Epithelial Cell Intrinsic and Extrinsic Effects of Wnt Disruption in Colorectal Tumorigenesis



Ester Gil Vázquez
Christ Church College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2021

Preface

This thesis is a collection of research work carried out by Ester Gil Vazquez in the Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, between October 2017 and October 2021, under the supervision of Professor Simon Leedham.

The research presented here is original, and the work of others is acknowledged and cited with appropriate references. This work has not previously been submitted for a degree at this or any other university.

Firstly, I would like to thank my supervisor Simon Leedham, for giving me the freedom to develop the skills and to explore the questions that interested me while offering his guidance, practical expertise and advise. Secondly, I would like to thank Hayley, who has been a central pillar during the course of my PhD. Thank you for all the patience you have treated me with, for all your teachings and for the long hours that you have invested in the work I was conducting. I would like to give a special thanks to Martijn, for all the mentoring, and the meaningful conversations that we shared. Thanks to all the members of the Leedham lab, Ryan, Sulo, Eoghan, Gabriel, Abby and Sam for all the help, insights, and for making the lab such a friendly and fun place to be working in. Finally, I would like to thank Nadia for all the experiences we shared inside and outside the lab, for being as silly as me when appropriate, and for inspiring me with her love for Science.

To my GMS cohort, Maria, Jonas, Ron, Dan and Adam, for the table tennis games, and for all the help and guidance offered during my PhD. To the members of the Church lab, Kasia, Charlotte, Noel, Matt, Yannick and everybody who kept the social gatherings going, for making the work atmosphere more friendly. To the LGBTQ+ committee members, Angela, Ash and Eoghan, for making the center a better place. To all the people within the Wellcome Center for Human Genetics who made my working days more pleasant: Claudia, thanks for cooking scrumptious food for me, and for all the fun talks. Jose, thanks for the after-work beers and the career guidance offered.

To Ross, Val, Athena, Johaness and Jonny, who were the first friends I made when I arrived to this dark country, and who have stood by my side throughout my years in Oxford. To Sergio, Nuria, Arno, Paul, Davide, and all the other Christ Church GCR members, for the fun times.

To my climbing pals, Alfie, Eddy, Josh, Ivan, Barbora, Alex, Michael, Sam, George, Jola, and all the cool people within the OUMC. It has been an absolute pleasure climbing with you, and I hope that life will bring us together for more fun times as children of the forest. Climbing has been more than a sport, it has been a source of meaningful connections and the oasis of peace that kept me sane in the toughest stages of my PhD.

To my Oxford family. Thank you all for the cooking sessions, the Plush outings, the BBQs with guitars in Port Meadow, for taking me to the hospital whenever I broke something, for supporting my rants, hugging me in the bad times and being kind of okay with me being the loudest in the room at any given time. I honestly do not think I could have done this without you. To Gil, my brother from a different

womb, thanks for making me feel understood in life, in science, and beyond. To Pablo, *el de letras*, thanks for keeping the kid alive, and for all the fun historical facts. To Alvaro, the philosopher, for being a calm sea in a corner of the maelstrom. To Cris, *la de la oida absoluta*, thanks for you perfect tuning of *Despacito*, and for making me see Oxford with a different set of eyes. To Javi, *el hombre a la sonrisa pegado*, thanks for being an inexhaustible source of good vibes. To Miren, the altruist, who was at so many places, but all the same always there. Thank you for that, and for the muffins! To Teresa, *la marabunta*, thanks for your commitment to making Oxford a funnier place.

To my friends back home, who have bared with me throughout my years abroad and been my rock when I needed them. To Lara, Geor and Roberto, for being that constant in my life that reminded me of who I was and where I came from. Thank you for your love and support, specially during the writing process. A special thanks to Geor, who had to endure all my bad days in this last stage, and who has taken the time and patience to comfort me when I needed it. I would aso like to thank Elisa, who was a constant behind the screen in times of Covid, who offered support and made my days entertaining. Finally, thanks to Julia for making this last few months easier and more fun.

Last but not least, I would like to thank my family, who have always believed in me, even when I lacked the confidence. Mum, thanks for encouraging me to sign up to *Joves I Ciencia* when I was just a kid. I never thought that Science could be such an important part of my life. Dad, thank you for letting me make mistakes without judgement, and for believing that I could stand up and recover from them. Laura, thanks for being such an important source of inspiration, and for all the support you have granted me throughout the years.

I would like to finish by thanking all the people who I have not named, but who have made my PhD and my stay in Oxford fun and pleasant. Thanks to all the staff from St. Annes and Christ Church College, from the Wellcome Center for Human Genetics, and from the University of Oxford.

One illusory stripe of a rainbow contains an infinite number of frequencies, but your concepts for "Red," "Blue," and other colors cause your brain to ignore the variability.

— Lisa Feldman Barrett, *How Emotions Are Made: The Secret Life of the Brain*

Abstract

Despite cancer initiation being attributed to specific pools of epithelial stem cells, tumorigenic transformation is a complex process that entails the co-evolution of the epithelial and stromal compartments. Here, I studied how the stem cell composition of a given malignancy shapes the interactions established between tissue compartments in Colorectal Cancer (CRC).

For many years, Lgr5+ Crypt Base Columnar (CBC) cells have been acknowledged as the cell of origin of CRC. However, alternative pools of intestinal progenitors, the Regenerative Stem Cells (RSC), have been described in settings of regenerations. In this piece of work, I analysed the contribution of these pools of stem cells to colorectal neoplasia. Using a RNA sequencing-based analytics pipeline, a range of mouse and human pre-malignant lesions and cancers could be classified based on their Stem Cell composition. This revealed that lesions with different molecular characteristics preferentially select for the expansion of specific progenitors. Furthermore, I narrowed down the molecular drive that may be conferring CBC and RSC characteristics. Finally, I demonstrated that these populations are not static, but tumours can change their stem cell composition to adapt to therapeutic selective pressure.

Using genetic models of colorectal disease, I investigated how alterations in epithelial-stromal interactions can drive colorectal transformation. In the *Vil1-Grem1* mouse, epithelial upregulation of *GREM1* was enough to cause the activation of fibroblast populations and modulate the production of niche factors within the intestinal stroma. This enabled the formation of polyps which presented pools of progenitors in ectopic locations. Importantly, these alterations could be reversed by a Gremlin-1 blocking antibody.

Despite WNT-driven immune evasion being an acknowledged phenomena in CRC, the molecular mechanisms by which it operates are yet not

completely understood. Here I used a range of models of WNT-driven intestinal transformation to investigate this phenomena. I found that WNT activation comes with the downregulation of pyroptosis, an immunogenic form of cell death. Furthermore, treatment of the *Lgr5-Rspo3* mouse with a WNT inhibitor was enough to drive re-infiltration of polyps by T lymphocytes, which came together with upregulation of the pyroptosis effector *Gsdmc4*. However, in APC^{Min} animals, lymphocytic infiltration after Lgr5+ cell ablation did not correlate with an increase in pyroptosis, suggesting that mouse models with different molecular drives/progenitor composition may possess alternative immune evasion mechanisms.

Overall, in this piece of work, I present a tool to infer the Stem Cell admixture within colorectal lesions, which could be used to inform clinical practice and to facilitate CRC research in the future. Moreover, I provided evidence for mechanisms by which epithelial cell-intrinsic characteristics can alter the fibroblasts and immune compartment within colorectal polyps.

Contents

1 Introduction **1**

1.1 Intestinal homeostasis 1
1.2 Intestinal Stem Cells 2
1.3 Immune Tolerance in the Intestine 3
1.3.1 Are Intestinal Stem Cells Immune Privileged? 3
1.4 Cell Plasticity 5
1.5 Organoids as Tools for the Study of Intestinal Stem Cells 6
1.6 Regulation of Intestinal Cell Fate 7
1.6.1 The Intestinal Stem Cell Niche 7
1.6.2 Wnt signaling 7
1.6.2.1 Canonical Wnt pathway 9
1.6.2.2 Non-canonical Wnt pathway 10
1.6.2.3 Lgr/R-spondin axis 11
1.6.3 BMP pathway 12
1.6.4 Inflammation 13
1.6.5 The Hippo-YAP/TAZ Pathway 15
1.7 Intestinal Cell Death 16
1.8 Colorectal Cancer 17
1.8.1 Molecular characterization of colorectal cancer 19
1.9 Cancer Stem Cells 20
1.10 Cell Fate Regulation in Colorectal Cancer 22
1.10.1 WNT Signaling in Colorectal Cancer 22
1.10.1.1 Tumorigenesis Initiated by R-spondin3 24
1.10.2 BMP Signaling in Colorectal Cancer 24
1.10.2.1 Tumorigenesis Initiated by Gremlin-1: HMPS and the *Vil1-Grem1* mouse model 25
1.10.3 YAP Signaling in Colorectal Cancer 25
1.10.3.1 YAP and WNT Interact to Regulate SC Identity . . 26

1.11 The Role of the Stroma in Colorectal Cancer Initiation 28
1.12 The Immune Compartment in CRC 29
1.12.1 Epithelial Cell Extrinsic Mechanisms of Immune Suppression in Colorectal Cancer 30
1.12.2 Epithelial Cell Intrinsic Mechanisms of Immune Suppression in Colorectal Cancer 31
1.12.2.1 WNT-driven Immune Exclusion 32
1.13 Objectives 33

2 Materials and Methods **35**
2.1 Mouse Procedures 35
2.1.1 Animal Models 35
2.1.2 Breeding 36
2.1.3 Earclips 36
2.1.4 Genotyping 36
2.1.5 Recombination 38
2.1.6 Cell Ablation 39
2.1.7 Administration of Therapeutic Compounds 39
2.1.7.1 Gremlin-1 Blocking Antibody 39
2.1.7.2 Procupine Inhibitors 39
2.1.8 Collection of Intestinal Sections 40
2.1.8.1 Collection of Fresh Polyps 40
2.1.8.2 Dissociation of the Epithelial and Stromal Compartments 40
2.2 Histology 41
2.2.1 Polyp Counts 41
2.2.2 Tissue Processing 41
2.2.3 Embedding 41
2.2.4 Sectioning 41
2.2.5 Staining Procedures 41
2.2.5.1 Haematoxylin and Eosin 41
2.2.5.2 Immunohistochemistry 42
2.2.5.3 *In Situ Hybridisation* 42
2.2.5.4 Multiplex immunoflouorescence 43
2.3 Nucleic Acids 44
2.3.1 RNA Extraction from Tissue 44

2.3.2 DNA Extraction from Tissue . . . 44
2.3.3 Quality Control of Nucleic Acids . . . 45
2.3.3.1 Nanodrop Spectophotometer . . . 45
2.3.3.2 Qubit Fluorometer . . . 45
2.3.3.3 TapeStation . . . 45
2.4 Sequencing . . . 46
2.4.1 RNA Sequencing . . . 46
2.4.2 Whole-Exome-Sequencing . . . 46
2.5 Data Analysis . . . 46
2.5.1 Pre-processing of Internal High Throughput Sequencing Data 46
2.5.1.1 RNA sequencing . . . 46
2.5.1.2 Whole-Exome-Sequencing . . . 47
2.5.2 Obtention of External Data . . . 47
2.5.2.1 Human Data . . . 47
2.5.2.2 Mouse Data . . . 47
2.5.3 Statistical Analysis . . . 48
2.5.3.1 Polyp Count . . . 48
2.5.3.2 Image Quantification of DAB stains . . . 48
2.5.3.3 Image Quantification of Fluorescent stains . . . 48
2.5.3.4 Stem Cell Index . . . 50
2.5.3.5 Pathway Analysis . . . 50
2.5.3.6 Single Nucleotide Variants . . . 51
2.5.3.7 Copy Number Variation . . . 51
2.5.3.8 Survival Analysis . . . 52
2.5.3.9 Therapy Response . . . 52

3 Assessing the Stem Cell Composition Across Colorectal Malignancies 53

3.1 Introduction . . . 53
3.2 Objectives . . . 54
3.3 Generation of a Stem Cell Index to Assess Stem Cell Dynamics in Colorectal Lesions . . . 54
3.4 *Lgr5* and *Anxa1* are Surrogate Markers of CBC and RSC Identity . . 55
3.5 Curation of the RSC Signature Using Single Cell RNAseq Data . . . 57
3.6 Stem Cell Composition Varies Across a Range of Mouse Intestinal Polyposis and Neoplasia Models . . . 58

3.7 Stem Cell Composition Varies Across a Range of Human Precursor Lesions and Tumour Molecular Subtypes . . . 60
3.8 Driver Genes and Pathways Associate with Stem Cell Molecular Phenotypes . . . 62
3.9 Application of a Selective Pressure Shifts the Equilibrium between CBC and RSC Poulations in APC^{Min} Mice . . . 66
3.10 Stem Cell Plasticity conditions Chemotherapy Resistance in Human Colorectal Cancer Patients . . . 68
3.11 Discussion . . . 69

4 Molecular Characterization and Wnt Requirements in Polyps Initiated by BMP Dysregulation 75
4.1 Introduction . . . 75
4.2 Objectives . . . 76
4.3 Histopathological Progression of the Vil1-*Grem1* Pathogenesis . . . 76
4.4 Ectopic Crypts are negative for CBC and RSC markers . . . 76
4.5 Grem1 Antibody Treatment Reverts the VG Phenotype . . . 77
4.6 Absence of Cell-Intrinsic WNT Mutations in VG Polyps . . . 79
4.7 Environmental Sources of WNT in VG Polyps . . . 81
4.8 Porcupine Inhibition Reverts Polyposis in VG Animals . . . 83
4.9 The Non-Canonical WNT Receptor Frizzled-5 is Profoundly Up-regulated in VG Polyps . . . 85
4.10 Escaper Cells Repopulate the Polyps of VG-VilER-Cre-Fzd5KO$^{-/-}$ Animals . . . 85
4.11 Characterization of Advanced VG Polyps . . . 89
4.12 Discussion . . . 96

5 Does Wnt disruption mediate immune exclusion? 100
5.1 Introduction . . . 100
5.2 Objectives . . . 101
5.3 T Lymphocytes are Distributed Unevenly Across the Crypt-Villus Axis 101
5.4 Animal Models of WNT Disruption . . . 104
5.5 Animal Models of WNT-driven Polyposis . . . 108
5.5.0.1 Phenotypic Characterization of the *Lgr5-Rspo3* Mouse 109
5.6 Immune Characterization of *Lgr5-Rspo3* and APC^{Min} Animals . . . 113
5.7 Lgr5+ Cell Ablation Mildly Changes the Cd3+ Composition of APC^{Min} Polyps . . . 116

5.8 Porcupine Inhibition Prevents Rspo3-induced Polyposis and Improves Survival of *Lgr5-Rspo3* Animals . . . 120
5.9 Porcupine Inhibition can Revert Immune Exclusion from *Lgr5-Rspo3* ECFs, but not Conventional Crypts . . . 123
5.10 Gasdermin4c is profoundly upregulated in *Lgr5-Rspo3* epithelium after RXC004 treatment . . . 123
5.10.1 *Gsdmc4* Upregulation Does not Mediate the Subtle T Cell Infiltrations Observed in *Lgr5-CreER^{ER}-DTR- APC^{Min}* Mice . . 126
5.11 Discussion . . . 129

6 Discussion **134**
6.1 Summary of Results . . . 134
6.2 Discussion . . . 135
6.2.1 Different Pools of Progenitors Select for Cell-Extrinsic or Cell-Intrinsic forms of WNT disruption . . . 135
6.2.2 Stem Cell Plastic Behaviour is a Useful Outcome of Therapy Response . . . 138
6.2.3 The Epithelial and Stromal Compartments co-evolve during Colorectal Tumorigenesis . . . 139
6.2.4 WNT-driven Stem Cell Identity may Include Immune Evasion Characteristics . . . 140
6.3 Concluding Remarks . . . 142

A Supplementary Figures and Tables **144**

Bibliography **160**

List of Figures

1 Schematic representation of the WNT pathway. 10
2 Schematic representation of the Hippo/YAP-TAZ pathway. 16
3 Schematic representation of the interactions between the Hippo/YAP-TAZ and WNT pathways. 27
4 The Stem Cell Index can determine the Stem Cell composition of a lesion. 55
5 Expression of the RSC and CBC genes/signatures in mouse and human intestines. 56
6 Distribution of CBC and RSC phenotypes across a range of CRC mouse models. 59
7 Distribution of CBC and RSC phenotypes across a range of human colorectal precursor lesions and cancers. 62
8 Signaling pathways associated with CBC and RSC phenotypes in mice. 63
9 Molecular characterization of human TCGA CRC in association with their SC phenotypes. 65
10 RSC temporally repopulate APC^{Min} mice after CBC ablation. 67
11 Stem Cell plasticity is associated with therapy resistance in a human data set. 68
12 Combinatorial therapies based on the Stem Cell composition of the tumour. 74
13 Histopathology of the Vil1-*GREM1* mouse model versus a WT animal. 77
14 Fluorescent co-stain for the CBC marker *Lgr5* and the RSC marker *Anxa1* in Vil1-*GREM1* polyps. 78
15 Treatment with a Gremlin-1 blocking antibody reverts the Vil1-*GREM1*(VG) phenotype. 80
16 Vil1-*GREM1* polyps do not show signs of strong Wnt activation. . . 81

17 Differences in the transcription of Wnt and Rspo ligands between Vil1-*GREM1* and WT animals. 82
18 Expression of *Wnt2b, Wnt5a* and *Rspo3* in WT and Vil1-*GREM1* animals. 84
19 *Wnt2b* and *Wnt5a* are co-expressed with the fibroblast markers *Gli* and *Wnt5a* respectively. 85
20 Treatment with a Gremlin-1 blocking antibody restores the levels of expression of Wnt ligands. 86
21 Treatment with a Porcupine inhibitor partially reverts the phenotype of Vil1-*GREM1*(VG) animals. 87
22 *Fzd5* is upregulated in the epithelium of Vil1-*GREM1* animals. . . . 88
23 Expression of *Fzd5* and *Fzd7* in VilER-Cre-Fzd5KO$^{-/-}$ and VilER-Cre-Fzd7KO$^{-/-}$ animals. 90
24 There are no difference in terms of polyp burden in VG-VilER-Cre-Fzd7KO$^{-/-}$ and VG-VilER-Cre-Fzd5KO$^{-/-}$ versus paired Vil1-*GREM1* controls. 91
25 Levels of long-term *Fzd5* and *Fzd7* downregulation in VG-VilER-Cre-Fzd7KO$^{-/-}$ and VG-VilER-Cre-Fzd5KO$^{-/-}$ animals. 92
26 Advanced Vil1-*GREM1* polyps acquire expression of CBC and RSC markers. 93
27 Advanced polyps acquire cell-intrinsic forms of Wnt activation. 94
28 The expression patterns of *Lgr5* and *Fzd5* are inversely correlated in Vil-GREM1 polyps. 95
29 Proposed model of the Vil1-*GREM1* pathogenesis. 98

30 CBC cells present an immune exclusion phenotype. 102
31 Immune processes are downregulated in the crypt compartment of WT animals. 104
32 *VilER-Cre-Apc$^{fl/fl}$* and Gli-Rspo3 animals present with an hyperproliferative phenotype. 105
33 WNT disruption drives the downrwegulation of immune pathways and pyroptosis. 106
34 Histopathology of *APC^{Min}* and *Lgr5-Rspo3* polyps. 108
35 Histopathology of *Lgr5-Rspo3* samples. 110
36 Remodelling of the proliferative and differentiated compartments in *Lgr5-Rspo3* animals. 111

37 There is an expansion of *Lgr4+,Fzd5+* cells in the intestines of *Lgr5-Rspo3* animals. . . . 112
38 T cell infiltration in mouse models of Ligand-Dependent and Ligand-Independent WNT disruption. . . . 114
39 Immune characterization of APC^{Min} and *Lgr5-Rspo3* . . . 115
40 CBC ablation in APC^{Min} mice drives a subtle T cell infiltration. . . . 117
41 Transcriptomic changes driven by the CBC ablation from APC^{Min}-$Lgr5^{DTR}$ polyps. . . . 118
42 Administration of Diphteria Toxin (DT) to APC^{Min}-$Lgr5^{DTR}$ animals causes CBC to die by apoptosis. . . . 119
43 High doses of RXC004 cause CBC ablation and intestinal collapse in *Lgr5-Rspo3* animals. . . . 121
44 Treatment with a Porcupine inhibitor reverts polyposis in *Lgr5-Rspo3* animals. . . . 122
45 Treatment with a Porcupine inhibitor drives T cell infiltration in ECF from *Lgr5-Rspo3* animals. . . . 124
46 Treatment with a Porcupine alters the immune microenvironment within *Lgr5-Rspo3* polyps. . . . 125
47 Treatment with a Porcupine inhibitor induces the expression of *Gsdmc4* in crypts from *Lgr5-Rspo3* animals. . . . 127
48 Administration of Diphteria Toxin (DT) to *Lgr5-*$CreER^{ER}$*-DTR-* APC^{Min} animals causes a transient downregulation of *Gsdmc4*. . . . 128
49 Treatment with a Porcupine inhibitor drives re-infiltration of T lymphocytes into the ECF from Lgr5-Rspo3 animals. . . . 131

50 Intestinal inter-compartmental interactions guide CRC . . . 135
51 Hypothesis: different populations of progenitors select for specific forms of WNT disruption. . . . 137
52 Distinct populations of Intestinal Epithelial Cells are subjected to different forms of Programmed Cell Death . . . 142

S1 Levels of long-term *Fzd5* VG-VilER-Cre-Fzd5KO$^{-/-}$ animals. . . . 157
S2 Caspase-3 is only sporadically expressed by cells in the ECF from RXC004-treated Lgr5-Rspo3 animals. . . . 158
S3 *Gsdmc4* is mostly expressed by Enterocyte progenitors. . . . 159

List of Tables

2.1 Mouse Models 37
2.2 Genotyping primers 38
2.3 IHC antibodies 42
2.4 ISH probes 43
2.5 Mouse Models Oxford and the Beatson Institute (Glasgow) 49
2.6 Immune Phenotypes 50
2.7 List of driver genes 51

A.1 Curated list of RSC genes 150
A.2 Statistics: Mouse autochtonous tumours 151
A.3 Statistics: Tumour-derived organoids 152
A.4 Statistics human pre-malignant lesions 152
A.5 Statistics human pre-malignant tumours 153
A.6 Number of alterations in top and bottom deciles. 156

List of Abbreviations

ANOVA: Analysis Of Variance

APC: Adenomatous Polyposis Coli

ATF3: Activating Transcription Factor 3

BATF3: Basic leucine zipper transcriptional factor ATF-like 3

BMP: Bone Morphogenetic Protein

BRAF: B-Raf

CAF: Cancer-Associated Fibroblasts

CK20: Cytokeratin 20

CBC: Crypt Base Columnar

CD: Cluster of Differentiation

CMS: Consensus Molecular Subgroup

CNV: Copy Number Variation

CRIS: Colorectal cancer Intrinsic Subtypes

CRC: Colorectal Cancer

CSC: Cancer Stem Cells

DAB: 3, 3'-diaminobenzidine

DAMPs: Damage-Associated Molecular Patters

DNA: Deoxyribonucleic acid

DC: Dendritic Cells

DCLK1: Doublecortin Like Kinase 1

DT: Diphtheria Toxin

DTR: Diphtheria Toxin Receptor

DSS: Dextran Sulfate Sodium

ECF: Ectopic Crypts

ECM: Extracellular Matrix

EGF: Epidermal Growth Facto

ER: Endoplasmic Reticulum

FOXP3: Forkhead box P3 protein

GFP: Green Fluorescent Protein

GSK3β: Glycogen Synthase Kinase 3 β

GSEA: Gene Set Enrichment Analysis

GSVA: Gene Set Variation Analysis

GTP: Guanosine Triphosphate

HMPS: Hereditary Mixed Polyposis Syndrome

H&E: Hematoxylin and Eosin

IAP: Inhibitor of Apoptosis

IBD: Inflammatory Bowel Disease

IEC: Intestinal Epithelial Cell

IELs: Intraepithelial Lymphocytes

IFN: Interferon

IL: Interleukin

IP: Immune Privilege

ISC: Intestinal Stem Cells

ISH: *In Situ Hybridization*

JAK: Janus kinase

JP: Juvenile Polyposis

KO: Knock-Out

KRAS: Kirsten Rat Sarcoma virus

LD: Ligand-Dependent

LEF: Lymphoid Enhancer Factor

LGR5: Leucinerich repeat-containing G-protein coupled Receptor 5

LI: Ligand-Independent

LRP: LDL Receptor Related Protein

MAPK: Mitogen-Activated Protein Kinase

MHC: Major Histocompatibility Complex

MPI: Multiplex Immunofluorescence

MSI: Multispectral images

MSS: Microssatellile Stable

MyD88: Myeloid differentiation primary response 88

PAMPs: Pathogen-Associated Molecular Patters

PBS: Phosphate-Buffered Saline

PCP: Planar Cell Polarity

PCR: Polymerase Chain Reaction

PGE2: Prostaglandin E2

PI3K: Phosphoinositide 3-Kinases

PRRs: Pattern Recognition Receptors

RA: Retinoic Acid

RNA: Ribonucleic acid

RNAseq:RNA sequencing

RNF43: Ring Finger Protein 43

RSC: Regenerative Stem Cells

SC: Stem Cell

SCI: Stem Cell Index

SMAD: Mothers Against Decapentaplegic homologue

SNV: Single Nucleotide Variant

TA: Transit Amplifying

TCGA: The Cancer Genome Atlas

TVA: Tubulo-villious Adenomas

SSL: Sessile Serrated Lesions

STAT: Signal Transducer and Activator of Transcription

TAP1: Antigen Peptide Transporter 1

TCF: T-Cell Factor

TCR: T-cell Receptor

TEAD: TEA Domain family member

TGF: Transforming Growth Factor

TLRs: Toll-like receptors

TNF: Tumour Necrosis Factor

Treg: regulatory T

TSA: Traditional Serrated Adenoma

VG: Vil1-Grem1

WES: Whole-Exome-Sequencing

WT: Wild Type

YAP: Yes-Associated Protein 1

ZNRF3: Zinc and Ring Finger 3

Chapter 1

Introduction

1.1 Intestinal homeostasis

With the surface area of a small pub ($32m^2$) and the mean fluid volume of a double whiskey (72-118ml), the intestine is a structurally complex organ[1, 2]. It is covered by a monolayer of columnar epithelium that folds within itself forming crypts and, in the case of the small bowel, villi. The crypt is the minimal functional unit of the intestine and it is regionally and functionally compartmentalized. At the bottom, we find a pool of adult Intestinal Stem Cells (ISC) that sustain the renewal of the tissue. The moment a cell leaves the stem cell (SC) pool, it migrates upwards to enter the Transit Amplifying (TA) area. Once there, cells rapidly proliferate and give rise to all the progenitors needed to maintain the fast turn-over of the tissue. Finally, as cells exit the TA area, they differentiate into the mature epithelial cell types. With the exception of Paneth cells in the small intestine, which migrate back to the bottom of the crypts, these descendants will remain in the villi or top of the colon crypts until they undergo programmed cell death and are shed off into the lumen[1].

The gut participates in two essential processes: it mediates nutrient uptake and acts as a barrier against pathogens. These roles are performed by the six different types of mature epithelial cells that cover the walls of the intestine. These cells can be divided into the secretory lineage, which includes goblet, Paneth, enteroendocrine and Tuft cells, and the absorptive lineage, which is formed by M cells and enterocytes[1].

At the bottom of crypts in the small intestine, Paneth cells produce antimicrobial products like α-defensins and lysozime, which protect cells locally but can also reach higher compartments. In addition, they secrete growth factors and BMP inhibitors as well as Notch and WNT ligands, which support the SC pool. In the colon, Reg4+ cells have been suggested to be the functional equivalent to Paneth cells[1, 3].

Towards the luminal aspect of the crypt, we find the rest of specialized cells. Within the secretory lineage, goblet cells produce the layer of mucus that shields the epithelium, whereas enteroendocrine cells mediate the secretion of hormones into the bloodstream (eg. serotonin, somatostatin) and Tuft cells specialize in protection against helminths. Within the absorptive lineage, M cells are localized in specialized lymphoid tissues (Peyer's Patches) where they uptake antigens from the lumen and deliver them to immune cells. Finally, enterocytes are the most abundant cells in the intestinal epithelium. They cover 80 per cent of the monolayer and are responsible for the absorption of water, ions and nutrients[1].

1.2 Intestinal Stem Cells

The intestinal epithelium is a very dynamic compartment. The moment a Intestinal Epithelial Cell (IEC) leaves the crypt bottom, it takes an average of 5 days to migrate upwards and be shed off into the lumen. This rapid turn-over is necessary to ensure tissue integrity given the fragile nature of a monolayer and the vast amount of insults the intestine withstands (eg. changes in pH, colonization by bacteria and mechanical damage). This dynamic is sustained by a pool of cells with two key characteristics, self-renewal capacity and multi-directional differentiation potential: the Intestinal Stem Cells (ISC)[1].

Even though the existence of a pool of ISC at the bottom of the intestinal crypts has been acknowledged for decades[4], the characterization of these cells required the development of specific markers that allowed to trace and modify them. Leucine-rich repeat-containing G-protein coupled receptor 5 (Lgr5) has been established as the hegemonic ISC marker. It was first described as a WNT target in colon cell lines[5], but latter validated as a surrogate of ISC identity by Barker et al. To prove this, they produced a mouse model with an EGFP-IRES-CreERT2 cassette under the control of the *Lgr5* promoter. Crossing this animal with a Rosa26-LacZ mouse enabled lineage tracing of Lgr5+ cells[6]. The authors reported ribbons of LacZ+ cells emerging from the crypt base and migrating to the side of the villi. These ribbons could still be observed 60 days after tamoxifen induction and contained enterocytes, Paneth and Goblet cells. Further studies contributed to characterize Lgr5+ cells at a transcriptomic and proteomic level, and revealed that they express high levels of SC genes like *Olfm4, EphB2* and *Ascl2*[7, 8, 9].

1.3 Immune Tolerance in the Intestine

There is plenty of evidence indicating that the intestinal mucosa is a location for immune tolerance induction. This is necessary to avoid eliciting an immune response against the wide range of harmless stimuli which this tissue is periodically exposed to[10]. Intestinal immune tolerance is achieved by the generation of regulatory cells able to contain the activation of antigen-specific effector T lymphocytes. This process involves the uptake of antigens by lamina propria Cluster of Differentiation (CD)103+ cDC2-like Dendritic Cells (DC), which migrate towards the local lymph nodes to meet naive T cells. Once there, they prime T cells to express gut-homing markers and drive their differentiation towards a regulatory Forkhead box P3 protein (Foxp3)+ phenotype[10, 11]. Subsequently, regulatory T (Treg) cells migrate into the intestine, where they maintain an immune suppressive milieu through cell-to-cell contact or secretion of immune-suppressive factors like interleukin(IL)-10 and Transforming Growth Factor (TGF)-β[10]. Interestingly, the levels of CD103+ tolerogenic DC in homeostatic settings are higher in the intestine than in other tissues, and their differentiation and function is regulated by epithelium and stroma-derived factors, such as thymic stromal lymphopoietin and Retinoic Acid (RA)[10, 12], which highlights the role of non-immune populations in tolerance maintenance.

In normal settings, we can find effector T cells that are specific for antigens present in the intestine, but they are repressed by the tolerogenic environment generated by DC and Treg cells[10]. However, when tolerance breaks, this results in the emergence of inflammatory disorders. A clear example of this is Inflammatory Bowel Disease (IBD), an umbrella term that encompasses a range of chronic inflammatory conditions occurring in the intestine, including Ulcerative Colitis and Crohn's Disease.

1.3.1 Are Intestinal Stem Cells Immune Privileged?

The capacity of tissues to withstand the introduction of antigens without inducing an immune response is known as Immune Privilege (IP). This old concept emerged from transplant experiments, where scientists observed that grafts where accepted in certain locations (eg.eye chamber, brain), but rejected in others. It has been suggested that this property may be in place to protect non-renewable tissues from the damaging side-effects associated to an immune response[13].

Effectively, IP can be mediated by a range of different mechanisms, from physical boundaries, like the blood-Central Nervous System barrier, to release of anti-inflammatory mediators by the surrounding parenchyma, like it happens in the testis[13].

In addition, IP can be acquired in strategic locations within immune-visible organs through cell-intrinsic mechanisms. Evidence for this comes from a publication by Agudo et al, where the authors showed that quiescent SC from the hair follicle and smooth muscle are able to avoid immune-mediated clearance via down-regulating Major Histocompatibility Complex (MHC)-I molecules[14].

Agudo et al investigated the existence of IP in the ISC compartment[14]. In order to do so, they used Jedi T cells, which recognize Green Fluorescent Protein (GFP) epitopes loaded into MHC-I complexes, and Lgr5-GFP mice, in which *GFP* was expressed by ISC. After being injected, Jedi T cells were activated through vaccination with GFP. One week after that, intestinal crypts showed signs of infiltration by T cells, and GFP-expressing ISC had been eliminated. From this evidence, the authors concluded that intestinal Lgr5+ cells were not immune privileged.

Similar lessons can be extracted from Graft-versus-Host-Disease, a condition in which donor T cells attack recipient tissues after an allogenic transplant. The intestine is one of the organs most commonly affected by this complication. In this context, the lamina propria at the crypt base is infiltrated by donor CD4+ T cells that can mediate ISC death through Interferon(IFN)-γ-dependent activation of Janus kinase(Jak) 1-Signal Transducer and Activator of Transcription(Stat) 1[15].

It is important to remark a common factor between the two aforementioned studies [15, 14]. In both cases, T cells were not activated by intestinal DC, but by artificial methods: either vaccination, or the introduction of donor T cells. Once the activated effector T cells infiltrated the intestinal mucosa, they introduced an inflammatory milieu that bypassed the pre-existing tolerogenic environment and favoured the establishment of an anti-host immune response. Interestingly, adoptive T cell therapy in Coloractal Cancer (CRC) has shown to have limited success, and numerous studies report different levels of T cell infiltration into tumours of treated individuals[16], suggesting that additional levels of regulation may be operating in colorectal malignancies.

The above evidence points towards a lack of IP in the ISC compartment. However, the homeostatic distribution of immune cells along the crypt-villus axis is not homogeneous. Intraepithelial Lymphocytes (IELs) comprise a group of CD8+ lymphocytes with cytotoxic activity which express pro-inflammatory mediators like IFN-γ and Tumour Necrosis Factor(TNF)-α upon stimulation[17]. Interestingly, despite Cd3+ cells being equally distributed throughout the lamina propria, IELs are more abundant in the villi compartment during homeostasis[15]. The exclusion of these

populations from the intestinal crypt may help maintain a tolerogenic environment in the surroundings of stem cell progenitors.

1.4 Cell Plasticity

In the past, it was thought that progenitor cells were subjected to a hierarchical and irreversible evolutionary path. Cell fate trajectories culminated in loss of stem properties and acquisition of a differentiated phenotype, which was fixated by means of epigenetic regulation[18]. Waddington conceptualized this process in his diagram as a series of branching decisions which set out cells on specific differentiation journeys[19]. However, more recent evidence suggests that stemness is not regional and temporally limited, but can be reacquired upon cell reprogramming. In 2006, Takahashi and Yamanaka proved that retroviral transduction of fibroblasts with 4 specific transcription factors (Oct3/4, Klf4, Sox2, and c-Myc) was enough to generate induced pluripotent stem cells *in vitro*[20]. Similar effects can be observed *in vivo* when the relationships between cells and their environment are altered, particularly in settings of damage and cell loss[18]. This capacity to reprogram in order to adapt to external cues is known as plasticity.

Cell plasticity is embedded within the intestinal epithelium even during homeostasis. Ristma et al used an *in vivo* lineage tracing system to assess stem cell dynamics in the intestinal crypts. They observed that cells located at the base of the crypts maintained their stem cell properties more often than those located at the edge after cell division. However, they were able to observe transmigration between these two compartments[21], suggesting that stem cell potential is not a cell-intrinsic property but can be conferred by cell positional external cues.

However, higher degrees of plasticity can only be observed when tissue homeostasis is lost. Intestinal damage can stir the default evolutionary paths of cells towards different directions[18]. In mice, ablation of Lgr5+ cells elicits an architectural remodelling of the intestine aimed to the replenishment of the crypts[22, 23]. Initially, it was thought that an alternative population of stem cells residing at the +4 position were responsible for the regeneration of the tissue[23, 24]. However, later studies showed that lineage-committed populations can respond to injury by reacquiring stem cell properties[22, 25, 26, 27, 28]. Indeed, even Krt20+ cells at the top of the colon crypts can repopulate the epithelium after Dextran Sulfate Sodium (DSS) treatment[29]. From an epigenetics perspective, intestinal epithelial cells keep their chromatin in an open conformation along the crypt-villus axis, which allows them to easily adapt to

changes in their surrounding environment despite of their level of differentiation[30, 31].

Several populations of cells, which some refer to as Regenerative Stem Cells (RSC), have been shown to activate and repopulate the intestine in settings of tissue repair[32, 33, 34]. Independent studies by Yui et al and Nusse et al contributed to the transcriptional characterization of RSC. The authors conducted bulk Ribonucleic Acid (RNA) sequencing in purified epithelial populations derived from mouse models of infection or tissue damage[33, 32]. Both reports concluded that conventional ISC markers (eg. *Olfm4, Lgr5*) were downregulated in RSC cells, whereas they reactivated the expression of fetal transcripts like *Sca1* and *Anxa1*. The reason for the existence of different pools of SC/progenitor cells is far for comprehended, but it is possible that distinct cell-intrinsic properties are selected for in settings of homeostasis and regeneration (see section 1.4.4 Inflammation).

1.5 Organoids as Tools for the Study of Intestinal Stem Cells

Organoid cultures exploit the capacity of ISC to self-organize in response to external cues[35]. This technique dates from 2009, when Sato et al established a protocol to grow tridimensional intestinal crypt-villus units, which they described as mini-guts, *in vitro* from single ISC[36]. The process starts with ISC forming villus-like cystic structures from which they extend crypt-like budding projections. Once formed, organoids contain SC as well as all the differentiated epithelial cell types found in the intestine. For mouse small bowel, the presence of Epidermal Growth Factor (EGF), Noggin and R-spondin3 is enough to sustain organoid culture, but colon organoids require extra Wnt3a, and the list grows even longer for human mini-guts, which require factors like TGF-β inhibitors and Prostaglandin E2(PGE2). Despite the complexity of this technique, in the adequate conditions, organoids can be maintained in culture unlimitedly. This makes them a useful tool to study the gradients that regulate SC fate, to model pathological settings *in vitro* and to screen new potential therapeutics[35].

1.6 Regulation of Intestinal Cell Fate

1.6.1 The Intestinal Stem Cell Niche

Within the intestinal crypts, there is a subset of cells that reside in proximity to the ISC and nurture them to ensure their good function. This framework is commonly referred to as the stem cell niche, and includes a combination of epithelial (Paneth cells) and stromal cells (fibroblasts, neurons, pericytes, endothelial, smooth muscle and immune cells)[1].

In addition to providing structural support, niche cells secrete molecules that confer cell identity (eg.like WNT3A or Gremlin-1). The four key pathways involved in this process are the Bone Morphogenetic Protein (BMP), Notch, EGF and WNT[1, 37]. Among them, in this piece of work I will focus on BMP and WNT, which generate opposing gradients from the base to the luminal side of the crypts. Stem-cell division and transit amplifying cell proliferation is driven by a high WNT/low BMP gradient in the lower third of the crypt whereas daughter cell differentiation and programmed cell death at the luminal surface is controlled by low WNT/high BMP expression[1].

1.6.2 Wnt signaling

The WNT family consist of nineteen secreted glycoproteins that act as morphogens and exert their functions by interacting with one of the ten cognate Frizzled receptors and an associate LDL Receptor Related Protein (LRP) 5/6 co-receptor. The production of Wnt proteins is a complex process involving post-translational modifications in the Endoplasmic Reticulum (ER) and Golgi Apparatus. Procupine is an acyltransferase that catalyses a palmitolation step in the ER essential for the maturation and secretion of the active form of WNT[38]. Upon binding of their receptors, WNT ligands regulate processes like cell fate determination, cell proliferation and asymetric cell division[39].

The wingless gene was first described in 1973 as a key director of limb formation in *Drosophila Melanogaster*. Since then, numerous studies highlighted the essential role of wingless and its murine orthologous, Wnt, during embryonic development and beyond[39]. During development, WNT acts by limiting cell differentiation as well as activating the expression of pluripotent transcription factors (eg. *OCT4, NANOG*)[40]. Among other tissues, WNT plays an essential role in intestinal development. Indeed, animals with a knock-out in the gene *Tcf7/2*, which codes for

the WNT downstream effector T-Cell Factor(Tcf)4, experience developmental abnormalities that are manifested in the lack of an ISC compartment and premature mortality[41].

After birth, the WNT pathway is one of the main regulators of intestinal homeostasis. In adult mice, defects in the WNT pathway compromise SC integrity and intestinal function to a greater or lesser extent depending on the magnitude of the disruption. When Pinto et. al expressed the secreted Wnt negative regulator *Dkk1* under the epithelial promoter *Villin*, they observed decreased proliferation and a lower number of crypts and shorter villi in the small intestine. Due to the patchy expression pattern of the *Villin* protomer, intestinal collapse was only found in confined regions, and these animals survived for 3 months without complications[42]. Adenoviral expression of *Dkk1* induced a more severe phenotype, which extended to the caecum and colon compartments, and lead to loss of intestinal architecture within seven days[43]. Deletion of the gene that codes for β-catenin, a core component of the Wnt pathway, has even more pronounced effects. In these animals, crypt loss occurred at day four, and the whole intestinal epithelium was lost six days after injection[9].

In addition, WNT is one of the main molecules that control intestinal cell plasticity. For example, *in vitro*, Delta Like Canonical Notch Ligand 1-expressing cells re-acquire stem cell properties upon exposure to Wnt3A[22]. Furthermore, reduction in WNT activity, either by treatment with a Porcupine inhibitor or by knockout of the WNT regulator Rala, hampers the ability of animals to recover from radiation injury[44, 45]. Ascl2 is a transcription target of Wnt, which expression is normally limited to the Lgr5+ SC compartment[46]. This transcription factor works in tandem with β-catenin and Tcf4 to activate genes that confer SC identity[47]. Murata et al recently demonstrated that ectopic expression of Ascl2 is essential for reprogramming of differentiated cells and replenishment of the ISC pool after Lgr5+ cell depletion mediated by γ-irradiation or diphtheria toxin[46].

There are several populations of epithelial and stromal cells involved in the production of intestinal WNT ligands. Among them, Wnt3a produced by Paneth cells has been shown to increase the efficiency of organoid culture *in vitro*[48]. In animal models, reduction in the number of Paneth cells translates into a decreased small intestine SC pool[48]. However, intestinal homeostasis is maintained when Porcupine is exclusively inhibited in the epithelial compartment, suggesting that stromal ligands are enough to sustain intestinal function[44].

In 2018, Shoshkes et al. and Degimenci et al. described two populations of fibroblasts that lie in close proximity to epithelial cells in the intestinal crypts. These

cells are characterized by the expression of the genes *Foxl1* and *Gli* and constitute essential sources of Wnt and Rspo ligands (eg. Wnt2b, Wnt5a, Rspo3)[49, 50, 51]. In mice, ablation of Wnt secretion from Foxl1+ cells resulted in reduced epithelial cell proliferation and collapse of the small bowel and colon compartments in only 3 days[49]. In the case of Gli+ cells, inhibition of Wnt production caused disappearance of stem cells and compromised tissue integrity in the colon after 3 weeks. However, effects in the small intestine were only observed when epithelial and stromal sources of Wnt were disrupted at the same time, which suggests that Paneth cells compensate for loss of Gli-derived WNT ligands in the small intestine[50].

From a mechanistic perspective, the Wnt pathway can be divided into the canonical and non-canonical arms.

1.6.2.1 Canonical Wnt pathway

The canonical arm of WNT signaling is centred around the stabilisation of β-catenin and the activation of its transcriptional programme[39]. β-catenin occupies two distinct locations within a cell. Firstly, it can be found at the cell membrane, forming the E-cadherin-catenin adhesion complex, which is involved in the maintenance of the cell architecture[52]. Secondly, β-catenin exists in a cytoplasmic/nuclear form that acts as the main intracellular mediator of WNT signaling.

In the absence of environmental WNT, cytoplasmic β-catenin is repressed by a destruction complex formed by Axin, Adenomatous Polyposis Coli (APC) and Glycogen synthase kinase 3 β (GSK3β). This complex operates at two levels. Firstly, β-catenin is sequestered and kept away from its direct targets, the TCF/Lymphoid Enhancer Factor (LEF) proteins. Secondly, β-catenin is phosphorylated by GSK3β and marked for ubiquitination by β-TrCP. In the absence of nuclear β-catenin, Groucho associates to TCF/LEF factors and promotes recruitment of histone deacetylases, which repress the expression of WNT target genes[39].

When WNT ligands interact with Frizzled receptors and LRP co-receptors, Dishevelled is recruited to the membrane and inactivates the destruction complex, thus preventing β-catenin degradation and allowing its translocation to the nucleus. Once there, β-catenin can displace Groucho and form an activating complex with TCF/LEF proteins. This allows for the recruitment of co-activators (eg. BCL9, CPB/p300) and induction of the Wnt transcriptional programme[39](Fig.1).

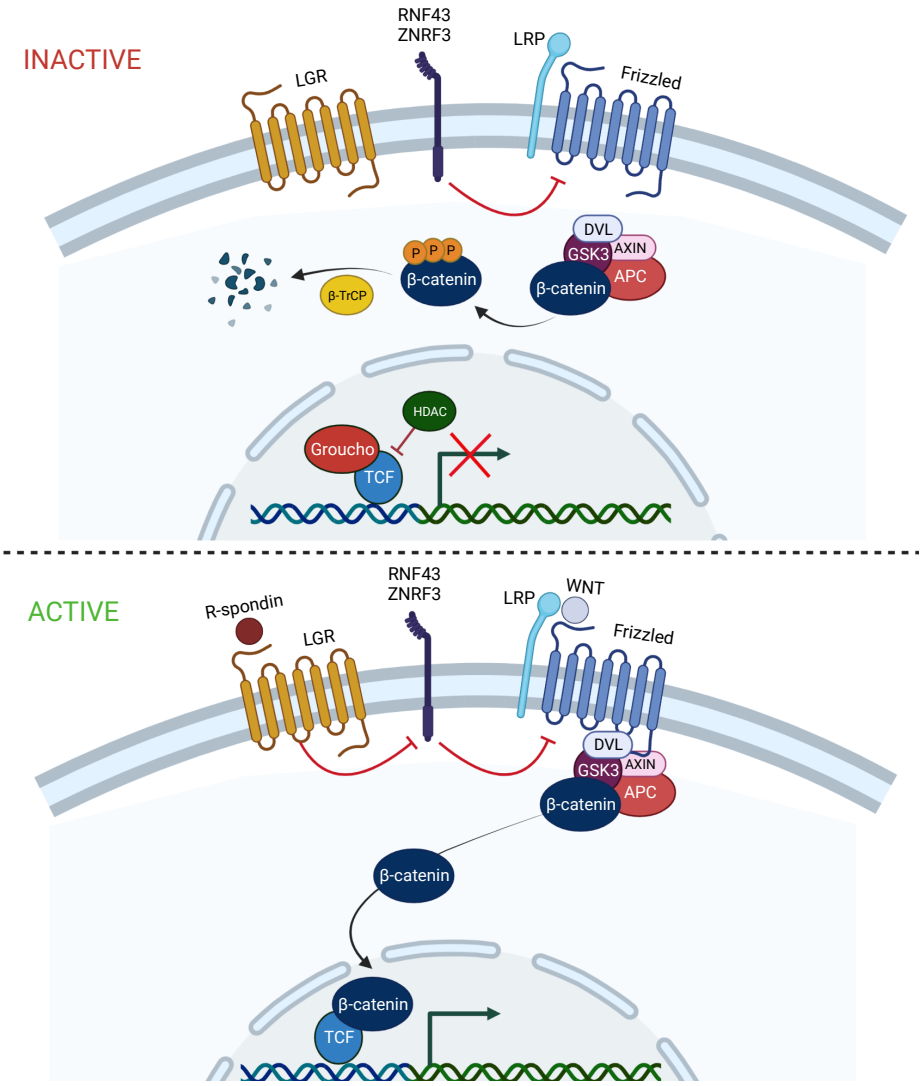

Figure 1: Schematic representation of the WNT pathway.

1.6.2.2 Non-canonical Wnt pathway

The non-canonical arm includes all the downstream effects of Wnt signaling that do not depend upon β-catenin and TCF/LEF. It is activated by a subset of Wnt ligands (eg. WNT5A, WNT11) which act through specific receptors (eg. FZD5, ROR2). Among the effectors of this arm, we find the Planar Cell Polarity (PCP) and calcium signaling (WNT-Ca) pathways[53].

The PCP acts through activation of the Guanosine Triphosphate(GTP)ase Ras-related C3 botulinum toxin substrate 1, the RAS homolog gene family member A and the c-Jun N-terminal Kinase. Among other functions, this pathways regulates cell polarity during development[54], but it has also been shown to contribute to

metastasis in the context of cancer[53].

The downstream effectors of the WNT-Ca pathway are the heterotrimeric G proteins and phospholipase C, which mediates the intracellular release of Ca2+ ions and consequent activation of genes involved in cell migration and cell fate[54].

Non-canonical WNT ligands have been proven to play a vital role in the recovery from large-scale intestinal damage. Miyoshi et al showed that Wnt5a, through activation on TGF-β, suppresses epithelial proliferation and induces the formation of clefts. This enables crypt fission, a process by which the intestine generates new crypt units[55].

In addition to ligand availability, WNT activation is subjected to further mechanisms of control. On the one hand, some WNT direct targets are negative regulators of the pathway that act by limiting the extent and duration of WNT activation (eg. *AXIN2, NOTUM*)[39]. On the other hand, the Lgr/R-spondin axis can modulate the sensitivity of cells to WNT.

1.6.2.3 Lgr/R-spondin axis

The R-spondin proteins belong to the superfamily of thrombospondin type 1-containing proteins. Among them, R-spondin 3 was the first one to be described in 2002. Soon after, reports about R-spondins 1,2 and 4 came to light[56]. At the time, experiments using cell lines suggested a link between R-spondin and the WNT pathway[57]. In addition, R-spondin was used as a complement to grow intestinal organoids *in vitro*[36], which suggested a role for these proteins in SC regulation. Nevertheless, R-spondin proteins reminded orphan until 2011, when several reports described Lgr4-6 as R-spondin1-4 receptors and confirmed their role as modulators of the WNT pathway[58, 59, 60].

The mechanism by which R-spondin ligands potentiate Wnt signaling involves two ubiquitin ligases located in the cell membrane: Ring Finger Protein 43 (RNF43) and Zinc and Ring Finger 3 (ZNRF3). In the absence of R-spondin ligands, these proteins constitutively ubiquitinate Frizzled receptors, targeting them for degradation. However, when a R-spondin ligand binds to an LGR receptor, they form a complex with RNF43 and ZNRF3 that inhibits their activity (Fig.1). This increases the availability of Frizzled receptors able to interact with WNT ligands, and consequently, the sensitivity to WNT of that particular cell[39]. In addition, by slowing the turn-over of Frizzled receptors, R-spondin can modify the extension of the WNT

gradient. As WNT is retained longer in the cell membrane, ligands can be transmitted to descendent cells and WNT can migrate further along the intestinal axis[1, 61].

The Lgr/R-spondin axis is an essential regulator of ISC homeostasis, and defects in this pathway results in intestinal stem cell loss. For example, combined deletion of Lgr4 and Lgr5 genes under the Villin promoter or treatment with a mixture of R-spondin2 and R-spondin3 antibodies lead to ablation of intestinal Lgr5+ cells in mouse models[58, 62], which is similar to what we observe in mice treated with Porcupine inhibitors[63].

This axis has also been implicated in cell fate regulation in the context of intestinal injury[51, 29]. In the mouse colon, DSS-mediated ablation of the stem cell compartment comes together with an increase of *Rspo3* expression in the mesenchymal compartment. In WT animals, Lgr4+ cells at the top of the crypts are able to de-differentiate and occupy crypt positions after DSS treatment. However, animals missing mesenchymal-derived Rspo3 experience crypt loss and a decrease in proliferation, which impedes their recovery[29].

The implication of R-spondin in SC regulation may be expected of a WNT modulator, but are the functions of WNT and R-spondin equivalent? To answer this question, Yan et.al used *in vivo* pharmacological perturbation of WNT and R-spondin signaling[64]. Their experiments showed that increased levels of R-spondin, but not WNT, resulted in an enlarged SC pool. Interestingly, Lgr5+ cells did not leave the crypt, an effect that the authors coined as crypt "trapping", and which co-occurred with increased proliferation in the TA compartment. Furthermore, their experiments showed that a basal level of WNT signaling was needed to prime the R-spondin response and, when used in combination, both types of agonists acted synergistically to induce a more pronounced crypt hyperplastic phenotype. However, one could not compensate for the loss of the other. This evidence suggests that both WNT and R-spondin ligands act in a co-operative, non-redundant fashion to regulate the ISC pool.

1.6.3 BMP pathway

In striking contrast to WNT, BMP is a master regulator of cell differentiation. Initially, BMP functions were described in the context of bone formation, but it later became apparent that its role expanded to many other tissues and was essential across stages of development as well as in adult homeostasis[65].

BMP proteins belong to the TGF-β superfamily. There are twenty BMP ligands, but BMP2 and BMP4 are the most prominent ones in the intestinal compartment. They are produced by the epithelium as well as mesenchymal cells intermingled between the crypts or located in the intervillus stroma, and exert their functions by interaction with BMP receptors type II[1].

When a BMP receptor type II binds an intestinal BMP ligand, it initiates a phosphorylation cascade involving activation of type I receptors, which in turn activate Mothers Against Decapentaplegic homologues (SMAD) 1, 5 or 8. These proteins form a complex with the co-SMAD4 and migrate into the nucleus, where they activate the expression of BMP target genes like *ID1*. In addition, Smad complexes bind and repress the transcription of SC markers like *Lgr5* and *Sox9*[65, 66].

BMP levels are tightly regulated across the crypt-villus axis. McCarthy et al described different populations of fibroblasts implicated in the regulation of this BMP gradient. On the one hand, Platelet Derived Growth Factor Receptor α (PDGFRA)hi telocytes secrete BMP ligands at the base of the villi, which enables the differentiation of the progenitors which enter this compartment. On the other, PDGFRlo CD81+ trophocytes at the base of the crypts secrete BMP inhibitors, like Gremlin-1, which limit levels of BMP signaling near the SC pool[67].

Furthermore, work from our laboratory highlights the importance of BMP inhibitors in the acquisition of plastic properties. Koppens et al used mouse models of Bmp4 and Gremlin-1 overexpression to investigate how the manipulation of the BMP gradient affected intestinal regeneration. Global Bmp4 overexpression resulted in a reduced Lgr5+ SC pool. Furthermore, specific Bmp4 overexpression under the secretory cell promoter *Atoh* altered the fate trajectory of secretory progenitors, which experienced reduced levels of proliferation and an accelerated terminal differentiation. Importantly, increased Bmp4 expression in this population was associated to a decreased regenerative potential upon DSS treatment. Contrarily, increased Gremlin-1 expression practically abolished DSS-mediated ulceration, although these animals developed an hyperplastic phenotype[68].

In homeostasis, gradients of secreted morphogens are enough to ensure an adequate SC function. However, in settings of injury or cell ablation, additional mechanisms enter the fray.

1.6.4 Inflammation

As mentioned before, in the context of the intestine, cell plasticity is normally restricted to settings of regeneration. Tissue damage is followed by the development of

an immune response, which will alter the characteristics of the tissue and facilitate the re-acquisition of SC properties[37].

Any immune response starts by the recognition of noxious stimuli by Pattern Recognition Receptors (PRRs). There is a vast array of PRRs, such as Toll-like receptors (TLR), which recognise specific Damage (eg.extracellular Deoxyribonucleic Acid(DNA), F-actine, ATP) and Pathogen (eg. lipopolysaccharides, peptidoglycan)-Associated Molecular Patters (DAMPs and PAMPs respectively). Generally, these receptors conduct their signalling through conserved intracellular mediators like Myeloid differentiation primary response 88 (MyD88). Rakoff-Nahoum et al. observed that animals defective for MyD88 or depleted from commensal bacteria were more susceptible to DSS, suggesting that the recognition of microbiota signals through TLR aids intestinal response to injury[69]. Similar results have been observed when elements down the TLR pathway were compromised, such as Nuclear Factor κ-light-chain-enhancer of activated B cells (Nf-κb) essential modulator or the inhibitor of Nf-κb kinase complex[70, 71].

When PRRs are activated, they elicit an immune response that entails the recruitment of immune cells and secretion of inflammatory mediators, some of which exert direct effects upon the Lgr5+ SC compartment. A study from Richmond et al showed how inflicting intestinal acute inflammation in mice models lead to the apoptosis of Lgr5+ progenitors at least partially through IFN-γ and TNF-α-dependent activation of the JAK-STAT1 pathway[72]. Further studies have confirmed the pro-apoptotic effects that inflammatory pathways exerts upon Lgr5+ ISC[15, 73, 74], either by promoting an increased expression of apoptotic ligands, or by secretion of molecules like Dkk1[73, 74]. However, in situations of inflammatory-driven Lgr5+ cell death, other populations of quiescent SC activate to replenish the SC pool[72]. In addition, *in vitro* treatment of IECs with activated immune populations or pro-inflammatory cytokines (i.e. IFN-γ) drives the down-regulation of conventional SC markers like *Olfm4* in favour of fetal transcripts like *Ly6a*[33, 75]. Based on this observations, in the context of inflammation, the expansion of a RSC population may be a mechanism aiming to ensure the continuity of the ISC compartment.

In addition to inflammation, WNT and BMP, other cues and signaling pathways have been proven to regulate intestinal stem cell fate in settings of tissue damage (eg. NOTCH, TGF-β, EGFR). Interestingly, many of these signals seem to converge in a molecular node that has been heavily implicated in plasticity and regeneration: the Hippo-Yes-Associated Protein 1(YAP)/TAZ pathway.

1.6.5 The Hippo-YAP/TAZ Pathway

Even though the Hippo-YAP/TAZ pathway is not required to maintain ISC homeostasis, it is pivotal to tissue repair. Yap is activated early after intestinal damage and limits stem cell apoptosis, thus contributing to long-term maintenance of the ISC compartment[76, 77, 78].

From a mechanistic point of view, the Hippo-YAP/TAZ pathway is reduced to a kinase cascade that involves the proteins Ste20-like kinases 1 and 2 (MST-1/2) and the large tumour suppressor kinase 1 and 2 (LATS-1/2) as well as their adaptor proteins. Activation of this node results in phosphorylation and cytoplasmic sequestration of two homologous transcription co-activators: YAP and TAZ. However, when this inhibition is relieved, YAP/TAZ can translocate into the nucleus, where they bind the transcription factors TEA Domain family member (TEAD)1-4. Then, YAP/TAZ-TEAD complexes interact with gene enhancers and chromatin-remodelling factors to induce or repress the expression of genes involved in proliferation, migration and stemness. Unlike other pathways, which are activated or repressed by ligand-receptor interactions, the Hippo-YAP/TAZ pathway constitutes an integrating node for a range of cell-intrinsic and cell-extrinsic signals[79] (Fig.2).

To start with, YAP and TAZ can be activated by a wide range of mechanical inputs, including cell shape, rigidity of the Extracellular Matrix (ECM) and shear stress. The mechanism by which these signals are transduced intracellularly has not been fully elucidated, but it involves Rho GTPases and the Focal Adhesion Kinase, and requires the structural re-organization of the F-actin cytoskeleton[80]. In line with this, *in vitro* culture of organoids in collagen, which is highly upregulated during intestinal repair, promotes the acquisition of RSC characteristics[32].

Coming back to inflammation, at least some of its effects upon regeneration are exerted through activation of YAP/TAZ. For example, TLR stimulation has been shown to activate EGFR signaling and induce expression of Cox-2 by IECs, with consequent production of PGE2 and YAP-dependent acquisition of plastic features[81, 82]. In addition, a recent report suggested that the pro-inflammatory cytokine Il-6 can activate Yap through interaction with the co-receptor gp130 and the consequent activation of Src Family Kinases[78].

Additional factors that modulate the activity of the Hippo-YAP/TAZ node include growth factors. Work by Pefani et al revealed a link between this pathway and TGF-β. They demonstrated that TGF-β-mediated degradation of Ras Association Domain Family Member 1 (RASSF1), facilitates the interaction between YAP and SMAD2.

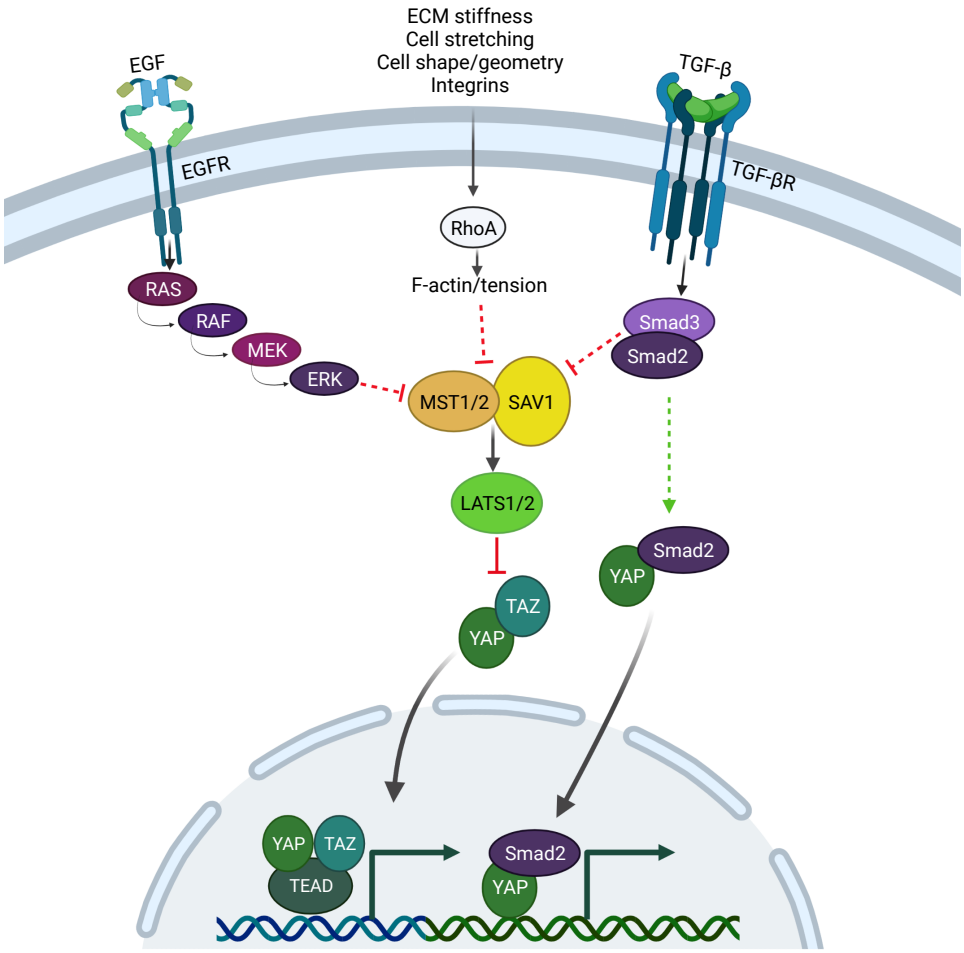

Figure 2: Schematic representation of the Hippo/YAP-TAZ pathway.

These proteins form a transcription complex that migrates into the nucleus to drive the activation of an oncogenic genetic programme[83].

Moreover, signaling through G-protein-coupled receptors and the metabolic state of the cells can also modulate the Hippo-YAP/TAZ pathway. Finally, YAP/TAZ cross-talks with many of the pathways involved in regulation of intestinal homeostasis, like WNT, Hedgehog and Notch[79].

1.7 Intestinal Cell Death

Intestinal cell death is tightly regulated along the crypt-villus axis, with different pools of cells using specific regulatory mechanisms. Within the ISC pool, the ARTS/X-linked inhibitor of apoptosis module has been shown to modulate the levels of apoptosis at the base of the crypt[84]. In homeostasis, caspases are inhibited by Inhibitor

of Apoptosis (IAP). However, upon cell death-inducing stimuli like irradiation or staurosporine, ARTS act on IAP to release that inhibition, leading to ISC death by apoptosis. Importantly, apoptosis is a controlled form of cell death which does not entail the exposure of damage signals or pro-inflammatory cytokines. As a consequence, apoptotic tissues can avoid immune detection with relative ease, which is the reason why apoptosis is commonly classified as a immune silent form of cell death[85].

In healthy intestines, the majority of IECs die when they reach the top of the villi by anchorage-dependent apoptosis, also known as Anoikis. This cell death mechanism is activated by pathogen-associated stress or by the mechanical forces to which the epithelium is subjected. It involves the reorganization of the actin cytoskeleton and the release of the extracellular contacts, which facilitates the migration of the cell towards the intestinal lumen[85].

However, when the PRR of IECs are engaged, IECs activate an alternative, gasdermin-dependent type of cell death called Pyroptosis[85, 86, 87, 88]. Gasdermins are a family of effector proteins that contain a N-terminal pore-forming domain. In steady state, this region is inhibited by interaction with their C-terminal domain, but caspase-mediated cleavage can release this repression, which causes membrane permeabilization and cell lysis[88]. Gasdermin cleavage commonly involves the activation of inflammatory caspases (1 and 4 in humans, 1 and 11 in mice), although it can also be catalysed by apoptotic caspases (caspase-8)[88]. Moreover, pyroptosis involves the activation of pro-inflammatory mediators (eg. Il-1β, Il-18) which are leaked into the extracellular matrix upon cell lysis. This release of inflammatory cues, added to the exposure of DAMPs, drives the recruitment of immune populations, which is the reason why pyroptosis is classified as an immunogenic type of cell death[85, 88].

The cell death rate in the villi tips is regulated by the turnover of the intestinal epithelium. Increase proliferation and migration of cells upwards the walls of the villi accelerates the extrusion of cells at the villi tips via mechanical forces[85]. An elevated degree of cell death which is not matched by an equal increase in proliferation rate leads to villi atrophy[89]. On the other hand, a reduced cell death rate can contribute to a net growth of the tissue. As an example, colorectal tumours with a p53 deficiency are protected against cell death, which facilitates their uncontrolled cell expansion[90].

1.8 Colorectal Cancer

CRC is the third most common malignancy globally, with almost two million patients being diagnosed every year, and its incidence is expected to more than double by 2035.

In terms of mortality, CRC is the fourth most common cause of cancer-related deaths, which translates into almost 900.000 deaths annually[91, 92].

In line with other complex disorders, CRC aetiology encompasses a range of environmental and genetic factors. The disease has an heritability of 12-35 per cent, but only between five and seven per cent of CRC cases have an underlying defined genetic syndrome[93]. Examples of this are the Familial Adenomatous Polyposis, caused by mutations in *APC*[94, 95] and Juvenile Polyposis (JP), which is initiated by mutations in *SMAD4* or *BMPR-1A*[96, 97].

The current model of colorectal tumorigenesis was proposed by Fearon and Vogelstein in 1990[98]. In this model, normal tissue transitions to benign lesions (adenomas) and, latter on, malignant tumours (carcinomas) by successive accumulation of driver mutations. Although the order of these events is highly variable, the majority of colorectal lesions acquire *APC* alterations early in the process, with activation of *RAS* and loss of *TP53* happening in later stages[98]. This sequence of events is known as the traditional adenoma–carcinoma pathway, and is observed in 70-90 percent of lesions. In the remaining 10-20 per cent of cases, the order of the mutational events tends to differ, with earlier activation of Kirsten Rat Sarcoma virus(KRAS) or B-Raf(BRAF) and acquisition of epigenetic instability. The early lesions characteristic of these pathways are referred to as Tubulovillous Adenoma (TVA) and serrated adenomas respectively[40].

It is estimated that 10-15 years pass from the emergence of an aberrant crypt to the development of CRC[40]. With this window of opportunity, the key to successful disease management lays with early detection. This is the reason behind the worldwide implementation of screening programmes aimed at detecting and endoscopically removing precancerous lesions[40]. In established tumours, available treatments include surgical removal of local lesions, radiotherapy, chemotherapy and targeted therapy. However, the prognosis of CRC is far from ideal, specially when the disease has reached advanced stages. Metastatic patients account for a 25 per cent of CRC diagnoses and have a 5-year survival rate of 12 per cent[92]. In this scenario, targeted therapies have emerged as a potential way to improve patient outcomes.

The first targeted therapies approved for CRC were the EGF receptor inhibitor cetuximab and the anti-angiogenesis biologic bevacizumab in 2004. Since then, other drugs have become first-line CRC therapeutical options. However, the repertoire of molecular mechanisms targeted by these drugs has barely expanded. With exception of immune checkpoint inhibitors (Ipilimumab, Nivolumab), which have been proven to

work in a subset of hypermutated CRCs[99], the rest of agents still target angiogenesis and the EGF pathway[92].

Various other pathways like TGF-β, Hedgehog, BMP and WNT constitute possible targets for new therapeutics. Several agents directed against these candidates have been tested in clinical trials, but none has been efficient enough to enter the clinic. One of the main challenges when developing targeted drugs is determining the subset of patients that will respond to the novel agent. Failure to deliver the drug to an adequate population will yield unsatisfactory results. For this reason, a comprehensive characterization of the molecular mechanisms underlying CRC pathogenesis is needed in order to develop more sensitive and specific biomarkers and more effective targeted therapies.

Finally, response to treatment is commonly limited by the existence of resistant clones within the tumour. Recently, genomic analysis of pre and post-treatment samples has been implemented to assess changes in the clonality of lesion in response to therapy[100]. This strategy aims to identify the genetic alterations that drive therapy resistance. Ultimately, pharmacological inhibition of the processes that mediate tissue adaptation to therapy could increase the number of patients that benefit from the available treatments.

1.8.1 Molecular characterization of colorectal cancer

Over the past 10 years, numerous researchers devoted themselves to the molecular classification of colorectal malignancies[101, 102, 103, 104, 105, 106]. In 2015, the CRC Subtyping Consortium lead a collaborative effort to integrate previous attempts aimed to the transcriptional subtyping of CRC. They used a compendium of datasets and published algorithms to find recurring patterns, thus increasing the robustness of the classifier[107].

This enterprise yielded four distinct molecular clusters, named Consensus Molecular Subgroup (CMS)1-4, with distinctive biological characteristics. The CMS1 subtype included hypermutated samples, was enriched for microsatellite instability and correlated with a strong anti-tumour immune response. CMS2, also known as the canonical subtype, included cases with strong activation of the WNT and MYC pathways. CMS3 tumours were characterized by KRAS activating mutations and consequent metabolic alterations. Finally, CMS4 was defined by an abundant stromal composition[107].

In addition to these molecular characteristics, belonging to a CMS subgroup carries important clinical implications: CMS4 patients present with a worse overall and

relapse-free survival, whereas CMS1 patients have a higher mortality after recurrence. Moreover, pre-clinical studies in cell lines and patient organoids have highlighted the potential of these CMS clusters to guide therapeutic intervention[108, 109].Based on this evidence, some clinical trials have been exploring the use of the CMS classification as a prognostic or predictive marker[110, 111].

One important characteristic of the CMS classifier is the impossibility to distinguish which cluster characteristics are specific to the epithelium and which are embedded in the stroma. This is of particular importance in the case of CMS4, because most of the genes that define this cluster have a stromal origin[112]. Isella et al found a way around this: they used patient-derived xenografts to find out which transcripts are specifically expressed by neoplastic cells, and defined a classifier purely based on epithelial traits: the Colorectal cancer Intrinsic Subtypes (CRIS).

The CRIS classifier comprises five clusters. To start with, CRIS A tumours are mostly microsatellite unstable and have mutations in KRAS or BRAF, whereas CRIS B contains dedifferentiated cancers with activation of the TGF-β pathway. On the other hand, CRIS C has chromosomal instability and EGFR activation, and CRIS D contains cancers with an hyperactive WNT pathway. Finally, CRIS E is characterized by a Paneth cell-like phenotype and mutations in *TP53*[113]. As expected, the authors reported little overlap between the CRIS and CMS classifiers. However, the contribution of the CMS clusters to the CRIS subtypes is not proportional to their prevalence. For example, when applying this classifier to The Cancer Genome Atlas (TCAG) data, CRIS B contained mostly cancers that were labelled as CMS1 or 4, but very few cases of the reminding two CMS subtypes.

1.9 Cancer Stem Cells

For a long time, researchers have cultivated the idea that cancers have a stem cell origin. Already in the 19th century, R. Virchow and J. Cohnheim discussed the possibility that tumours arose from embryonic tissue[114, 115]. In 1907, Askanazy, M. proposed that ovarian teratomas developed from either a multipotent single cell or group of progenitors[116]. However, the study of the role of stem cells in human malignancy laid low until the end of the century.

In the 1990s, John E. Dick et al. were able to demonstrate the tumour-initiating properties of cells isolated from leukaemia patients by transplanting them into Severe Combined Immune-Deficient mice[117, 118], which reignited interest in the topic. Latter, these findings were extended to solid tumours[119, 120]. These studies strongly

suggested that cancer initiation and growth relied on a population of cancer cells with self-renewal capacity and multi-directional differentiation potential: the so called Cancer Stem Cells (CSC).

In 2007, several groups were able to isolate colorectal CSC from patient-derived tumours. Interestingly, these cells expressed conventional ISC markers, like CD133 and CD44[121, 122, 123]. The work of Vermeulen et al pointed out that colorectal CSC were characterized by activation of a β-catenin-driven transcriptional profile[124], a trait which is also shared with ISC in healthy crypts[8].

Building upon this work, Barker et al demonstrated that the introduction of *APC* mutations in Lgr5+ ISC, but not in TA or differentiated cells, resulted in β-catenin activation and adenoma formation in mice within 3 to 5 weeks[125]. In addition to Lgr5+ progenitors, other populations of cells with stem cell properties, like Bmi1+ and Leucine-rich repeats and immunoglobulin-like domains protein 1-expressing cells, have been shown to possess cancer-initiating potential[126, 127, 128]. These studies fuelled the bottom-up theory of colorectal tumorigenesis, which sustains that colorectal adenomas originate from conventional stem cells at the crypt compartment.

However, and despite all of the above evidence, the origin of CRC reminds controversial. Early studies noticed that, in some human malignancies, dysplasia traits and mutational changes localized to the luminal side of the lesions, whereas Crypt Base Columnar (CBC) cells reminded normal[129]. Furthermore, in some precursor lesions, namely Traditional Serrated Adenomas (TSA), we can find ectopic epithelial invaginations that grow perpendicularly to the crypt-villus axis[130]. These structures were coined by Haramis et al as ectopic crypts, and originate when the gradients that regulate ISC homeostasis are perturbed [131, 132]. Observations like these sparked a different school of thought which argues that CRC can be initiated in areas populated by a priori differentiated cells: the top-down model for adenoma morphogenesis[129]. But how can differentiated cells acquire the characteristics needed to succeed in cancer initiation?

Even though acquisition of plastic characteristics is normally controlled and limited to situations of regeneration, it can be co-opted in malignant settings. In their work, Schwitalla et al used a mouse model with activating mutations in β-catenin and NF-κB to study the interaction between WNT and inflammation in stem cell regulation and tumour initiation[133]. They reported that the NF-κB target RelA/p65 could potentiate β-catenin activity via recruitment of the transcriptional co-activator CREB-binding protein. This strong WNT stimulation reactivated the expression of stem cell markers in the villi compartment (eg. *Lgr5, Ascl2*), which resulted in the

appearance of ectopic crypts populated by Lgr5+ progenitors. This evidence proved that differentiated cells in the intestinal epithelium can re-acquire stem properties and act as cancer-initiating cells when alterations in WNT and inflammatory pathways co-occur. Nevertheless, the possibility that multiple mutations accumulate in the short lifespan of differentiated cells is low.

A way to circumvent this limitations was pointed out by the work of Westphalen et al. They studied a population of quiescent, long-lived Doublecortin Like Kinase(Dclk)1+ Tuft cells that have a differentiated phenotype, but can contribute to regeneration after epithelial injury. As they showed, deletion of *APC* under the tamoxifen inducible Dclk1 promoter failed to fuel the generation of adenomas even 18 months after injection. However, upon DSS treatment, these animals developed invasive colon carcinomas[26]. Here, a more feasible scenario emerges: the lifespan of Dclk1+ cells allows for the accumulation of mutations over time, whereas inflammation drives the reactivation of proliferation and acquisition of tumour-initiating properties.

Overall, there is still work to do before we fully understand the contribution of different pools of SC/progenitors to malignancies in general, and CRC in particular.

1.10 Cell Fate Regulation in Colorectal Cancer

1.10.1 WNT Signaling in Colorectal Cancer

The first link between Wnt and cancer was observed in the mammary glands, where activation of Wnt1 lead to hyperplasia and tumour development[134]. Nevertheless, more than a decade passed before Wnt was shown to be implicated in the pathogenesis of CRC. In 1991, the work from Groden et al. and Kinzler et al. revealed that the hereditary syndrome Familiar Adenomatous Polyposis was caused by mutations in the *APC* gene[135, 94]. These patients present in the clinic with early intestinal polyposis and have a higher predisposition to develop colorectal malignancies. At the time, mutations in *APC* were also described in sporadic forms of CRC[94, 95].By 1997, activating WNT mutations were recognized as initiation events in colorectal tumorigenesis[136, 137].

Since then, studies aiming to characterize the role of WNT signaling in CRC have increased in complexity. In 2012, a comprehensive molecular study by the TCGA Network revealed that more than 94 percent of human CRCs had alterations in the Wnt pathway[138]. Among them, the majority consisted of single nucleotide variants (eg.*APC, CTNNB1*), but occasional fusion events (eg.*RSPO3, TCF7L2*)[138, 139]

and targeted hyper-methylation of negative regulators (eg. *AXIN2, NKD1*)[140] were also observed. Moreover, Wnt equilibrium can be altered by uncontrolled production of stromal ligands (eg. R-spondin)[140].

Alterations in the WNT pathway can be divided in Ligand-Dependent (LD) or Ligand-Independent (LI) based on whether they occur upstream or downstream the Frizzled receptors respectively[140]. Interestingly, the distribution of LD and LI alterations is not homogenous across colorectal polyp subtypes: TVAs tend to acquire activating mutations in *APC* or *CTNNB1*, whereas serrated lesions have common *RNF43* mutations or RSPO3 fusions[141]. In addition, LD and LI forms of WNT disruption are mutually exclusive in human colorectal tumours[140]. Collectively, this evidence suggests that, from a pre-cancerous stage, there is a selective pressure for specific forms of WNT dysruption which translates into different evolutionary paths.

As previously mentioned, WNT dysregulation can act as an initiating event in colorectal transformation[136, 137]. In animal models, intestinal epithelial WNT disruption in the canonical pathway or the R-spondin axis leads to the generation of adenomas[128, 142]. Sansom et al showed that the introduction of *APC* mutations in the intestinal epithelium under the Cyp1A promoter increases cell proliferation and inhibits differentiation and migration. This results in the generation of hyperplastic crypts populated by progenitor cells that are unable to exit the compartment[128]. Taking into account that colorectal cancer relies on the acquisition of multiple mutations over a long time [143], expansion of the stem cell pool and retention of stem cell identity via activation of WNT may be permissive factors for the emergence and development of intestinal tumours.

Furthermore, WNT has been proven to be crucial in already established malignancies. Indeed, restoring normal levels of WNT activation could induce cell differentiation and, in some cases, tumour regression[40]. For example, R-spondin-driven tumorigenesis can be reverted by using Porcupine inhibitors or an anti-RSPO3 antibody[144, 62]. Experiments by Dow et al showed that, even in invasive APC-driven lesions with additional mutations in *Kras* and *Tp53*, reversion of *APC* disruption caused tumour disappearance and restoration of intestinal homeostasis[145].

Despite this promising evidence, targeting WNT is not straightforward. For starters, LI tumours do not respond well to Pocupine inhibitors and R-spondin blocking antibodies. Even in LD tumours, the success of this strategy relies in finding a therapeutic window between the levels of WNT needed to sustain homeostasis and those that support tumour growth. Providing a tailored level of inhibition is vital in

order to minimize the side effects intrinsic to WNT disruption. So far, Porcupine inhibition is the most established strategy to therapeutically target WNT signaling[39], but a better understanding of the WNT pathway and its heterogeneity in disease settings is important to determine which patients will respond to treatment and to develop more specific WNT therapeutics.

1.10.1.1 Tumorigenesis Initiated by R-spondin3

The increased expression of R-spondin proteins in malignancies can be a result of genomic rearrangements in epithelial cells, like *PTPRK-RSPO3* translocations[139] or aberrant stromal expression of R-spondin ligands[140, 146]. Hilkens et al developed a conditional Rspo3 overexpressing mouse model by placing the *Rspo3* gene below the control of the inducible Lgr5-GFP-CreERT2 recombinase. Within two months from the recombination, these animals displayed epithelial hyperplasia and sporadic appearance of adenomas and adenocarcinomas. This came together with expansion of Lgr5+ stem cells, Paneth cells and Lgr4+ cells. Upon lineage tracing of Lgr5+ cells, the authors observed that these progenitors did rarely leave the base crypts, and suggested that a Lgr5- cells might be the main players behind Rspo3-driven transformation[142]. Importantly, cancers which do not have mutations downstream the WNT or LGR/RSPO axis, but have aberrant expression of R-spondin ligands, respond well to WNT or R-spondin inhibition[62, 144].

1.10.2 BMP Signaling in Colorectal Cancer

As in the case of WNT, the first link between BMP and CRC was found in patients of a rare hereditary disease known as Juvenile Polyposis. JP is an autosomal dominant polyposis syndrome that confers an increased risk to develop CRC. Approximately 50 percent of the JP patients have mutations in either *SMAD4* or the BMP receptor *BMPR1A*. In mice, ectopic epithelial expression of the Bmp inhibitor Noggin or mutations in *Bmpr1a* phenocopy the pathological alterations observed in JP patients[131, 147]. Collectively, this data pointed to a role of the BMP in hereditary CRC.

Further studies confirmed the involvement of this pathway in sporadic forms of this disease[148]. Generally, BMP tends to be inactivated in CRC. This can be achieved through different mechanisms, including mutations in downstream effectors (eg.BMPR2, SMAD4), methylation of BMP ligands (BMP3) and overexpression of secreted BMP inhibitors (Gremlin-1)[132, 98, 149, 138, 148, 150]. However, BMP can increase cancer aggressiveness in advanced disease stages. In mesenchymal tumours,

BMP ligands induce invasion and metastasis[151, 152], which is at least partially mediated by activation of Notch[153].

1.10.2.1 Tumorigenesis Initiated by Gremlin-1: HMPS and the *Vil1-Grem1* mouse model

As mentioned before, Gremlin-1 expression is confined to a thin layer of fibroblasts located below the crypts, and it is upregulated in inflammatory settings[68]. However, this pattern of expression can be modified during tumorigenesis. *GREM1* expression has been reported to be increased in colorectal lesions, which entails the disruption of the homeostatic gradients that regulate cell identity, and a net increase of mutationally susceptible progenitors. In particular, the stroma of some established cancers, specially CMS4 tumours, presents higher levels of *GREM1* expression, whereas this protein is aberrantly expressed by the epithelium of some early lesions, namely TSA polyps and a rare hereditary form of CRC, the Hereditary Mixed Polyposis Syndrome (HMPS), which will be discussed shortly[132, 154]. Furthermore, high GREM1 expression is a poor prognostic factor in rectal cancers[155].

Aberrant expression of Gremlin-1 is enough to initiate colorectal tumorigenesis. The HMPS is an hereditary form of CRC caused by a duplication above the *GREM1* promoter which results in ectopic expression of this gene in the epithelial compartment. As a consequence, patients develop a mixed polyposis characterized by the appearance of ectopic crypts[156]. Davis et al recreated this phenotype in mice by expressing *GREM1* under the epithelial promoter *Villin*[132], thus generating a $\text{Grem1}^{\text{VillOE}}$ animal which I will refer as the *Vil1-Grem1* (VG) mouse. Interestingly, transcriptomic analysis of polyps formed in VG animals revealed that they express a TA programme, suggesting that these lesions may not originate from CBC cells[132]. Although the genomic alterations which underpin this condition are very well understood, the molecular mechanisms activated by aberrant Gremlin-1 expression await to be clarified.

1.10.3 YAP Signaling in Colorectal Cancer

Although mutations in Hippo-YAZ/TAZ are rare, this pathway is activated generally across cancers and particularly in CRC[157], which suggest that indirect YAP/TAZ activation may be evolutionary selected. For example, mutations in *KRAS* and *p53* can drive YAP/TAZ activation and the conversion of cells towards a RSC phenotype[158, 159, 160]. In addition, methylation silencing of Hippo-YAP/TAZ compo-

nents and/or modulators, like *RASSF1*, may contribute to activation of the pathway in neoplastic settings[161].

It is worth noting that YAP and TAZ are transcriptional targets of SMAD2/3, which sit downstream the TGF-β receptor. In organoid models of CRC, treatment with TGF-β elicited the nuclear translocation of YAP/TAZ and increased the chromatin accessibility to TEAD-binding motifs. This lead to the up-regulation of a fetal transcriptional programme and the expansion of a population of RSC[158].

Tumours experience changes in their ECM composition with respect to normal tissue, which are mainly mediated by the production of ECM components and modifiers by Cancer-Associated Fibroblasts (CAF). In this settings, an increased ECM stiffness and the consequent reorganization of the epithelial cytoskeleton induce a mechanotransduction cascade which culminates in epithelial activation of YAP. Interestingly, TEAD transcriptional targets include ECM-modifying enzymes, integrins and F-actin remodelling factors. This suggests that, upon YAP/TAZ activation, cells can in turn tweak their internal structure as well as the mechanical properties of their environment, which may be important in the context of metastasis[157].

Furthermore, inflammatory cues can also modulate the Hippo-YAZ/TAZ axis in the context of cancer. Work by Nakanishi et al demonstrated that defects in Protein kinase C drove the emergence of an intestinal inflammatory milieu, which resulted in YAP activation and enabled serrated tumorigenesis independent from defects in *Braf* and *Kras*[162]. Moreover, treatment with antibiotics has been shown to reduce inflammation, decrease the levels of YAP activation, and limit tumorigenesis in models of serrated cancer[162, 160]. This suggests that the microbiota may be regulating cell identity at least partially through modulation of the Hippo-YAZ/TAZ module.

Overall, both cell-intrinsic (eg. mutations) as well as cell-extrinsic factors (eg. stromal cues, mechanical signals) can activate the YAP/TAZ pathway. Recently, Dow et. al showed that transient exposure to TGF-β and long-term disruption of p53 were necessary to maintain a RSC phenotype in organoid models of CRC[158]. This suggest that, at least in cancer settings, a combination of epithelial cell-intrinsic and extrinsic factors may be needed to sustain YAP activation and RSC characteristics.

1.10.3.1 YAP and WNT Interact to Regulate SC Identity

As described in previous sections, WNT and YAP regulate cell fate in a context-dependent manner. In homeostasis, WNT activates the expression of conventional ISC markers (eg. *Lgr5, Axin2*)[1], whereas in regeneration settings, Yap represses the expression of these genes and, instead, induces a regenerative programme (eg. *Clu,*

Ereg)[77]. However, WNT and YAP can act synergistically in CRC settings[77, 82]. The reason why these observations are difficult to piece together is that YAP and WNT interact at different levels. Non-canonical WNT ligands like Wnt5a/b, have been shown to activate Yap/Taz through a β-catenin-independent mechanism[163]. However, the interactions between canonical WNT and YAP are far more complex.

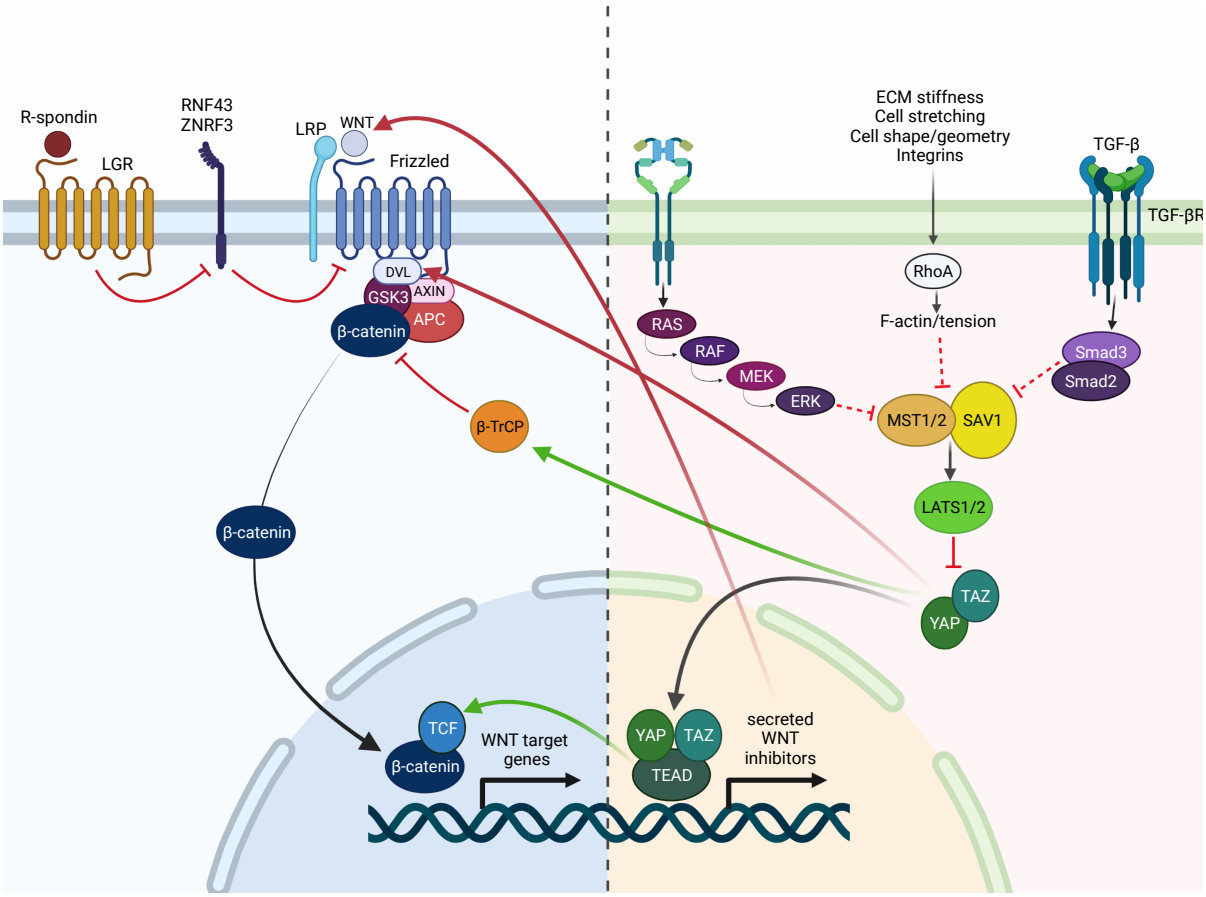

Figure 3: Schematic representation of the interactions between the Hippo/YAP-TAZ and WNT pathways.

On the one hand, YAP can inhibit WNT either by secretion of inhibitors like *DKK1*[163], or by cytoplasmic interaction with β-catenin regulatory elements. In the absence of Wnt ligands, Yap/Taz is associated to the β-catenin destruction complex and recruits the ubiquitin-ligase β-TrCP, which targets β-catenin for degradation. However, upon Wnt stimulation, Lrp6 competitively displaces Yap/Taz from the complex and releases it into the cytoplasm[164]. In parallel to this mechanism, Hippo can inhibit β-catenin through Taz-mediated phosphorylation of Dvl[165] (Fig.3).

On the other hand, *in vitro* experiments point out to a possible co-operation between Wnt and Yap/Taz at a transcriptional level to drive the expression of ISC

genes like *LGR5* and *AXIN2*[166, 167]. In this scenario, TEAD4 binds to TCF4 and stabilizes its association to the chromatin. Interestingly, disrupting this interaction is sufficient to decrease the number of polyps in APC^{Min} mice by more than half[167]. In addition to TCF4, YAP and beta-catenin can form a complex with the transcription factor T-Box Transcription Factor (TBX)-5, which promotes proliferation and anchorage-independent growth of cancer cell lines[168](Fig.3), and TBX-2/3, which activate survival and invasive genes[169]. Lastly, the quaternary complex YAP-TEAD-β-catenin-TCF3 contributes to stemness by binding the *Oct4* enhancer [170]. Overall, the diversity of pathways activated in a particular cell will ultimately dictate the way in which the available transcription factors interact.

Based on this evidence, we can begin to understand how the context determines the nature of the interactions between YAP and WNT. In regeneration, the up-regulation of non-canonical WNT ligands activates Yap/Taz, which results in inhibition of β-catenin though secretion of Wnt inhibitors or cytoplasmic interactions. However, in CRC models where β-catenin activation occurs downstream the destruction complex, like APCmin animals, cells could bypass cytoplasmic Yap-mediated Wnt inhibition. In addition, nuclear effectors of these two pathways can cooperate to induce the expression of genes related to oncogenic properties, like cell survival and stemness[77, 82]. This raises important questions relating to the nature of SC in different types of malignancies. For example, if canonical WNT drives a CBC phenotype whereas YAP pushes cells towards a RSC fate, how will the synergy of these two pathways affect the SC phenotype in cancers with LI WNT mutations? Furthermore, will acquisition of LD WNT alterations favour an RSC phenotype? A better characterization of the CRC SC compartment is needed in order to address these questions.

1.11 The Role of the Stroma in Colorectal Cancer Initiation

As we saw before, an increased fibroblastic infiltrate has been associated with poor prognosis in CRC[112]. I already discussed the role of the microenvironment conferring stem cell identity, but now we'll see how these mechanisms can be co-opted in CRC.

In homeostasis, WNT activation is limited to the base of the crypts, which are surrounded by nurturing pericryptal fibroblasts. Similarly, within a tumour, the strongest levels of WNT activation are found at regions in contact with stromal

cells[124, 171]. Within this populations, fibroblasts can promote SC maintenance via secretion of WNT activating signals (eg. Hepatocyte Growth Factor, stromal-derived factor 1α and IL-17)[124, 172, 173]. Furthermore, stromal cues like PGE2 and TGF-β can re-instate stem properties in differentiated cells[82, 158].

Not only stem cell identity, but also function seems to be dependent on the microenvironment. Despite expression of stem cell markers throughout a tumour, biosynthetic and proliferative potential is limited to cells located in specific positions, namely those in close contact with the stroma[174]. However, a study using CRC xenografts showed that ectopic expression of the myofibroblast-secreted protein osteopontin by tumour cells homogenized their proliferative potential regardless of spatial localization[171]. This suggests that aberrant expression of key molecular cues may bypass stromal dependence in colorectal neoplasia.

1.12 The Immune Compartment in CRC

The molecular characterization efforts conducted over the last decade revealed the existence of a great immune diversity across colorectal malignancies[107]. Importantly, the immune characteristics of colorectal tumours have been shown to have implications in the prognosis and treatment outcome of patients. In particular, the presence of an adaptive Th1 anti-cancer immune response, and the consequent infiltration by CD8+ T lymphocytes, are major positive prognostic factors[175, 176, 177].

An increased cytotoxic T cell infiltrate is characteristic of hypermutated CRC, which include microsatellite unstable and *POLE* mutant tumours. As a result of their increased number of mutations, these cancers have a high neoantigenic burden, which has been proven to induce immune surveillance[178]. Shortly, neoantigens are captured by DC and presented in MHC molecules in the surface of these cells. Then, DC migrate into local lymph nodes, where they activate naive T cells and prime them to travel into the tumour and elicit a response against the neoplastic cells. In these circumstances, cancer cells up-regulate immune checkpoint ligands, which interact with their receptors in the surface of immune cells and send an "off" signal that protects cancer cells from immune-mediated killing. This makes hypermutate cancers good candidates for therapy with immune checkpoint blockers[179, 180, 181].

On the other hand, Microssatellile Stable (MSS) cancers, which have a lower number of mutations, tend to have limited levels of immune infiltration[182]. Although mutational burden is a clear factor contributing towards this phenotype, it alone can

not explain the high immunogenic variability observed between malignancies. For example, despite having higher neoantigenic loads than mesotheliomas or sarcomas, the response of colorectal MSS tumours to immune checkpoint inhibitors is reduced with respect to these neoplasias. This disparity suggests that other immune-modulatory mechanisms must be mediating immune evasion in colorectal malignancies[183].

Cancer immune evasion mechanisms can be divided in two groups depending on whether they prevent T cell entrance into the tumour or mediate dysfunction of infiltrating lymphocytes[184]. Jiang et al developed a computational framework which enabled the inference of T cell dysfunction and exclusion in human samples based on a small panel of transcripts[184]. Based on this work, Tang et al developed a Tumour Immune Microenvironment classifier which distinguished between four types of CRC according to their immune transcriptomic profiles. Among the low infiltrated cancers, the authors observed that only some of them were enriched for immune suppressive populations, whereas the rest were described as "immune-inactive", suggesting that intrinsic properties of the tumours where hampering the emergence of an anti-cancer immune response[185].

Importantly, although dysfunctional T cells could potentially be re-activated through strategies like immune checkpoint inhibition[183], immune-excluded cancers would benefit from additional approaches aiming to increase the lymphocytic infiltrate. If we want to overcome cancer tolerance and elicit an effective immune response, we need a better understanding of the molecular forces mediating immune suppression.

1.12.1 Epithelial Cell Extrinsic Mechanisms of Immune Suppression in Colorectal Cancer

In the context of cancer, homeostatic tolerogenic mechanisms may be exploited to avoid immune-mediated elimination. As described by Tang et al[185], the microenvironment of some CRC contains immune suppressive populations that can keep T cell responses in check, including M2 macrophages, myeloid-derived suppressor cells and cancer-associated fibroblasts[186, 184]. In CRC, both high density of myeloid cells and fibroblasts have been linked to poor prognosis[112, 186], which may be at least partially explained by the contributions of these cells to immune evasion[186]. For example, stromal-derived TGF-β can compromise the acquisition of an adaptive anti-tumour immune response by inhibiting Th1 CD4+ lymphocytes. In mice models of CRC, administration of the TGF-β inhibitor Galunisertib increases recruitment of CD4+ T cells, enhances activation of cytotoxic lymphocytes, and reduces tumour size and incidence of liver metastasis. In a human CRC dataset, Th1 activation was

inversely correlated with TGF-β expression at the transcriptomic level, suggesting that this mechanism may also be operating in human cancers[108].

In addition to fibroblasts and innate immune populations, the adaptive immune system also contributes towards immune evasion. In malignant settings, FOXP3+ Treg cells can suppress potential immune responses against neoplastic cells, and hence correlate with poor clinical outcomes[187]. However, the role of FOXP3+ cells in CRC has been a subject of controversy[188, 189, 190]. Several reports suggested that FOXP3+ cell infiltration had either no effect on clinical outcome, or was a positive prognostic factor in CRC patients. However, FOXP3 expression constitutes a poor readout for the levels of Treg cells, which may have confounded the interpretation of these results. A publication by Saito et al described a population of FOXP3low T cells which was found in inflamed CRC tumours and did not show immune suppressive activity in vitro, but instead produced pro-inflammatory cytokines (IFN-γ, IL-17). In CRC patients infiltrated by this population, *FOXP3* expression correlated with good clinical outcomes. Contrarily, infiltration by immune-suppressive FOXP3high Treg cells was a poor prognosis factor[191], which is in agreement with the role that these cells play in other malignancies[187].

1.12.2 Epithelial Cell Intrinsic Mechanisms of Immune Suppression in Colorectal Cancer

Immune evasion can sometimes be driven by intrinsic properties of the tumour. In many cases, neoplastic cells escape immune killing by hiding from cytotoxic effectors. For example, latent cancer cells from human lung and breast carcinomas survive in metastatic locations by acquiring quiescent properties, which include the downregulation of death receptors (eg. FAS, TRAILR1) and UL16-binding proteins (eg. ULBP2, ULBP4). This allows them to escape NK cells-mediated killing[192]. Furthermore, glioma stem cells express lower levels of MHC-I molecules and antigen-presenting machinery (eg. Antigen Peptide Transporter 1-TAP1-, LMP7) than their non-stem counterparts, and disruption of IFN-γ signaling results in downregulation of MHC-I molecules in melanoma, which protects cancer cells from immune-mediated elimination[193, 194].

The list of cell-intrinsic factors that can contribute to immune evasion in CRC is endless. To start with, the Human Leukocyte Antigen haplotype of the patient and the clonality of the immunogenic mutations can influence the capacity of cancer cells to efficiently hide from the immune system[183]. In addition, immunogenic forms of programmed cell death (eg. necroptosis, pyroptosis) result in production of DAMPS

to a greater extent than non-immunogenic forms (eg. apoptosis), which translates into higher priming and recruitment of T cells[185, 195]. Interestingly, WNT activation has been linked to immune exclusion across a wide range of malignancies, including CRC, but the mechanisms behind this correlation are only partially understood.

1.12.2.1 WNT-driven Immune Exclusion

Immune exclusion is defined as the capacity to keep immune cells out of the vicinity of the tumours[196]. The first link between epithelial WNT activation and T cell exclusion was described in the context of melanoma. Spranger at al showed that WNT signaling inversely correlated with T cell infiltration in biopsies from melanoma patients and that the disruption of β-catenin in animal models of melanoma caused a lack of T cell infiltration into the tissue. They noted that Wnt activation in epithelial cells lead to repression of *Ccl4* by the transcription factor Activating Transcription Factor 3(Atf3). Ccl4 mediates the recruitment of Basic leucine zipper transcriptional factor ATF-like 3(Batf3)+ DC, a subset of antigen presenting cells specialized in cross-presentation, a mechanisms by which tumour-derived antigens can be exposed to CD8+ T lymphocytes. As a consequence of reduced Ccl4 production, Batf3+ DC failed to enter the tumour, which resulted in a deficient priming of naive T cells and low recruitment of effector T cells into the tissue. In these models, both immune checkpoint inhibitors and T-cell adoptive transfer were rendered inefficient unless they were combined with the introduction of Batf3+ DC[197, 198]. In addition to its effects on the DC compartment, WNT can hamper the ability of SC to present antigen. For example, in glioma stem cells, β-catenin acts through c-Myc to downregulate the expression of MHC-I and other antigen presenting proteins like TAP1 and Endoplasmic Reticulum Aminopeptidase 1[194].

Subsequent studies have consolidated the notion that WNT activation compromises T cell infiltration across cancers, and particularly in CRC[199]. Molecular characterization of primary colorectal cancers revealed that WNT activation is inversely correlated with the levels of lymphocytic infiltration after correction for microsatellite instability status[182]. However, the molecular mechanisms behind this widespread phenomenon are far from comprehended. The work from Xiao et al described a process by which APC-mediated epithelial up-regulation of *DKK2* suppresses the activation of Natural Killer and CD8+ T cells. According to this report, Dickkopf2 acts directly upon immune cells and reduces the nuclear levels of phosphorylated STAT5, which downregulates the expression of cytotoxic genes (eg. *GZMB*) in a β-catenin-independent manner[200]. Whether recruitment of Batf3+ DC, defects in

the antigen presenting machinery or novel WNT-driven mechanisms may be driving immune exclusion in CRC requires further investigation.

Importantly, only ligand-dependent WNT activation can be targeted pharmacologically with agents like Porcupine inhibitors, which could be combined with immunotherapy to improve its efficacy in immune excluded cancers. Ganesh et al used a Dicer-substrate small interfering RNA against *CTNNB1* to target this pathway in syngeneic models of melanoma, neuroblastoma, renal and mammary carcinoma. This treatment resulted in increased CD8+ T cell infiltration and sensitized non-inflamed tumours to immune checkpoint blockers[201]. Interestingly, the use of Porcupine inhibitors was enough to drive CD8+ infiltration in MMTV-Wnt1 mammary tumours, which had acquired ligand-dependent activation of the WNT pathway. Despite ongoing clinical trials aimed to test the potential of this approach, a better understanding of the mechanisms behind WNT-driven immune exclusion could help target these therapies to the subsets of patients that are going to benefit the most, both across malignancies and specifically in CRC.

1.13 Objectives

A bast body of bibliography describes how alterations in the gradients of morphogens that regulate cell identity result in cancer. Despite CBC being recognize as paradigmatic cancer-initiating cells in the context of the intestine, other progenitor populations have the intrinsic abilities to fuel tumorigenesis. In Chapter 3, I investigated to what extent different SC/progenitor pools contribute to CRC.

The regulation of homeostatic signaling gradients is complex, and relies on intercompartmental interactions between epithelial cells and the surrounding stroma, which is the reason why these compartments co-evolve during cancer formation. In Chapter 4 and 5, I studied how epithelial cell-intrinsic mutations drive changes in the friborblastic and immune compartments in models of early colorectal transformation.

The main objectives of this piece of work are:

- To assess the SC composition across mouse and human intestinal pre-malignant lesions and neoplasia.
- To study the molecular mechanisms behind the changes driven by aberrant epithelial Gremlin-1 overexpression.

- To assess the levels of immune infiltration across mouse models of WNT activation, and to investigate the mechanisms by which WNT signaling may be hampering immune recruitment.

Chapter 2

Materials and Methods

2.1 Mouse Procedures

All procedures were carried out in accordance to Home Office UK regulations and the Animals (Scientific Procedures) Act (1986). All mice were housed in individually ventilated cages within a specific-pathogen-free (SPF) facility, either the animal unit either from the Functional Genetics Facility within the Wellcome Centre for Human Genetics (University of Oxford), or the Beatson Institute (Glasgow). Throughout their lives, animals were granted unrestricted access to food and water. Prior to the experiments described in this report, mice had not been involved in any procedures. The strains used in this study were maintained on a C57BL/6J background for $\geq$6 generations. All procedures were carried out on mice of at least 6 weeks of age. Assistance and advice on animal procedures was provided by Hayley Davis (an experienced Post-doctorate researched within the Leedham Laboratory).

2.1.1 Animal Models

The Vil1-Grem1 (VG) mouse model was established by Davis et al in the Leedham laboratory[132]. The Lgr5$^{\mathrm{tm1(cre/ERT2)Cle}}$[6], Gli1-CreER$^{\mathrm{T2}}$[202], Lgr5$^{\mathrm{tm2(DTR/EGFP)Fjs}}$[23] (Lgr5^{DTR}), APC^{Min}[203, 204], Apc$^{\mathrm{tm1Tno}}$[205], Fzd5$^{\mathrm{tm2Nat}}$/J[206] and Vil1-creER$^{\mathrm{T2}}$[207] mice were obtained from the Jackson Laboratories. C57Bl/6J$^{\mathrm{Fzd7}fl}$[208] animals were obtained from Owen Sansom's group at the Beatson Institure.

FVB.129P2-Gt(Rosa)26Sortm$^{6(\mathrm{CAG\text{-}Rspo3)Nki}}$/A[142] (Rspo3inv) animals were provided by Hilkens et al (The Netherlands Cancer Institute, Amsterdam) on a FVB background, and were backcrossed for six generation with C57BL/6J Wild Type (WT) mice. Then, they were crossed to Lgr5$^{\mathrm{tm1(cre/ERT2)Cle}}$ and Gli1-CreER$^{\mathrm{T2}}$ mice to obtain

Lgr5$^{tm1(cre/ERT2)Cle}$-Rspo3inv (Lgr5-Rspo3) and Gli1-CreER^{T2}-Rspo3inv (Gli-Rspo3) strains.

B6;129-Fzd5^{tm2Nat}/J and C57Bl/6J^{Fzd7fl} were bred with VilER-Cre mice to obtain VilER-Cre-Fzd5KO$^{-/-}$ and VilER-Cre-Fzd7KO$^{-/-}$ strains respectively. Afterwards, these mice were crossed to VG mice to generate the VG-VilER-Cre-Fzd5KO$^{-/-}$ and VG-VilER-Cre-Fzd7KO$^{-/-}$ models.

Lgr5^{DTR} were bred to C57BL/6J-APC^{Min} to generate Lgr5^{DTR}-APC^{Min} animals.

Finally, Apc^{tm1Tno} animals were bred to VilER-Cre mice to generate the VilER-Cre-Apc$^{fl/fl}$ model.

A summary of the mouse models used throughout this study can be found in Table 2.1.

2.1.2 Breeding

Breeding pairs were set up from an age of 6 weeks (sexual maturity) onwards. When elevated numbers of animals were needed for experimental purposes or to increase the probabilities to obtain useful descendants, trios composed by two females and one male were set up.

2.1.3 Earclips

Once weaning had been completed, at three weeks of age, ear punches were taken from mice to provide material for genotyping. These earclips were also used for identification purposes, following a system by which each animal had a different number of earclips, or these were differently distributed across both ears.

2.1.4 Genotyping

To extract the DNA from the tissue, earclips were incubated in 50μl of HotSHOT reagent (25 mM NaOH, 0.2 mM disodium Ethylenediaminetetraacetic acid, pH 12) at 100°C for 30 min. The reaction was terminated by adding 50μl of neutralisation buffer (40 mM Tris-HCL, pH 5) as described by Truett et al[209].

Afterwards, genomic regions of interest were amplified to determine the genotype of the animals using the primers described in Table 2.1. Polymerase Chain Reaction(PCR) mastermix, $MgCl_2$ (25mM), Deoxynucleoside Triphosphate mix (10mM) and Taq polymerase were purchased from Bioline. Betanine was purchesed from Merck. For DNA amplification, these reagents and the primers (100μM) were added to 2ul of DNA in a 96-well plate. The PCR reaction was conducted in a total volume

Allele name	**Description**	**Reference**
Lgr5$^{\text{tm1(cre/ERT2)Cle}}$	Tamoxifen-inducible Cre recombinase that drives the deletion of *loxP*-flanked sequences in cells expressing the epithelial promoter *Lgr5*	Barker et al (2007)[6]
Gli1-CreER$^{\text{T2}}$	Tamoxifen-inducible Cre recombinase that drives the deletion of *loxP*-flanked sequences in cells expressing the fibroblast promoter *Gli*	Ahn et al (2004)[202]
Vil1-creER$^{\text{T2}}$	Tamoxifen-inducible Cre recombinase that drives the deletion of *loxP*-flanked sequences in cells expressing the epithelial promoter *Vil1*	el Marjou et al (2004)[207]
Vil1-Grem1	Drives the constitutive expression of *Grem1* under the epithelial promoter *Vil1*	Davis et al (2015)[132]
Lgr5$^{\text{tm2(DTR/EGFP)Fjs}}$	Drives the constitutive expression of the Diphteria Toxin receptor under the epithelial promoter *Lgr5*	Tian et al (2011)[23]
APC^{Min}	Harbours a constitutive heterozygous mutation in *Apc* that renders it more susceptible to sporadic colorectal cancer	Moser et al (1990, 1995)[203, 204]
Apc$^{\text{tm1Tno}}$	Cre-inducible model with a heterozygous deletion in *Apc* that renders it more susceptible to sporadic colorectal cancer	Shibata el at (1997)[205]
Fzd5$^{\text{tm2Nat}}$/J	Conditional frizzled 5 allele flanked by loxP sites	Liu et al (2008)[206]
C57Bl/6J$^{\text{Fzd7}\mathit{fl}}$	Conditional frizzled 7 allele flanked by loxP sites	Yu et al (2012)[208]

Table 2.1: Mouse Models

Primer	Sequence
Generic Cre (Fw)	GGTCGATGCAACGAGTGATGAGG
Generic Cre (Rev)	CCAGAGACGAAATCCATCGCTCG
Fzd5tm2Nat (oIMR9005)	GGGATGCCTAGACTCCAGCTGTTTC
Fzd5tm2Nat (oIMR9006)	CTGGGAACAAACCTTGGCTCAGG
Fzd5tm2Nat (oIMR9007)	TCAGGCCATCCAGCCACCACCTACAGCCCA
Rspo3inv (Fw)	CGCCATTAATCGATCCCG
Rspo3inv (Fw)	CCTATCTGCTTCATGCCAATCC
Gli1-CreERT2 (oiHR1084)	GCGGTCTGGCAGTAAAAACTATC
Gli1-CreERT2 (oiHR1085)	GTGAAACAGCATTGCTGTCACTT
Gli1-CreERT2 (oiHR1088)	GGGATCTGTGCCTGAAACTG
Gli1-CreERT2 (oiHR1089)	CTTGTGGTGGAGTCATTGGA
Fzd7fl (Fw)	GCACCATCATGAAGCACG
Fzd7fl (Rev)	CACAGTTAGCATCGTCCTGC
APC (P3)	GTTCTGTATCATGGAAAGATAGGTGGTC
APC (P4)	CACTCAAAACGCTTTTGAGGGTTG
VG(Fw)	GAGGTCGAGGCTAAAGAAGA
VG(Rev)	ACCTAGTGTCGGGCATCTCC
Apctm1Tno (580S Fw)	GTTCTGTATCATGGAAAGATAGGTGGTC
Apctm1Tno (580S Rev)	CACTCAAAACGCTTTTGAGGGTTG
Lgr5-DTR (F2)	AGGCAGATCTGGACCTTTTGA
Lgr5-DTR (R2)	GCAGACAGACAGATGACAGC
ApcMin (9)	GCCATCCCTTCACGTTAG
ApcMin (15)	TTCCACTTTGGCATAAGGC
ApcMin (MT)	TGAGAAAGACAGAAGTTA

Table 2.2: Genotyping primers

of 25ul and consisted of an initial denaturation step (95°C, 5 min), 35 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min, and a final extension step (72°C, 10 min). PCR products were run on a 2% agarose gel supplemented with 2% Ethidium bromide (Merck) at 120mV for 30 min to allow for amplified bands to separate based on molecular size before being visualized using a Bio-rad ChemiDoc MP imaging system.

2.1.5 Recombination

In conditional models, recombination was induced by intraperitoneal injection of tamoxifen. Tamoxifen (Merck) was resuspended in corn oil supplemented with 100% ethanol (10%). Two different regimes were used:1) daily injection of 1mg of tamoxifen for 5 days 2) a single dose of 80mg/kg followed by a second dose of an equivalent

concentration 6 weeks later. After each injection, animals were monitored for signs of pathology.

2.1.6 Cell Ablation

Lgr5+ cell ablation in $Lgr5^{DTR}$-APC^{Min} animals was induced with a single intraperitoneal injection of Diphtheria Toxin (DT) (322326-1mg, Merck) in saline (50μg x kg^{-1}). After each injection, animals were monitored for signs of pathology.

2.1.7 Administration of Therapeutic Compounds

2.1.7.1 Gremlin-1 Blocking Antibody

A Gremlin-1 blocking antibody was provided by UCB. Two distinct injection regimes were used in this study. In the early treatment regime, VG animals were subcutaneously injected weekly with 30mg/kg of blocking antibody or mouse Immunoglobulin G1 isotype control antibody (UCB) from an age of 6 weeks and for a total of 6 weeks. In the late treatment regime, VG animals were subcutaneously injected weekly with 30mg/kg of blocking antibody or mouse Immunoglobulin G1 isotype control antibody (UCB) from an age of 21 weeks and for a total of 1, 2, 4, 6 or 10 weeks.

2.1.7.2 Procupine Inhibitors

Two different WNT blocking compounds were used throughout this study: LGK974 (Active Biochem, A-1400) and RXC004 (Redex). They are orally administered Porcupine inhibitors that bind to their target in the endoplasmic reticulum and block the secretion of Wnt ligands by preventing their post-translational acylation[210]. LGK974 was administered at a concentration of 5 mg/kg, in 0.5% Methylcellulose (MC) and 0.5% Tween-80 twice daily by oral gavage. LGK974 treatment of VG animals was conducted by an experienced member of Owen Sansom's team at the Beatson Institute (Glasgow). The therapeutic administration of this compound started at an age of 45 days, a time by which VG animals had developed mild signs of pathology, and it was continued until day 120.

RXC004 was administered at a concentration of 1mg/kg in 0.5% Hydroxypropyl Methylcellulose by oral gavage following two different regimes. For the survival experiment, Lgr5-Rspo3 animals were gavaged for a total of 48 days, which was the maximum number of administrations allowed by our license, following a pattern of 5 days on, 2 days off. For the re-infiltration experiment, the compound was administered for a total of 5 days.

2.1.8 Collection of Intestinal Sections

Animals were sacrificed via cervical dislocation either when they were manifesting signs of pathology, such as anaemia, diarrhoea, hunching and lack of activity, or at an experimental timepoint. Then, the intestines were removed and divided in four regions corresponding to the proximal (SB1), central (SB2) and distal (SB3) small bowel and the colon. These intestinal fragments were washed with Phosphate-buffered saline (PBS) and opened longitudinally using a gut preparation device described by Rudling et al [211]. At this stage, fresh polyps and tissue samples were collected when required (Sections 2.1.6.1 and 2.1.6.2). Afterwards, the opened intestines were pinned to a filter paper, fixed overnight in 10% Neutral Buffered Formalin and then transferred to 70% ethanol prior to processing (Section 2.2.2) and embedding (Section 2.2.3).

2.1.8.1 Collection of Fresh Polyps

Once the intestine had been collected and opened longitudinally, polyps were identified by eye and collected using a scalpel, with care to avoid contamination by the underlying tissues. Half of the tissue was embedded whereas the other half was snap frozen briefly in liquid nitrogen and transferred to a -80 °C freezer. At the time of tissue processing, samples were thaw and introduced in the adequate buffers, ALT buffer with 10% of proteinase K (Quiagen) and RLT lysis buffer (Qiagen) supplemented with 1% β-mercaptoetanol, for DNA and RNA extraction respectively (Sections 2.3.1 and 2.3.2).

2.1.8.2 Dissociation of the Epithelial and Stromal Compartments

Once the intestine had been collected and opened longitudinally, the surface of the lumen was scrapper to discard faeces and dead cells using a coverslip. To separate the villi and crypt compartments when required, the lumen was scrapped three more times in order to collect the villi, and 1cm of the reminding colon was cut using a scalpel. Fresh tissue was placed in 3 ml of dissociation media (30 mM Ethylenediaminetetraacetic acid, 0.5 mM Dithiothreitol and 2% RNAlater(Qiagen) in DMEMDulbecco's Modified Eagle Medium) for mild digestion. After a 30 min incubation, samples were transferred to PBS previously treated with Diethyl pyrocarbonate. Tubes were shaken vigorously to loosen the tissues and placed in a 90 mm Petri dish under the magnifying glass. Then, glass needles were used to pick up the epithelium by capillarity,

and it was placed in 100μl of RLT lysis buffer (Qiagen) supplemented with 1% β-mercaptoetanol prior to RNA extraction (Section 2.3.1), whereas the stroma was discarded.

2.2 Histology

2.2.1 Polyp Counts

After fixation, intestines were stained with 0.2% Methylene Blue for 10s before being washed in PBS for half an hour. Then, the tissue was placed under a magnifying glass and polyps were counted at 3x magnification. Lesions were categorized according to size (<1mm, 1-2mm, 2-3mm or >3mm) and location (SB1, SB2, SB3 or colon).

2.2.2 Tissue Processing

Prior to embedding, fixed intestinal sections were rolled into swiss rolls, using pins to hold them in position. Then, samples were placed into histology cassettes, which were introduced in a Histomaster Tissue Processor (Bavimed). Gut rolls were processed with the following protocol: 70% ethanol for 45 min, 95% ethanol for 30 min, 95% ethanol for 45 min, 100% ethanol for 30 min, 100% ethanol for 45 min, 100% ethanol for 45 min, Histoclear for 30 min (3 times), wax for 30 min, wax for 45 min (2 times).

2.2.3 Embedding

After tissue processing, gut rolls were embedded in paraffin wax using a Modular Tissue Embedding Center (EG1150H, Leica).

2.2.4 Sectioning

Paraffin blocks were cut at a width of 4μm using a HistoCore Rotary Microtome (RM2255, Leica).

2.2.5 Staining Procedures

2.2.5.1 Haematoxylin and Eosin

Tissue sections were dewaxed in xylene for 5 min. Then they were rehydrated using a decreasing gradient of ethanols (100%, 90%, 70%, 2 min each) followed by a 2 min incubation in tap water. Then, sections were submerged in Haematoxylin for 30s before being washed in water for 15s and briefly dipped in acid alcohol (70%

ethanol, 5% HCl 5M). Afterwards, samples were left under running water for 5 min and stained with Eosin (3 min). Next, sections were briefly washed in H2O, dehydrated through an increasing gradient of ethanols (70%, 90%, 100%, 2 min each) and mounted with coverslips using Dibutylphthalate Polystyrene Xylene mounting medium (X1525, ThermoFisher).

2.2.5.2 Immunohistochemistry

4μm tissue section were mounted on slides and baked overnight at 60 °C before being dewaxed in xylene for 5 min. Then they were rehydrated using a decreasing gradient of ethanols (100%, 90%, 70%, 2 min each) followed by a 2 min incubation in tap water. Antigen retrieval was done by pressure cooking in 10 mmol/L citrate buffer (pH 6.0) for 5 minutes. Endogenous peroxidase activity was blocked by incubating in 3% hydrogen peroxidase for 20 minutes. Next, sections were blocked with Superblock buffer (ThermoFisher, 37515) for 30 min and incubated with primary antibodies for 2h at room temperature. The sections were then washed twice in PBS (5 min) and incubated with appropriate secondary antibodies (1:200 in PBS) for 30 minutes at room temperature. For chromogenic visualization, sections were incubated with ABC (VectorLabs) for 30 min and stained using a 3, 3'-diaminobenzidine (DAB) solution (VectorLabs). Afterwards, they were counterstained with hematoxylin, dehydrated through an increasing gradient of ethanols (70%, 90%, 100%, 2 min each) and mounted with coverslips using Dibutylphthalate Polystyrene Xylene mounting medium (X1525, ThermoFisher). Antibodies used are listed in Table 2.3. Quantification of DAB stains was performed as indicated in Section XXX

Antibody	**Manufacturer**	**Dilution**
β-catenin (horse)	BD Biosciences (610154)	1:50
Caspase-3 (rabbit)	R&D Systems (AF835)	1:800
Cd3 (rat)	Abcam (ab11089)	1:500
Cytokeratin-20 (rabbit)	Abcam (ab118574)	1:200
Ki67 (rabbit)	Cell Signaling Technology (12202S)	1:500
Lysozime (rabbit)	DAKO (EC3.2.1.17)	1:500

Table 2.3: IHC antibodies

2.2.5.3 *In Situ Hybridisation*

4μm tissue section were mounted on slides and baked at 60 °C for 1 hour before dewaxing in xylene (2 incubations of 5 min each). Then samples were washed twice in

100% ethanol. Fluorescent *In Situ Hybridization* (ISH) was then performed using the RNAscope 2.5 HD Reagent Kit—BROWN (Bio-techne) or the RNAscope Fluorescent Multiplex Reagent Kit (Bio-techne) in accordance with the supplier's guidelines. Probes were purchased from Bio-techne(Table 2.4). Quantification of DAB stains was performed as indicated in Section 2.5.3.2.

Probe	**Reference**	**Species**	**Channel**
Hs-ANXA1	Bio-techne(465319)	Human	1
Hs-CLU	Bio-techne(606241)	Human	1
Hs-PLAUR	Bio-techne(542701)	Human	1
Hs-LGR5-C2	Bio-techne(311021-C2)	Human	2
Mm-Anxa1	Bio-techne(509291)	Mouse	1
Mm-Foxl1	Bio-techne(407401)	Mouse	1
Mm-Fzd5	Bio-techne(404911)	Mouse	1
Mm-Fzd7-O1	Bio-techne(534101)	Mouse	1
Mm-Gli1	Bio-techne(311001)	Mouse	1
Mm-Gsdmc4	Bio-techne(434691)	Mouse	1
Mm-Rspo3	Bio-techne(402011)	Mouse	1
Mm-Wnt2b	Bio-techne(405039)	Mouse	1
Mm-Wnt5a	Bio-techne(316791)	Mouse	1
Mm-Wnt11	Bio-techne(405021)	Mouse	1
Mm-Lgr5-C2	Bio-techne(312171-C2)	Mouse	2
Mm-Ly6a-C2	Bio-techne(427571-C2)	Mouse	2
Mm-Wnt2b-C2	Bio-techne(405031-C2)	Mouse	2
Mm-Wnt5a-C2	Bio-techne(316791-C2)	Mouse	2
Mm-Clu-C3	Bio-techne(427891-C3)	Mouse	3

Table 2.4: ISH probes

Fluorescent *In situ hybridization* images were imaged using a Leica SP8 Confocal Microscope and processed using the software ImageJ[212].

2.2.5.4 Multiplex immunoflouorescence

Multiplex immunofluorescence staining was carried out by the Translational Histopathology Laboratory (Old Road Campus Research Building, Oxford) using the OPAL protocol (AKOYA Biosciences) on the Leica BOND RXm autostainer (Leica, Microsystems). Six consecutive staining cycles were performed using the following primary Antibody-Opal fluorophore pairings; 1) Ly6G (BD Pharmingen, 551459, 1:300) – Opal 540, 2) CD4 (Abcam, ab183685, 1:500) – Opal 520, 3) CD8 (Cell signalling, 98941, 1:800) – Opal 570, 4) CD68 (Abcam, ab125212, 1:1200) – Opal 620, 5) FoxP3

(Cell signalling, 126553, 1:400) – Opal 650, 6) E-Cad (Cell signalling, 3195, 1:500) – Opal 690.

Prior to antibody incubations, antigen retrieval was performed at 100°C for 20 min with Epitope Retrieval Solution 1 or 2, as indicated in the Leica protocol. Then, primary antibodies were incubated for 60 minutes and detected using the BOND Polymer Refine Detection System (DS9800, Leica Biosystems) as indicated by the manufacturers (Akoya Biosciences) but substituting the DAB for the Opal fluorophores, with a 10-minute incubation time, and removing the haematoxylin step. Afterwards, sections were incubated for 10 minutes with DAPI (FP1490, Akoya Biosciences) and mounted with VECTASHIELD Vibrance Antifade Mounting Medium (H-1700-10, Vector Laboratories). Multispectral images (MSI) were generated using the AKOYA Biosciences Vectra Polaris. Batch analysis of the MSIs from each sample was performed using the inForm 2.4.8 software. Finally, MSIs were fused in HALO (Indica Labs) to reconstruct whole tissue images.

2.3 Nucleic Acids

2.3.1 RNA Extraction from Tissue

RNA extraction was performed using the RNeasy micro kit (Qiagen 74004) following the manufacturer's instructions. The RNA was eluted in 12ul of RNase-free water by carefully adding the volume to the center fo the column before centrifuging it (>8.000g, 1min). This last step was repeated to increase the RNA yield.

RNA was DNase treated using the DNA-free kit (Life Technologies AM1906) as indicating in the manual. Volume of reaction was limited to 10ul (8ul of RNA + 1ul rDNAse I + 1ul DNAse I buffer). Reactions were conducted in a 96-well plate. Then, the quality of the samples was assessed using Nanodrop (Section 2.3.3.1) and Qubit (Section 2.3.3.2).

2.3.2 DNA Extraction from Tissue

Samples were incubated overnight at 65°C in 100μl of ATL buffer with 10% of proteinase K (Qiagen 56304) for tissue digestion. Then, DNA was extracted using the QIAamp DNA Micro Kit (Qiagen 56304) following the manufacturer's instructions. DNA was eluted in 12ul of TE buffer by carefully adding the volume to the center fo

the column before centrifugin it (>8.000g, 1min). This last step was repeated to increase the DNA yield. Then, the quality of the samples was assessed using Nanodrop (Section 2.3.3.1) and Qubit (Section 2.3.3.2).

2.3.3 Quality Control of Nucleic Acids

2.3.3.1 Nanodrop Spectophotometer

The purity of the nucleic acids samples was assessed using a Ultraviolet-visible spectrophotometer (Nanodrop One, ThermoScientific). To do so, 1μl of samples was deposited on the optical reader and absorption was measured at 260nm and 280nm. The ration of these readings (260/280) is an established measure of the purity of the nucleic acids. DNA and RNA samples with ratios of $\approx$ 1.8 and $\approx$2 respectively qualify as pure, with deviations from these values indicating sample contamination, namely by reagents used during the extraction (eg. phenol, guanidine). Ratios close to the reference values were achieved for all of the samples sent to sequence (Section 2.4).

2.3.3.2 Qubit Fluorometer

Nucleic acid concentration was estimated using a Qubit 2.0 Fluorometer (Invitrogen) with the Qubit dsDNA High Sensitivity Assay (ThermoFisher Scientific, Q32851) and the Qubit RNA High Sensitibvity Assay kits (ThermoFisher Scientific, Q32852) according to the manufacturer's instructions. Briefly, 1μl of sample was added to 199μl of Qubit working solution (Qubit High Sensitibvity reagent diluted 1:200 in Qubit High Sensitibvity buffer). Prior to the reading, samples were vortexed and incubated at room temperature for 2min. The concentration of the samples was automatically inferred from a curve fit with two Qubit standards of known concentration (ThermoFisher Scientific, Q32851 and Q32852), which were used to calibrate the machine.

2.3.3.3 TapeStation

The size, quantity, and integrity of the samples was assessed using the High Sensitibvity 1000 DNA TapeStation kit (Agilent) either by Nadia Nasreddin (a DPhil student within the Leedham laboratory)or by the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (Oxford).

2.4 Sequencing

2.4.1 RNA Sequencing

Transcriptomic profiling was conducted by two different sequencing methods: Poly-A RNA sequencing (RNAseq) and 3'RNAseq. Poly-A libraries (Illumina) were produced and processed by the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (Oxford). Libraries were sequenced using a HiSeq 4000 Sequencing System (Illumina).

3'RNAseq libraries were prepared using the QuantSeq 3' mRNA-Seq Library Prep Kit (Lexogen, Austria) by Nadia Nasreddin. Library quality and quantity were assessed using TapeStation (Section 2.3.3.3) and the Qubit Fluorometer (Section 2.3.3.2), before pooling and sequencing on an Illumina HiSeq4000 instrument (Illumina).

2.4.2 Whole-Exome-Sequencing

Library preparation was performed using the SureSelect Library Preparation Kit (Agilent) and Whole-Exome-Sequencing was carried out on an Illumina HiSeq4000 instrument (Illumina). All procedures were performed by the High-Throughput Genomics Group at the Wellcome Centre for Human Genetics (Oxford).

2.5 Data Analysis

2.5.1 Pre-processing of Internal High Throughput Sequencing Data

2.5.1.1 RNA sequencing

Raw sequence reads underwent adapter trimming using BBduk (BBTools ver. 38.46). Trimmed reads were aligned to the GRCm38 build of the mouse reference using STAR (ver. 2.7.0f). Ensembl 96 annotations were used for alignment and subsequent quantifications. Gene counts were generated using featureCounts (ver. 1.6.4). All these steps were performed either by Gabriel Valbuena (a post-doctoral researcher within the Leedham laboratory) (Chapter 3) or by me (Chapter 5). Analysis of the counts was performed using the statistical software R[213] (Section 2.5.3.5).

2.5.1.2 Whole-Exome-Sequencing

Fastq files were aligned to the GRCm38 ENSEMBL reference genome using the Burrows-Wheeler Aligner-MEM algorithm[214] and samtools[215]. Binary Alignment Map (BAM) files were processed following the Genome Analysis Toolkit best practices[216]. Filtering criteria used was as follows: minimum variant allele frequency = 0.1, minimum number of supporting total reads at a variant site $\geq$10, and minimum number of variant reads = 2. The described pipeline was set up by Ryan Schenk (a DPhil student within the Leedham laboratory). Subsequent analysis of the Single Nucleotide Variants (SNV) was performed using the statistical software R[213] (Section 2.5.3.6).

2.5.2 Obtention of External Data

2.5.2.1 Human Data

In addition to the data generated in the lab, RNA sequencing, Copy Number Variation and Single Nucleotide Variant data from the TCGA Colorectal adenocarcinoma project (TCGA-COADREAD) was also analysed in this study. RSEM-normalized RNAseq files, Mutation Annotation Format files generated by SomaticSniper and mean segment copy number calls for TCGA-COADREAD were downloaded from the Genomic Data Commons data portal.

Furthermore, human samples were collected from 4 cohorts: S:CORT cancer cohort, Oxford BRC cancer cohort, pre-cancer polyps cohort and colitis cohort. All human samples were obtained following ethical approval and individual informed consent (Ethics No 16/NI/0030 and 15/EE/0241 for S:CORT cohorts, REC: 17/NW/0252 BRC cancer cohort, and Oxford GI biobank ethics 18/SL/JE/Early-detection and OCHRe:18/A11 for the colitis cohort). Samples were subject to histopathological review.

Data from the FOXTROT trial, developed by the National Cancer Research Institute Colorectal Cancer Clinical Studies Group (NCT00647530) was also used. For each individual, pre-treatment biopsies and a post-treatment resection sample were collected for genetic and transcriptomic characterization.

2.5.2.2 Mouse Data

The study comprised the use of two mouse cohorts: an internal mouse cohort (Oxford) and the ACRCelerate cohort. The ACRCelerate cohort data was provided by Owen

Sansom's laboratory (Beatson Institute, Glasgow). The data was derived from the animal models in Table 2.5.

2.5.3 Statistical Analysis

2.5.3.1 Polyp Count

Polyp counts were visualized using stacked barplots generated using the package *ggplot2* within the R environment[213].

2.5.3.2 Image Quantification of DAB stains

Slides stained using DAB were scanned at 20x magnification using an Aperio CS2 Digital Pathology Scanner (Leica Biosystems). Image annotation and analysis were performed using the software QuPath[220]. Cells within regions of interest were detected using an in-built algorithm. Then, they were classified as positive or negative using their average intensity value. Thresholds were determined for each stain and experiment based on the background of the stain and the intensity of the signal. Then, quantitative data was exported for subsequent analysis in R[213]. Two-way Analysis Of Variance (ANOVA) and pairwise t-test were used for comparison of the % of positive cells between experimental groups. Differences were considered significant when the p-value of the pairwise t-test was < 0.05.

2.5.3.3 Image Quantification of Fluorescent stains

HALO (Indica Labs) was used to analyse the multispectral images. First, the in-built Random Forest classifier was trained to recognise epithelium and stroma by feeding it with manually annotated tissue regions. Then, regions of interest were annotated for analysis. Within those regions, cell detection was performed by using the cytoplasmic Epcam staining in the case of the epithelium, or by the in-built cell detection algorithm. Parameter were optimized for each experiment.

Next, thresholds for each channel were set up by eye in order to provide an accurate and consistent recognition of cell types between samples. Cell types were defined based on their positivity and negativity for a given combination of markers as defined in Table 2.6.

Then, Positive Cell Counting and Tissue Classification were performed in the regions of interest and quantitative data was exported for subsequent analysis in R[213]. Differences between groups were tested using ANOVA and pairwise t-test.

Allele name	**Description**	**Reference**
Vil1-CreER^{T2};Apc$^{fl/+}$	Cre-inducible model with a constitutive heterozygous truncating mutation in *Apc* gene under the epithelial promoter *Vil1*	Oshima et al(1995)[217]
Vil1-CreER^{T2};Apc$^{fl/+}$; Kras^{G12D}	Cre-inducible model with a constitutive heterozygous truncating mutation in *Apc* gene and constitutively active form of *KRAS* expressed under the epithelial promoter *Vil1*	Sakai et al(2018)[218]
Vil1-CreER^{T2};Apc$^{fl/+}$; Kras^{G12D};p53$^{fl/fl}$; Tgfbr1$^{fl/fl}$	Cre-inducible model with a constitutive heterozygous truncating mutation in *Apc*, a constitutively active form of Kras and an homozygous deletion in the genes *p53* and *Tgfbr1*, all expressed under the epithelial promoter *Vil1*	Sakai et al(2018)[218]
Vil1-CreER^{T2}; Braf^{V600E}; Tgfbr1$^{fl/fl}$	Cre-inducible model with a constitutively active form of Braf and an homozygous deletion in *Tgfbr1* expressed under the epithelial promoter *Vil1*	Leach et al(2021)[160]
Vil1-CreER^{T2}; Braf^{V600E}; p53$^{fl/fl}$;Rosa26$^{N1icd/+}$	Cre-inducible model with constitutively active BRAF and NOTCH signaling, and an homozygous deletion in *p53* expressed under the epithelial promoter *Vil1*	Leach et al(2021)[160] Jackstadt et al(2019)[219]
Vil1-CreER^{T2}; Kras^{G12D}; p53$^{fl/fl}$	Cre-inducible model with a constitutively active form of Kras and an homozygous deletion in *p53* expressed under the epithelial promoter *Vil1*	Jackstadt et al(2019)[219]
Vil1-CreER^{T2}; Kras^{G12D}; p53$^{fl/fl}$;Rosa26$^{N1icd/+}$	Cre-inducible model with active BRAF and NOTCH signaling and an homozygous deletion in *p53* expressed under the epithelial promoter *Vil1*	Jackstadt et al(2019)[219]

Table 2.5: Mouse Models Oxford and the Beatson Institute (Glasgow)

Cell Type	Cd4	Cd8	Cd68	Foxp3	Epcam
Regulatory T cells	+	-	-	+	-
Cd4+ T lymphocytes	+	-	-	-	-
Cd8+ T lymphocytes (stroma)	-	+	-	-	-
Cd8+ T lymphocytes (IELs)	-	+	-	-	+
Macrophages	-	-	+	-	-

Table 2.6: Immune Phenotypes

Differences were considered significant when the p-value of the pairwise t-test was $<$ 0.05. Plots were generated using the *ggplot2* package.

2.5.3.4 Stem Cell Index

The CBC signature is the Lgr5 intestinal stem cell signature described by Muñoz et al[7], and the RSC gene signature was derived from the repair signature in Yui et al[32] and the spheroid upregulated gene list from Mustata et al[221]. Cell type-associated expression was characterized using hierarchical clustering of signature genes in a publicly available single cell RNAseq dataset from colorectal tumors and normal colon tissue[222]. Genes from clusters that represented epithelial-expressed genes with high stromal expression, stromal-expressed genes with no distinctive expression in other cell types, and genes with no clear cell-type specific expression, were excluded from the signature. The resulting gene list of 265 genes is provided in the supplementary Table A.1. This curation process was carried out by Gabriel Valbuena.

Single sample enrichment scores for the RSC and CBC signatures were calculated using Gene Set Variation Analysis (GSVA) [223] from mean per gene values in microarray data or from normalized Counts Per Million values in RNAseq data. Then, to obtain a relative measure of the SC composition of the tumours, a Stem Cell Index was calculated by subtracting the CBC score from the RSC score.

2.5.3.5 Pathway Analysis

Once counts had been produced, the *cpm* function from the *edgeR* package was used for normalization. Single sample enrichment scores were calculated using GSVA[223] from mean per gene values in microarray data or from normalized Counts Per Million values in RNAseq data, and correlation scores were calculated using the *cor.test* function in R[213]. Gene Set Enrichment Analysis was performed using the *fgsea* package[224]. Heatmaps were plotted using the *heatmap.2* function from the *gplots* package.

The signatures used were previously published: CBC[7], Yap[77], Wnt[128, 5], Kras[225], IFN-γ[226], TGF-β[227]. Mouse genes were mapped to the most confident human orthologs, based on the Ensembl annotation (March 2020).

2.5.3.6 Single Nucleotide Variants

SNVs were filtered for pathogenicity using the *SIFT* and *Polyphen* predictors. Driver mutations were identified using a list of common CRC drivers (Table.2.5) prepared by Ryan Schenk by combining genes from the Intogen Colorectal Cancer (COREAD) Driver Gene list and genes of interest for the Leedham laboratory. SNV frequencies were plotted using the R function *barplot*.

Driver Genes					
ACO1	*ACSL6*	*ACVR1B*	*AKAP9*	*APC*	*ARID1A*
ARNTL	*ASPM*	*ATM*	*ATRX*	*AXIN2*	*BCOR*
BMPR2	*BPTF*	*BRAF*	*BRWD1*	*CAD*	*CASP8*
CDC73	*CDK12*	*CDKN1B*	*CEP290*	*CHD4*	*CHD9*
CLSPN	*CNOT1*	*CREBBP*	*CTCF*	*CTNNB1*	*CUL1*
DIS3	*DNMT3A*	*EGFR*	*ELF3*	*FBXW7*	*FN1*
FOXP1	*FXR1*	*GATA3*	*GNAS*	*GOLGA5*	*GSK3B*
IDH2	*ITSN1*	*KRAS*	*LGR4*	*LGR5*	*LGR6*
MAP2K1	*MAP3K4*	*MECOM*	*MED12*	*MED24*	*MGA*
MSR1	*NF1*	*NR2F2*	*NR4A2*	*NRAS*	*NTN4*
NUP107	*NUP98*	*PCBP1*	*PIK3CA*	*PIK3R1*	*POLA1*
POLA2	*POLB*	*POLD1*	*POLD2*	*POLD3*	*POLD4*
POLDIP2	*POLDIP3*	*POLE*	*POLE2*	*POLE2*	*POLE3*
POLE4	*POLR2B*	*PPP2R1A*	*PTEN*	*PTGS1*	*PTPN11*
PTPRU	*RAD21*	*RBM10*	*RNF43*	*RTN4*	*RUNX1*
SF3B1	*SMAD2*	*SMAD4*	*SMC1A*	*SOS2*	*SOX9*
SRGAP3	*STAG2*	*SYNCRIP*	*TAF1*	*TBX3*	*TCF12*
TCF7L2	*TGFBR2*	*TP53*	*TP53BP1*	*TRIO*	*WIPF1*
WT1	*ZC3H11A*	*ZC3H11A*	*ZNRF3*		

Table 2.7: List of driver genes

2.5.3.7 Copy Number Variation

For Copy Number Variation (CNV) analysis, CNV values smaller than -0.3 were categorized as a loss (-1), whereas values above 0.3 were annotated as gains (+1). Gains and losses were plotted using the *plotFreq* function from the R package *copynumber*.

2.5.3.8 Survival Analysis

Kapplan-Meier curves were plotted using the *ggsurvplot* function from the *survminer* package within the R environment[213]. Differences in the survival distribution between experimental groups were calculated using a logrank test.

2.5.3.9 Therapy Response

Response to treatment of patients was determined by histological assessment of the pre- and post-treatment samples. Patients with no regression (no change) or mild regression (residual cancer cells are higher than fibrosis) were labelled as non-responders, and those with marked regression (presence of fibrosis with rare residual cancer cells) or moderate regression (increased presence of residual cancer cells but still high fibrosis) as non-responders. No patients with complete response (complete absence of residual cancer) were found within the cohort.

The stem cell index was calculated from transcriptomic data for each sample. Next, samples with positive and negative scores were classified as RSC and CBC respectively. Patients were then divided into "plastic" or "static" based on the changes in SCI from the pre-treatment to post-treatment samples: those where pre- and post-treatment samples had the same SC category where labelled as "static", whereas those where the tumour shifted from one category to the other were classified as "plastic". Differences in treatment response between "static" and "plastic" stem cell phenotypes were evaluated using a Fisher's exact test. This analysis was conducted together with Gabriel Valbuena and Nadia Nasreddin.

Chapter 3

Assessing the Stem Cell Composition Across Colorectal Malignancies

3.1 Introduction

Intestinal homeostasis is maintained by a pool of Lgr5+ Stem Cells (SC) that constantly divide to sustain the fast turn-over of the tissue[6]. These cells reside at the base of the intestinal and colon crypts, which is why they are commonly referred to as Crypt Base Columnar (CBC). Furthermore, a range of Intestinal Epithelial Cells (IEC) retain the capacity for de-differentiation in a permissive environment (eg. inflammation, cell ablation)[32, 33]. This adaptive cell reprogramming is essential for regeneration, and entails a transcriptional re-wiring, with expression of fetal cell markers and a return to stem cell function. This alternative type of induced progenitors have been coined Regenerative Stem Cells (RSC)[32, 33, 34]. Several publications contributed to the transcriptomic characterization of CBC and RSC cells. These studies revealed that, unlike CBC cells, which are predominately characterized by the expression of WNT target genes (eg. *Lgr5, Ascl2*), RSC progenitors re-activate the expression of a gene module normally restricted to developmental stages, which includes the transcripts *Sca1* and *Anxa1* [33, 32].

For over a decade now, the potential of CBC to initiate tumorigenic transformation has been acknowledged[125], and they have been established as the paradigmatic CRC cell of origin. However, several studies highlighted the importance of Lgr5- progenitors in animal models of serrated colorectal polyposis[132] and inflammation-driven transformation [133]. Furthermore, the up-regulation of fetal markers has been described in human colorectal serrated polyps[228, 229], which the authors reframed

within the context of gastric reprogramming, but which could be reflecting an expansion of RSC progenitors in these early lesions. Finally, RSC have been described in scenarios of therapy resistance and metastasis[77, 157, 158, 230].

Despite the above evidence, the precise role that RSC play in CRC has not yet been fully elucidated, and many questions remain unanswered: to what extend do CBC and RSC populations contribute to different types of colorectal cancer (CRC)? Are these SC populations associated with the presence of certain mutational events or activation of specific signaling pathways? Can the SC composition of a tumour be of clinical use to inform therapeutic choices? How plastic are CBC and RSC? Finally, can interconversion between phenotypes be therapeutically induced? This chapter presents a framework to infer epithelial SC admixture in murine and human colorectal pre-malignant lesions and tumours, which can be used to investigate these open questions.

Disclaimer: the work in this chapter is the result of a collaboration between Nadia Nasreddin (a DPhil student within the Leedham laboratory) and myself. We both contributed equally to the bulk RNA sequencing analysis and *In situ hybridization* staining of the sections.

3.2 Objectives

- To generate a pipeline that permits evaluating the SC phenotype of mouse and human intestinal samples.
- To assess the SC composition across mouse and human intestinal pre-malignant lesions and neoplasia.
- To investigate the association of SC phenotypes with specific genomic and transcriptomic alterations.
- To test whether interconversion between SC categories can be therapeutically induced, and whether this influences the clinical response to treatment.

3.3 Generation of a Stem Cell Index to Assess Stem Cell Dynamics in Colorectal Lesions

The first step to investigate the SC composition of CRC was to develop a set of tools that permits tracing this populations across samples. Work by Munoz et al

[7] and Yui et al [32] described transcriptomic signatures derived from FACS sorted epithelial Lgr5+ and Sca1+ progenitors respectively, and which could potentially be exploited to assess the proportion of CBC and RSC in bulk transcriptomic samples. These signatures were used to calculate single sample enrichment scores using gene set variation analysis[223]. In order to obtain a measure of the relative abundance of these two populations, a Stem Cell Index (SCI) was generated by subtracting the CBC score from the RSC score, which enabled the positioning of each individual samples across a spectrum of CBC to RSC phenotypes (Fig.4).

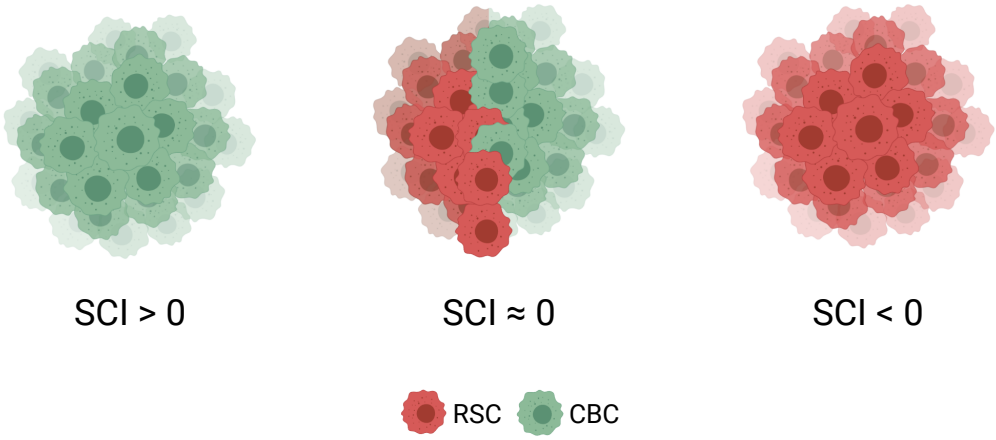

Figure 4: The Stem Cell Index can determine the Stem Cell composition of a lesion.

3.4 *Lgr5* and *Anxa1* are Surrogate Markers of CBC and RSC Identity

Lgr5 is widely recognized as the *bona fide* marker for homeostatic CBC cells[6], whereas the fetal marker *Anxa1* is expressed by populations of regenerative progenitors[32]. Hence, the expression of these genes was proposed as a morphological marker for the CBC and RSC identity respectively.

To validate the use of *Lgr5* and *Anxa1* as specific markers of the neoplastic epithelium, their regional compartmentalization in steady-state intestines of mice and humans was assessed by fluorescent *In situ hybridisation*(ISH). As expected, expression of *Lgr5* was limited to the epithelial progenitors at the base of the crypts in human and mice samples. In the case of *Anxa1*, transcripts were completely absent

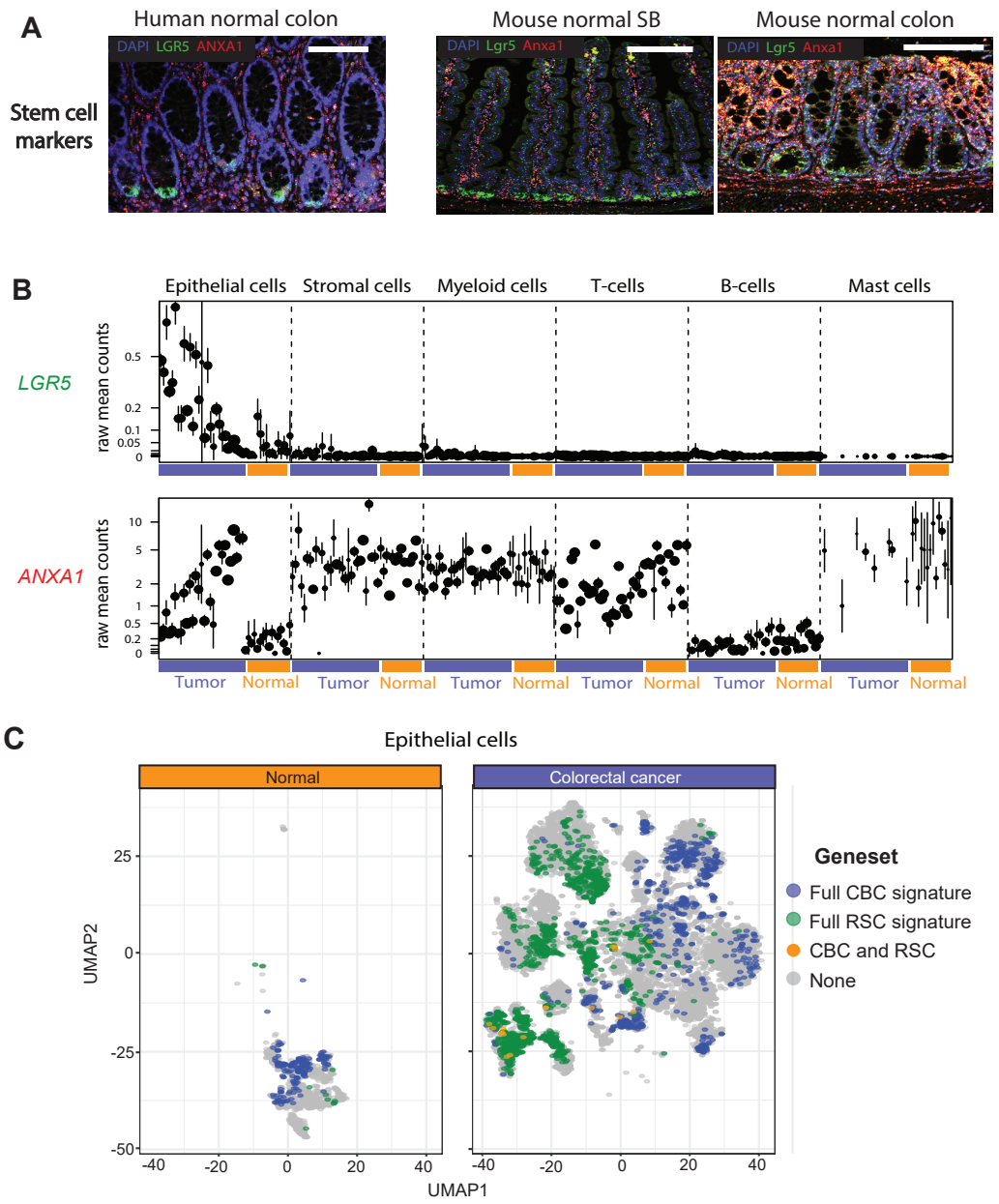

Figure 5: Expression of the RSC and CBC genes/signatures in mouse and human intestines. A. Fluorescent *In situ hybridization* for the RSC marker *Anxa1* (red) and the CBC marker *Lgr5* (green) in small bowel and colon of human and mouse. B. *Lgr5* and *Anxa1* expression in normal (yellow) and CRC (blue) single cell RNAseq samples. C. Uniform Manifold Approximation and Projection plot of single epithelial cells from human normal and colorectal cancer samples showing cells enriched for CBC (blue), RSC (green) and mixed CBC and RSC gene expression (yellow). Cells with no enriched stem cell signature expression are grey.

from the epithelium of human colon and murine small bowel, although some expression was detected in the colon of WT mice. However, *Anxa1* was strongly produced by stromal populations in human and mouse intestines (Fig.5A).

Next, publically available human single cell RNAseq data was used by Gabriel Valbuena, a post-doctorate bioinformatician within the Leedham laboratory, to provide a more precise characterization of the cell distribution of these two transcripts. As observed in the ISH stains, *Lgr5* expression was restricted to epithelial cells in homeostasis, whereas *Anxa1* was expressed in stromal, myeloid, mast cell and T-cell compartments. However, despite the constitutive expression of *Anxa1* in non-epithelial cells, it was strongly and specifically upregulated in neoplastic epithelium, a trend which is also followed by *Lgr5*. This behaviour, added to the ease of distinguishing epithelial cells in intestinal sections, established *Lgr5* and *Anxa1* as useful histopathological markers to track neoplastic epithelial populations. Interestingly, the expression of *Lgr5* and *Anxa1* in this compartment appeared to be inversely correlated in the fore-mentioned single cell RNAseq dataset, suggesting that CBC and RSC may exist in a dynamic equilibrium within human malignancies (Fig.5B).

Overall, despite the promiscuous expression of *Anxa1*, its specific upregulation in neoplastic cells validated its use as a surrogate marker of RSC identity. Furthermore, ISH and single cell RNA sequencing analysis of *Lgr5* highlighted its specificity as a marker for CBC. Consequently, these two transcripts were selected to morphologically track CBC and RSC populations in sections from intestinal neoplasia.

3.5 Curation of the RSC Signature Using Single Cell RNAseq Data

In addition to *Anxa1*, many of the genes within the signature published by Yui et al[32] were co-expressed by non-epithelial cell populations, which suggests that this gene list my be capturing stromal processes associated with the acquisition of fetal-like traits. In order to generate a good read-out of the epithelial characteristics within a tumour, independent of the microenvironment, single cell RNAseq data was used to curate this signature and eliminate clusters of genes heavily expressed by non-epithelial populations and which were not reflective of epithelial processes. This curation was conducted by Gabriel Valbuena (see methods). The resulting gene list can be found in Supplementary Table.A1.

Using the curated RSC signature and the original CBC gene list as described by Munoz et al [7], samples from the single cell RNAseq dataset could be labelled as

CBC, RSC, or cells with an intermediate RSC-CBC phenotype, although this last category was a minority. This validated the use of these gene lists to infer the SC phenotypes from single cell samples, but their applicability to bulk transcriptomic data remained to be demonstrated (Fig.5C).

3.6 Stem Cell Composition Varies Across a Range of Mouse Intestinal Polyposis and Neoplasia Models

To test the use of this transcriptomic pipeline in bulk RNA sequencing data, a range of samples capturing the complexity of colorectal disease were required. Mouse models are experimental tools that permit the controlled accumulation of genomic alterations, which allows to study the contribution of specific mutations towards clinical phenotypes. Here, bulk transcriptomic data from a diversity of polyposis and CRC mouse models was used. The range of phenotypes, which included both early and advanced models of colorectal transformation, permitted to study how the sequential accumulation of mutations conditioned the SC composition of the polyps/tumours. The data was provided by Owen Sansom's group, at the Beatson Institute (Glasgow). Descriptions and references of the animal models used in this analysis can be found in Table 2.5 (Chapter 2).

To start with, RNAseq counts were processed to generate SCI scores, which were plotted to observe the distribution of the samples across a spectrum of SC phenotypes ranging from CBC to RSC (Fig6.A). Importantly, samples of the same genetic background clustered together, indicating that the degree of SC admixture is consistent within genotypes, although some intra-group variance existed, which may be attributed to distinct evolutionary paths. As one may expect, some key driver mutations produced a notable skew in cell admixture. For example, models initiated by truncation of the *Apc* gene, like the *Vil1-CreER^{T2}; Apc$^{fl/+}$*(A) or *APC^{Min}* mice were enriched for CBC phenotypes. On the other hand, RSC scores were relatively higher in models with activating mutations in *Kras/Braf* or disruption of the TGFβ signaling through mutation of *Tgfbr1*, like the *Vil1-CreER^{T2};Braf^{V600E};Tgfbr1$^{fl/fl}$*(BT) mouse. Complex models with an accumulation of several mutated alleles, like the *Vil1-CreER^{T2};Apc$^{fl/+}$; Kras^{G12D}; p53$^{fl/fl}$; Tgfbr1$^{fl/fl}$* (AKPT) presented intermediate SCI scores, suggesting that an admixture of both populations may co-exists in these samples.

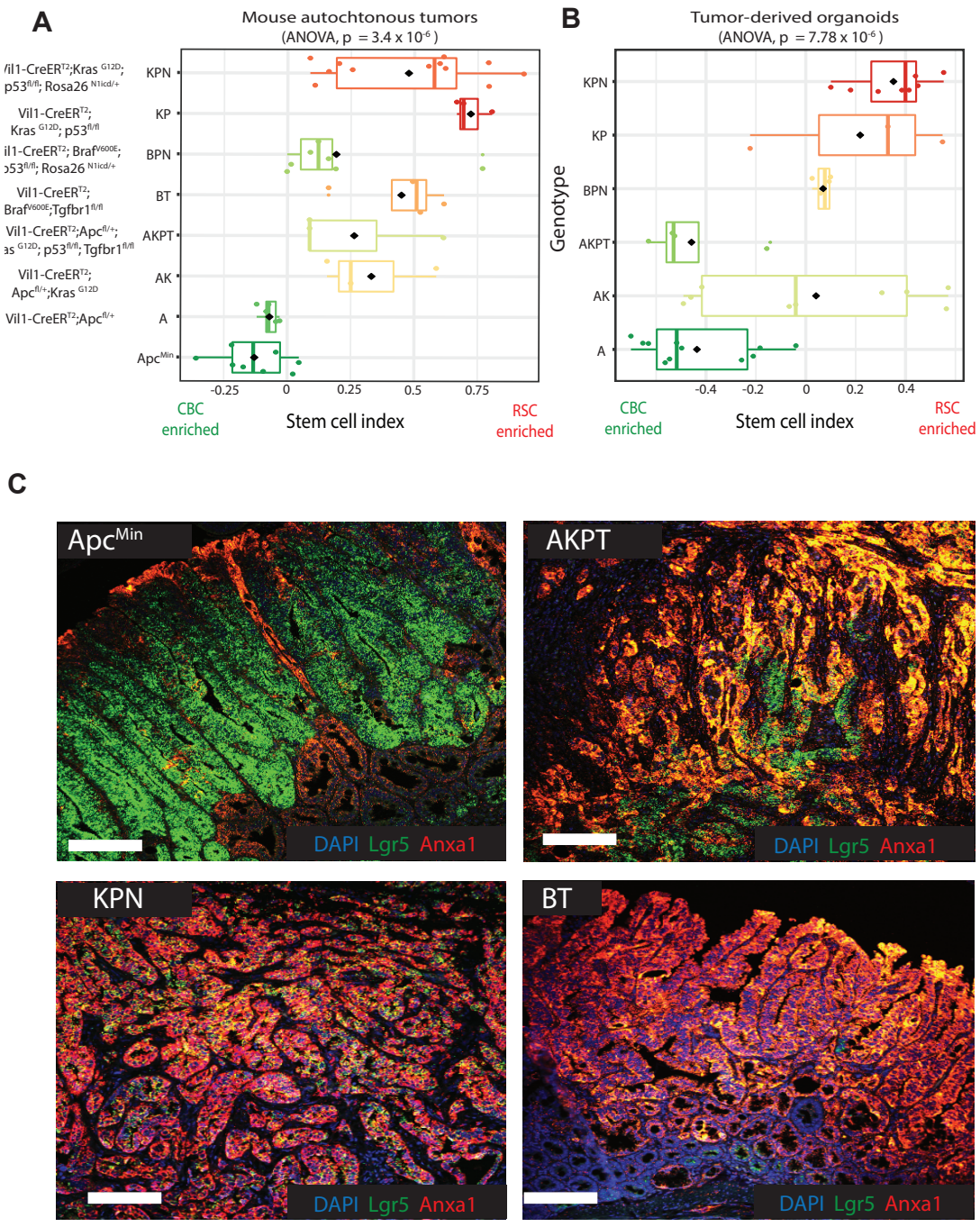

Figure 6: Distribution of CBC and RSC phenotypes across a range of CRC mouse models. A,B. Boxplots showing the Stem Cell Index scores for mouse autochtonous tumours (A) and matched derived organoids (B). Statistical significance was assessed using an ANOVA test and a Wilcoxon test. Differences between groups can be found in Supplementary Table A.2-3. C. Fluorescent *In situ hybridization* for the CBC marker *Lgr5* (green) and the RSC marker *Anxa1* (red). Driver alleles initialisation - A is Apcfl/+, ApcMin is ApcMin, B is BrafV600E, K is KrasG12D, P is p53fl/fl, T is Tgfbr1fl/fl, N is Rosa26N1icd/+.

Despite the consistency of the observed results with established literature, there was a possibility that the curation process might not have been stringent enough, in which case the RSC gene lists might be picking up signals from non-epithelial compartments. To address this, the same analysis was conducted in CRC organoids, which are devoid of stromal and immune cells. Reassuringly, the distribution of SC phenotypes obtained in this analysis was comparable to that of whole-tissue samples, although there was a slight shift of the scores towards the CBC end of the spectrum. This could be a consequence of the standard organoid culture procedures, which are optimised to maintain survival of the CBC cell population. However, the observed change was comparable across phenotypes, and it had no effect on the relative distribution of the samples along the spectrum of SC phenotype (Fig.6B). All in all, this analysis proved that the signals detected in bulk RNA sequencing data were reflective of the epithelial SC biology, and not a surrogate of the stromal composition of the tumours.

Finally, fluorescent ISH was performed in paraffin-embedded sections of tumours derived from representative mouse models to confirm the observed trends and assess the regional distribution of the CBC and RSC populations. As observed in the bulk RNAseq data, polyps from APC^{Min} animals showed abundant *Lgr5* expression, although some *Anxa1+* areas could be observed towards the top and bottom regions of the polyps. On the contrary, BT polyps were mostly populated by *Anxa1+* cells. Sitting between these two extremes, tumours from AKPT mice exhibited a mixed phenotype, with yellow areas which resulted from the co-expression of *Lgr5* and *Anxa1* transcripts (Fig.6C).

Shortly, this data evidences the capacity of the proposed framework to pick up epithelial signals and to map malignancies along a spectrum of SC phenotypes. The next step was to use this methodology in human samples.

3.7 Stem Cell Composition Varies Across a Range of Human Precursor Lesions and Tumour Molecular Subtypes

To apply the previous analysis to human samples, CBC and RSC genes were mapped to the most confident human orthologs based on the Ensembl annotation (March 2020). Sporadic CRC originates from precursor lesions which can be classified based on their histopathological traits as Tubulo-villious Adenomas (TVA), Traditional Serrated Adenomas (TSA) or Sessile Serrated Lesions (SSL). In addition, some cancers

arise in the context of pre-existing chronic inflammation, and their precursor lesions are known as IBD dysplasias. CBC are well established as the cell of origin of TVA[231], which are commonly initiated by cell-intrinsic forms of WNT disruption. However, the contribution of additional progenitors to the establishment of pre-malignant lesions should not be disregarded. To investigate this, the previous pipeline was used to infer the SC composition in colorectal polyps.

The results revealed that differences in the SC composition of transformed tissues can be detected from an early stage. In accordance with their role in tissue repair, RSC were enriched in IBD dysplasias, whereas TVA sat towards the CBC end of the spectrum (Fig.7A). Interestingly, serrated lesions presented intermediate scores and were located in between of these two types of progenitors. These observations were confirmed by ISH, which showed abundance of *LGR5* expression in TVA and a clear dominance of *ANXA1+* populations in IBD dysplasias. Contrarily, the levels of expression of these transcripts in TSA and SSL were fairly limited, but localized expression of *ANXA1* was found at the luminal side of the polyps. This suggests the presence of a regenerative response in early serrated lesions prior to desmoplastic stroma generation.

To investigate how cancer progression affected these baseline differences, the same pipeline was applied to human colorectal tumours, which were labelled according to their Consensus Molecular Subtypes (CMS)[107] and Colorectal Cancer Intrinsic Subtypes (CRIS)[113]. As in the previous analysis, categories characterized by Wnt activation, like CMS2 and CRIS C,D and E subtypes, were enriched for CBC gene expression, whereas CMS1,4 and CRIS B were enriched for RSC genes, suggesting that processes upregulated in these types of malignancy favour the acquisition of a regenerative-like phenotype. Consistent with these results, histopathological assessment of CMS2 malignancies showed a clear predominance of *LGR5+* populations, whereas CMS1,3 and 4 sections were mostly populated by *ANXA1+* cells(Fig.7B-D).

These results validate the use of the CBC and RSC signatures to infer SC phenotypes from human bulk transcriptomic data. Interestingly, differences in the SC composition could be detected from early stages of transformation, although they were more accentuated in established malignancies. As observed in mouse data, certain molecular traits seemed to be associated with CBC (eg. WNT activation) and RSC phenotypes (eg. inflammation), but a comprehensive characterization of the forces driving the preferential expansion of each progenitor type was still missing.

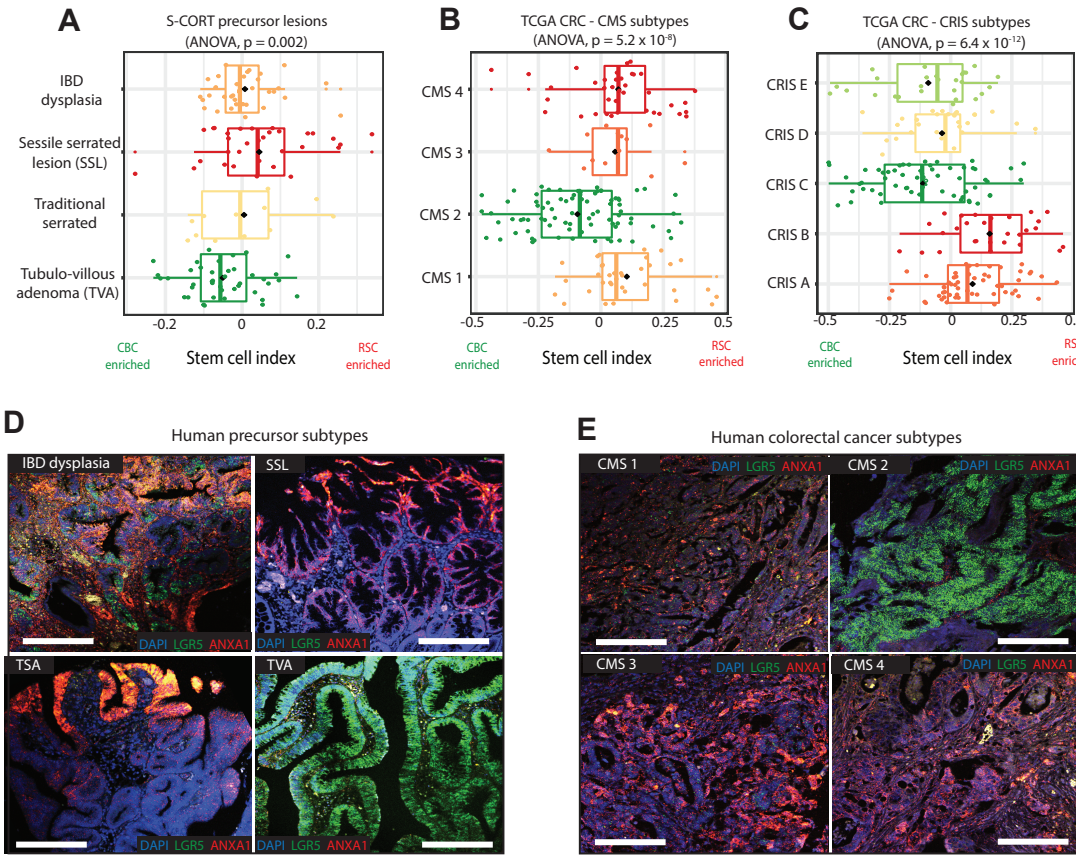

Figure 7: Distribution of CBC and RSC phenotypes across a range of human colorectal precursor lesions and cancers. A-C. Boxplots showing the Stem Cell Index scores for human colorectal precursor lesions (A) and cancers (B-C). Statistical significance was assessed using an ANOVA test and a Wilcoxon test. Differences between groups can be found in Supplementary Table A.4-5. D-E. Fluorescent *In situ hybridization* for the CBC marker *Lgr5* (green) and the RSC marker *Anxa1* (red) in samples from human colorectal precursor lesions (D) and cancers (E).

3.8 Driver Genes and Pathways Associate with Stem Cell Molecular Phenotypes

In addition to WNT activation and inflammation, other processes have been reported to determine the acquisition of a CBC or RSC phenotype, like mutations in *TP53* or activation of the Tumour Grouth Factor-β (TGF-β) and Yes-Associated Protein (YAP) pathways[77, 158]. Here, the proposed framework was used to investigate the processes that correlate with CBC and RSC enrichment in colorectal malignancies.

To start with, a set of publicly available signatures for pathways of interest were

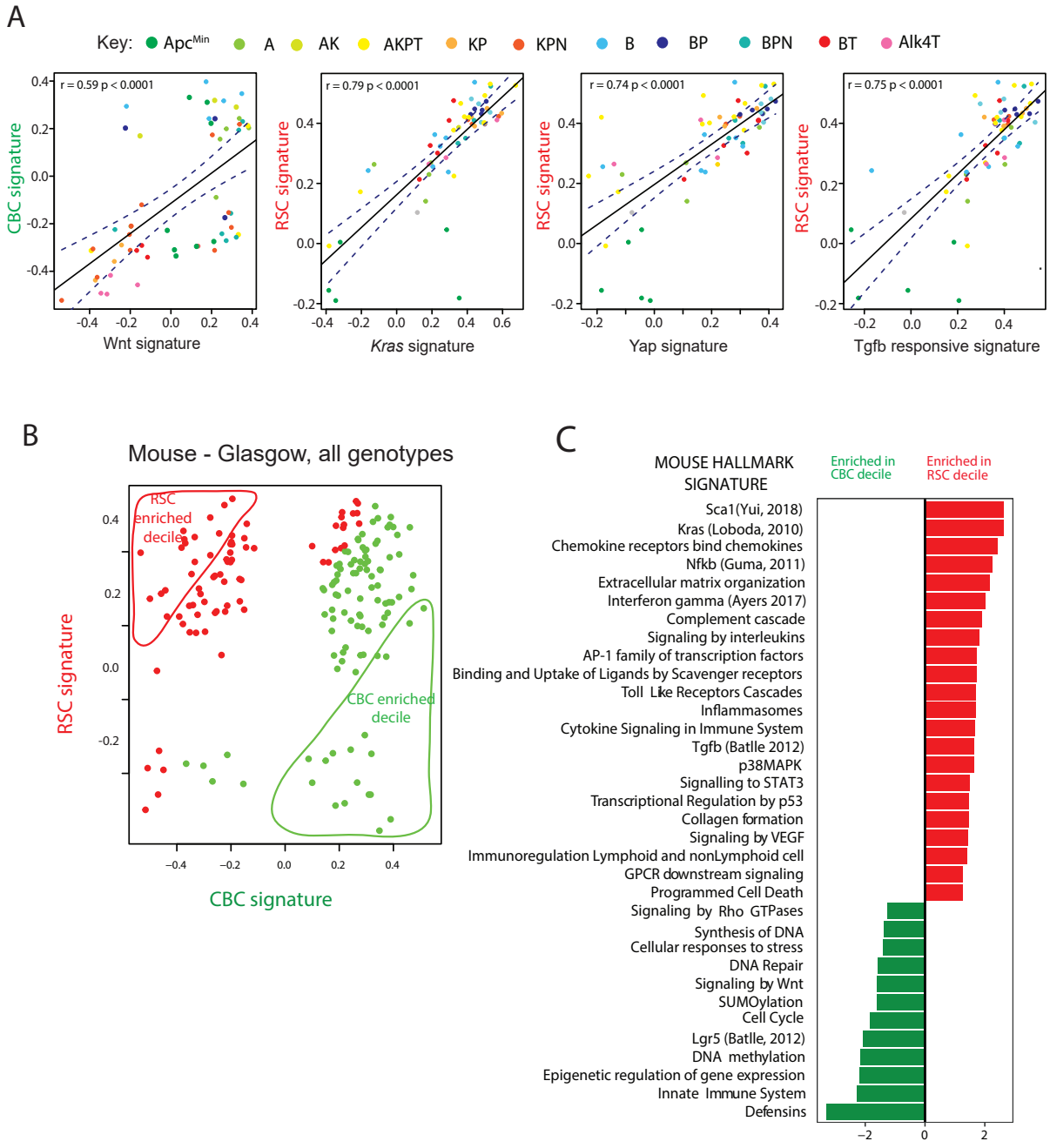

Figure 8: Signaling pathways associated with CBC and RSC phenotypes in mice. A. Correlation of the RSC and CBC scores with a range of publicly available signatures representative of WNT, KRAS, YAP and TGF-β activation. Different genotypes are identified by colors as determined by the key. Correlation was tested for using the Pearson method. B. Mouse CRC samples segregated by CBC and RSC scores. The predominant expression signature in each tumor is defined by color (CBC green, and RSC red). C. Gene Set Enrichment Analysis comparing the top and bottom deciles defined based on the Stem Cell Index score. Pathways were ranked by Negative normalized Enrichment Score. Statistical differences were assessed by the fast algorithm[232] and significance threshold was set at an adjusted p.value < 0.05.

correlated with the CBC and RSC scores in mouse samples (Fig.8A). Furthermore, an agnostic approach was applied to the same set of samples, which were segregated into top and bottom deciles based on their SCI scores to compare tumours with extreme CBC and RSC phenotypes (Fig.8B). Consistent with the previous results, CBC scores were associated with WNT activation[128], whereas RSC scores correlated with multiple pathways: KRAS activation[225], YAP[77], TGF-β[227], and inflammatory pathways like Interferon-γ (IFN-γ) and Nuclear Factor kappa-light-chain-enhancer of activated B cells (NF-κB) (Fig.8C). The extent of pathways that correlated with the RSC phenotype seems to indicate that a range of cell-intrinsic (eg. *KRAS* mutations) and cell-extrinsic (eg. microenvironmental TGF-β) signals may be driving this phenotype, which is in accordance with published studies[158].

Following a similar strategy, TCGA colorectal tumours were divided into top and bottom deciles (Fig.9A). The comprehensive characterization of TCGA samples allowed for an assessment of the Single Nucleotide Variants (SNV), Copy Number Alterations (CNA) and transcriptional pathways which were particularly enriched in each of these subgroups. In terms of SNVs, ligand-independent mutations in the WNT pathway (eg. *APC*, *CTNNB1*) were present in 84% of CBC tumours versus only 35% of RSC cancers, which presented a higher proportion of ligand-dependent WNT mutations (eg. *RNF43*) and *RSPO2/3* fusions (Supplementary Table A.6). Other pathways which accumulated mutations at a different rate between groups were the TGF-β superfamily and the Mitogen-Activated Protein Kinase (MAPK)/Phosphoinositide 3-Kinases (PI3K) pathway. Interestingly, although the percentage of *KRAS* mutations was comparable in both groups, *BRAF* was four times more commonly mutated in RSC cancers, suggesting that not all MAPK activating mutations may have a similar effect on the SC phenotype (Fig9B-E).

Regarding CNA, CBC cancers presented higher numbers globally, but the distribution of events along the genome was different between these two groups, with some regions being more comonly altered in RSC cancer and vice versa (Fig.9F). Finally, Gene Set Enrichment Analysis yielded similar results to those observed in mice data: WNT signaling was enriched in CBC tumours whereas RSC malignancies showed higher activity of TGF-β, YAP, KRAS and inflammatory pathways (Fig.9G).

In summary, these analysis allowed for the assessment of genomic and transcriptomic characteristics which correlated with the acquisition of CBC and RSC characteristics. However, the functional validation of these mechanisms requires further investigation.

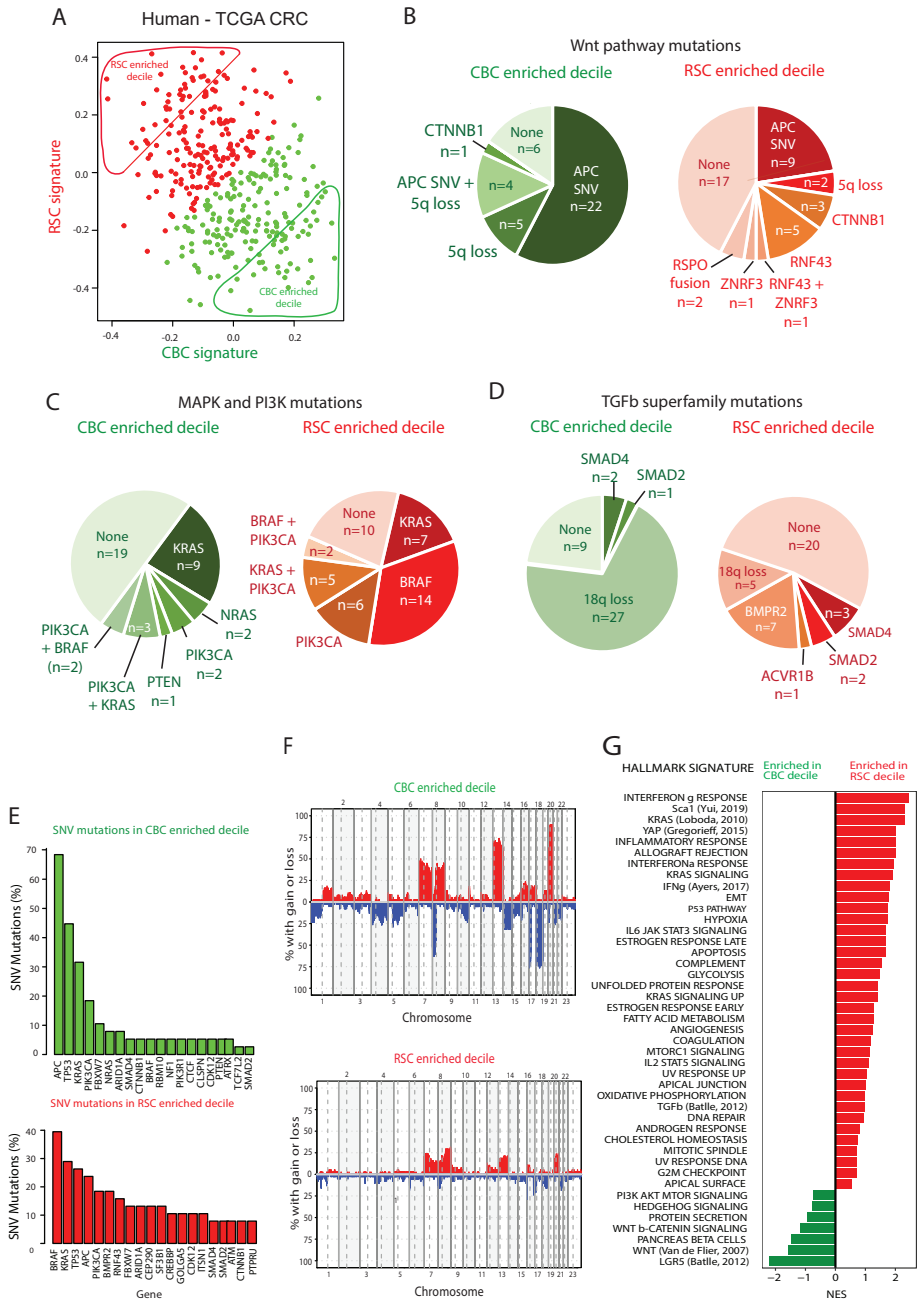

Figure 9: Molecular characterization of human TCGA CRC in association with their SC phenotypes. A. Human TCGA CRC samples segregated by CBC and RSC scores. The predominant expression signature in each tumor is defined by color (CBC green, and RSC red). Top and bottom deciles were defined based on the Stem Cell Index score. B-D Comparison of type and prevalence of disrupting mutations in the WNT(B), MAPK and PIK3CA(C) pathways and TGF-β(D) superfamily in top and bottom deciles. E-G. Twenty most common Single Nucleotide Variations (E), Copy Number Alterations (F) and Gene Set Enrichment Analysis(GSEA)(G) comparing the top and bottom deciles. For GSEA, pathways were ranked by Negative normalized Enrichment Score. Statistical differences were assessed by the fast algorithm[232] and significance threshold was set at an adjusted p.value < 0.05.

3.9 Application of a Selective Pressure Shifts the Equilibrium between CBC and RSC Poulations in APC^{Min} Mice

Previous research demonstrated the different sensitivity to selective sources of pressure (eg. irradiation) of distinct intestinal progenitors[126]. This intrinsic differences could give tumours the capacity to buffer therapeutic effects by compensating the loss of one SC phenotype with the expansion of another. In order to investigate this, APC^{Min} animals were crossed with a *Lgr5-CreER*^{ER}*-DTR* mouse[23], which expressed the Diphtheria Toxin Receptor (DTR) under the Lgr5 promoter to yield *Lgr5-CreER*^{ER}*-DTR-* APC^{Min} (APC^{Min}*-Lgr5*^{DTR}) mice. In this model, injection with Diphtheria Toxin (DT) drives apoptosis in cells expressing the DTR. Animals were allowed to age for 3 months, a time by which they had developed polyps. Then, Lgr5+ cell ablation was induced by DT administration, and animals were sacrificed 24 hours, 3 and 5 days after the injection (Fig.10A). All samples were processed for histopathological analysis. In addition, RNA was extracted from intestine of control (day 0) and treated animals 1 and 5 days post-injection. These samples were submitted for bulk RNA sequencing to the Genomic facilities at the Wellcome Center for Human Genetics.

Analysis of the SC composition in these samples revealed a clear but transient shift in the SCI scores, with values drifting towards the CBC end of the spectrum 24 hours after DT administration and returning to the pre-injection values by day 5 (Fig.10B). Fluorescent ISH for *Lgr5* and *Anxa1* confirmed these findings: one day after Lgr5 cell depletion, tumours from APC^{Min}*-Lgr5*^{DTR} animals were completely taken over by *Anxa1+* cells, but *Lgr5+* clones re-appeared by day 5 (Fig.10C). These results suggest that *Anxa1+* cells not only repopulate the tumours after CBC have been eliminated by external sources of pressure, but are also plastic enough to allow restitution of the Lgr5+ SC compartment within 5 days. Interestingly, the INF-γ and YAP pathways were upregulated 24 hours after DT injection, suggesting that these processes may be playing a role in the SC changes observed in this scenario (Fig.10D). Importantly, despite the shift in the stem cell populations, CBC ablation did not alter the mean polyps size nor cell proliferation in treated animals (Fig.10E,F).

All in all, these results demonstrate that therapeutic selective pressure can transiently alter the SC composition of colorectal lesions. In this scenario, the observed changes had no apparent effects in tumour burden according to conventional markers of therapy response (ie. tumour size and number of proliferative cells). However,

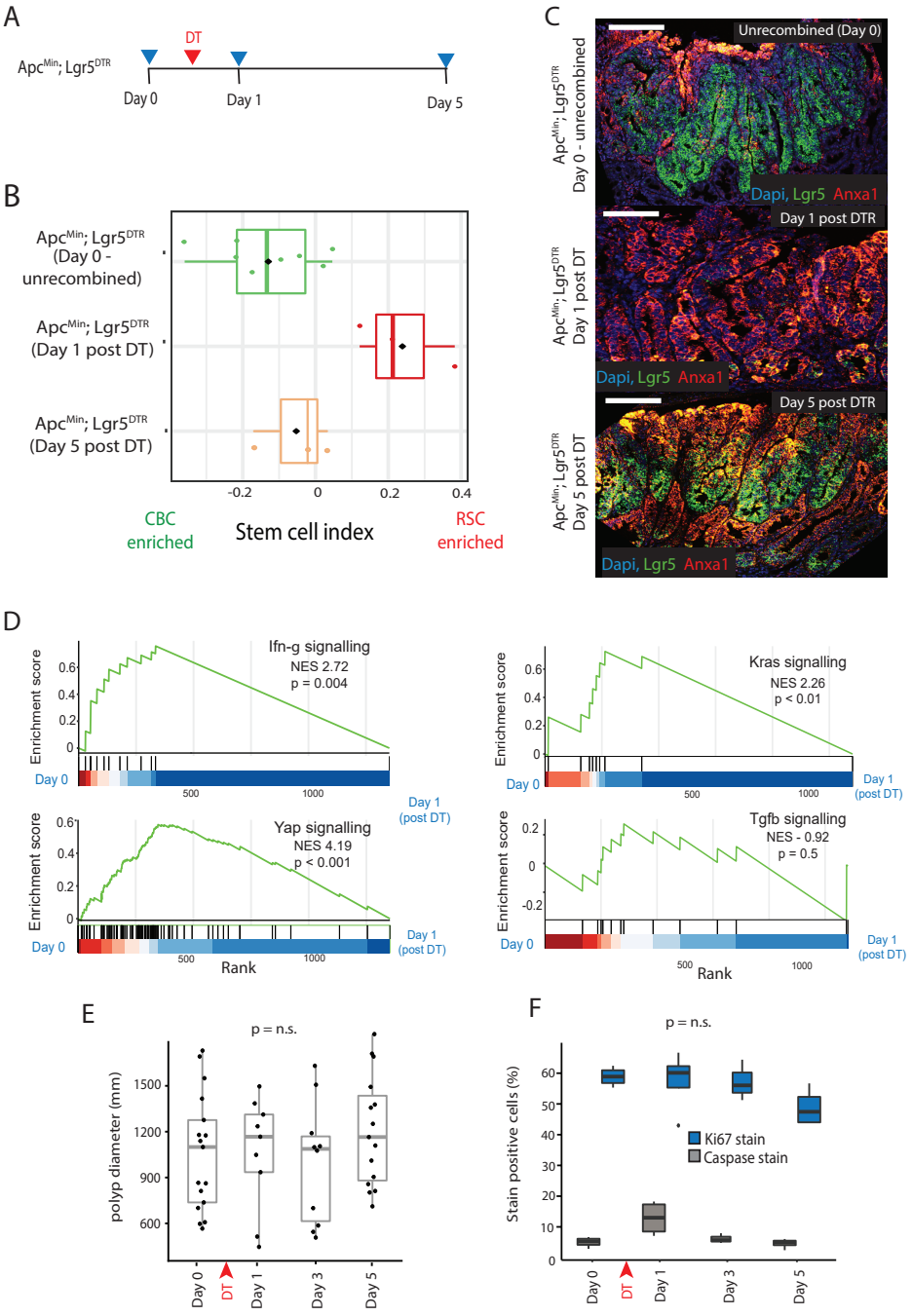

Figure 10: RSC temporally repopulate APC^{Min} mice after CBC ablation. A. Experimental design: mice were injected at day0 and samples collected at days 1, 3 and 5. B-D. Stem Cell Index scores (B), fluorescent *In situ hybridization* for *Lgr5* and *Anxa1* (C) and Gene Set Enrichment Analysis(D) in samples collected on days 1 and 5 versus control. E,F. Polyp diameter (E) and percentage of Ki67+ and caspase3+ cells in samples from days 1,3 and 5 versus control. Differences between groups were tested for using a Wilcoxon test. Complete stats can be found in Sup. Table 2.

whether this plasticity affected therapy response in clinically relevant scenarios remained to be determined.

3.10 Stem Cell Plasticity conditions Chemotherapy Resistance in Human Colorectal Cancer Patients

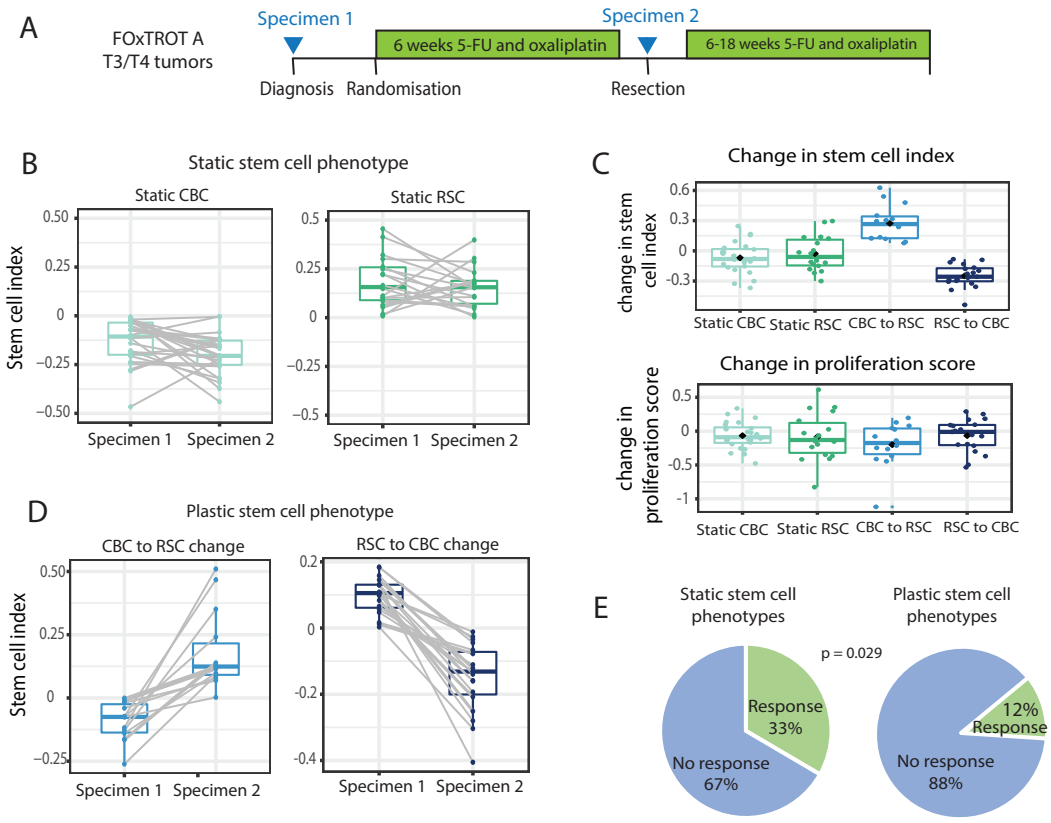

Figure 11: Stem Cell plasticity is associated with therapy resistance in a human data set. A. FOxTROT (track A) trial schedule showing specimen acquisition before (specimen 1) and after 6 weeks (specimen 2) of 5-FU and oxaliplatin chemotherapy. B. Ladder plots showing GSVA (RSC-CBC signature) analysis of human tumor samples before and after chemotherapy, with patients grouped as "static" or "plastic" dependent on directional signature change following therapy. C. Changes in the Stem Cell Index and proliferation scores of "static" and "plastic" patients. D. Responsive and resistant patients within the "Static" and "plastic" groups represented as pie charts. Differences were tested using a Fisher's exact test.

To see whether the plastic nature of the CBC and RSC populations had any clinical implications in treatment response, the Fluoropyrimidine, Oxaliplatin & Targeted

Receptor pre-Operative Therapy for colon cancer (FOxTROT) trial data set (track A) was analysed together with Nadia Nasreddin and Gabriel Valbuena. Patients from these cohort were randomized to neoadjuvant oxaliplatin and 5-FU treatment between diagnostic biopsy and surgical resection, which enables temporal analysis of these samples at a transcriptomic level (Fig.11A). Individuals were classified as "plastic" or "static" depending on the degree of change in their SCI scores before and after therapy, and as "responders" or "non-responders" upon histological assessment. Patients with no regression (no change) or mild regression (residual cancer cells higher than fibrosis) were classified as non-responders, while those with marked regression (presence of fibrosis with rare residual cancer cells) or moderate regression (increased presence of residual cancer cells but still high fibrosis) were classified as responders. Comparison between SC phenotypes revealed that, in this particular cohort, "plastic" patients were less likely to respond to chemotherapy (Fig.11B-D).

In order to rule out that differences in cell proliferation were causing the observed changes, transcriptome based cell proliferation scores were calculated for each individual, but no changes were observed between "plastic" and "static" patients (Fig.11C). Overall, these results seem to indicate that there is an association between the degree of adaptive change experienced by a tumour in response to therapy and the clinical impact of the treatment.

3.11 Discussion

Intestinal homeostasis is maintained by a pool of adult Lgr5+ ISC sitting at the base of the crypts, which constantly divide to renew the intestinal epithelium[1]. Moreover, in settings of tissue injury, a second pool of fetal-like Lgr5- progenitors de-differentiates to enable tissue regeneration, the so-called RSC. The role of Lgr5+ CBC in CRC initiation is well established[125], but the contribution of RSC to colorectal neoplasia was unknown. Here, intrinsic differences in the transcriptomic profiles of these SC populations were used to assess their relative abundance across a range of CRC mouse and human tumours, and to morphologically track adaptive change in the SC composition of tumours in response to therapeutic selective pressure. Furthermore, the spatial distribution of CBC and RSC within lesions was investigated using the transcripts *Lgr5* and *Anxa1* as surrogate markers respectively.

The use of the RSC and CBC signatures in combination adds a new dimension to the study of cancer SC, which in the past was exclusively focused on the Lgr5+

progenitors. Indeed, the set of tools presented in this chapter permitted the segregation of murine and human pre-malignant lesions and tumours based on their SC compositions, which to my knowledge had not been achieved in the past. Importantly, this classification was consistent for lesions with shared molecular characteristics (i.e. within mouse models and CMS/CRIS subtypes), which validates the consistency of the method.

The fact that many of the genes within the RSC signature were shared by non-epithelial cells posed a risk for RSC scores to be a surrogate of the stromal composition of the tumour. Consequently, a curation was conducted by an experienced bioinformatician. The capacity for the resulting gene list to segregate CRC organoids was indicative that the curated signature was capturing a signal from the epithelial compartment. Importantly, the distribution of SCI scores was conserved between organoids and whole-tissue samples, which supports the use of this tool in bulk transcriptomic data.

In this study, differences in SC admixture were observed already in pre-malignant settings, which indicates that the preferential expansion of specific progenitor can occur from early stages in colorectal transformation. These differences were consistent with the cell of origin of the lesion: CBC in the case of TVA and RSC for inflammatory-driven dysplasias. The progenitor from which serrated lesions arise has not yet been established, but the ectopic crypt structures observed within the differentiated areas of these polyps suggests that they may not have a CBC cell of origin[130, 129]. Consistent with this hypothesis, SSL and TSA lesions presented clusters of *Anxa1+* cells at their luminal side, which could potentially be acting as cancer-initiating cells. This is in accordance with a recent report from Chen et al [233], where the authors used single cell transcriptomic techniques to investigate the evolutionary trajectories of TVA and SSL. According to their results, SSL originate from absorptive progenitors with an acquired gastric metaplasia transcriptome, in opposition to the CBC cell of origin of adenocarcinomas. It is worth noting that some of the genes within this "gastric metaplasia" signature, like *ANXA1* or *ANXA10*, are shared by fetal progenitors. The extend to which the processes reflected by the metaplasia [233] and RSC signatures overlap remains to be determined. All in all, this data points towards the existence of a Lgr5- cell of origin for serrated lesions.

Whether epithelial cell-intrinsic mutations are cause or effect of the SC composition of early lesion is a question open to debate. However, recent work by Flanagan et al suggests that Apc mutations may be selected within the confines of intestinal crypts, because they confer a competitive advantage to mutated CBC versus their

non-mutant neighbours[234]. Contrarily, *KRAS/BRAF* mutations occurring outside of the crypt could increase YAP signaling and prime an RSC response[160], which would generate a pool of de-differentiated progenitor cells susceptible to accumulate additional mutations.

Importantly, the proposed set of tools enables tracing of the dynamic changes experienced by stem cell populations within a tumour. In APC^{Min} animals, polyps were repopulated by *Anxa1+* RSC within 24h after CBC ablation. These changes did not alter current clinical read-outs of treatment efficacy, like tumour size or number of proliferative cells, but translated into a shift in the transcriptomic profile of the samples that could be quantified using the SCI. These changes were transient and tumours recovered *Lgr5* expression and their original SCI scores within 5 days, which contrasts with the permanent alteration of the SC phenotype observed in previous studies [158]. The reason why some adaptive processes persist over others is not clear, but it may be related to the molecular background of the lesion and the genetic and epigenetic mechanisms involved in the reprogramming of the cell. Importantly, the plastic behaviour of CBC and RSC may have implications in patient monitoring and treatment design.

In the FOxTROT human cohort, tumours which experienced shifts in their SC composition upon treatment were less likely to present a histopathological response to chemotherapy. In this scenario, plasticity could enable cells to adapt to external sources of pressure which might buffer the effect of the treatment, as observed in APC^{Min}-$Lgr5^{DTR}$ animals. However, this adaptation may shift the sensitivity of the lesion to alternative forms of therapy. In a recent study by Han et al[158], exposure of CRC organoids to a WNT inhibitor drove the appearance of resistant clones. This adaptation implied the permanent conversion of cells from a CBC to a RSC phenotype, which occurred though epithelial YAP activation. Consequently, these organoids were sensitive to YAP inhibition, which could be exploited therapeutically. This indicates that, even when treatments are inefficient according to current clinical read-outs, they may drive SC populations through an evolutionary bottleneck, which can be exploited in combinatorial therapy to accomplish the annihilation of a given lesion (Fig.12). Hence, clinical practitioners could benefit from a measure of the SC phenotype (eg. SCI) when deciding upon and assessing the effects of therapeutic regimes.

The molecular characterization of murine and human colorectal lesions unveiled evolutionary conserved mechanisms associated with the acquisition of RSC and CBC

characteristics. In particular, WNT activation skewed cells towards a CBC phenotype, whereas a range of mutations and signaling pathways were associated with extreme RSC scores. This study revealed that RSC selected for preferential acquisition of mutations in *BRAF*, whereas the percentage of *KRAS* mutations was comparable between both groups. Interestingly, V600E BRAF mutations have been shown to inhibit MST1 in the context of thyroid carcinoma [235]. Further understanding of this interaction in the context of CRC could shed light into the aforementioned observation. Theoretically, BRAF-mediated activation of the Hippo-YAP/TAZ pathway could at least partially explain the preferential BRAF disruption observed in RSC cancers.

It is worth noting that some of these molecular drives arise from cell-intrinsic perturbations (eg. *KRAS* mutation), whereas others are consequence of a reactive stroma (eg. increased TGF-β production). Moreover, the composition of the extracellular matrix is a key element conferring SC identity, which is evidenced by the induction of RSC characteristic in organoids cultured in collagen[32]. It is worth noting that many of these signaling pathways converge at the Hippo/YAP-TAZ node, which was also strongly associated to RSC characteristics in the analysed datasets. In the work of Han et al [158], simultaneous mutation of *SMAD4, KRAS* and *TP53* was insufficient to confer plastic properties to CRC organoids presenting *RSPO3* fusions, but pre-treatment of these samples with TGF-β drove the emergence of RSC clones upon exposure to external sources of pressure (i.e. WNT inhibition). Collectively, this evidence highlights the importance of inter-compartmental interactions in acquisition of a given SC phenotype, and suggests that a plethora of mechanisms may be weighing into this cell fate decision. Functional validation of the fore-mentioned pathways, and study of how the epithelial and stromal compartment co-evolve to sustain a given SC landscape, will require the development of models that permit specific tracing and modification of RSC populations.

In addition to therapy resistance, RSC progenitors have been proposed to play a role in metastasis. According to the work by Jacco et al, cancer dissemination to secondary locations is initiated by Lgr5- progenitor, which re-acquiring a Lgr5+ phenotype when metastatic seeds re-activate and expand[230]. Consequently, therapies targetting the conversion between SC phenotypes may have an important application not only to prevent the development of resistant clones, but also metastasis.

When it comes to WNT disruption, human cancers with a predominant CBC component tended to present LI WNT mutations (eg.*CTNNB1, APC*), whereas RSC tumours had a higher percentage of LD WNT alterations (eg. *RNF43* mutations and

RSPO2/3 fusions). The reason for the differential distribution of WNT-disrupting mechanisms is not clear, but it may be related to the interactions between WNT and the pathways that drive the RSC phenotype. For example, YAP might limit WNT activation in colorectal lesions with LD WNT mutations by interacting with the β-catenin destruction complex[164, 165], which could facilitate the acquisition of a RSC phenotype. Contrarily, in cancers with LI forms of WNT disruption, cytoplasmic YAP-driven WNT inhibition might not be effective, which could help to consolidate the commitment to a CBC phenotype. This would explain why, despite the YAP activation observed in APC^{Min} polyps by Gregorieff et al[77], these lesions were mostly populated by CBC cells. The study of the interaction between WNT and other common plasticity drives may improve our understanding of the differential selection of WNT alterations across malignancies.

Overall, the presented work highlights that inference of changes in the SC molecular phenotype can inform clinical practice, namely the monitoring and scheduling of therapeutic regimes. Moreover, development of novel combinatorial therapies will be needed to exploit the sensitivity of metastatic and therapy resistant clones to pathways that promote cell plasticity. One important limitation of this study lies within the inherent complexity of any given cell type, which is neglected by the assumption made here that the stem cell phenotypes can be characterized as an admixture of two extreme categories. This is evidenced by the fact that not all cells expressing a given marker (eg. Lgr5) will be functionally equivalent[174, 236]. However, the possibility to use the SCI on bulk transcriptomic samples and its potential to detect shifts in the SC composition of a tumour renders this as a practical tool for its use in the clinic.

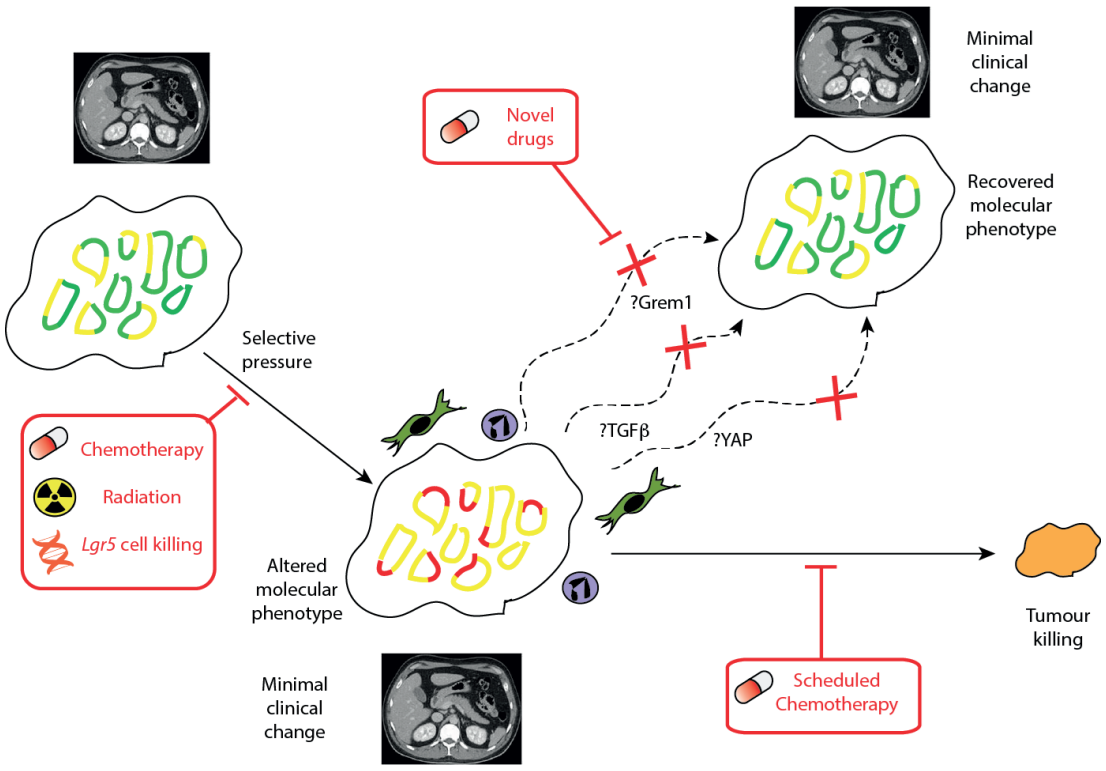

Figure 12: Combinatorial therapies based on the Stem Cell composition of the tumour. Treatment of tumours with conventional methods (eg. chemotherapy, radiotherapy) can drive cell through an evolutionary bottleneck which completely alters the Stem Cell composition of the tumours while rendering them dependent on pathways that enable cell plasticity (eg. TGF-β, YAP). Combinatorial therapy aimed to inhibit these plasticity drives could culminate in the successful eradication of the tumour.