Learning and Segmentation Toolkit*. Proceedings, 230–233. Retrieved from [www.ilastik.org](http://www.ilastik.org).
- Sottoriva, A., Kang, H., Ma, Z., Graham, T. A., Salomon, M. P., Zhao, J., ... Genet, N. (2015). A Big Bang model of human colorectal tumor growth HHS Public Access Author manuscript. *Nature Genetics*, 47(3), 209–216.
- Srinivasan, T., Than, E. B., Bu, P., Tung, K.-L., Chen, K.-Y., Augenlicht, L., ... Shen, X. (2016). Notch signalling regulates asymmetric division and inter-conversion between lgr5 and bmi1 expressing intestinal stem cells. *Scientific Reports*, 6, 26069.
- Steele, C. W., Karim, S. A., Leach, J. D. G., Barry, S. T., Sansom, O. J., Morton, J. P. (2016). CXCR2 Inhibition Profoundly Suppresses Metastases and Augments Immunotherapy in Pancreatic Ductal Adenocarcinoma. *Cancer Cell*, 29, 832–845.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, 102(43), 15545–15550.
- Suda, T., Takubo, K., Semenza, G. L. (2011). Metabolic regulation of hematopoietic stem cells in the hypoxic niche. *Cell Stem Cell*, 9(4), 298–310.
- Sun, G.-G., Wang, Y.-D., Cui, D.-W., Cheng, Y.-J., Hu, W.-N. (2014). Epithelial membrane protein 1 negatively regulates cell growth and metastasis in colorectal carcinoma.

- noma. *World Journal of Gastroenterology*, 20(14), 4001–4010.
- Sun, G. G., Wang, Y. D., Lu, Y. F., Hu, W. N. (2014). EMP1, a member of a new family of antiproliferative genes in breast carcinoma. *Tumour Biology*, 35, 3347–3354.
- Sun, G., Zhao, G., Lu, Y., Wang, Y., Yang, C. (2014). Association of EMP1 with gastric carcinoma invasion, survival and prognosis. *International Journal of Oncology*, 45, 1091–1098.
- Suzuki, M., Mose, E. S., Montel, V., Tarin, D. (2006). Dormant cancer cells retrieved from metastasis-free organs regain tumorigenic and metastatic potency. *The American Journal of Pathology*, 169(2), 673–681.
- Szymczak, A. L., Vignali, D. A. (2005). Development of 2A peptide-based strategies in the design of multicistronic vectors. *Expert Opinion Biology Therapy*, 5(5), 627–638.
- Takaku, K., Oshima, M., Miyoshi, H., Matsui, M., Seldin, M. F., Taketo, M. M. (1998). Intestinal tumorigenesis in compound mutant mice of both *Dpc4* (*Smad4*) and *Apc* genes. *Cell*, 92(5), 645–656.
- Takeda, N., Jain, R., Leboeuf, M. R., Wang, Q., Lu, M. M., Epstein, J. A. (2011). Inter-conversion between intestinal stem cell populations in distinct niches. *Science*, 334(6061), 1420–1424.
- Tanaka, T., Kohno, H., Suzuki, R., Hata, K., Sugie, S., Niho, N., ... Wakabayashi, K. (2006). Dextran sodium sulfate strongly promotes colorectal carcinogenesis in *Apc* *Min/+* mice: Inflammatory stimuli by dextran sodium sulfate results in development of multiple colonic neoplasms. *International Journal of Cancer*, 118(1), 25–34.
- Tarlow, B. D., Finegold, M. J., Grompe, M. (2014). Clonal tracing of Sox9+ liver progenitors in oval cell injury. *Hepatology*, 60(1), 278–289.
- Tata, P. R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B. M., ... Rajagopal, J. (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature*, 503(7475), 218–223.
- Tauriello, D. V. F., Calon, A., Lonardo, E., Batlle, E. (2016). Determinants of metastatic competency in colorectal cancer. *Molecular Oncology*, 11(1), 97–119.
- Tauriello, D. V. F., Palomo-Ponce, S., Stork, D., Berenguer-Llargo, A., Badia-Ramentol, J., Iglesias, M., ... Batlle, E. (2018). TGF $\beta$  drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature*, 554(7693), 538–543.
- Taylor, V., Welcher, A. A., Program, A. E., Suter, U. (1995). Epithelial membrane protein-1, peripheral myelin protein 22, and lens membrane protein 20 define a novel gene family. *The Journal of Biological Chemistry*, 270(48), 28824–28833.
- TCGA. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407), 330–337.
- Team, R. D. C. (2008). R: a language and environment for statistical computing. In R Foundation for Statistical Computing, Vienna. Retrieved from <https://www.r-project.org/>

- Tetteh, P. W., Basak, O., Farin, H. F., Wiebrands, K., Kretzschmar, K., Begthel, H., ... Clevers, H. (2016). Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. *Cell Stem Cell*, 18(2), 203–213.
- Tian, H., Biehs, B., Chiu, C., Siebel, C., Wu, Y., Costa, M., ... Klein, O. D. (2015). Opposing activities of Notch and Wnt signaling regulate intestinal stem cells and gut homeostasis. *Cell Reports*, 11(1), 33–42.
- Tian, H., Biehs, B., Warming, S., Leong, K. G., Rangell, L., Klein, O. D., de Sauvage, F. J. (2011). A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature*, 478(7368), 255–259.
- Till, J. E., McCulloch, E. A., Siminovitch, L. (1964). A stochastic model of stem cell proliferation, based on the growth of spleen colony-forming cells. *PNAS*, 51(1), 29–36.
- Todaro, M., Gaggianesi, M., Catalano, V., Benfante, A., Iovino, F., Biffoni, M., ... Stassi, G. (2014). Cell Stem Cell Article CD44v6 Is a Marker of Constitutive and Reprogrammed Cancer Stem Cells Driving Colon Cancer Metastasis. *Cell Stem Cell*, 14, 342–356.
- Valiente, M., Obenauf, A. C., Jin, X., Chen, Q., H-F Zhang, X., Lee, D. J., ... Massagué, J. (2014). Serpins Promote Cancer Cell Survival and Vascular Cooption in Brain Metastasis. *Cell*, 156(5), 1002–1016.
- van de Wetering, M., Francies, H. E., Francis, J. M., Bounova, G., Iorio, F., Pronk, A., ... Clevers, H. (2015). Prospective Derivation of a Living Organoid Biobank of Colorectal Cancer Patients. *Cell*, 161(4), 933–945. 3
- van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., ... Clevers, H. (2002). The  $\beta$ -Catenin/TCF-4 Complex Imposes a Crypt Progenitor Phenotype on Colorectal Cancer Cells. *Cell*, 111(2), 241–250.
- van der Flier, L. G., Clevers, H. (2009). Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annual Review of Physiology*, 71, 241–260.
- van der Flier, L. G., Haegerbarth, A., Stange, D. E., van de Wetering, M., Clevers, H. (2009). OLFM4 Is a Robust Marker for Stem Cells in Human Intestine and Marks a Subset of Colorectal Cancer Cells. *Gastroenterology*, 137(1), 15–17.
- van der Flier, L. G., van Gijn, M. E., Hatzis, P., Kujala, P., Haegerbarth, A., Stange, D. E., ... Clevers, H. (2009). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell*, 136(5), 903–912.
- van Es, J. H., Sato, T., van de Wetering, M., Luybimova, A., Gregorieff, A., Zeinstra, L., ... Clevers, H. (2012). Dll1 marks early secretory progenitors in gut crypts that can revert to stem cells upon tissue damage. *Nature Cell Biology*, 14(10), 1099–1104.
- van Wyk, H. C., Park, J., Roxburgh, C., Horgan, P., Foulis, A., McMillan, D. C. (2015). The role of tumour budding in predicting survival in patients with primary operable colorectal cancer: a systematic review. *Cancer Treatment Reviews*, 41(2), 151–159. 7
- Verheijen, R., Kuipers, H. J. H., Schlingemann, R. O., Boehmer, A. L. M., Van Driel, R., Brakenhoff, G. J., Ramaekers, F. C. S. (1989). Ki-67 detects a nuclear matrix-associated proliferation-related antigen I. Intracellular localization during interphase. *Journal of*

Cell Science, 92, 123–130.

Vermeulen, L., De Sousa E Melo, F., van der Heijden, M., Cameron, K., de Jong, J. H., Borovski, T., ... Medema, J. P. (2010). Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nature Cell Biology*, 12(5), 468–476.

Vermeulen, L., Todaro, M., de Sousa Mello, F., Sprick, M. R., Kemper, K., Perez Alea, M., ... Medema, J. P. (2008). Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. *PNAS*, 105(36), 13427–13432.

Vlachogiannis, G., Hedayat, S., Vatsiou, A., Jamin, Y., Fernández-Mateos, J., Khan, K., ... Valeri, N. (2018). Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science*, 359(6378), 920–926.

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., ... Bos, J. L. (1988). Genetic Alterations during Colorectal-Tumor Development. *New England Journal of Medicine*, 319(9), 525–532.

Wang, D., Sun, H., Wei, J., Cen, B., Dubois, R. N. (2017). CXCL1 is critical for pre-metastatic niche formation and metastasis in colorectal cancer. *Cancer Research*, 77(13), 3655–3665.

Warren, R. S., Yuan, H., Matli, M. R., Gillett, N. A., Ferrara, N. (1995). Regulation by Vascular Endothelial Growth Factor of Human Colon Cancer Tumorigenesis in a Mouse Model of Experimental Liver Metastasis. *Journal of Clinical Investigation*, 95, 1789–1797.

Wculek, S. K., Malanchi, I. (2015). Neutrophils support lung colonization of metastasis-initiating breast cancer cells. *Nature*, 528(7582), 413–417.

Weeber, F., Van De Wetering, M., Hoogstraat, M., Dijkstra, K. K., Krijgsman, O., Kuilman, T., ... Voest, E. E. (2015). Preserved genetic diversity in organoids cultured from biopsies of human colorectal cancer metastases. *PNAS*, 112(43), 13308–13311.

Weisenberger, D. J., Siegmund, K. D., Campan, M., Young, J., Long, T. I., Faasse, M. A., ... Laird, P. W. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature Genetics*, 38(7), 787–793.

Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G., Cunningham, S. J. (1999). Weka: Practical machine learning tools and techniques with Java implementations. . . Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems, 192–196.

Yamamoto, T., Kawada, K., Itatani, Y., Inamoto, S., Okamura, R., Iwamoto, M., ... Sakai, Y. (2017). Loss of SMAD4 Promotes Lung Metastasis of Colorectal Cancer by Accumulation of CCR1+ Tumor-Associated Neutrophils through CCL15-CCR1 Axis. *Clinical Cancer Research*, 23(3), 833–844.

Yamashita, Y., Leanne Jones, D., Fuller, M. (2003). Orientation of Asymmetric Stem Cell Division by the APC Tumor Suppressor and Centrosome. *Science*, 301(5639), 1547–1550.

- Yan, K. S., Chia, L. A., Li, X., Ootani, A., Su, J., Lee, J. Y., ... Kuo, C. J. (2011). The intestinal stem cell markers *Bmi1* and *Lgr5* identify two functionally distinct populations. *PNAS*, 109(2), 466–471.
- Yang, J., Weinberg, R. A. (2008). Developmental Cell Review Epithelial-Mesenchymal Transition: At the Crossroads of Development and Tumor Metastasis. *Developmental Cell*, 14, 818–829.
- Yarden, Y., Shilo, B.-Z. (2007). SnapShot: EGFR Signaling Pathway. *Cell*, 131, 1018.
- Yonenaga, Y., Mori, A., Onodera, H., Yasuda, S., Oe, H., Fujimoto, A., ... Imamura, M. (2005). Absence of smooth muscle actin-positive pericyte coverage of tumor vessels correlates with hematogenous metastasis and prognosis of colorectal cancer patients. *Oncology*, 69(2), 159–66.
- Yui, S., Nakamura, T., Sato, T., Nemoto, Y., Mizutani, T., Zheng, X., ... Watanabe, M. (2012). Functional engraftment of colon epithelium expanded in vitro from a single adult *Lgr5*<sup>+</sup> stem cell. *Nature Medicine*, 18(4), 618–623.
- Zernicka-Goetz, M., Morris, S. A., Bruce, A. W. (2009). Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nature Reviews Genetics*, 10(7), 467–477.
- Zhang, R., Ma, M., Lin, X.-H., Liu, H.-H., Chen, J., Chen, J., ... Chen, R.-X. (2018). Extracellular matrix collagen I promotes the tumor progression of residual hepatocellular carcinoma after heat treatment. *BMC Cancer*, 18(1), 901.
- Zhang, R., Zhang, Z., Zhang, C., Zhang, L., Robin, A., Wang, Y., ... Chopp, M. (2004). Stroke Transiently Increases Subventricular Zone Cell Division from Asymmetric to Symmetric and Increases Neuronal Differentiation in the Adult Rat. *The Journal of Neuroscience*, 24(25), 5810–5815.
- Zhang, S., Balch, C., Chan, M. W., Lai, H.-C., Matei, D., Schilder, J. M., ... Nephew, K. P. (2008). Identification and characterization of ovarian cancer-initiating cells from primary human tumors. *Cancer Research*, 68(11), 4311–20.
- Zhang, Y., Davis, C., Shah, S., Hughes, D., Ryan, J. C., Altomare, D., Peña, M. M. O. (2017). IL-33 promotes growth and liver metastasis of colorectal cancer in mice by remodeling the tumor microenvironment and inducing angiogenesis. *Molecular Carcinogenesis*, 56(1), 272–287.
- Zheng, X., Carstens, J. L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., ... Kalluri, R. (2015). Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature*, 527(7579), 525–530.
- Zlobec, I., Hädrich, M., Dawson, H., Koelzer, V. H., Borner, M., Mallaev, M., ... Lugli, A. (2014). Intratumoural budding (ITB) in preoperative biopsies predicts the presence of lymph node and distant metastases in colon and rectal cancer patients. *British Journal of Cancer*, 110(4), 1008–1013.
- Zlobec, I., Lugli, A. (2010). Epithelial mesenchymal transition and tumor budding in aggressive colorectal cancer: tumor budding as oncotarget. *Oncotarget*, 1(7), 6516–6561.

Part VIII  
APPENDIX



## APPENDIX

## A.1 SUPPLEMENTARY TABLES

Knock-in	PDO	iRFP+ cells (day 3)	Reporter+ cells (day 20)	Total nucleofection efficiency	On-target clones	Clones without off-targets
LGR5 <sup>EGFP</sup>	PDO6	62.3%	10.3%	6.02%	76.3% (42/55)	39% (5/13)
	PDO7	46.1%	8.8%	4.56%	85.3% (35/41)	84.6% (11/13)
AAVS1 <sup>BFPTOM</sup>	PDO7	27.1%	22.2%	6.02%	47.83% (11/23)	20% (2/10)
LGR5 <sup>CreERT2</sup>	PDO7	36.4%	—	36.4%	16.67% (10/60)	50% (4/8)
	AAVS1#20					
KI67 <sup>RFP</sup>	PDO7	9%	0.4%	3.6%	76.66% (23/30)	100% (17/17)
	LGR5#1					
KRT20 <sup>CreERT2</sup>	PDO7	12.9%	0.5%	6.45%	69.44% (25/36)	
	LGR5#2					
KRT20 <sup>CreERT2</sup>	PDO7	12.8%	—	12.8%	16.67% (13/78)	75% (6/8)
	AAVS1#20					
KRT20 <sup>iRFP</sup>	PDO7	6.7%	—	6.7%	3.2% (1/31)	—
	LGR5/KI67					
EMP1 <sup>EGFP/Cre</sup>	PDO7	26.9%	—	26.9%	13.95% (6/43)	EGFP 50% (6/6)
	AAVS1#20					Cre 16.67% (1/6)
EMP1 <sup>EGFP/Cre</sup>	PDO6	9.5%	—	9.5%	13.97% (4/29)	
EMP1 <sup>iCTOM</sup>	PDO7	14%	—	14%	14.8% (4/27)	—
EMP1 <sup>iCTOM</sup>	MTO93	15.4%	7.3	1.12%	73.5% (36/49)	—

**Table A.1: CRISPR/Cas9 knock-in efficiencies in PDO6, PDO7 and MTO93.** Percentage of cells sorted at short term (iRFP+) and long term (Fluorescent reporter+, EGFP or TOM respectively) post nucleofection. Single cell clone derivation efficiency for the different modifications was assessed by integration PCR and Southern Blot. — indicates that long term sorting of Southern blot were not performed for that particular knock-in generation.

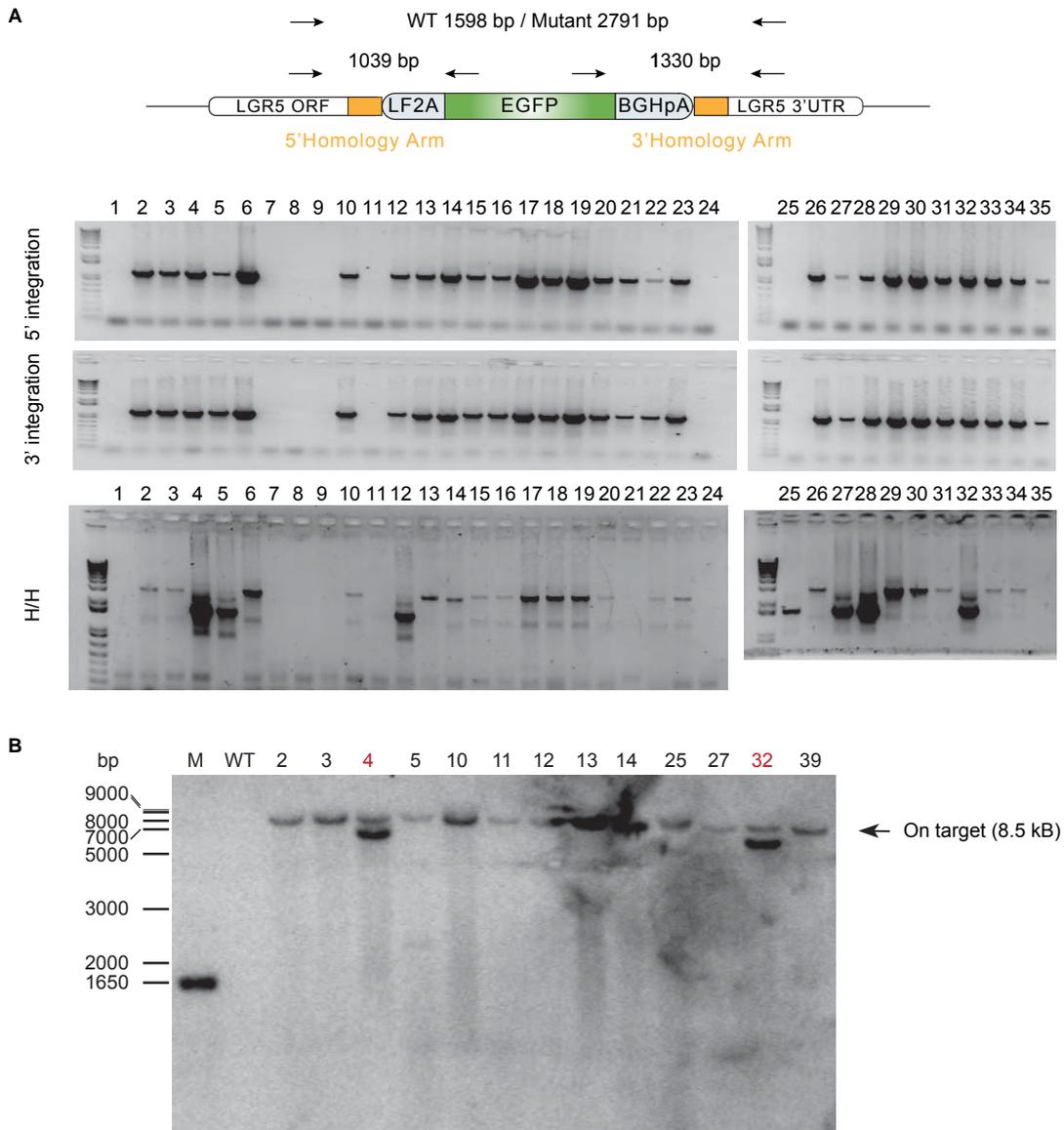
Hallmark	Size	ES	NES	FDR q-value
Interferon gamma response	192	0.650585	2.719787	0
Interferon alpha response	89	0.696786	2.612567	0
TNFA signaling via NFkB	192	0.615489	2.556763	0
E2F targets	191	0.606391	2.531585	0
Apoptosis	157	0.564595	2.318576	0
P53 pathway	192	0.528742	2.223051	0
G2M checkpoint	193	0.50642	2.125659	0
Complement	190	0.50174	2.080819	0
Epithelial mesenchymal transition	194	0.487056	2.039825	0
Mitotic spindle	198	0.475098	2.014934	0
Estrogen response late	194	0.469901	1.986545	0
UV response up	153	0.472744	1.943232	2.33E - 04
KRas signaling up	193	0.458492	1.919834	3.16E - 04
Estrogen response early	195	0.455094	1.910546	2.94E - 04
Coagulation	134	0.458058	1.844924	6.09E - 04
Inflammatory response	198	0.429894	1.80745	8.030E - 04
TGF-beta signaling	52	0.502367	1.73278	0.002379
IL6 JAK/STAT3 signaling	86	0.465063	1.723764	0.002398
Reactive oxygen species pathway	45	0.524041	1.714168	0.002437
Unfolded protein response	107	0.440978	1.690779	0.003222
Allograft rejection	196	0.382302	1.581632	0.010346
IL2 STAT5 signaling	195	0.366784	1.531935	0.016177
Protein secretion	94	0.403728	1.526831	0.016705
DNA repair	137	0.373303	1.508522	0.018971
Hypoxia	195	0.337511	1.416315	0.04329
UV response dn	139	0.353102	1.415583	0.041773
Angiogenesis	35	0.45246	1.413359	0.045205
Heme metabolism	187	0.333701	1.398766	0.044239
Apical surface	44	0.418618	1.937356	0.044239
MTORC1 signaling	191	0.334717	1.391453	0.045076

**Table A.2: Gene Ontology hallmarks enriched in PDO7 EMP1-High cells.** Detailed description of the gene size (SIZE), enrichment score (ES), normalized enrichment score (NES) and false discovery rate q-value (FDR q-val) for the GSEAs comparing EMP1-High vs -Negative gene expression microarrays against broad GO hallmarks.

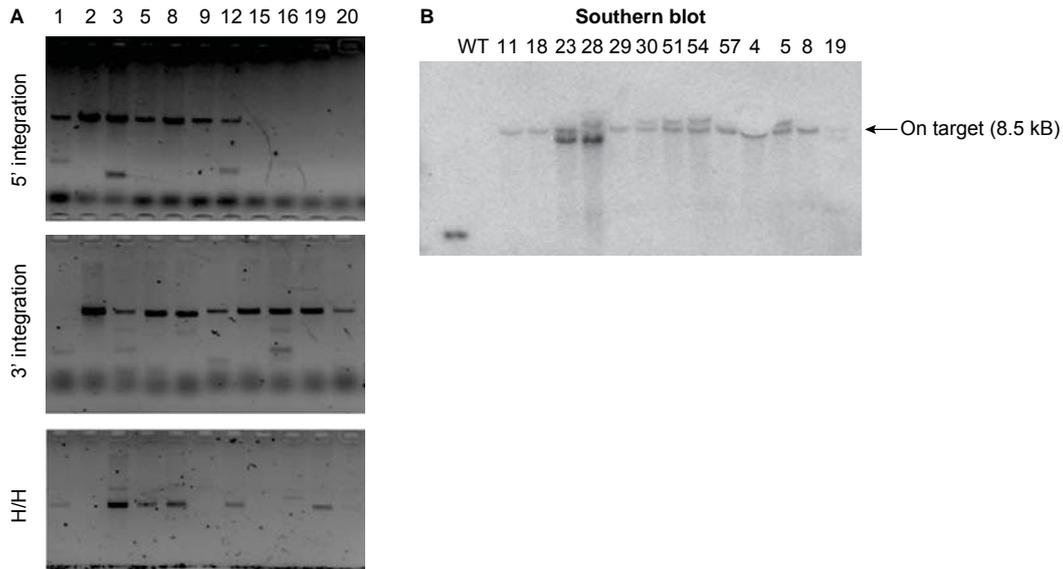
Gene Symbol	Fold Enrichment
<i>MMP7</i>	66.766
<i>PPBP/CXCL7</i>	32.412
<i>KLK10</i>	27.443
<i>CTSE</i>	19.279
<i>KRT6A</i>	13.807
<i>GABBR1</i>	12.554
<i>EMP1</i>	12.133
<i>ANXA</i>	12.128
<i>CXCL1</i>	12.121
<i>PI3</i>	11.824
<i>BIRC3</i>	9.711
<i>CASP9</i>	8.633
<i>IFI44</i>	8.624
<i>TM4FS20</i>	8.463
<i>CXCL8</i>	8.159
<i>CXCL1</i>	8.018
<i>CAV2</i>	7.974
<i>IFI27</i>	7.951
<i>RARRES3</i>	7.168
<i>TNNC1</i>	7.162

**Table A.3: Marker genes of the EMP1-High population.** Fold upregulation of the top 20 enriched genes in EMP1-High profiled cells compared to EMP1-Negative from PDO7. In green, proteases and chemoattractants.

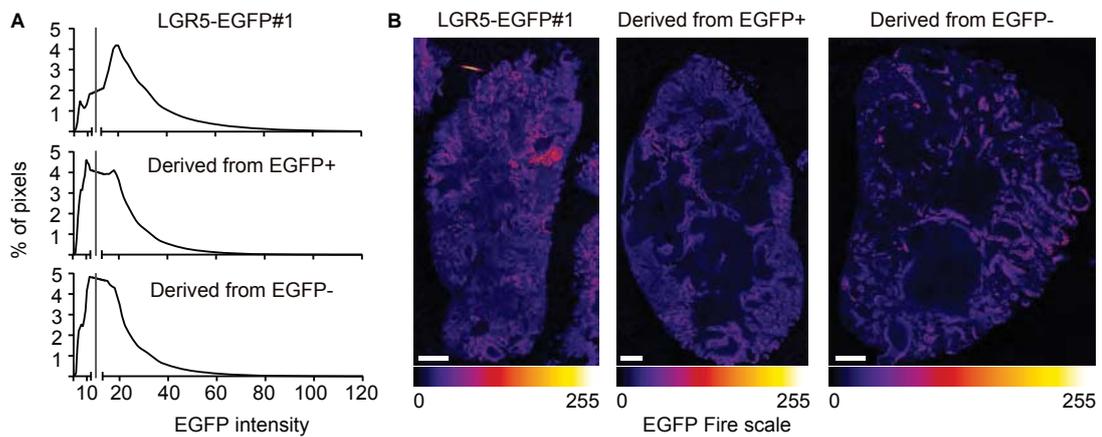
A.2 SUPPLEMENTARY FIGURES



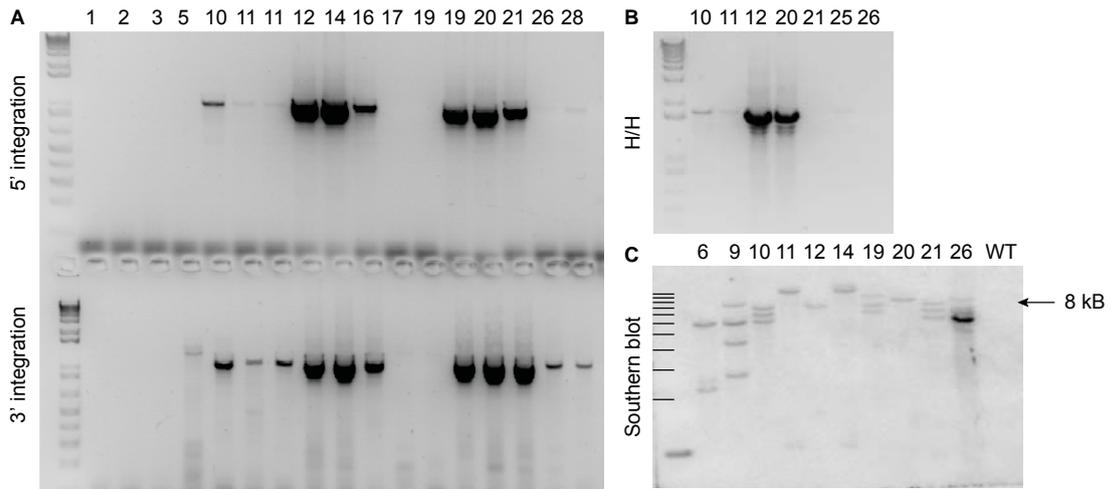
**Figure A.1: LGR5-EGFP clone screening strategy in PDO7.** **A.** Integration PCRs amplifying the 5' and 3' ends of the LGR5-EGFP knock-in (5' correct band: 1039 bp / 3' correct band: 1330 bp) as well as whole locus amplification for homozygous or heterozygous integration (wt band: 1598 bp / mutant band: 2791 bp) in PDO7 LGR5-EGFP clones. **B.** Southern Blot against correctly targeted clones to find random EGFP integrations (on-target band: 8.5 kB). Adapted from Cortina, Turon et al., 2017



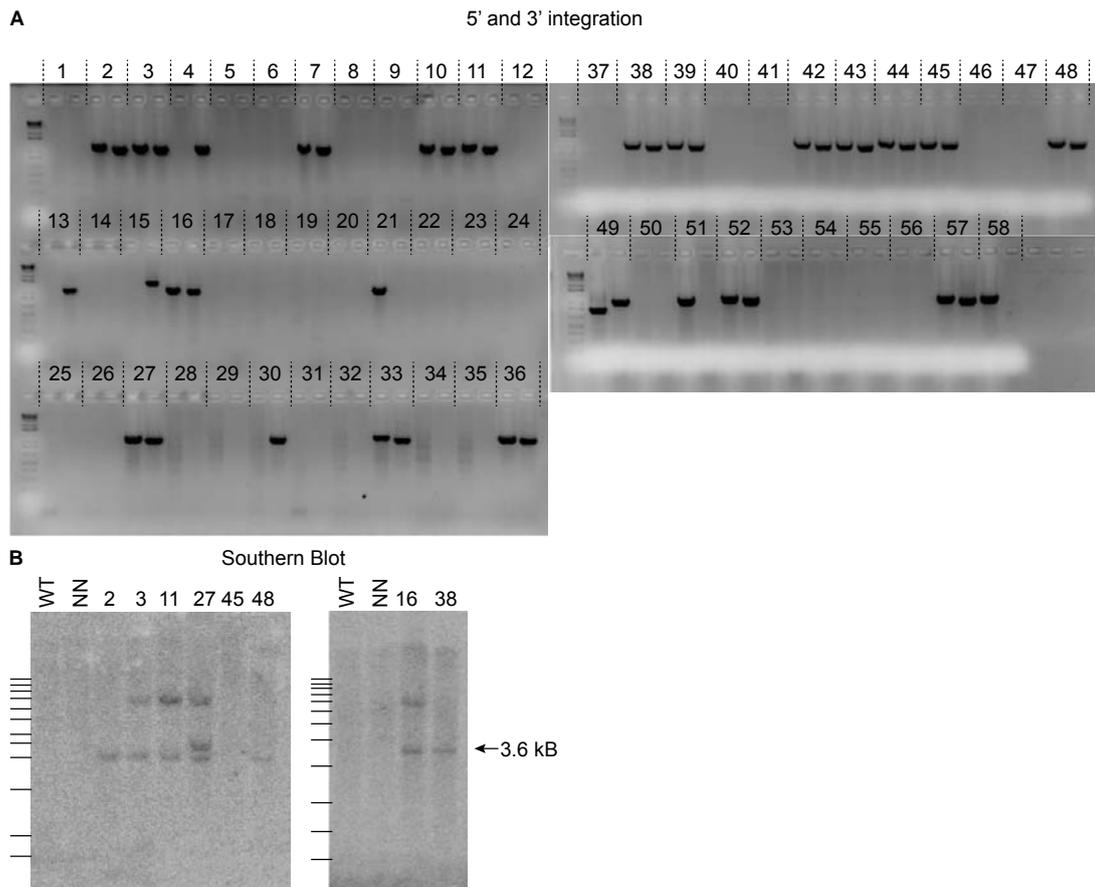
**Figure A.2: PDO6 LGR5-EGFP clone screening.** **A.** 5' and 3' integration PCRs (5' correct band: 1039 bp / 3' correct band: 1330 bp) as well as whole locus amplification (wt band: 1598 bp / mutant band: 2791 bp) of LGR5-EGFP knock-in PDO6 clones. **B.** Southern Blot against correctly targeted clones to find random EGFP integrations (on-target band: 8.5 kB).



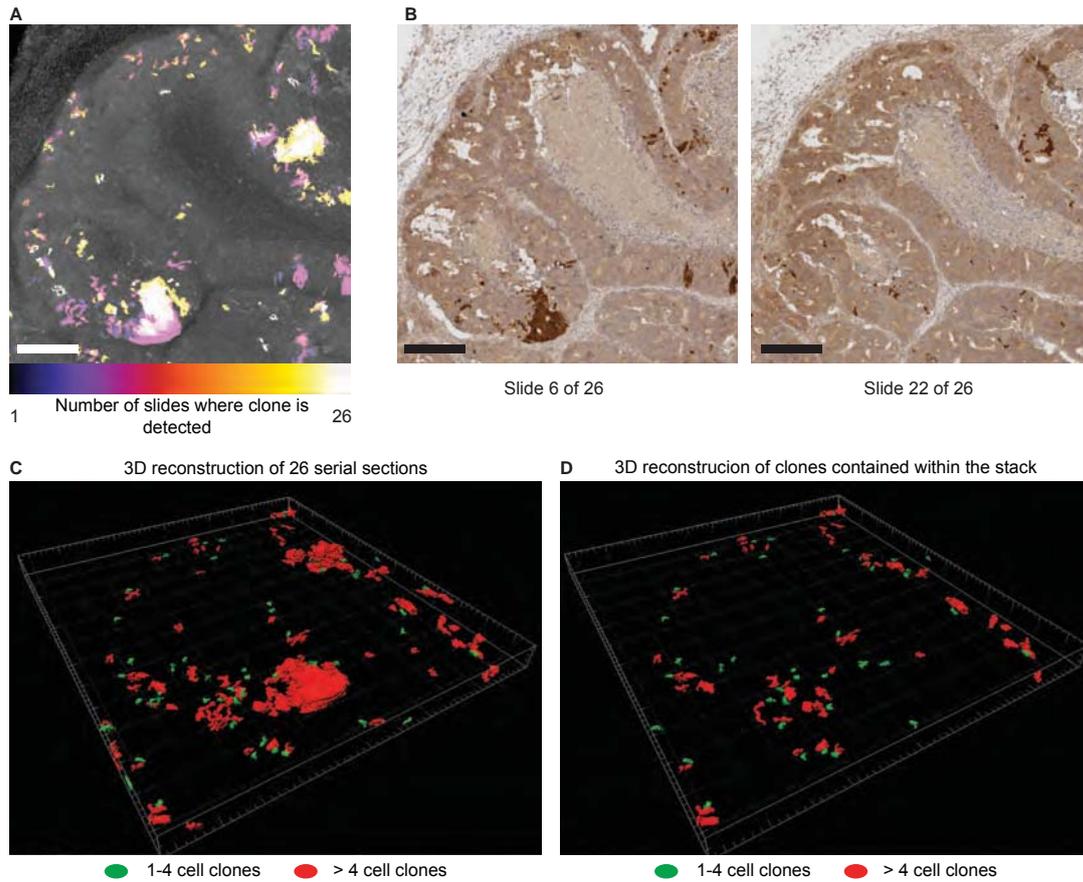
**Figure A.3: Tumor initiation-derived xenografts reproduce the parental LGR5-EGFP pattern.** **A.** Distribution of EGFP staining intensity in PDO7#1 and in xenografts derived from EGFP+ and EGFP- cells. Gray line indicates background fluorescence levels. **B.** Representative example of EGFP immunofluorescence on whole tumors formed by PDO7#1 or by EGFP+ and EGFP- cells in TIC experiments. The intensity of the signal is indicated by the Fire lookup table. Scale bars indicate 1mm. Adapted from Cortina, Turon et al., 2017



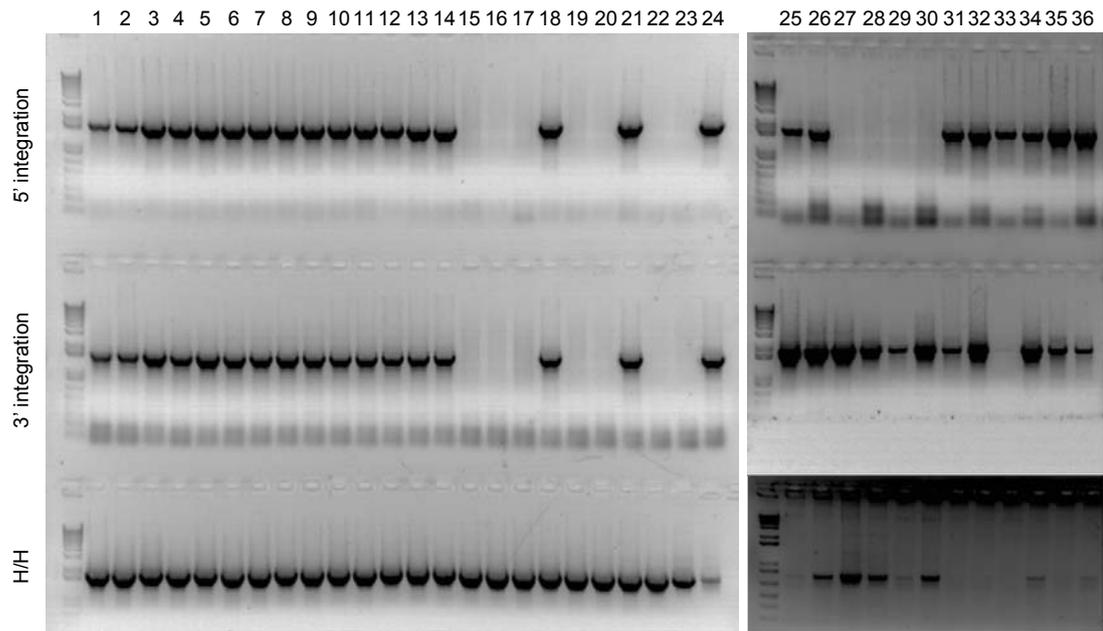
**Figure A.4: PDO7 AAVS1-BFPTOM clone screening.** **A.** 5' and 3' integration PCRs (5' correct band: 1202 bp / 3' correct band: 1084 bp) of AAVS1-BFPTOM PDO7 clones **B.** Whole locus amplification (wt band: 1720 bp / mutant band: not amplified) of the integrated clones. **C.** Southern Blot against correctly targeted clones to detect off-target integrations (on-target band: 8 kB).



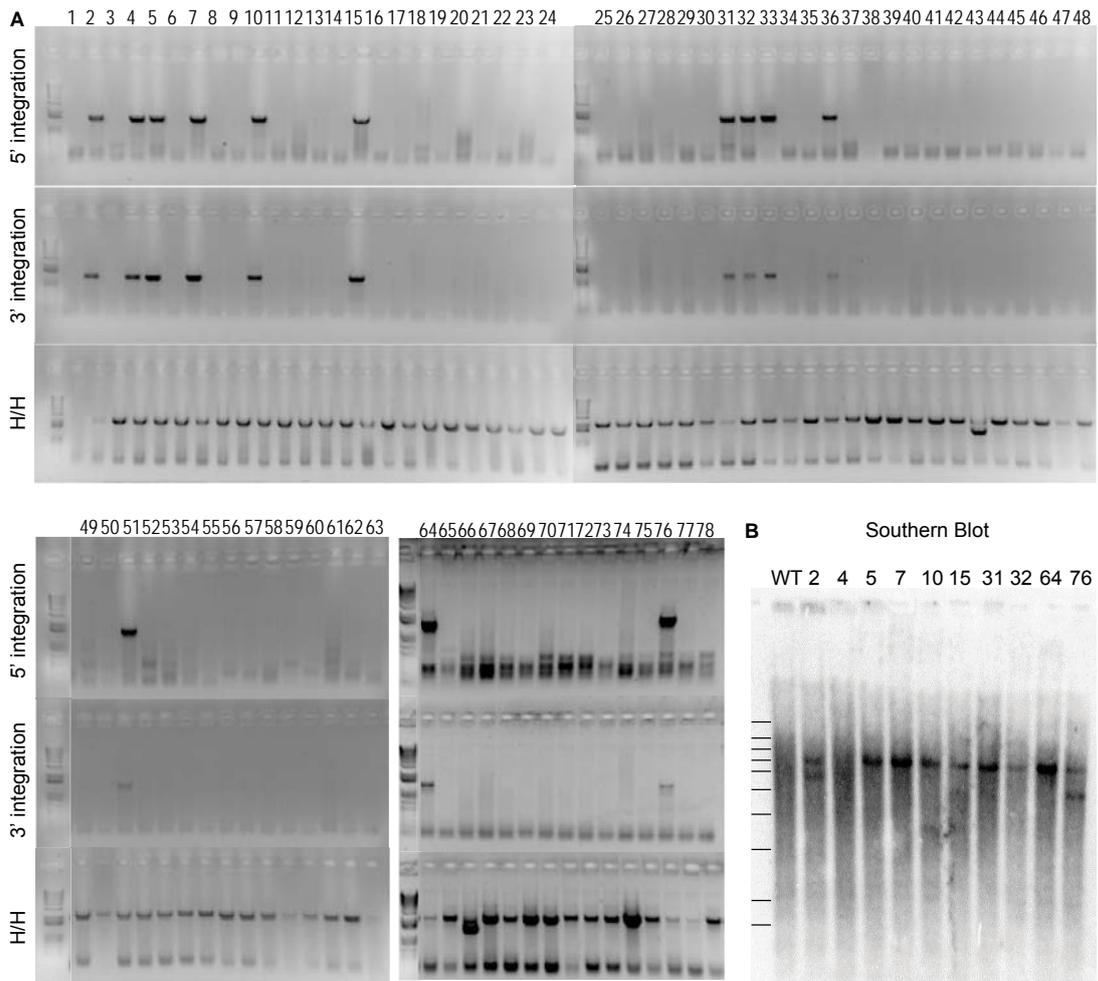
**Figure A.5: PDO7 AAVS1-BFPTOM#20 LGR5-CreERT2 clone screening.** **A.** 5' and 3' integration PCRs (5' correct band: 1039 bp / 3' correct band: 1330 bp) as well as whole locus amplification (wt band: 1598 bp / mutant band: 2791 bp) of LGR5-CreERT2 PDO7 clones. **B.** Southern Blot against correctly targeted clones to detect off-target integrations (on-target band: 3.6 kB). NN: non-nucleofected control



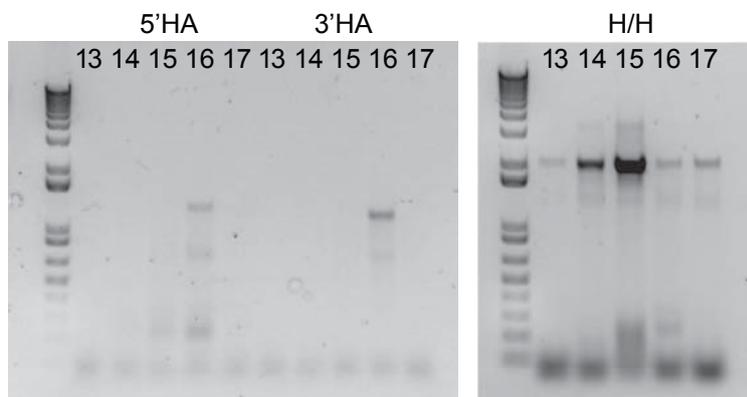
**Figure A.6: 3D reconstruction of 1-month grown *LGR5* clones.** **A.**Z-projection of clones in 26 serial sections of subcutaneous PDO7 *LGR5*-Cre#1 xenografts 1 month after tamoxifen injection. Color scale denotes the number of sections occupied by a given clone. Scale bar is 200 $\mu$ m. **B.** Example of two slides stained for Tomato used for 3D reconstruction. Scale bars are 200 $\mu$ m. **C.** Projection of the 3D reconstructed clones. Scale bar indicates 200 $\mu$ m. **D.** Projection of the 3D reconstructed clones that do not contact borders and therefore are fully embedded within the Z plan analyzed. Scale bar indicates 200 $\mu$ m. Adapted from Cortina, Turon et al., 2017



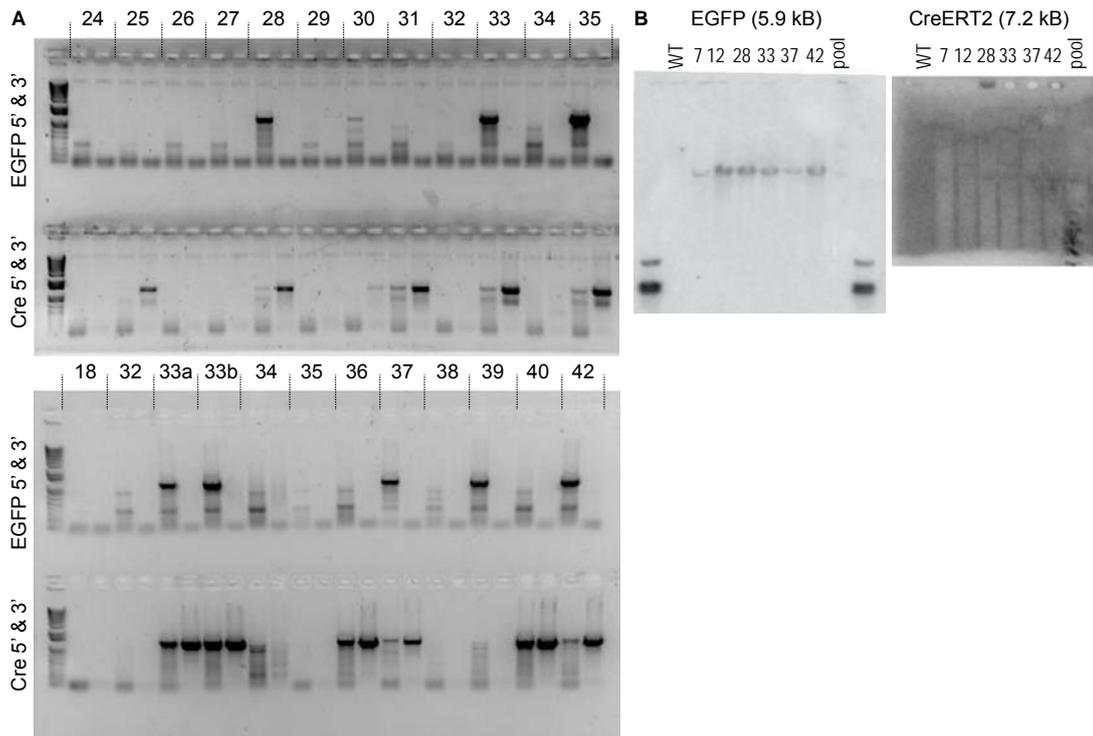
**Figure A.7: LGR5-EGFP#2/KI67-RFP clone screening.** 5' and 3' integration PCRs (5' correct band: 1633 bp / 3' correct band: 1612 bp) and WT band detection (1769 bp) of PDO7 LGR5-EGFP#2 KI67-RFP clones.



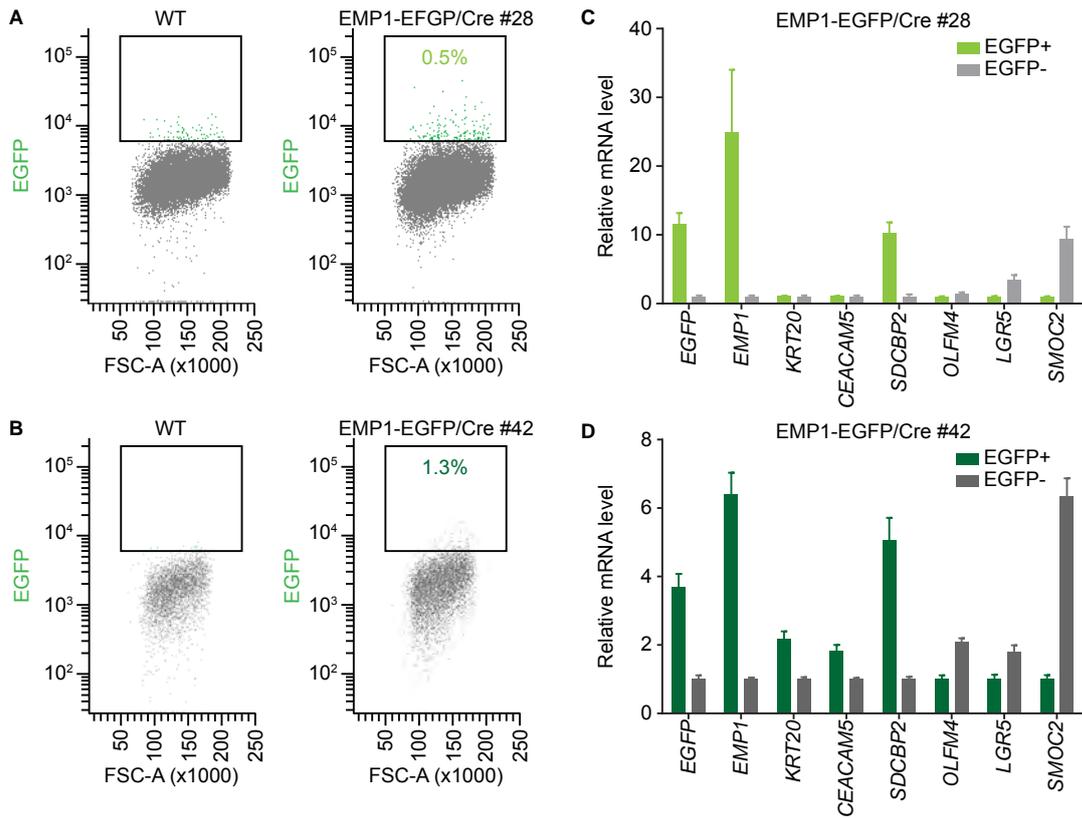
**Figure A.8: PDO7 AAVS1-BFPTOM#20 KRT20-CreERT2 clone screening.** **A.** 5' and 3' integration PCRs (5' correct band: 1606 bp / 3' correct band: 1328 bp) and WT band detection (4494 bp) of PDO7 KRT20-CreERT2 clones. **B.** Southern blot of the correctly integrated clones. On-target band: 6.8 kB.



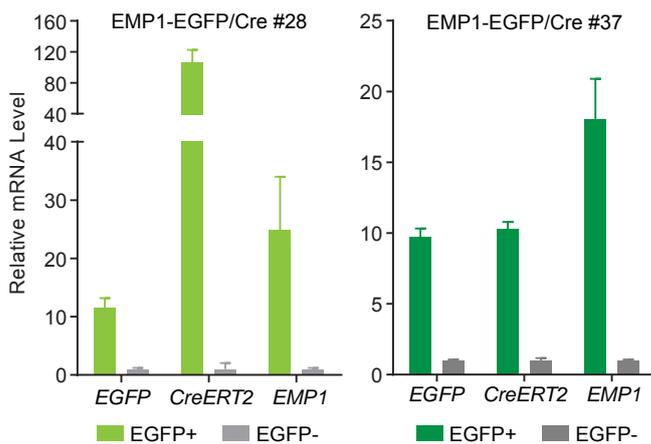
**Figure A.9: PDO7 KRT20-iRFP clone screening.** **A.** 5' and 3' integration PCRs and WT band detection (2055 bp) of PDO7 KRT20-iRFP clones.



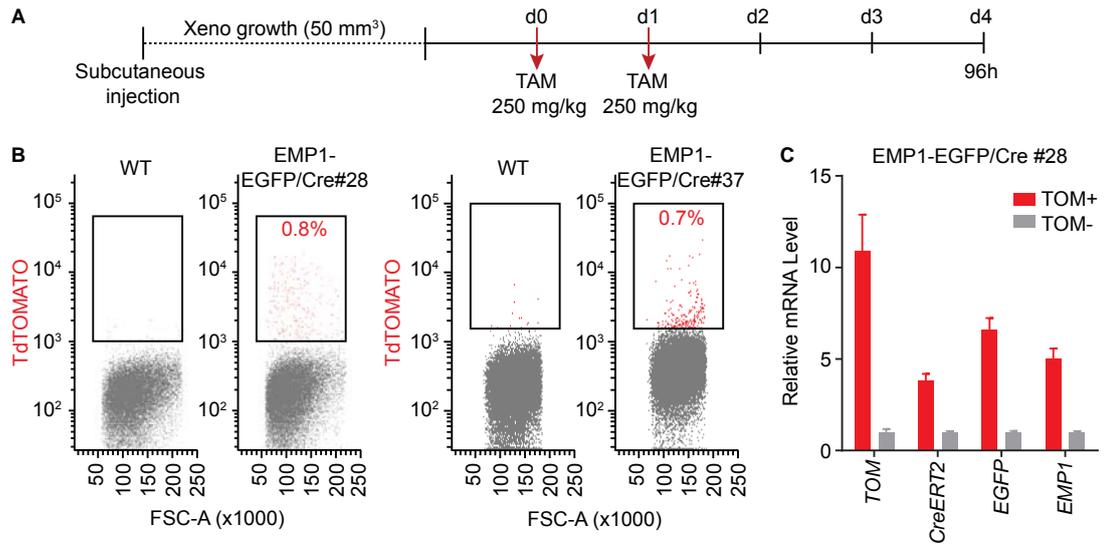
**Figure A.10: PDO7 AAVS1-BFPTOM#20 EMP1-EGFP/CreERT2 clone screening.** A. 5' and 3' integration PCRs for EGFP (5' correct band: 1204 bp / 3' correct band: 1263 bp) and CreERT2 (5' correct band: 1393 bp / 3' correct band: 1267 bp) of PDO7 EMP1-GFP/Cre clones. B. Southern blot of the correctly integrated clones for both the EGFP and the CreERT2 knock-ins. On-target band: 5900 kb for EGFP insertion and 7200 Kb for CREERT2 insertion



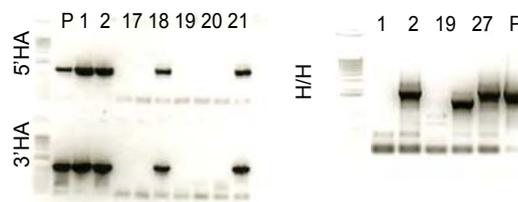
**Figure A.11: EMP1 is a differentiated-like non-stem population.** **A and B.** FACS analysis of the EPCAM+ purified cells from disaggregated subcutaneous xenografts of PDO7 EMP1-EGFP/Cre #28 and #42 respectively. **C and D.** mRNA expression analysis by qrtPCR of differentiated and stem genes in the EGFP-High and -Negative sorted populations from disaggregated subcutaneous xenografts of PDO7 EMP1-EGFP/Cre #28 and #42 respectively. Data represents mean $\pm$ s.d. of 3 technical replicates.



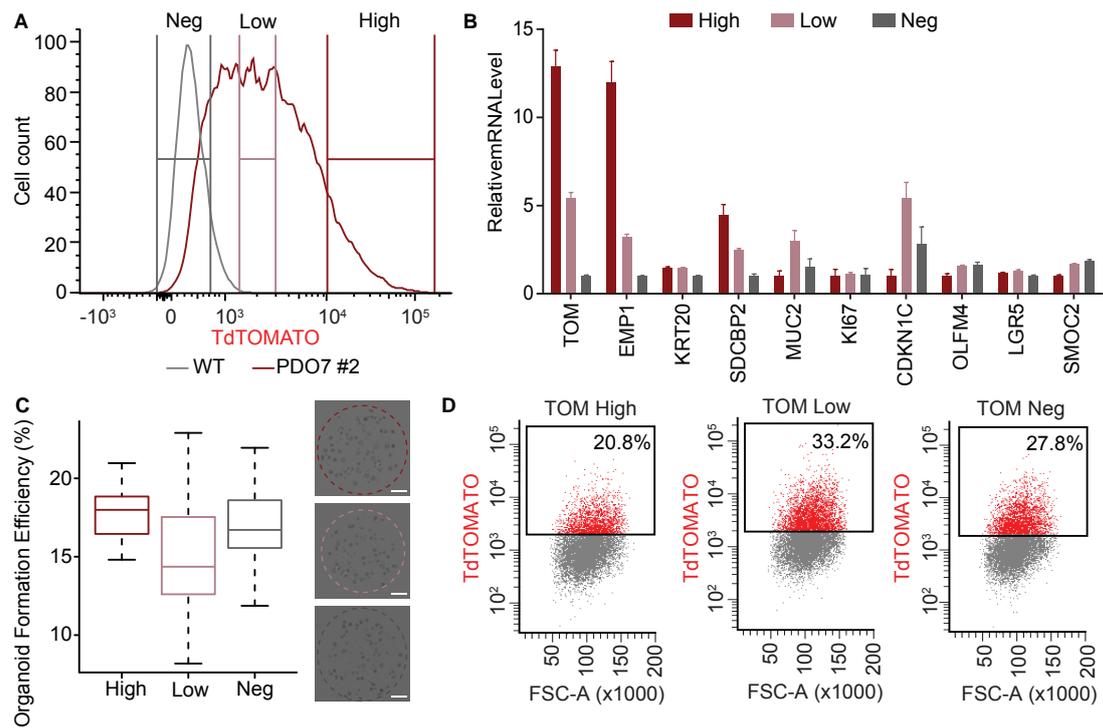
**Figure A.12: EMP1 mRNA expression correlates with CreERT2 and EGFP expression levels in PDO7 EMP1-EGFP/Cre clones.** RT-qPCR analysis of the indicated gene expression levels in FACS-isolated EGFP+ and EGFP- populations from disaggregated subcutaneous xenografts derived from PDO7 EMP1-EGFP/Cre#28 and #37. EMP1 and CreERT2 mRNA levels are specifically enriched in the EGFP+ fraction of the tumors, confirming the correct CRISPR/Cas9 knock-in. Data is represented as mean $\pm$ s.d. of three technical replicates.



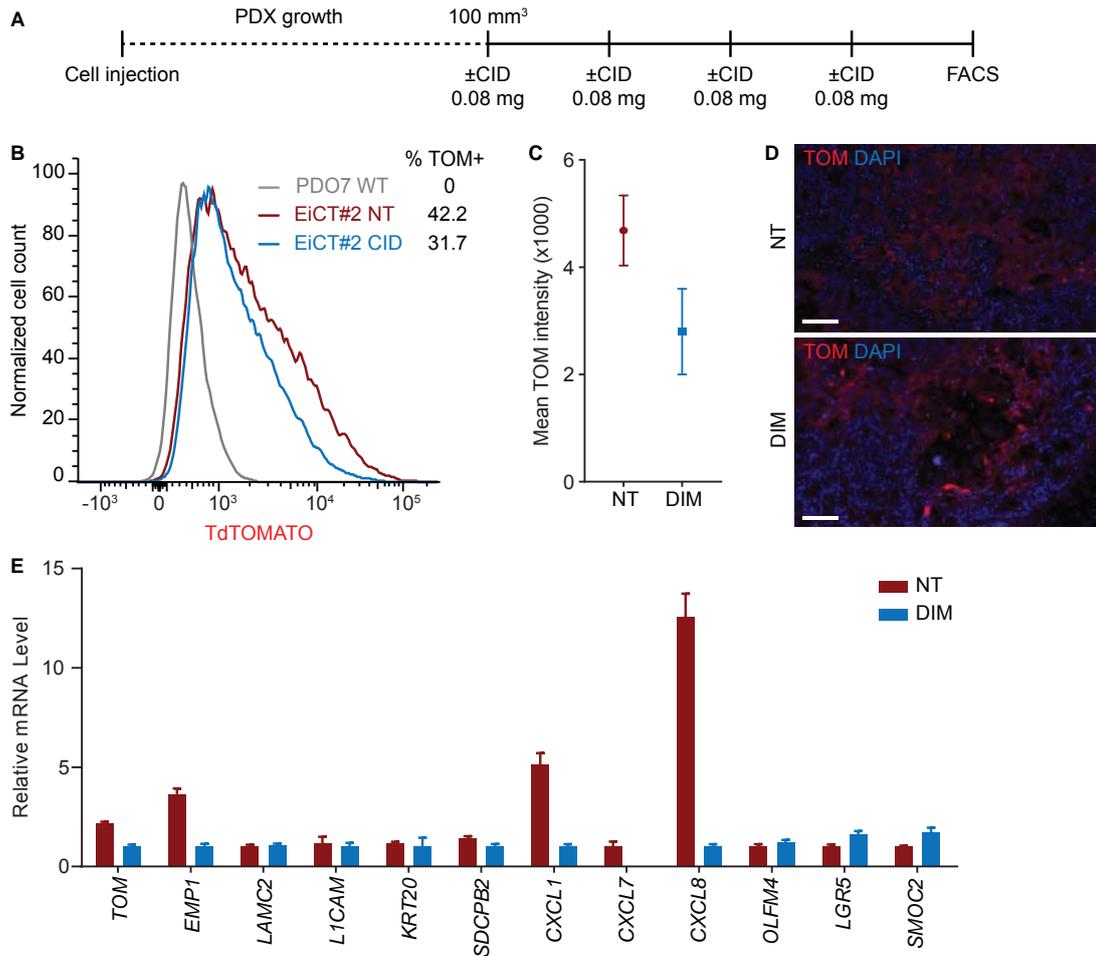
**Figure A.13: Tamoxifen induction generates a TOM+ EMP1+ population in PDO7 xenografts.** **A.** Schematic representation of the *EMP1* lineage tracing experimental design. **B.** FACS analysis of disaggregated PDO7 *EMP1*-EGFP/Cre subcutaneous xenografts induced for 96h with intraperitoneal Tam. Only EPCAM+DAPI- cells are plotted. **C.** Relative mRNA expression levels of the indicated genes by RT-qPCR in TOM+ and TOM- sorted populations 96h post-induction. Data is represented as mean $\pm$ s.d. of three technical replicates.



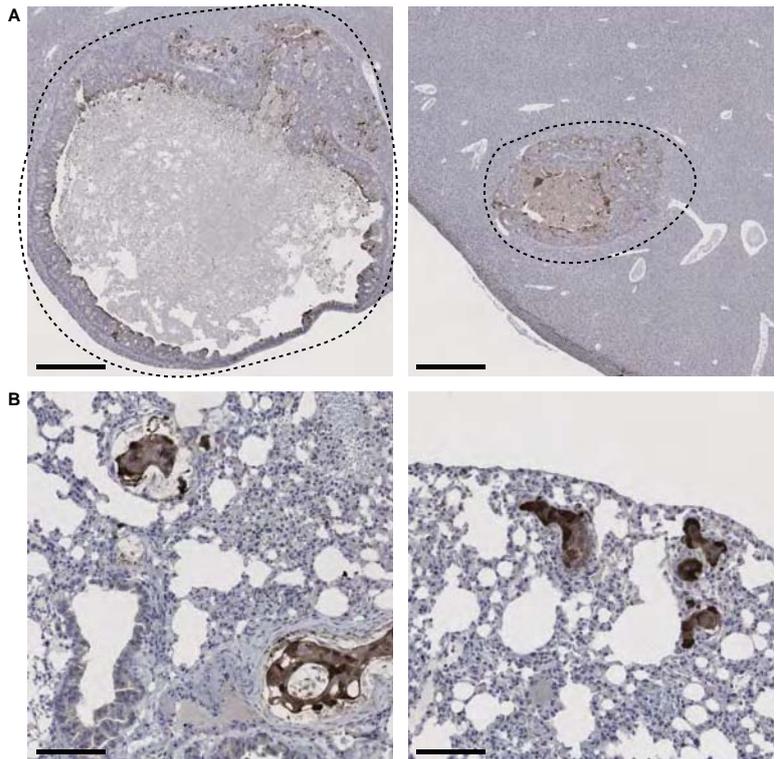
**Figure A.14: PDO7 *EMP1*-iCaspase9-TOM clone screening.** 5' and 3' integration PCRs for *iCaspase9*-TOM knock-in (5' correct band: 1235 bp / 3' correct band: 996 bp) and WT band detection (1927 bp) of PDO7 *EMP1*-iCasp-TOM clones.



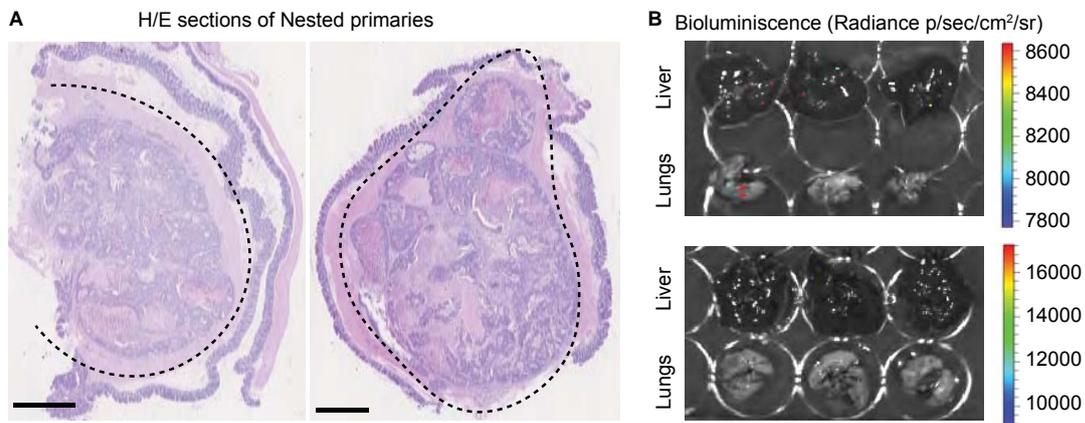
**Figure A.15: PDO7 EMP1-iCT#2 xenograft characterization.** **A.** FACS analysis of EPCAM+ cells from disaggregated EMP1-iCT#2 PDX. **B.** qPCR reporting mRNA expression levels of TOM+ vs TOM- sorted populations. Data is represented as mean $\pm$ s.d. of 3 technical replicates. **C.** Organoid formation efficiency of TOM-High, -Low and -Negative FACS-purified cells from EMP1-iCT#2 derived xenograft, and representative images of each condition. 1500 cells were plated per well, and each experiment assessed the growth of 4 wells per condition. Results are represented as mean $\pm$ sd of 3 experiments. Scale bars indicate 1mm. **D.** Representative FACS profile of the sorted cells 15 days after *in vitro* growth.



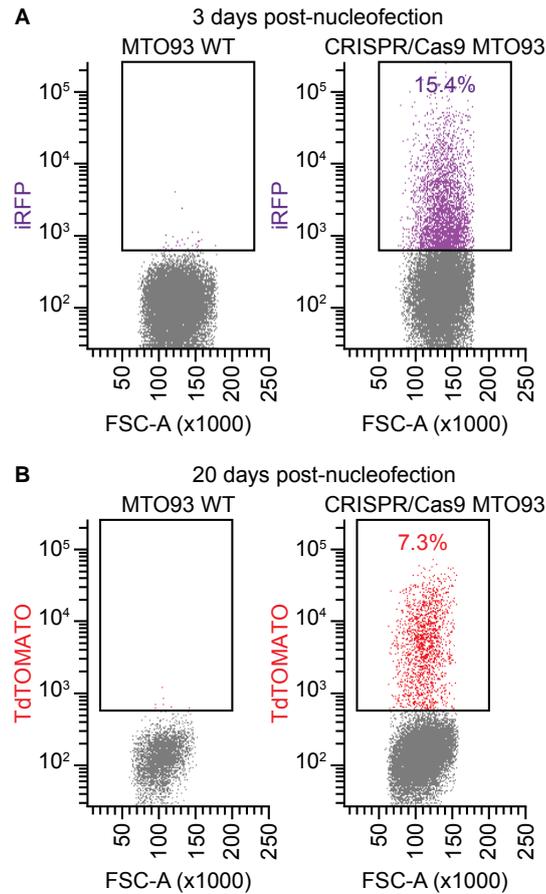
**Figure A.16: PDO7 EMP1-iCT#2 xenografts respond mildly to acute CID treatment.** **A.** Experimental setup of the acute CID treatment in subcutaneous xenografts. **B.** Flow cytometry representation of the TOM levels in DAPI-EPCAM<sup>+</sup> cells from non-treated (NT) and dimerized (CID) disaggregated EMP1-iCT#2 xenografts. **C.** Mean TOM fluorescent intensity in DAPI-EPCAM<sup>+</sup> cells. Results show mean±s.d. of 2 xenografts per condition. **D.** Representative immunofluorescence against TOM in paraffin sections of NT and CID xenografts. Scale bar indicates 100µm. **E.** Relative mRNA expression by qrtPCR of EPCAM<sup>+</sup>DAPI- sorted cells from NT vs CID group. *EMP1*, as well as the invasive markers (*LAMC2*, *L1CAM*, chemokines) are reduced with treatment. Data is represented as mean±s.d. of three technical replicates.



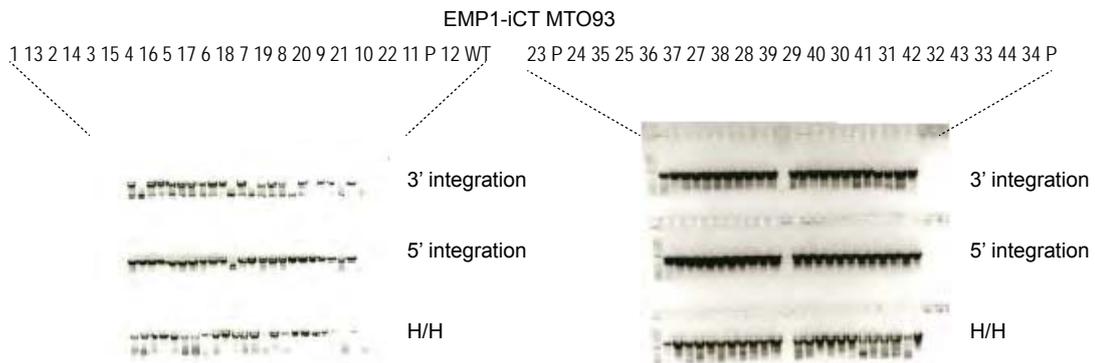
**Figure A.17: PDO7 metastatic nodules express EMP1.** A. TOM IHC in liver metastatic nodules derived from PDO7 EMP1-iCT#2 primary tumors. Few cells retain *EMP1* expression. Dashed lines encircle the nodules. Scale bar indicates 500 $\mu$ m. B. TOM immunostaining of lung micrometastasis, highly positive for *EMP1*. Dashed lines encircle the nodules. Scale bar indicates 100 $\mu$ m.



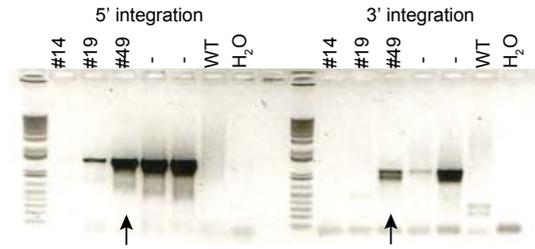
**Figure A.18: Nesting technique allows faster primary growth but does not increase metastatic potential.** A. Representative H/E IHC of primary tumors derived from nested pieces 60 days after surgery. The glandular structure resembles that of subcutaneous xenografts and of orthotopic primaries grown from *in vitro* injected cells. Dashed lines delimit the tumor mass. Scale bars indicate 1mm. B. *Ex-vivo* bioluminescent activity in livers and lungs of animals injected via nesting with PDO7 tumor pieces. Virtually no signal is detected in any of the samples.



**Figure A.19: EMP1-iCT CRISPR/Cas9 modification of mouse tumor organoids.** **A.** Flow cytometry purification of iRFP+ cells 3 days after MTO93 nucleofection with EMP1-iCT+guide plasmids. **B.** Long-term FACS-sorting (day 20) of Emp1-TOM+ cells for single cell seeding.



**Figure A.20: MTO93 EMP1-iCT single cell clone PCR screening.** 5' and 3' integration PCRs for EMP1-iCasp9-TOM modification (5' correct band: 1338 bp / 3' correct band: 1123 bp) and WT band detection (2159 bp) in MTO93 EMP1-iCT clones.



**Figure A.21: MTO93 EMP1-iCT clones LGR5-EGFP pool knock-in PCR assessment.** 5' and 3' integration PCRs for LGR5-ires-EGFP modification (5' correct band: 1332 bp / 3' correct band: 1048 bp). Correctly integrated clone #49 is indicated with arrows.

### A.3 PUBLISHED ARTICLES

*Article* A genome editing approach to study cancer stem cells in human tumor.

*Authors* Carme Cortina, Gemma Turon, Diana Stork, Xavier Hernando-Momblona, Marta Sevillano, Mònica Aguilera, Sébastien Tosi, Anna Merlos-Suárez, Camille Stephan-Otto Attolini, Elena Sancho and Eduard Batlle.

*Journal* EMBO Molecular Medicine.

*Type* Research article.

*Stage* Published.

*Citation* Cortina, Turon et al. (2017).

*Note* The article and the supplementary information are available at <https://doi.org/10.15252/emmm.201707550>.

#### ABSTRACT

The analysis of stem cell hierarchies in human cancers has been hampered by the impossibility of identifying or tracking tumor cell populations in an intact environment. To overcome this limitation, we devised a strategy based on editing the genomes of patient-derived tumor organoids using CRISPR/Cas9 technology to integrate reporter cassettes at desired marker genes. As proof of concept, we engineered human colorectal cancer (CRC) organoids that carry EGFP and lineage-tracing cassettes knocked in the LGR5 locus. Analysis of LGR5-EGFP<sup>+</sup> cells isolated from organoid-derived xenografts demonstrated that these cells express a gene program similar to that of normal intestinal stem cells and that they propagate the disease to recipient mice very efficiently. Lineage-tracing experiments showed that LGR5<sup>+</sup> CRC cells self-renew and generate progeny over long time periods that undergo differentiation toward mucosecreting- and absorptive-like phenotypes. These genetic experiments confirm that human CRCs adopt a hierarchical organization reminiscent of that of the normal colonic epithelium. The strategy described herein may have broad applications to study cell heterogeneity in human tumors.

---



# A genome editing approach to study cancer stem cells in human tumors

Carme Cortina<sup>1,†</sup>, Gemma Turon<sup>1,†</sup>, Diana Stork<sup>1</sup>, Xavier Hernando-Momblona<sup>1</sup>, Marta Sevillano<sup>1</sup>, Mònica Aguilera<sup>1</sup>, Sébastien Tosi<sup>1</sup>, Anna Merlos-Suárez<sup>1</sup>, Camille Stephan-Otto Attolini<sup>1</sup>, Elena Sancho<sup>1</sup> & Eduard Batlle<sup>1,2,\*</sup>

## Abstract

The analysis of stem cell hierarchies in human cancers has been hampered by the impossibility of identifying or tracking tumor cell populations in an intact environment. To overcome this limitation, we devised a strategy based on editing the genomes of patient-derived tumor organoids using CRISPR/Cas9 technology to integrate reporter cassettes at desired marker genes. As proof of concept, we engineered human colorectal cancer (CRC) organoids that carry EGFP and lineage-tracing cassettes knocked in the LGR5 locus. Analysis of LGR5-EGFP<sup>+</sup> cells isolated from organoid-derived xenografts demonstrated that these cells express a gene program similar to that of normal intestinal stem cells and that they propagate the disease to recipient mice very efficiently. Lineage-tracing experiments showed that LGR5<sup>+</sup> CRC cells self-renew and generate progeny over long time periods that undergo differentiation toward mucosecreting- and absorptive-like phenotypes. These genetic experiments confirm that human CRCs adopt a hierarchical organization reminiscent of that of the normal colonic epithelium. The strategy described herein may have broad applications to study cell heterogeneity in human tumors.

**Keywords** cancer stem cells; colorectal cancer; CRISPR/Cas9; LGR5

**Subject Categories** Cancer; Digestive System; Stem Cells

**DOI** 10.15252/emmm.201707550 | Received 9 January 2017 | Revised 7 April 2017 | Accepted 11 April 2017

2007; Vermeulen *et al.*, 2008, 2010; Merlos-Suarez *et al.*, 2011). These findings have led to the notion that CRCs retain a hierarchical organization reminiscent of that of the normal intestinal mucosa, with only cancer stem cells being capable of self-renewal and of sustaining long-term tumor growth (Zeuner *et al.*, 2014). To a large extent, this model has emerged from experiments of tumor cell transplantation. Typically, putative stem and non-stem cell populations are isolated from patient samples using combinations of surface markers, and then, each cell population is inoculated into immunodeficient mice. The capacity to generate xenografts and to reproduce some of the traits of the tumor of origin are used as read-outs of stemness. These assays, however, only provide a snapshot of the state of the cells in the moment they were isolated. It is also unclear to what extent experimental manipulations influence the tumor-initiating capacity of purified cells (Clevers, 2011). Furthermore, the requirement of antibodies against surface markers to isolate tumor cells from patient samples imposes limitations to explore the diversity of cell phenotypes within cancers. Alternatively, the existence of tumor stem cells has been confirmed in mouse adenomas through genetic fate-mapping experiments (Schepers *et al.*, 2012; Kozar *et al.*, 2013). Yet, these lesions are benign and contain few mutations compared to human CRCs. To overcome these restrictions, we combined CRC patient-derived organoids (PDOs) with CRISPR/Cas9 technology to label defined tumor cell populations and perform fate-mapping experiments *in vitro* and *in vivo*.

## Introduction

Most cancers are amalgams of phenotypically distinct tumor cell populations, which display marked differences in their behaviors and fates. In colorectal cancer (CRC), a subpopulation of cells with elevated tumorigenic potential expresses a gene program similar to that of intestinal stem cells (ISCs). These ISC-like tumor cells give rise to differentiated-like progeny, which is poorly tumorigenic (Dalerba *et al.*, 2007, 2011; O'Brien *et al.*, 2007; Ricci-Vitiani *et al.*,

## Results

### Generation of LGR5-EGFP knock-in human CRC organoids

The expression of the Leucine-rich repeat-containing G-protein-coupled receptor 5 (LGR5) marks adult ISCs in mice and humans (Barker *et al.*, 2007; Jung *et al.*, 2011). Knock-in mice engineered to carry EGFP and CreERT2 cassettes integrated into the LGR5 locus have been instrumental to visualize and track ISCs in the healthy mucosa and in tumors (Barker *et al.*, 2007, 2009; Schepers *et al.*,

<sup>1</sup> Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

<sup>2</sup> Institut Catalana de Recerca i Estudis Avançats (ICREA) and CIBER-Onc, Barcelona, Spain

\*Corresponding author. Tel: +34 934039008; E-mail: eduard.batlle@irbbarcelona.org

<sup>†</sup>These authors contributed equally to this work

2012). In contrast, the analysis of LGR5<sup>+</sup> cell populations in human cancers has been hampered by the lack of good commercial antibodies that recognize this protein at the cell surface. We thus designed a strategy based on CRISPR/Cas9-mediated homologous recombination to mark LGR5<sup>+</sup> cells in human CRCs. We made use of CRC PDOs, which are good surrogates of the disease *in vitro* and *in vivo* (Calon *et al*, 2015; van de Wetering *et al*, 2015). For these experiments, we initially selected a PDO derived from a stage IV CRC that displayed a prototypical combination of genetic alterations in major driver pathways including activation of the WNT pathway by APC loss of function, activation of EGFR signaling by KRAS G13D mutations, and loss of TGF-beta-mediated tumor suppression by inactivating mutations in SMAD4 (PDO#7 in Appendix Table S1). The targeting strategy is summarized in Fig 1A and detailed in the Materials and Methods section. In brief, we designed Cas9 guide RNAs complementary to sequences overlapping the stop codon of the LGR5 locus and generated a donor vector that contained LGR5 homology arms flanking an EGFP reporter cassette positioned immediately upstream of the stop codon. We added a LF2A self-cleavage peptide (de Felipe *et al*, 2010) fused to EGFP so that LGR5-EGFP locus was expressed as a single mRNA, whereas the resulting polypeptide was cleaved in the two encoded proteins, LGR5 and EGFP (Fig 1A). Next, we nucleofected organoid cells with the donor vector together with a guide-RNA-Cas9 encoding plasmid in a 3:1 proportion, and 48 h after, we sorted cells that had incorporated the Cas9 vector (IRFP<sup>+</sup> cells). About 1 in 11 IRFP<sup>+</sup> cells expressed EGFP after 20 days in culture (Fig 1B). Subsequently, we generated single cell-derived organoid cultures and assessed integration of the EGFP reporter cassette by PCR (examples in Appendix Fig S1A and B) and Southern blot (examples in Appendix Fig S1C and D). These analyses showed that 41.7% of the clones had correctly integrated the EGFP reporter in the LGR5 locus (Appendix Table S2). Equivalent LGR5-EGFP knock-in experiments in a PDO grown from a different patient sample (PDO#6) (Fig EV1A) rendered a frequency of correct integrations of 84.6% (Appendix Table S2). In these single cell-derived knock-in PDO cultures, every organoid was composed by an admixture of cells expressing distinct EGFP levels (Figs 1C and D, and EV1B and C). LGR5-EGFP-hi cells isolated by FACS expressed highest LGR5 mRNA levels confirming that EGFP reported endogenous LGR5 expression (Figs 1E and EV1D). Staining with KRT20 or MUC2 antibodies revealed complementary expression patterns of these differentiation markers with EGFP implying that LGR5<sup>+</sup> CRC cells generated differentiated progeny *in vitro* (Fig 1C).

To demonstrate the broad applicability of this approach, we also engineered PDO#7 expressing TagRFP2 fused to endogenous KI67 protein (Fig 1F). The KI67 antigen is a nuclear protein which is expressed in all active phases of the cell cycle (G1, S, G2, and mitosis), but is absent in resting cells (G0) (Scholzen & Gerdes, 2000). In a previous study, knock-in mice expressing a KI67-RFP fusion protein were used to isolate cycling (KI67-RFP<sup>+</sup>) and non-cycling differentiated cells (KI67-RFP<sup>-</sup>) from the intestinal epithelium (Basak *et al*, 2014). Targeting efficiency for this knock-in construct in human CRC organoids was similar to that observed for LGR5-EGFP knock-in organoids (Appendix Table S2). KI67-TagRFP2 was visualized in the nucleus of organoid cells (Fig 1G) and of xenografts derived from these organoids (Fig 1H). Cell cycle profiling of epithelial cells isolated using FACS from dissociated xenografts (Fig 1I) demonstrated that TagRFP2<sup>+</sup> cells were distributed in all cell cycle phases, whereas the TagRFP2<sup>-</sup> population was largely enriched in cells at the G1/G0 phase (Fig 1J and K).

#### Characterization of human LGR5<sup>+</sup> CRC cells *in vivo*

To study LGR5<sup>+</sup> cells *in vivo*, we initially used an LGR5-EGFP expressing organoid (clone #1) that by exome sequencing revealed few acquired mutations compared to the parental population, none of which affected known cancer driver genes (Appendix Table S3). Clone #1 neither contained mutations in the top off-target sequences predicted through bioinformatics for the CRISPR guide sequence (Appendix Table S4). The LGR5-EGFP knock-in PDO was inoculated into immunodeficient mice of the NOD/SCID strain. Xenografts displayed a glandular organization and prominent stromal recruitment. LGR5-EGFP expression labeled a substantial proportion of the epithelial component of the tumor yet cells showed a wide range of EGFP levels (Figs 2A and EV2). In contrast, the EGFP<sup>-</sup> compartment overlapped largely with the expression domain of the pan differentiation marker KRT20<sup>+</sup> (Fig 2B). We also observed LGR5<sup>-</sup>/MUC2<sup>+</sup> cells with goblet-like morphology intermingled between LGR5<sup>+</sup> and LGR5<sup>-</sup> compartments throughout the tumor (Fig 2C). These tumor cell populations displayed equivalent distributions in xenografts produced by a different single cell-derived clone from PDO#7 (Fig EV3A–C) or by PDO#6 (Fig EV1E–G). Overall, this cellular organization is reminiscent of that of the normal intestinal epithelium as previously proposed by several laboratories (Dalerba *et al*, 2011; Merlos-Suarez *et al*, 2011).

In flow cytometry analysis, LGR5-EGFP-high cells represented about 3–4% of the epithelial component (EPCAM<sup>+</sup>) of dissociated

**Figure 1. LGR5-EGFP and KI67-TagRFP2 knock-in PDOs.**

- A Design of LGR5-EGFP donor and CRISPR/Cas9 sgRNA vectors. Blue circle represents the CRISPR/Cas9 protein complex and the yellow box underneath the guide RNA.
- B Flow cytometry profiles at day 20 post-nucleofection.
- C Immunofluorescence for DAPI, EGFP, and KRT20 or MUC2 in *in vitro* cultured PDO#7-LGR5-EGFP#1. Scale bars indicate 100  $\mu$ m.
- D FACS profiles showing EGFP-high (green), -low (blue), and -negative (gray) cells in PDO#7-LGR5-EGFP#1 and #2 organoids.
- E Relative mRNA expression level by real-time qPCR in cells expressing distinct levels of EGFP isolated from PDO#7-LGR5-EGFP#1 and #2 knock-in organoids. Values show mean  $\pm$  s.d. of three measurements.
- F Design of KI67-TagRFP2 donor and CRISPR/Cas9 sgRNA vectors. Blue circle represents the CRISPR/Cas9 protein complex and the yellow box underneath the guide RNA.
- G Images of PDO#7-KI67-TagRFP2#1 organoids. Scale bars indicate 100  $\mu$ m.
- H PDO#7-KI67-TagRFP2#1 xenograft. TagRFP2 co-localizes with DAPI nuclear staining. Scale bars indicate 25  $\mu$ m.
- I Flow cytometry analysis of EPCAM<sup>+</sup>/DAPI<sup>-</sup> cell population of PDO#7-LGR5-EGFP/KI67-TagRFP2#1 from disaggregated xenografts.
- J Cell cycle analysis of KI67-TagRFP2-positive and KI67-TagRFP2-negative cells from PDO#7-LGR5-EGFP/KI67-TagRFP2#1 disaggregated xenografts. X-axis shows DNA content and y-axis EdU incorporation.
- K Quantification of the frequencies of KI67<sup>+</sup> versus KI67<sup>-</sup> cells found in each cell cycle phase.

xenografts (Fig 2D). We isolated EGFP-high and EGFP-low/negative cells by FACS (for simplicity we termed them EGFP<sup>+</sup> and EGFP<sup>-</sup>) and analyzed their global gene expression profiles. LGR5-EGFP<sup>+</sup> cells expressed over 10-fold higher levels of ISC marker genes LGR5 and SMOC2 than EGFP<sup>-</sup> cells (Fig 2E). LGR5-EGFP<sup>-</sup> cells expressed genes that characterize differentiated cells of the intestinal epithelium such

as EFN2, KRT20 or MUC2 (Fig 2E). We validated these results using a second LGR5-EGFP knock-in clone derived from PDO#7 (Fig EV3D and E). Microarray profiling followed by gene set enrichment analysis (GSEA) confirmed that mouse and human intestinal stem cell gene expression signatures were upregulated in LGR5-EGFP<sup>+</sup>, whereas the differentiation program of colon epithelium was enriched in

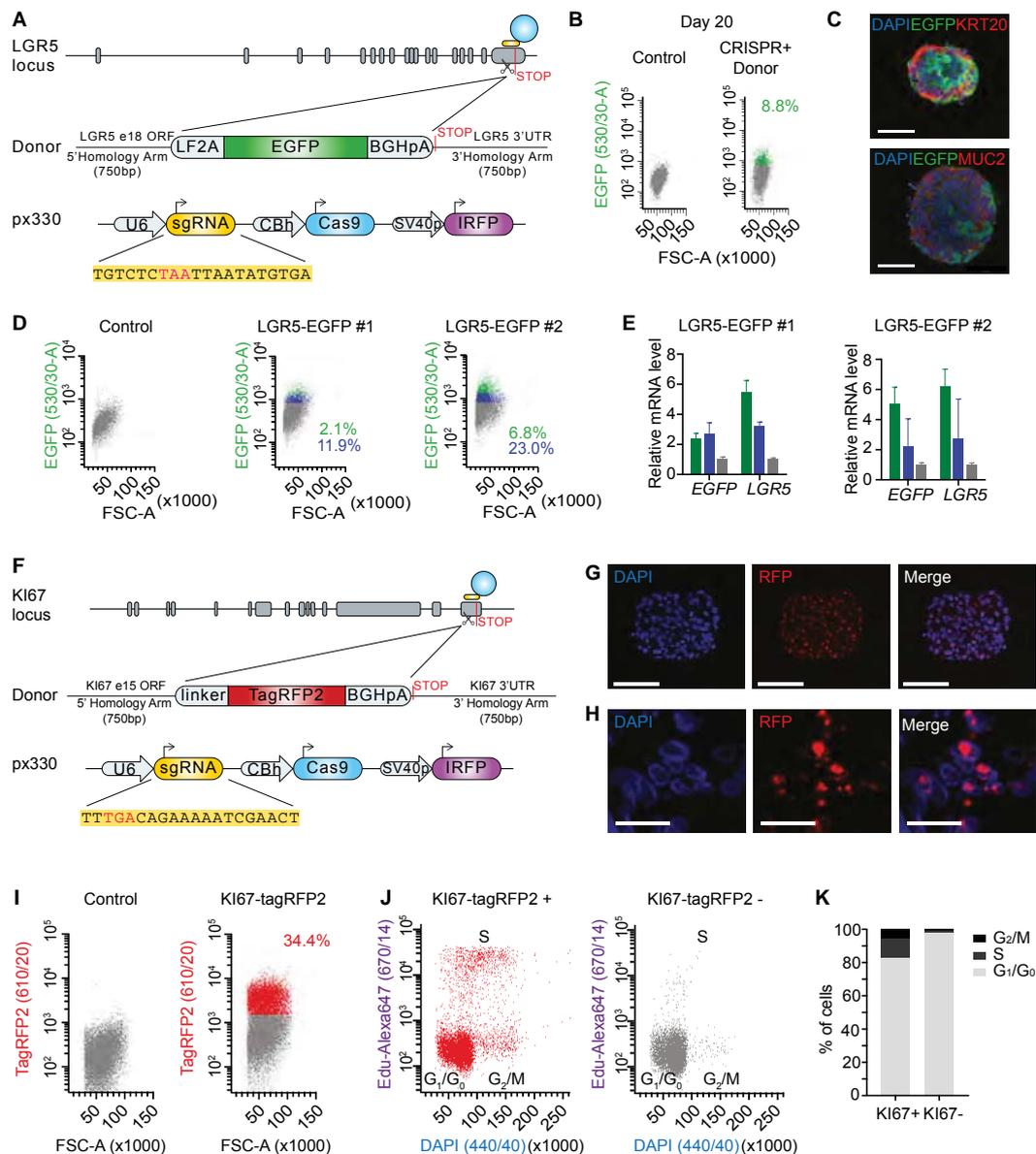


Figure 1.

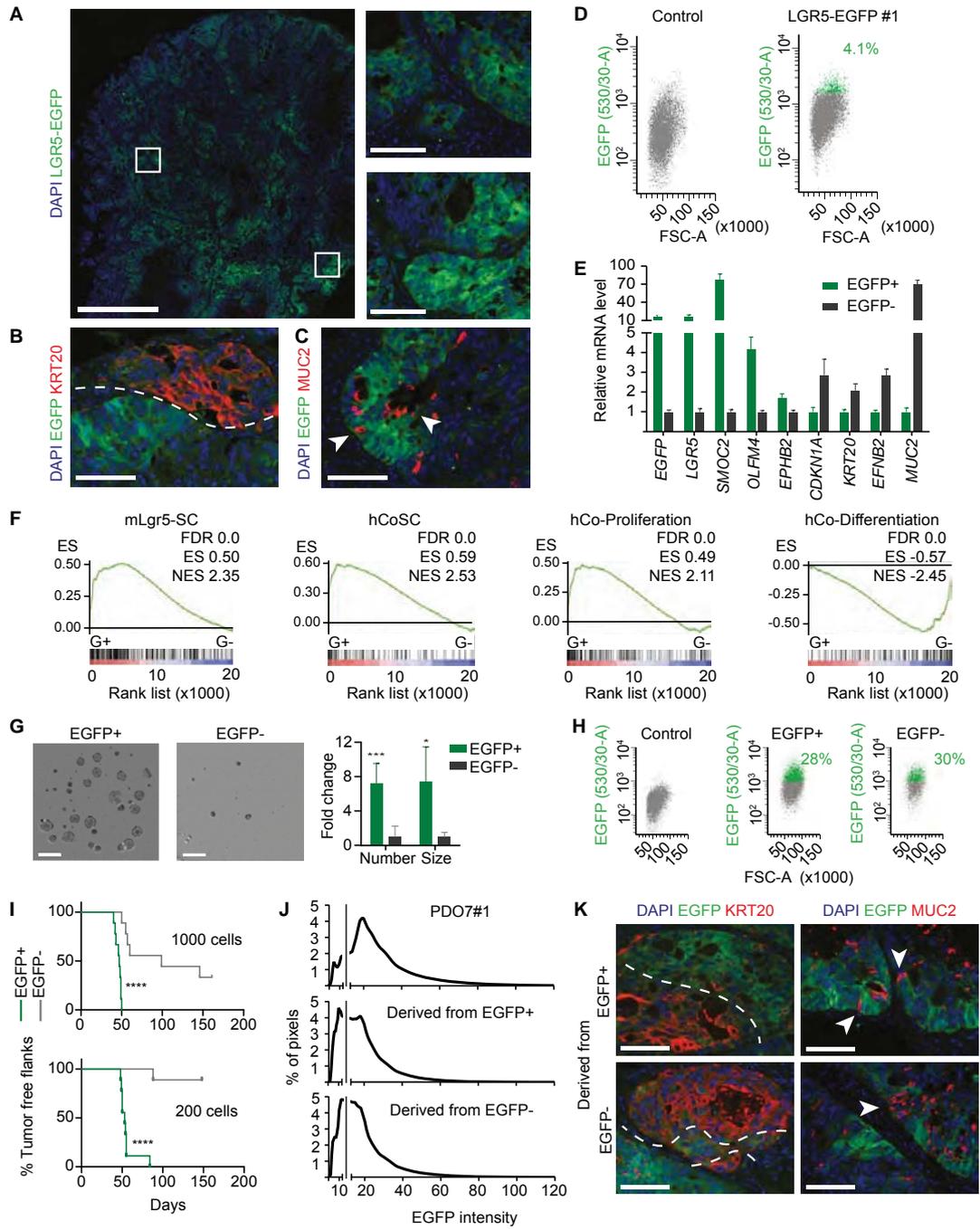


Figure 2.

**Figure 2. Characterization of human LGR5<sup>+</sup> CRC cells *in vivo*.**

- A Representative images of EGFP by immunofluorescence on a section of PDO#7-LGR5-EGFP#1-derived subcutaneous xenograft. White squares indicate the position of the insets. Scale bars indicate 1 mm for the whole xenograft and 100  $\mu$ m for the insets.
- B Dual immunofluorescence on paraffin sections for KRT20 and LGR5-EGFP showing complementary expression domains. Dashed line delimits expression domain in adjacent glands. Scale bar indicates 100  $\mu$ m.
- C Dual immunofluorescence on paraffin sections of clone #1 for MUC2 and LGR5-EGFP. White arrows point to LGR5-EGFP<sup>-</sup>/MUC2<sup>+</sup> cells. Scale bar indicates 100  $\mu$ m.
- D Representative FACS profiles of EGFP<sup>+</sup> and EGFP<sup>-</sup> in EPCAM<sup>+</sup>/DAPI<sup>-</sup> subpopulation from disaggregated xenografts.
- E Relative mRNA expression level of intestinal stem and differentiation genes for the sorted EGFP<sup>+</sup> and EGFP<sup>-</sup> populations. Values show mean  $\pm$  standard deviation (s.d.) of three measurements.
- F GSEA comparing the expression of signatures of mouse LGR5<sup>+</sup> cells (MmLgr5-SC), human colon stem cells (hCoSCs), differentiated cells (hCo differentiation), or proliferative crypt cells (Jung et al, 2011) in profiled LGR5-EGFP<sup>+</sup> versus LGR5-EGFP<sup>-</sup> cells.
- G Representative images and quantification of organoid formation assays from cells purified from PDO#7-LGR5-EGFP#1-derived subcutaneous xenograft (n = 4 wells per condition). Data is represented as mean  $\pm$  s.d. Scale bars indicate 1 mm.
- H Representative flow cytometry analysis of 15-day grown organoids from the EGFP<sup>+</sup> and EGFP<sup>-</sup> sorted populations.
- I *In vivo* tumor initiation capacity of 1,000 and 200 sorted cells from PDO#7-LGR5-EGFP#1-derived subcutaneous xenografts. Graphs show Kaplan–Meier plots (n = 9 xenografts).
- J Distribution of the EGFP staining intensity in PDO#7#1 and in xenografts derived from EGFP<sup>+</sup> and EGFP<sup>-</sup> cells. Gray line indicates background fluorescence levels.
- K Dual immunofluorescence for KRT20/EGFP and MUC2/EGFP on paraffin sections of xenografts generated by EGFP<sup>+</sup> and EGFP<sup>-</sup> sorted populations respectively. Dashed lines mark stem-like and differentiated-like compartments. White arrows point to secretory cells intermingled in the LGR5<sup>-</sup> compartment. Scale bars indicate 100  $\mu$ m.

Data information: Differences in organoid formation assay were assessed with Student's t-test and in tumor initiation assay by log-rank (Mantel–Cox) test:

\*P-value < 0.05, \*\*\*P-value < 0.005, \*\*\*\*P-value < 0.0001. The exact P-values are specified in Appendix Table S5.

LGR5-EGFP<sup>-</sup> CRC cells (Fig 2F). We next assessed the clonogenic potential of LGR5-EGFP CRC cells. LGR5-EGFP<sup>+</sup> cells purified from xenografts displayed several fold higher organoid forming capacity than LGR5-EGFP<sup>-</sup> cells (Figs 2G and EV3F). Organoids generated by LGR5-EGFP<sup>+</sup> cells contained both EGFP<sup>+</sup> and EGFP<sup>-</sup> tumor cells in a proportion similar to that of the PDO of origin (Figs 2H and EV3G). We obtained similar results using LGR5-EGFP knock-in cells generated from PDO#6 (Fig EV1H–K).

Finally, to assess the capacity of these tumor cell populations to propagate the disease to mice, we inoculated 200 or 1,000 LGR5-EGFP<sup>+</sup> or LGR5-EGFP<sup>-</sup> epithelial tumor cells isolated from xenografts into secondary hosts. These experiments showed that the EGFP<sup>+</sup> cell population was largely enriched in tumor-initiating cells compared to their differentiated EGFP<sup>-</sup> counterparts (Figs 2I and EV3H). Tumors generated by LGR5-EGFP<sup>+</sup> cells were populated with stem-like (EGFP<sup>+</sup>/KRT20<sup>-</sup>) and differentiated-like (EGFP<sup>-</sup>/KRT20<sup>+</sup>) tumor cells in similar proportions than the primary xenografts from which they were purified (Fig 2K) thus implying that LGR5-expressing CRC cells undergo self-renewal and differentiation during tumor expansion. Of note, xenografts generated by LGR5-EGFP<sup>-</sup> cell population were also formed by EGFP<sup>+</sup> and EGFP<sup>-</sup> cells with equivalent intensities and proportions to those observed in xenografts derived from LGR5-EGFP<sup>+</sup> cells (Fig 2J). The expression pattern of the differentiation markers KRT20 and MUC2 was also similar in xenografts arising from the two cell populations (Fig 2K).

#### Lineage tracing of human LGR5<sup>+</sup> CRC cells

As discussed in the introduction, currently it is not possible to perform cell fate-mapping experiments in human cancers similar to those performed in mouse models. To overcome this limitation, we used CRISPR/Cas9 to engineer PDOs containing a lineage-tracing system. We first introduced a Cre recombinase-inducible reporter into the neutral AAVS1 locus (Fig 3A). This reporter consisted of a constitutive Ubiquitin C (UBC) promoter driving the expression of blue fluorescent protein mTagBFP2. This cassette was flanked by LoxP sites so that expression of a downstream tdTomato (TOM) reporter remains blocked until the mTagBFP2 cassette is excised by

Cre recombinase activity. Following the approach described for LGR5-EGFP targeting, we selected long-term mTagBFP2-expressing cells after nucleofection and expanded single cell-derived organoids. Subsequently, we generated a second genomic edition consisting in an LF2A-CreERT2 cassette recombined upstream of the LGR5 stop codon (Fig 3B). The frequencies of correct integrations for these cassettes were 47.8 and 1.78%, respectively (Appendix Table S2). We further confirmed correct integrations of these constructs by PCR as well as by Sanger sequencing of genomic DNA. To test their functionality, we induced PDOs with 4-hydroxytamoxifen (4-OHT) *in vitro*, which demonstrated conversion of mTagBFP2<sup>+</sup> cells into TOM<sup>+</sup> cells (Figs 3C and D, and EV4A). We also tested the utility of these constructs *in vivo* by inoculating double-edited PDOs in mice. Analysis of xenografts 96 h after induction with tamoxifen revealed the appearance of a TOM<sup>+</sup> side population, which retained expression of LGR5 mRNA (Fig EV4B and C) supporting tracing from the LGR5<sup>+</sup> cell population. In contrast, we did not observe TOM<sup>+</sup> cells in xenografts growing in untreated mice. Based on a frequency of about 2–4% LGR5-EGFP<sup>hi</sup> cells in xenografts (Figs 2D and EV3D), and on the number of TOM<sup>+</sup> cells arising 96 h post-tamoxifen administration (Fig EV4B), we roughly estimated that recombination occurred in 1 in every 10–20 LGR5-EGFP<sup>+</sup> cells.

Next, we mapped the fate of LGR5<sup>+</sup> CRC cells over an extended period of time. The experimental setup is described in Fig 3E. In brief, a cohort of mice bearing edited PDO-derived xenografts were given tamoxifen once tumors were palpable. Mice were sacrificed and tumors analyzed at the indicated time points over 28 days. After this period, tumor pieces were transplanted into secondary recipient mice and xenografts were analyzed for further 4 weeks. We already observed the emergence of TOM<sup>+</sup> individual cells scattered throughout the tumor glands 96 h after tamoxifen induction (Fig 3F). About 75% of these marks corresponded to isolated cells and the rest to two cell clones (Fig 3G) implying that these experimental conditions enable tracing from individual tumor cells. Quantification of clone size revealed heterogeneity in the growth dynamics of LGR5<sup>+</sup> CRC cells. Whereas some clones expanded steadily over time, a substantial proportion of LGR5<sup>+</sup> divided slowly or even remained as individual entities over extended periods (56 days

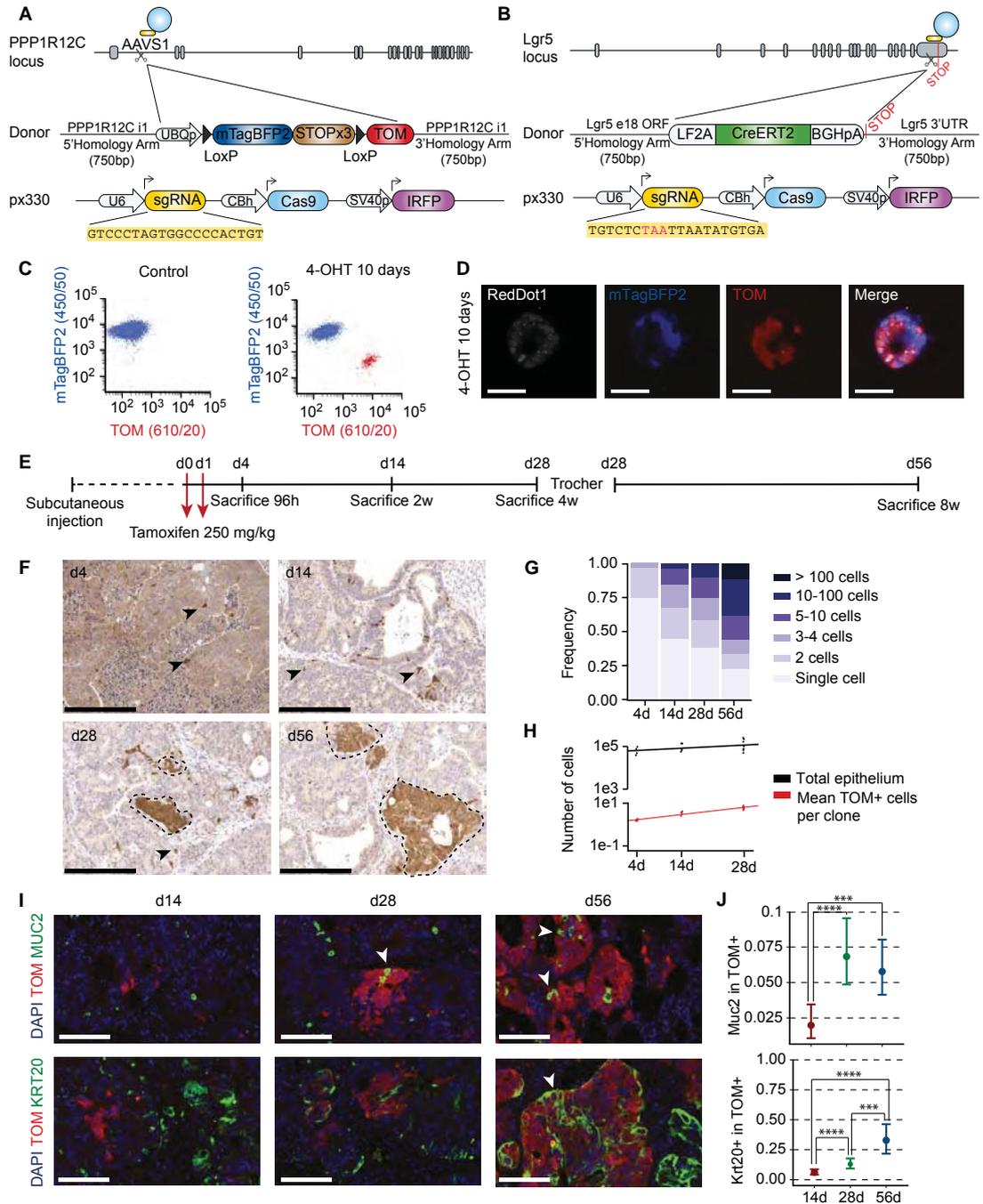


Figure 3.

**Figure 3. Lineage tracing of LGR5<sup>+</sup> CRCs in human colorectal xenografts.**

- A Design of the donor vector containing lineage-tracing cassette and AAVS1 homology arms.  
 B Design of LGR5-CreERT2 donor and CRISPR/Cas9 sgRNA vectors.  
 C Flow cytometry analysis of double knock-in PDO#7 carrying AAVS1-LSL-TOM and LGR5-CreERT2 cassettes. Organoids were treated *in vitro* with 1  $\mu$ M 4-hydroxytamoxifen (4-OHT). About 3.6% of the cells recombined the stop cassette (i.e., expressed low levels of mTagBFP2) and gained expression of TOM.  
 D Confocal imaging of double knock-in organoids 10 days after *in vitro* 1  $\mu$ M 4-OHT addition. Scale bars indicate 50  $\mu$ m. Note that recombined organoids switch mTagBFP2 to TOM expression.  
 E Experimental setup used for lineage-tracing experiments.  
 F Representative immunohistochemistry using anti-Tomato antibodies on paraffin sections of the four time points after tamoxifen treatment. Arrowheads point to single and two cell clones. Dashed lines delimit large clones. Scale bars indicate 250  $\mu$ m.  
 G Clone size frequency per time point according to number of cells. Number of clones quantified was 878 for day 4, 2,424 for day 14, 6,940 for day 28, and 6,940 for day 56.  
 H Correlation of number of epithelial cells per xenograft and number of cells per clone over time ( $n = 4$  xenografts for 4 days time point,  $n = 5$  xenografts for 14 days time point,  $n = 8$  xenografts for 28 days time point,  $n = 8$  xenografts for 56 days time point).  
 I Expression domains of TOM and differentiation markers MUC2 and KRT20. White arrowheads indicate double-positive cells. Scale bars indicate 100  $\mu$ m.  
 J Quantification of the number of MUC2<sup>+</sup> and KRT20<sup>+</sup> cells within TOM<sup>+</sup> clones at each time point. Data is represented as the 95% confidence intervals of the measurements. Number of clones assessed was 872 (4 days), 372 (day 14), and 69 (day 28) for KRT20 and 387 (day 4), 611 (day 14), and 130 (day 28) for MUC2. The  $P$ -value was calculated using a generalized linear model with binomial response. \*\*\* $P$ -value < 0.005, \*\*\*\* $P$ -value < 0.0001. The exact  $P$ -values are specified in Appendix Table S5.

in these experiments) (Fig 3F and G). 3D reconstruction from multiple serial tissue sections confirmed the existence of many isolated 1–4 cell clones at day 28 after tamoxifen induction (Fig EV5A–D and Movie EV1). Quantification of clone number over time showed that the number of cells generated by LGR5<sup>+</sup> cells was directly proportional to the expansion kinetics of the tumor epithelial compartment (Fig 3H). The scaling pattern of the LGR5<sup>+</sup> cell output is compatible with the hypothesis that tumor growth is largely the result of LGR5<sup>+</sup> cell activity. Of note, we observed few MUC2<sup>+</sup> and KRT20<sup>+</sup> cells in clones during the first 2 weeks of tracing, whereas the frequency of differentiated cells increased after this period (Fig 3I and J). Therefore, in CRC, the progeny of LGR5<sup>+</sup> cells remains undifferentiated during extended periods of time.

#### Marking of quiescent LGR5<sup>+</sup> CRC cells

The observation that a proportion of LGR5<sup>+</sup> cell in lineage-tracing experiments produced few progeny may reflect a quiescent state. Indeed, we found that about half of LGR5<sup>+</sup> cells stained negative for KI67 (Fig 4A and B). To further characterize this cell population, we generated LGR5-EGFP PDOs that expressed TagRFP2 fused to endogenous KI67 protein following the approach described in Fig 1. Analysis of xenografts derived from LGR5-EGFP/KI67-TagRFP2

PDOs confirmed that a large proportion of LGR5-EGFP<sup>+</sup> cells did not express KI67-TagRFP2 (Fig 4C). In independent xenografts and clones, the fraction of LGR5-EGFP<sup>+</sup>/KI67-TagRFP2<sup>-</sup> ranged from 20 to 50%. LGR5-EGFP<sup>+</sup>/KI67-TagRFP2<sup>-</sup> cells purified from xenografts displayed cell cycle profiles that indicated arrest in G1/G0 phase (Fig 4D). Using FACS, we purified KI67-TagRFP2<sup>+</sup> (K<sup>+</sup>) and KI67-TagRFP2<sup>-</sup> (K<sup>-</sup>) cells within the LGR5-EGFP<sup>+</sup> (L<sup>+</sup>) and LGR5-EGFP<sup>-</sup> (L<sup>-</sup>) gates and compared their gene expression profiles. The L<sup>-</sup>/K<sup>-</sup> cell population showed downregulation of proliferation genes, upregulation of the cell cycle inhibitor CDKN1A, and expression of markers of terminal differentiation KRT20 implying that they represent mature differentiated CRC cells (Fig 4E). L<sup>-</sup>/K<sup>+</sup> cells displayed low levels of ISC marker genes and upregulated genes characteristic of early absorptive differentiation such as FABP1 and SI (Fig 4E). By analogy with the normal intestinal epithelium, we hypothesize that this cell population resembles proliferative progenitors undergoing differentiation toward an enterocyte-like phenotype. Our analysis also showed that the L<sup>+</sup>/K<sup>-</sup> cell population was characterized by downregulation of key genes involved in proliferation and cell cycle progression such as KI67, AURKB, FOXM1, and UBE2C (Fig 4E and F) but retained elevated levels of ISC marker genes including LGR5 and SMOC2 (Fig 4E–G). GSEA further demonstrated an overall downregulation of the proliferative genes in L<sup>+</sup>/K<sup>-</sup> cells (Fig 4G). A

**Figure 4. Dual LGR5 and KI67 knock-in PDOs enable separation of quiescent and cycling LGR5<sup>+</sup> CRC cells.**

- A Representative immunofluorescence image of PDO#7-LGR5-EGFP#1 stained with KI67 antibody. White arrowheads point to double-positive cells; yellow arrowheads point to LGR5<sup>+</sup>/KI67<sup>-</sup> cells. Scale bars indicate 100  $\mu$ m.  
 B Quantification of KI67<sup>+</sup> cells within the LGR5<sup>+</sup> and LGR5<sup>-</sup> compartments ( $n = 2,749$  LGR5<sup>+</sup> cells, 1,798 LGR5<sup>-</sup> cells assessed). Data is represented as the 95% confidence intervals of the measurements.  
 C Representative FACS profiles from PDO#7-LGR5-EGFP/KI67-TagRFP2 disaggregated xenografts. Only EPCAM<sup>+</sup>/DAPI<sup>-</sup> cells are shown. The four represented populations are: LGR5-EGFP<sup>-</sup>, KI67-RFP<sup>-</sup> (gray), LGR5-EGFP<sup>-</sup>, KI67-RFP<sup>+</sup> (orange), LGR5-EGFP<sup>+</sup>, KI67-RFP<sup>+</sup> (red) and LGR5-EGFP<sup>+</sup>, KI67-RFP<sup>-</sup> (green).  
 D Cell cycle analysis LGR5-EGFP<sup>+</sup> and KI67-RFP-positive or KI67-RFP-negative sorted populations. 5,363 and 5,398 cells were analyzed in each case.  
 E RT-qPCR analysis of proliferation, stem, and differentiation marker genes in the cell populations defined by EGFP/TagRFP levels. K indicates KI67, and L indicates LGR5. Values show mean  $\pm$  standard deviation (s.d.) of three measurements.  
 F Volcano plot representing gene expression profile of L<sup>+</sup>/K<sup>+</sup> versus L<sup>-</sup>/K<sup>-</sup> purified populations from LGR5-EGFP/KI67-TagRFP2 PDO#7 clone #1. Green dots indicate genes belonging to the human colon stem cell signature, orange dots represent genes belonging to the differentiated cell signature, and blue dots depict genes of the crypt proliferative progenitor signature. Well described genes involved in proliferation are indicated.  $P$ -values and fold changes computed by fitting a linear model with the R package limma.  
 G GSEA comparing LGR5-EGFP<sup>+</sup> cells positive or negative for KI67-RFP for the signatures used in Fig 2 as well as for signatures derived from mouse crypt LGR5-high/KI67-high or LGR5-high/KI67-low (Basak et al, 2014). Note that the only signatures that are differentially expressed between the two populations correspond to proliferative cells derived from either human crypts or from KI67-RFP mice. In contrast, the signature of mouse LGR5-EGFP<sup>+</sup>/KI67-RFP<sup>-</sup> is significantly enriched in LGR5-EGFP<sup>+</sup>/KI67-RFP<sup>-</sup> tumor cells.

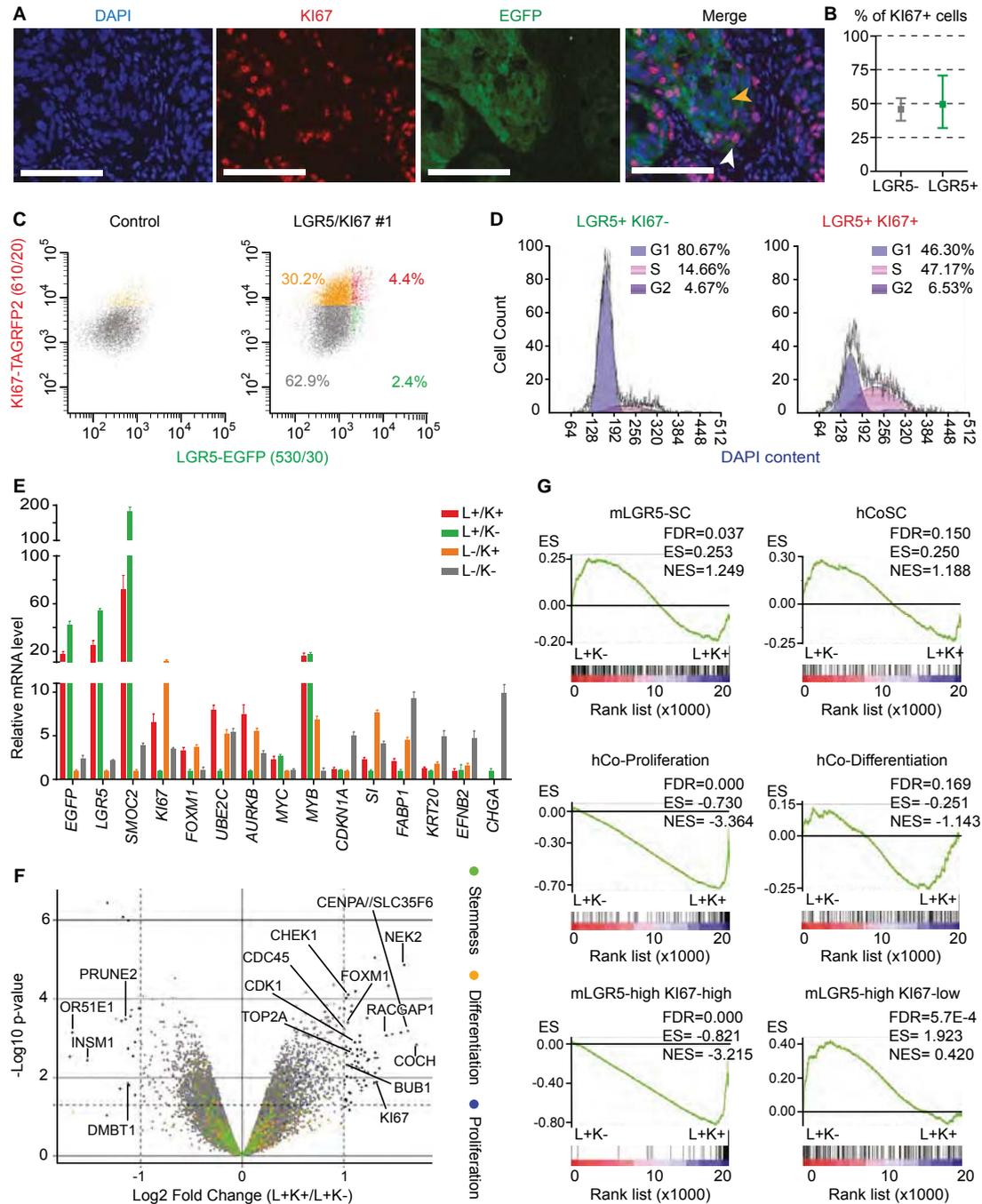


Figure 4.

previous work used KI67-RFP knock-in mice to show that a small subset of LGR5<sup>+</sup> cells in the healthy mucosa downregulate the expression of KI67 (Basak *et al*, 2014). We used transcriptomic datasets from these mice to identify genes up- and downregulated in normal LGR5<sup>+</sup>/KI67<sup>+</sup> and LGR5<sup>+</sup>/KI67<sup>-</sup> crypt cells (Basak *et al*, 2014). Our analyses showed striking enrichment of these gene sets in L<sup>+</sup>/K<sup>+</sup> versus L<sup>+</sup>/K<sup>-</sup> CRC cells (Fig 4G). In contrast, the oncogenes MYC and MYB, the expression of which is driven by the WNT pathway in CRC (van de Wetering *et al*, 2002), remained elevated in quiescent and proliferating LGR5<sup>+</sup> tumor cell populations (Fig 4E).

## Discussion

The combination of organoid and CRISPR/Cas9 technology described herein opens up the study of human tumors through genetic approaches that had only been feasible in animal models. This advance is particularly well suited to analyze phenotypic diversity of cell populations within cancers as it enables labeling and tracing of distinct tumor cells through specific marker genes, which are not necessarily expressed at the cell surface. In contrast, its utility to study genomic heterogeneity is limited, as the current method requires cloning of individual tumor cells to guarantee the accuracy of the genomic insertions. Therefore, tumors generated from edited organoids reflect the behavior of a single stem cell lineage in a genetically homogenous mutational background. To ensure that edited organoids are good surrogates of the parental population, we selected those displaying mutational profiles that overlapped with that of the organoid of origin. Still, although unlikely, we cannot rule out that the few private mutations identified in these monoclonal organoids or other epi-genetic alterations may confer differential properties. Despite these caveats, the possibility of performing cell fate-mapping experiments in human cancers represents a substantial advance. For the first time, this approach enables the analysis of cell lineage relationships in intact tumors and will help address how distinct cell populations contribute to growth, dissemination and resistance to therapy.

Colorectal cancer stem cells had been previously isolated from patient samples using distinct surface markers including CD44, CD133, or EPHB2, which enrich in populations of tumor-initiating cells (O'Brien *et al*, 2007; Dalerba *et al*, 2007; Ricci-Vitiani *et al*, 2007; Merlos-Suarez *et al*, 2011). In normal colonic mucosa, these markers are expressed broadly throughout the stem and transient amplifying compartments (Zeilstra *et al*, 2008; Snippert *et al*, 2009; Jung *et al*, 2011). In contrast, the expression domain of LGR5 is restricted to ISCs (Barker *et al*, 2007) yet the analysis of LGR5-expressing cells in human CRCs had not been possible due to the lack of good reagents. Our work shows that human LGR5<sup>+</sup> CRC cells express the gene program of normal ISCs, are clonogenic *ex vivo*, and display robust tumor-initiating capacity in xenograft assays. We also performed for first time experiments of lineage tracing in human CRC, which demonstrate that LGR5<sup>+</sup> tumor cells produce progeny over long periods of time, which undergo differentiation to distinct lineages. Hence, our work reinforces the notion that despite the accumulation of multiple genetic alterations, human CRCs are governed by a cell hierarchy reminiscent of that present in the normal intestinal epithelium. Our observations revealed two other interesting aspects. First, the kinetics of differentiation of tumor cells in CRC appears to be a relatively slow process compared to the

normal epithelium, where the progeny of LGR5<sup>+</sup> ISCs undergoes differentiation 2–3 days after they leave the crypt base (Clevers, 2013). In contrast, clones produced by LGR5<sup>+</sup> CRC cells were largely devoid of differentiated cells, which only started to accumulate after 2 weeks approximately. This delayed differentiation fits in well with the observation that LGR5<sup>+</sup> and KRT20<sup>+</sup> tumor cells reside in complementary compartments rather than intermingled in the same area and may suggest that distinct tumor niches facilitate stem or differentiation states. Second, whereas the vast majority of normal ISCs remain in a proliferative state (Schepers *et al*, 2012; Basak *et al*, 2014), a substantial proportion of LGR5<sup>+</sup> CRC cells contribute with few progeny according to the lineage-tracing data. This subset of inactive LGR5<sup>+</sup> cells likely represent LGR5<sup>+</sup>/KI67<sup>-</sup> cells identified in double-reporter knock-in PDOs. These data further support previous clonal analysis of CRC using lentiviral marking of patient samples, which revealed the existence of dormant cells that can be reactivated upon passaging or chemotherapeutic treatment (Dieter *et al*, 2011; Kreso *et al*, 2013). Finally, the finding that the progeny of LGR5<sup>+</sup> tumor cells scales with the total number of epithelial cells fits in well with the hypothesis that CRC growth is the result of the activity of multiple LGR5<sup>+</sup> tumor stem cells. Nevertheless, our data does not rule out that LGR5<sup>-</sup> cells could contribute equally to tumor growth. In the normal intestinal epithelium, differentiated cells can opportunistically replace LGR5<sup>+</sup> ISCs through plasticity (van Es *et al*, 2012; Tetteh *et al*, 2016), implying that the ISC phenotype is not hardwired but rather is induced by the niche. Thus, it is likely that LGR5<sup>+</sup> and LGR5<sup>-</sup> tumor phenotypes are also plastic. Our observation that the xenografts generated by LGR5<sup>-</sup> cells display cellular patterns equivalent to those produced by LGR5<sup>+</sup> cells may indicate interconversion of the two cell populations in these transplantation assays, yet confounding effects such as suboptimal isolation of the LGR5-EGFP<sup>-</sup> population could as well explain our results. Proper assessment of cell plasticity will require mapping the fate of LGR5<sup>-</sup> cells in intact tumors through genetic strategies equivalent to those described herein.

## Materials and Methods

### Organoid cultures

PDO#6 and PDO#7 have been previously described (Calon *et al*, 2015). In brief, the tumor sample used to expand PDO#6 was obtained from an individual treated at Hospital de la Santa Creu i Sant Pau, under informed consent and approval of the Tumor Bank Committees according to Spanish ethical regulations. The study followed the guidelines of the Declaration of Helsinki, and patient identity for pathological specimens remained anonymous in the context of this study. Tumor cells were grown as organoids embedded in BME2 (basement membrane extract 2, AMSbio) using a modification of the media described by the Clevers laboratory (van de Wetering *et al*, 2015) (Advanced DMEM/F12, 10 mM HEPES, 1× Glutamax; 1× B-27 without retinoic acid, 20 ng/ml bFGF (basic fibroblast growth factor); 50 ng/ml EGF (epidermal growth factor), 1 μM LY2157299, 10 μM Y27632, and recombinant Noggin (100 ng/ml). PDO#7, a kind gift from G. Stassi (University of Palermo), was obtained from the dissociation of whole CRCs in suspension as described elsewhere (Lombardo *et al*, 2011). Upon arrival to our laboratory, they were cultured with

the medium described above. All cells were tested weekly for mycoplasma contamination with negative results.

### Xenograft assays

All experiments with mouse models were approved by the Animal Care and Use Committee of Barcelona Science Park (CEEA-PCB) and the Catalan government. We inoculated 150,000 cells (PDO#7) or 2 million cells (PDO#6) in a format of 5- to 7-day grown organoids subcutaneously into NOD/SCID female mice in 50% BME2-HBSS. Generally, a maximum of 4 xenografts were generated per animal. Tumor volume was measured with manual calipers. For tumor initiation assays, viable single human cells (EPCAM-positive; DAPI-negative) from disaggregated xenografts were sorted according to their EGFP levels and subsequently transplanted into recipient mice in 100  $\mu$ l of BME2:HBSS (1:1).

### Lineage tracing and clonal analysis

Cohorts of NOD/SCID mice were inoculated with organoids as described above. When tumors were palpable, mice were given two consecutive doses of tamoxifen (250 mg/kg) to maximize recombination. Mice were sacrificed at indicated time points and tumors were processed for histological analysis. Clone sizes over time were determined in histological sections and scored by manual counting or image analysis software. We averaged measures from distinct sections and xenografts at each time point. Size of clones present at 4-day post-tamoxifen induction was assessed manually. For subsequent time points, we analyzed images using Interactive Learning and Segmentation Toolkit, Ilastik software ([www.ilastik.org](http://www.ilastik.org)). We set the algorithm parameters so that adjacent clones or cells that were not in contact computed as independent clones. A full description of the methodology used for clonal analysis is included in the Appendix Supplementary Methods.

### Transcriptomic profiling

RNA from selected tumor cell populations isolated from xenografts by FACS (1,000–5,000 cells per sample) was amplified using pico-profiling (Gonzalez-Roca *et al*, 2010) and subsequently hybridized on Primeview arrays (Affymetrix). Gene expression was analyzed using standard methodology as described in the Appendix Supplementary Methods. Data have been deposited at Gene Expression Omnibus (GSE92960 and GSE92961).

A detailed description of the methods is included in the Appendix Supplementary Methods.

**Expanded View** for this article is available online.

### Acknowledgements

We thank all members of the Batlle laboratory for support and discussions. We are grateful for the excellent assistance by IRB Barcelona core facilities for histology, functional genomics, biostatistics/bioinformatics, and advanced digital microscopy, as well as the flow cytometry and animal facilities of the UB/PCB. This work has been financed by the European Research Council (ERC advanced grant 340176). Work in the laboratory of EB is also supported by Fundació Botín and Banco Santander, through Santander Universities. IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from the MINECO.

### The paper explained

#### Problem

The study of stem cell hierarchies and of other sources of cellular diversity in human cancers has been largely based on experiments of tumor cell isolation from dissociated patient samples. These experiments impose a number of limitations; first, the requirement of antibodies against specific membrane proteins to label particular cell populations limits the repertoire of cell phenotypes that can be analyzed. Second, the necessity to dissociate the sample impedes the examination of tumor cell populations in an intact environment (i.e., in a growing tumor).

#### Results

We combine two novel methodologies—patient-derived tumor organoids (PDOs) and genome editing tools—and apply them to study the cellular heterogeneity of CRC without the limitations described above. To illustrate the utility of this approach, we studied LGR5<sup>+</sup> cells in human CRC, the analysis of which has been hampered by lack of good commercial reagents to recognize this protein. Using CRISPR-mediated targeting, we engineered PDOs carrying an EGFP reporter cassette recombined at LGR5 locus. We discovered that the LGR5<sup>+</sup> tumor cell population expresses a gene program similar to that of normal ISCs. In xenograft experiments, human LGR5<sup>+</sup> tumor cells propagated the disease to recipient mice with high efficiency implying that this cell population is largely enriched in tumor-initiating cells. Furthermore, we generated PDOs bearing a lineage-tracing cassette and subsequently mapped the fate of LGR5<sup>+</sup> cells in intact tumors. We found that LGR5<sup>+</sup> cells display long-term self-renewal and multilineage differentiation capacity. Finally, by generating dual LGR5-EGFP/Ki67-RFP knock-in PDOs, we characterized a population of quiescent stem cell-like cells in human CRCs.

#### Impact

The approach described herein has broad applicability to analyze the phenotypic diversity of human tumors. In essence, it brings the power of mouse genetic tools to study human cancer. Among many other possibilities, it enables lineage-tracing experiments that can help to elucidate the behavior of distinct tumor cell populations during growth, dissemination, and resistance to therapy.

### Author contributions

EB conceived the study and wrote the manuscript. CC and GT designed experimental work, executed experiments, and helped with manuscript writing. DS provided technical support with targeting vector generation and genome editing of organoids. XH-M performed mice work. AM-S and ES helped designing and conceptualizing the study. MS provided crucial help with immunohistochemistry and organoid cultures. CS-OA performed all statistical analyses. MA and ST contributed to image analysis and 3D reconstruction of serial sections.

### Conflict of interest

The authors declare that they have no conflict of interest.

### References

- Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, van den Born M, Danenberg E, Clarke AR, Sansom OJ, Clevers H (2009) Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* 457: 608–611
- Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ *et al* (2007) Identification of

- stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* 449: 1003–1007
- Basak O, van de Born M, Korving J, Beumer J, van der Elst S, van Es JH, Clevers H (2014) Mapping early fate determination in *LGR5*<sup>+</sup> crypt stem cells using a novel *Ki67-RFP* allele. *EMBO J* 33: 2057–2068
- Calon A, Lonardo E, Berenguer-Llargo A, Espinet E, Hernando-Momblona X, Iglesias M, Sevillano M, Palomo-Ponce S, Tauriello DV, Byrom D et al (2015) Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet* 47: 320–329
- Clevers H (2011) The cancer stem cell: premises, promises and challenges. *Nat Med* 17: 313–319
- Clevers H (2013) The intestinal crypt, a prototype stem cell compartment. *Cell* 154: 274–284
- Dalerba P, Dylla SJ, Park IK, Liu R, Wang X, Cho RW, Hoey T, Gurney A, Huang EH, Simeone DM et al (2007) Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci USA* 104: 10158–10163
- Dalerba P, Kalisky T, Sahoo D, Rajendran PS, Rothenberg ME, Leyrat AA, Sim S, Okamoto J, Johnston DM, Qian D et al (2011) Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* 29: 1120–1127
- Dieter SM, Ball CR, Hoffmann CM, Nowrouzi A, Herbst F, Zavidij O, Abel U, Arens A, Weichert W, Brand K et al (2011) Distinct types of tumor-initiating cells form human colon cancer tumors and metastases. *Cell Stem Cell* 9: 357–365
- van Es JH, Sato T, van de Wetering M, Lyubimova A, Yee Nee AN, Gregorieff A, Sasaki N, Zeinstra L, van den Born M, Korving J et al (2012) *Dll1*(+) secretory progenitor cells revert to stem cells upon crypt damage. *Nat Cell Biol* 14: 1099–1104
- de Felipe P, Luke GA, Brown JD, Ryan MD (2010) Inhibition of 2A-mediated 'cleavage' of certain artificial polyproteins bearing N-terminal signal sequences. *Biotechnol J* 5: 213–223
- Gonzalez-Roca E, Garcia-Albeniz X, Rodriguez-Mulero S, Gomis RR, Kornacker K, Auer H (2010) Accurate expression profiling of very small cell populations. *PLoS One* 5: e14418
- Jung P, Sato T, Merlos-Suarez A, Barriga FM, Iglesias M, Rossell D, Auer H, Gallardo M, Blasco MA, Sancho E et al (2011) Isolation and *in vitro* expansion of human colonic stem cells. *Nat Med* 17: 1225–1227
- Kozar S, Morrissey E, Nicholson AM, van der Heijden M, Zecchini HI, Kemp R, Tavare S, Vermeulen L, Winton DJ (2013) Continuous clonal labeling reveals small numbers of functional stem cells in intestinal crypts and adenomas. *Cell Stem Cell* 13: 626–633
- Kreso A, O'Brien CA, van Galen P, Gan OI, Notta F, Brown AM, Ng K, Ma J, Wienholds E, Dunant C et al (2013) Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* 339: 543–548
- Lombardo Y, Scopelliti A, Cammareri P, Todaro M, Iovino F, Ricci-Vitiani L, Gulotta G, Dieli F, de Maria R, Stassi G (2011) Bone morphogenetic protein 4 induces differentiation of colorectal cancer stem cells and increases their response to chemotherapy in mice. *Gastroenterology* 140: 297–309
- Merlos-Suarez A, Barriga FM, Jung P, Iglesias M, Cespedes MV, Rossell D, Sevillano M, Hernando-Momblona X, da Silva-Diz V, Munoz P et al (2011) The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell* 8: 511–524
- O'Brien CA, Pollett A, Gallinger S, Dick JE (2007) A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 445: 106–110
- Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, De MR (2007) Identification and expansion of human colon-cancer-initiating cells. *Nature* 445: 111–115
- Schepers AG, Snippert HJ, Stange DE, van den Born M, van Es JH, van de Wetering M, Clevers H (2012) Lineage tracing reveals *LGR5*<sup>+</sup> stem cell activity in mouse intestinal adenomas. *Science* 337: 730–735
- Scholzen T, Gerdes J (2000) The *Ki-67* protein: from the known and the unknown. *J Cell Physiol* 182: 311–322
- Snippert HJ, van Es JH, van den Born M, Begthel H, Stange DE, Barker N, Clevers H (2009) *Prominin-1/CD133* marks stem cells and early progenitors in mouse small intestine. *Gastroenterology* 136: 2187–2194.e1
- Tetteh PW, Basak O, Farin HF, Wiebrands K, Kretschmar K, Begthel H, van den Born M, Korving J, de Sauvage F, van Es JH et al (2016) Replacement of lost *Lgr5*-positive stem cells through plasticity of their enterocyte-lineage daughters. *Cell Stem Cell* 18: 203–213
- Vermeulen L, De Sousa EMF, van der Heijden M, Cameron K, de Jong JH, Borovski T, Tuynman JB, Todaro M, Merz C, Rodermond H et al (2010) *Wnt* activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat Cell Biol* 12: 468–476
- Vermeulen L, Todaro M, de Sousa Mello F, Sprick MR, Kemper K, Perez Alea M, Richel DJ, Stassi G, Medema JP (2008) Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. *Proc Natl Acad Sci USA* 105: 13427–13432
- van de Wetering M, Francies HE, Francis JM, Bounova G, Iorio F, Pronk A, van Houdt W, van Gorp J, Taylor-Weiner A, Kester L et al (2015) Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* 161: 933–945
- van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Battle E, Coudreuse D, Haramis AP et al (2002) The *beta*-catenin/*TCF-4* complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 111: 241–250
- Zeilstra J, Joosten SP, Dokter M, Verwiel E, Spaargaren M, Pals ST (2008) Deletion of the *WNT* target and cancer stem cell marker *CD44* in *Apc* (*Min*<sup>+</sup>) mice attenuates intestinal tumorigenesis. *Cancer Res* 68: 3655–3661
- Zeuner A, Todaro M, Stassi G, De Maria R (2014) Colorectal cancer stem cells: from the crypt to the clinic. *Cell Stem Cell* 15: 692–705



**License:** This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.