Molecular and Histological Characterisation of Colitis-associated Colorectal Cancer



Nadia Nasreddin

Wolfson College

Nuffield Department of Medicine

University of Oxford

This thesis is submitted for the degree of

Doctor of Philosophy in Clinical Medicine

Michaelmas 2022

"Considerate la vostra semenza:

fatti non foste a viver come bruti

ma per seguir virtute e canoscenza."

Divina Commedia, Inferno, Canto XXVI, I. 118-20

Dante

"Consider well the seed that gave you birth:

you were not made to live your lives as brutes,

but to be followers of virtue and knowledge. "

Divine Comedy, Inferno, Canto XXVI, I. 118-20

Dante

Abstract

Colitis-associated colorectal cancer (CA-CRC) is phenotypically, biologically and molecularly different from sporadic colorectal cancer (S-CRC), and its carcinogenesis is poorly understood. In this thesis, I explored the biological and molecular differences between S-CRC and CA-CRC, with the goal of better elucidating the carcinogenesis of CA-CRC and improving patient management. Three specific aspects were investigated: the stem cell dynamics in colorectal cancer (CRC), the WNT signalling landscape and the early detection of CA-CRC.

During homeostasis, perpetual self-renewal of the intestinal epithelium is maintained by the crypt base columnar stem cells (CBC). However, upon damage a dynamic regeneration response, characterised by dedifferentiation leads to the emergence of a reservoir of regenerative stem cells (RSC). This is essential for regeneration, but can also drive neoplastic conversion. Using a combined histological and molecular investigative approach, the stem cell dynamics in human lesions and mouse models of CRC was assessed. This led to the identification of variable expression of stem cell markers across different lesion subtypes and mouse models. The degree of CBC/RSC admixture was associated with specific epithelial mutation signatures and microenvironmental signalling disruption. Furthermore, the CBC/RSC equilibrium of a lesion was found to be adaptive, with adaptability being associated with poor response to treatment.

While aberrant WNT activation is a hallmark of S-CRC, the role of the WNT signalling pathway in CA-CRC is yet to be fully understood. Herein, a molecular and histological characterisation of colitis-associated low grade dysplasia (LGD) suggested a potential selection for ligand-dependent (LD) alterations in colitis-associated lesions. To determine the reason behind such potential selection, the cell dynamics in LD lesions, particularly those carrying *Rspo3* alteration, was examined in murine models of WNT disruption. This revealed that *Rspo3* overexpression causes an increase in Paneth cells, with *Rspo3* overexpression in Atoh1-positive cells skewing secretory cell fate towards Paneth cells. These are an important source of antimicrobials and WNT ligands, thus an increase in Paneth cells may encourage regenerative healing.

Due to their increased cancer risk, patients with longstanding IBD are offered endoscopic surveillance with concomitant histopathological assessment, aimed at identifying dysplasia as precursor lesion. However, this strategy is beset with difficulties and limitations. Herein, three novel stratification methods were assessed in progressor and nonprogressor LGD: a) A histological classification of LGD using a newly outlined criteria, followed by assessment of the association between nonconventional dysplasia and progression; b) A high-resolution assessment and comparison of copy number alteration (CNA) burden between the two patient groups; c) A digital pathology analysis investigating cellular features of progressor and nonprogressor LGD. From the strategies assessed, CNA burden was the only cancer risk predictor capable of discriminating between colitis-associated LGD likely to progress and those that are not.

Acknowledgments

Working on this thesis provided me with the great opportunity to meet new people, learn from and collaborate with incredible scientists, pathologists, physicians and support staff, while having the love and support of my family and friends, all of whom I would like to acknowledge.

Thank you to my supervisor Simon Leedham for providing me with the opportunity to pursue a DPhil. For allowing me to explore my ideas and scientific curiosities. For his support throughout my DPhil, including sharing his office with me so I could have a quiet place to write in. Thank you for being a reachable PI, one you can send a quick email or knock on the door for a quick meeting. Thank you to my second supervisor James East for his support, particularly for providing me with financial support and stability for me to be able to undertake my DPhil without worrying.

Thank you to past and present members of the Leedham lab. Thank you to Gabriel for his incredible support. From his advice on data analysis and writing, to advice on how to respond to situations and emails, to the "GURL, pull yourself together" interventions, I have been able to count on him and his addition to the lab has enriched my DPhil. Thank you to Hayley for introducing me to the Leedham lab, keeping the lab running allowing me to focus on the experiments, for all the ordering she very patiently did for me, for teaching me the different histology techniques and introducing me to the world of mouse work. Thank you to Eoghan for his help in the lab and his bubbly personality. Thank you to Ester for her infectious passion for science and for all the crazy fun memories. Thank you to Martijn for patiently teaching me various techniques, while sharing all his tips and tricks, for checking on me and all his advice even after having left the lab. Thank you to Ryan for the fun memories and lab get-togethers organised in his time in Oxford.

Thank you to the fantastic pathologists, who not only provided me with their expertise and knowledge, but also gave me the opportunity to improve my limited pathology knowledge from my undergraduate lectures: Maurice, Lai Mun, Alistair, Marnix, Viktor, Manuel, Marco and Morgan. A particularly special thank you to Maurice and Lai Mun for going above and beyond, for their advice and patience. Thank you to Jan Hudecek for allowing me to use his online reviewer for the pathology-based study and for being extremely helpful when I was setting up the study. Thank you to Myrtle J.van der Wel for her advice on the statistical analysis of the pathology-based study.

Thank you to Trevor and the collaborators in the Graham lab, particularly the previous members (Ibrahim, Kit, Annie, Will, Chris and Jacob) who had made me feel part of the group in my time there. A particular thank you to Ibrahim, with whom I began by NGS journey. I have had the opportunity to learn from the best, whilst

sharing moments of scientific and random discussion, horrible lab radio choices, great food (I now know my way through Barbican) and same level of paranoid three times labelling of each tube. Thank you to Kit for her help with data analysis, Kane for his help and Salpie.

Thank you to the other collaborators who were involved in the stem cell work, including members of the Sansom lab, members of the Dunne lab, Sabine, Petros, Pratyaksha, David, the SCORT consortium, particularly Andrew and James for their patience and help with giving access to the data. A particular thank you to Philip, for his great support, and Keira, for all her help.

Thank you to the Genomics Birmingham core facility for their help with the RNA fusion panel, particularly to Celina and Andrew.

Thank you to the technicians in the FGF Unit for all their hard-work. Thank you for the weekends spent weighting my DSS mice, for the all the help and guidance. Thank you also for all the triple checking of the mice they indulged me with for my peace of mind. A special thank you to Luke, for all his help and advice, the laughs (mostly at my expense), and for turning mouse work from terrifying to fun. A special thank you also to Doug, for his supportive words, Steve, Claudia, Jade and Heather.

Thank you to the cores at the Wellcome Centre for Human Genetics. To the OGC and particularly Simon Engleglow for coordinating all my NGS projects with high efficiency and for always being helpful. To James and Ed for making microscopy accessible and for their help with all my imaging work. Thank you to the Flow Cytometry Facility and Ruddy for all his help and advice on FACS.

Thank you to Felicia for her help with fluorescent IHC, the scientific discussions and being a friendly face at the Wellcome Centre. Similarly, thank you to all the other friendly faces at the Wellcome Centre, who made it a nicer, warmer place to work in: Annie, Rachel, Emma, Santiago, Mike, Alicia, Kasia, Ruth and Natalie...

Thank you to the TGU for allowing me to aquire the material necessary for my studies, and particularly to the people who facilitated this: Michael, Yukari, James, and Andrew. A particular thank you to Andrew, who has always been very nice, very kind and helpful. Thank you to the OCHRe, particularly Stephanie and Alicia, for their help in getting samples from the Oxford Trust.

Thank you to my friends, particularly my best friend Sulo for all her support. For spoiling me with chocolates, tea, flowers and every other pampering item she has sent on my way to make this journey, and particularly the writing processes enjoyable. For the experiments run together, mental health tea-breaks (all those biscoffs), conferences attended and the initiatives we dragged each other to. Thank you for Daniel for having to share Sulo with me and having to (very patiently) deal with the both of us. Thank you to Noeline for the humour over the shared PhD tribulations and her motivational texts. Thank you to Raquel for the lovely memories at the Mad hatter and Cafe Baba, her support and for being such an inspiration as an incredible young woman in research.

Thank you to my family for all their love, support and patience. Thank you to my Mum for being such an amazing role model throughout my life and for playing an active role in supporting anything I put my mind into (from attending recitals, sports' matches and cerimonies, to helping with projects). Thank you also for all the cakes baked and Habesha teas made as a thesis writing treat. Thank you to my Dad for believing in me and supporting me in every way he could, from treats, to supportive words and encoragements, and for making sure to check on me from wherever around

the world. To my little brother, Nadal, who was (unlucky enough) to be here with me to witness the ups and downs of writing a thesis, thank you for his love, patience and endless encouragements (as well as for still being my affectionate Nadalino even as a teenager). To my sister, Dalia, for believing in me, for the memories of us growing up together through our move to the UK and academic years, and for the early morning/evening phone calls to keep me company during my walks to/from work.

Thank you to my big, fat and beautiful Raffetto family for always being there for me. To my uncle Mauro, my aunties Rita and Alba for their support, checking-ins and Sunday messages. My cousins, who are more like older siblings, Lella, Maura and Silvio for being my support system since moving to the UK. To Lella, for all the times she did London-Oxford to be there for me, including rushing me back to Oxford for a sick mouse, to Maura, for inspiring me to always point towards greater things, and Silvio, for being my study, travel, fun and procastination buddy. Thank you to David, my acquired English cousin, for showing interest in what I do and always being supportive, to Arnaud, my protective acquired cousin for always being super protective and my typo-checker, and Ian, for his support and all the Cosmopolitans over the years. Thank you to the little humans, Amélie, Éloïse and Alex, who have been a source of joy and a pleasant distraction from the stresses of a PhD. Thank you to aunty Negiat for her support.

To Matt, thank you for being the light in the tunnel. I could not have had a better person to share this journey with. Through the ups and downs of a DPhil, your hard-working nature, passion for science and meticulousness have inspired me to hold my work to a high standard, whilst your love, support and kindness have made me feel 100% supported, listened and understood. Thank you for all your advice, the times you had to listen to me rant, for being there for me at all times, all the wonderful cooking you have treated me to after a long day of work (including making Engera), for all the reasurring you did, for helping me set visual code for thesis and allowing me to use your precious PC for the large QuPath analyses, for being the reason behind my better work-life balance, and much, much, more. You and your ability to always make me laugh have made this journey a lot easier.

Lastly, thank you to anyone I might have failed to mention who has contributed to this journey.

Preface

This thesis is a collection of research work carried out by the author in the Gastrointestinal Stem Cell Biology Laboratory, at the Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, between October 2018 and April 2022, under the supervision of Prof Simon Leedham. The research described herein is original. Many aspects of this thesis involved collaborating with other researchers, accordingly work of others is acknowledged and cited with appropriate references within the text.

The study reported in chapter 3, "The stem cell population dynamics in sporadic and colitis-associated colorectal cancer", was co-developed with Dr Ester Gil Vazquez as part of our DPhil projects and the chapter contains work published in the following paper:

- Vazquez, E., **Nasreddin, N.,** Valbuena, G., Mulholland, E., Belnoue-Davis, H., Eggington, H., Schenck, R., Wouters, V., Wirapati, P., Gilroy, K., Lannagan, T., Flanagan, D., Najumudeen, A., Omwenga, S., McCorry, A., Easton, A., Koelzer, V., East, J., Morton, D., Trusolino, L., Maughan, T., Campbell, A., Loughrey, M., Dunne, P., Tsantoulis, P., Huels, D., Tejpar, S., Sansom, O. and Leedham, S., 2022. Dynamic and adaptive cancer stem cell population admixture in colorectal neoplasia. Cell Stem Cell, 29(8), 1213-1228.e8.
  - *Journal article*

The study reported in chapter 5, "The early detection of colitis-associated colorectal cancer" was part of a large multicentre collaboration and comprised three strands.

Strand B: Early detection of colitis-associated colorectal cancer via copy number alteration burden, was developed with Dr Ibrahim Al-Bakir and Dr Kathleen Curtius and is part both of Dr Al-Bakir's and my own doctoral theses. Some of the work described in strand B was published in the following conference abstracts:

- Al Bakir, I., Curtius, K., **Nasreddin, N.,** Clarke, T.S.O., Moorghen, M., Rodriguez-Justo, M., Jansen, M., Wright, N.A., Leedham, S.J., Hart, A.L. and Graham, T.A., 2019, September. A Quantitative Evolutionary Approach Utilising High Resolution Chromosomal Copy Number Analysis Accurately Stratifies Patients with Ulcerative Colitis and Low Grade Dysplasia by Future Colorectal Cancer Risk. In JOURNAL OF PATHOLOGY (Vol. 249, pp. S12-S12). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

  - *Conference abstract*

- Curtius, K., Al Bakir, I., Baker, A., Clarke, T., **Nasreddin, N.,** Kopczynska, M., Agnew, M., Smith, K., Moorghen, M., Rodriguez-Justo, M. and Jansen, M., 2019. Quantifying evolution of early dysplastic lesions in ulcerative colitis predicts future colorectal cancer risk. Gastroenterology, 156(6 (S1)).

  - *Conference abstract*

Furthermore, the Notum staining in human premalignant lesions reported in strand B of chapter 4 of the thesis has been published in the following paper:

- Flanagan, D., Pentinmikko, N., Luopajärvi, K., Willis, N., Gilroy, K., Raven, A., Mcgarry, L., Englund, J., Webb, A., Scharaw, S., **Nasreddin, N.,** Hodder, M., Ridgway, R., Minnee, E., Sphyris, N., Gilchrist, E., Najumudeen, A., Romagnolo, B., Perret, C., Williams, A., Clevers, H., Nummela, P., Lähde, M., Alitalo, K., Hietakangas, V., Hedley, A., Clark, W., Nixon, C., Kirschner, K., Jones, E., Ristimäki, A., Leedham, S., Fish, P., Vincent, J., Katajisto, P. and Sansom, O., 2021. NOTUM from Apc-mutant cells biases clonal competition to initiate cancer. Nature, 594(7863), 430-435.

– *Journal article*

I have additionally contributed to the following publications that are not covered in this thesis:

- Koppens, M., Davis, H., Valbuena, G., Mulholland, E., **Nasreddin, N.,** Colombe, M., Antanaviciute, A., Biswas, S., Friedrich, M., Lee, L., Wang, L., Koelzer, V., East, J., Simmons, A., Winton, D. and Leedham, S., 2021. Bone Morphogenetic Protein Pathway Antagonism by Grem1 Regulates Epithelial Cell Fate in Intestinal Regeneration. Gastroenterology, 239-254.e9.

  – *Journal article*

- Heino, S., Fang, S., Lähde, M., Högström, J., Nassiri, S., Campbell, A., Flanagan, D., Raven, A., Hodder, M., **Nasreddin, N.,** Xue, H., Delorenzi, M., Leedham, S., Petrova, T., Sansom, O. and Alitalo, K., 2021. Lef1 restricts ectopic crypt formation and tumor cell growth in intestinal adenomas. Science Advances, 7(47).

  – *Journal article*

Collaborators

Dr Ester Gil Vazquez - Assistance with study design of cell of origin study and wet lab sample processing as part of their DPhil

Dr Gabriel N Valbuena - Assistance with bioinformatics analysis as postdoctoral researcher in the Leedham laboratory

Dr Hayley L Belnoue-Davis - Injection of mouse models, assistance with mouse work and sample preparation for FACS as postdoctoral laboratory manager in the Leedham laboratory

Ms Sulochana Omwenga - Assistance with wet lab sample processing as research assistant in the Leedham laboratory

Dr Eoghan J Mulholland - Assistance with quantification of immunohistochemistry samples as postdoctoral researcher in the Leedham laboratory

Dr Ibrahim Al Bakir - Study design of early detection of colitis-associated colorectal cancer by copy number alteration and assistance with wet lab sample processing of the discovery colitis cohort as part of their PhD

Dr Kathleen Curtius - Study design of early detection of colitis-associated colorectal cancer by copy number alteration and bioinformatics analysis of low pass whole genome sequencing data as postdoctoral researcher in Prof Trevor Graham laboratory

Mr Kane Smith - Targeted sequencing panel library preparation and assistance with wet lab sample processing as research technician in Prof Trevor Graham laboratory

Dr Salpie Nowinski - Bioinformatics analysis targeted panel sequencing data as postdoctoral researcher in Prof Trevor Graham laboratory

Dr Dustin Flanagan - Collaboration on Notum paper

Dr Lai Mun Wang - Histopathological assessment

Dr Maurice Loughrey - Identification of cases of interest from the Belfast and histopathological assessment

Dr Marnix Jansen - Histopathological assessment

Dr Manuel Rodriguez-Justo - Histopathological assessment

Dr Marco Novelli - Histopathological assessment

Dr Viktor Koelzer - Histopathological assessment

Dr Alistair Easton - Histopathological assessment

Dr Morgan Moorghen - Histopathological assessment

Professor James E East - Second supervisor

Professor Simon J Leedham - Mentorship, direct supervision and study design

Contents

Abstract i

Acknowledgments ii

Preface v

Collaborators viii

Acronyms xix

List of Tables xxvi

List of Figures xxviii

1 General introduction 1

1.1 The intestinal tract . . . 1

1.1.1 Intestinal architecture . . . 2

1.1.2 Intestinal cellular dynamics . . . 4

1.1.3 The stem cell niche . . . 5

1.2 Signalling pathways in intestinal homeostasis . . . 6

1.2.1 The WNT signalling pathway . . . 6

1.2.1.1 WNT signalling in physiological conditions . . . 9

1.2.2 The BMP signalling pathway . . . 9

1.2.2.1 BMP signalling pathway in physiological conditions . . . 10

1.2.3 The Notch signalling pathway . . . 10

1.2.3.1 Notch signalling pathway in physiological conditions . 11
1.2.4 The TGFB signalling pathway 12
1.2.4.1 TGFB signalling pathway in physiological conditions . 13
1.2.5 The Hippo-YAP/TAZ signalling pathway 13
1.2.5.1 Hippo-YAP/TAZ signalling pathway in physiological conditions 14
1.3 Disruption of homeostasis 14
1.3.1 Inflammatory Bowel Disease 16
1.3.1.1 Epidemiology 16
1.3.1.2 Aetiology 17
1.3.1.3 Pathophysiology 18
1.4 Inflammation, Wound healing and Cancer 18
1.5 Colorectal Cancer 19
1.5.1 Epidemiology 19
1.5.2 Aetiology 19
1.5.3 Signalling pathways in intestinal carcinogenesis 20
1.5.4 Pathogenesis 21
1.5.5 Colorectal cancer classification 23
1.5.5.1 Aetiology based classification 24
1.5.5.2 Histological classification 24
1.5.5.3 Molecular classification 26
1.5.6 Disease models of colorectal cancer 28
1.5.6.1 Mouse models 29
1.5.6.2 Organoids 30
1.6 Colitis-associated colorectal cancer 30
1.6.1 Pathogenesis: chronic inflammation and colitis-associated colorectal cancer 32
1.6.2 Colitis-associated colorectal cancer versus sporadic colorectal cancer 34
1.7 Thesis objectives 37

2 Materials and Methods **39**

2.1 General Materials and Methods . . . 39

2.1.1 Human . . . 39

2.1.1.1 Human cohorts and ethical approval . . . 39

2.1.2 Mouse . . . 40

2.1.2.1 Housing . . . 40

2.1.2.2 Mouse strains and models . . . 40

2.1.2.3 Breeding . . . 40

2.1.2.4 Earclip . . . 42

2.1.2.5 Genotyping . . . 42

2.1.2.6 Induction of recombination . . . 43

2.1.2.7 Collection . . . 43

2.1.2.8 Processing . . . 44

2.1.2.9 Embedding . . . 45

2.1.3 Sectioning . . . 45

2.1.4 Histology . . . 45

2.1.4.1 Haematoxylin and Eosin staining . . . 45

2.1.4.2 Haematoxylin staining . . . 46

2.1.4.3 *In situ* hybridisation . . . 46

2.1.4.4 Immunohistochemistry . . . 47

2.1.4.5 Imaging and analysis . . . 48

2.1.5 Nucleic acids . . . 48

2.1.5.1 Epithelial enrichment for DNA extraction . . . 48

2.1.5.2 DNA extraction from FFPE samples . . . 48

2.1.5.3 Assessment of DNA quantity and quality . . . 49

2.1.5.4 Tissue enrichment for RNA extraction . . . 49

2.1.5.5 RNA extraction from FFPE samples . . . 50

2.1.5.6 RNA extraction from fresh frozen tissue . . . 50

2.1.5.7 Assessment of RNA quality and quantity . . . 51

2.1.6 3′ RNA sequencing . . . 51

2.1.6.1 3′RNA sequencing library preparation for FFPE samples 51
2.1.6.2 3′RNA sequencing library preparation for fresh frozen samples . . . 52
2.1.6.3 3′RNA sequencing library pooling and sequencing . . . 52
2.1.6.4 3′RNA sequencing raw data processing . . . 53
2.2 Chapter 3 specific materials and methods . . . 53
2.2.1 Sample selection . . . 53
2.2.2 Derivation of molecular signatures . . . 54
2.2.3 RSC signature curation . . . 54
2.2.4 Identification of signature markers and assessment of expression of markers . . . 55
2.2.5 Stem cell index calculation . . . 55
2.2.6 Assessment of stem cell marker expression . . . 55
2.2.7 Transcriptomic profiling . . . 55
2.2.8 Stem cell sorting by FACS and clonogenicity analysis . . . 56
2.2.9 Analysis of driver genes associated with variable stem cell molecular phenotype . . . 58
2.2.10 Analysis of pathways associated with variable stem cell molecular phenotype . . . 59
2.2.11 Determination of stem cell phenotype dynamicity in *in vivo* murine model . . . 59
2.2.12 Analysis of associations of stem cell phenotype to therapy response 60
2.2.13 Assessment of intratumor heterogeneity in the BOSS study . . . 61
2.3 Chapter 4 specific materials and methods . . . 62
2.3.1 Phenotyping of the WNT signalling landscape in colitis-associated human samples . . . 62
2.3.1.1 Sample selection . . . 62
2.3.1.2 Sample processing . . . 62
2.3.1.3 Assessment of the expression of the biomarker *AXIN2* . 62

2.3.1.4 Mutational phenotyping for WNT genes of colitis-associated LGD . . . 63
2.3.1.5 Assessment of *RSPO* expression . . . 63
2.3.1.6 Validation of *RSPO* expression by fusion panel and *in situ* hybridisation . . . 63
2.3.2 Assessment of WNT alterations in mouse models of WNT disruption . . . 64
2.3.2.1 Mouse models . . . 64
2.3.2.2 *In situ* hybridisation for *NOTUM* . . . 65
2.3.2.3 Assessment of the effects of Rspo3 on cell fate determination . . . 65
2.3.2.4 Paneth cells and goblet cells characterisation by *in situ* hybridisation and fluorescent immunohistochemistry . . . 66
2.4 Chapter 5 specific materials and methods . . . 67
2.4.1 Sample Selection . . . 67
2.4.2 Histopathological assessment of colitis-associated progressor and nonprogressor dysplasia . . . 68
2.4.2.1 Slides staining and scanning . . . 68
2.4.2.2 Assessors and assessments . . . 69
2.4.2.3 Assessment of diagnostic agreement . . . 69
2.4.2.4 Assessment of association between nonconventional dysplasia and progression . . . 71
2.4.3 Molecular assessment of colitis-associated progressor and non-progressor dysplasia . . . 72
2.4.3.1 DNA isolation . . . 72
2.4.3.2 Sonication of lowly fragmented DNA for lpWGS library preparation . . . 72
2.4.3.3 lpWGS library preparation, pooling and sequencing . . . 72
2.4.3.4 lpWGS raw data processing . . . 74
2.4.3.5 Analysis of copy number alterations . . . 74

2.4.4 Digital cytohistological assessment of colitis-associated progressor and nonprogressor dysplasia . . . 76
2.4.4.1 Haematoxylin staining and scanning . . . 76
2.4.4.2 HALO analysis . . . 76
2.4.4.3 HALO data analysis . . . 77

3 The stem cell population dynamics in sporadic and colitis-associated colorectal cancer 79
3.1 Introduction . . . 79
3.2 Study rationale, design study and aims . . . 84
3.3 Results . . . 85
3.3.1 Sample selection . . . 85
3.3.2 Derivation of stem cell molecular signatures and markers . . . 86
3.3.3 Stem cell marker expression across human premalignant and malignant subtypes . . . 89
3.3.4 Stem cell expression across murine intestinal neoplasia models . 92
3.3.5 Stem cell potential of murine CBC and RSC cell populations . . . 94
3.3.6 Driver genes and pathways associated with CBC and RSC stem cell molecular phenotypes . . . 97
3.3.7 Stem cell phenotype dynamics in *in vivo* murine model . . . 101
3.3.8 Clinical implication of the adaptive response to selective pressure of cancer stem cells . . . 105
3.4 Key findings . . . 108
3.5 Discussion . . . 108
3.6 Limitations and future work . . . 114

4 The WNT signalling pathway in colitis-associated colorectal cancer 116
4.1 Introduction . . . 116
4.2 Study rationale, design study and aims . . . 122
4.3 Results . . . 123
4.3.1 Phenotyping of colitis-associated precursor lesions . . . 123

4.3.1.1 Assessment of *AXIN2* expression to predict ligand dependency in colitis-associated precursor lesions 124
4.3.1.2 Mutational phenotyping for WNT genes of colitis-associated LGD 125
4.3.1.3 Assessment of *RSPO* alterations 126
4.3.1.4 Comprehensive assessment of the multifaceted profiling of the colitis-associated LGD 130
4.3.2 Assessment of WNT alterations in mouse models of WNT disruption 132
4.3.2.1 *NOTUM* expression pattern across human and mouse lesions bearing WNT alterations 133
4.3.2.2 Characterisation of intestinal phenotype in murine models of WNT disruption 135
4.3.2.3 Assessment of the effects of Rspo3 on Paneth cells . . . 138
4.3.2.4 Assessment of the effects of Rspo3 on goblet cells 140
4.3.2.5 Assessment of the autocrine/paracrine nature of Rspo3's effect on cell fate determination 142
4.3.2.6 Paneth cells in human colitis-associated lesions 143
4.4 Key findings 145
4.5 Discussion 146
4.6 Limitations and Future work 150

5 The early detection of colitis-associated colorectal cancer 152
5.1 Introduction 152
5.1.1 Stratification based on lesion macroscopic appearance 154
5.1.2 Stratification based on lesion microscopic appearance 155
5.1.3 Stratification based on lesion molecular phenotype 156
5.2 Study rationale, design study and aims 156
5.3 Results 158

5.3.1 Histopathological assessment of colitis-associated progressor and nonprogressor dysplasia . . . 158
5.3.1.1 Demographics and lesion details . . . 159
5.3.1.2 Histological assessment and diagnostic agreement . . . 161
5.3.1.3 Diagnostic agreement in the assessment of sporadic regions . . . 164
5.3.1.4 Diagnostic agreement in the assessment of colitis regions 166
5.3.1.5 Association between nonconventional dysplasia and progression . . . 168
5.3.1.6 Association between progression and high-risk histologies . . . 171
5.3.2 Molecular assessment of colitis-associated progressor and nonprogressor dysplasia . . . 174
5.3.2.1 Demographics and lesion details . . . 174
5.3.2.2 Copy number profiles in progressor and nonprogressor patients . . . 176
5.3.2.3 Copy number events . . . 180
5.3.3 Digital cytohistological assessment of colitis-associated progressor and nonprogressor dysplasia . . . 185
5.3.3.1 Demographics and lesion details . . . 185
5.3.3.2 Digital pathology analysis results . . . 187
5.3.3.3 Exploratory analysis of digital pathology data . . . 187
5.3.3.4 Analysis of the histological features of the epithelium . 189
5.4 Key findings . . . 193
5.5 Discussion . . . 193
5.6 Limitations and Future work . . . 197

6 Thesis discussion **199**

6.1 Summary of results and clinical relevance . . . 199
6.2 Main discussion . . . 203

6.2.1 Colitis-associated colorectal cancer: a molecularly heterogeneous group . . . 203
6.2.2 Colitis-associated colorectal cancer: cancer stem cell and cancer cell of origin . . . 204
6.2.3 WNT signalling alterations selfish vs altruistic model . . . 205
6.2.4 From *TP53* mutation to aneuploidy . . . 208
6.2.5 Colitis-associated colorectal cancer carcinogenesis: how does work from this thesis shape our current understanding . . . 210
6.3 Final remarks . . . 211

Appendix A **212**
A.1 Chapter 3 specific materials and methods: Mouse cohort . . . 212

Appendix B **214**
B.1 The stem cell population dynamics in sporadic colorectal cancer and colitis-associated colorectal cancer . . . 214

Appendix C **224**
C.1 The WNT signalling pathway in colitis-associated colorectal cancer: Phenotyping of colitis-associated precursor lesion . . . 224

Bibliography **226**

Acronyms

3'RNAseq 3'RNA sequencing.

Ahr aryl hydrocarbon receptor.

Alpi alkaline-phosphate intestine.

Anapc1 anaphase-promoting complex subunit 1.

Anxa1 annexin 1.

Apc adenomatous polyposis coli.

Apc2 apc regulator of WNT signalling pathway 2.

Apcdd1 apc down-regulated 1.

Atoh1 atonal bHLH transcription factor 1.

Axin2 axin 2.

BAM binary alignment map.

Bcl9 Bcl9 transcription coactivator.

Bcl9l Bcl9-like.

Bmi1 polycomb complex protein 1.

Bmp bone morphogenetic protein.

Bmpr1 bone morphogenetic protein type I receptor.

BOSS biopsies of surgical specimens.

Braf B-Raf proto-oncogene, serine/threonine kinase.

BRC biomedical research centre.

bTrCP1 beta-transducin repeat containing E3 ubiquitin protein ligase pseudogene 1.

CA-CRC colitis-associated CRC.

CBC crypt base columnar.

CCA/D crypt cell atypia.

CCD crypt cell dysplasia.

CD Crohn's disease.

CD31 cluster of differentiation 31.

CD45 cluster of differentiation 45.

Cdkn1a cyclin dependent kinase inhibitor 1A.

Cdkn1b cyclin dependent kinase inhibitor 1B.

Cdkn2b cyclin dependent kinase inhibitor 2B.

CIMP CpG island methylator phenotype.

CIN chromosomal instability.

CMS consensus molecular subtypes.

CN copy number.

CNA copy number alteration.

CPM count per million.

CRC colorectal cancer.

Crebbp Creb binding protein.

CRIS colorectal cancer intrinsic subtypes.

CSL CBF1, suppressor of hairless, lag-1.

Csnk1a1 casein kinase 1 alpha 1.

Ctnnb1 catenin beta 1.

DAPI 4',6-diamidino-2-phenylindole.

DEPC diethyl pyrocarbonate.

Dkk1 dickkopf WNT signalling pathway inhibitor 1.

Dlc1 Dlc1 Rho GTPase activating protein.

Dlck1 doublecortin like kinase 1 protein.

Dll1 delta-like1.

DNA deoxyribonucleic acid.

dNTP deoxynucleoside triphosphate.

DPD dysplasia with increased Paneth cell differentiation.

dsDNA double-stranded DNA.

DTR diphtheria toxin receptor.

EDTA ethylenediaminetetraacetic acid.

Egf epidermal growth factor.

Egfr epidermal growth factor receptor.

EMT epithelial-mesenchymal transition.

EpCAM epithelial cellular adhesion molecule.

Ephb2 EPH receptor B2.

F-CRC familial CRC.

FACS fluorescence-activated cell sorting.

FBS fetal bovine serum.

FFPE formalin-fixed paraffin-embedded.

FMO fluorescence minus one.

FOxTROT Fluoropyrimidine, Oxaliplatin and Targeted-Receptor pre-Operative Therapy.

Fzd frizzled.

GCD goblet cell deficient.

GDC genomic data commons.

GI gastrointestinal.

GSEA gene set enrichment analysis.

Gsk3 glycogen synthase kinase.

Gsk3b glycogen synthase kinase 3 beta.

GSVA gene set variation analysis.

H-CRC hereditary CRC.

H&E haematoxylin and eosin.

HDAC histone deacetylases.

Hes1 hes family bHLH transcription factor 1.

HGD high grade dysplasia.

HH Hedgehog.

IBD inflammatory bowel disease.

Id inhibitor of DNA binding.

Igf2 insulin like growth factor 2.

IHC immunohistochemistry.

Il-22 interleukin 22.

Il22r interleukin 22 receptor.

Infg interferon gamma.

ISH *in situ* hybridisation.

Kras KRAS proto-oncogene.

Lats large tumour suppressor kinase.

LCM laser-capture microdissection.

LD ligand-dependent.

Lef1 lymphoid enhancer factor 1.

LGD low grade dysplasia.

Lgr4 leucine-rich repeat-containing G-protein coupled receptors 4.

Lgr5 leucine-rich repeat-containing G-protein coupled receptors 5.

LI ligand-independent.

lpWGS low pass whole genome sequencing.

LRC label-retaining cells.

Lrig1 leucine-rich repeats and immunoglobulin-like domains 1.

Lrp LDL receptor related protein.

Ly6a lymphocyte antigen 6 complex.

MAML mastermind-like proteins.

Mapk mitogen-activated protein kinase.

MLH1 MutL homolog 1.

MMR mismatch repair.

Mob1a MOB kinase activator 1a.

Mob1b MOB kinase activator 1b.

MSI microsatellite instability.

MSS microsatellite stable.

Myc MYC proto-oncogene, bHLH transcription factor.

NCID notch intracellular domain.

Nfkb1 nuclear factor kappa B.

NGS normal goat serum.

NHS national health service.

NOS not otherwise specified.

Notum notum, palmitoleoyl-protein carboxylesterase.

NSAIDS non-steroidal anti-inflammatory drugs.

Osmr oncostatin M receptor.

OxFP fluorouracil and oxaliplatin.

PBS phosphate-buffered saline.

PCA principal component analysis.

PCR polymerase chain reaction.

Pigr polymeric immunoglobulin receptor.

Pik3ca phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha.

PLAUR plasminogen activator, urokinase receptor.

PSC primary sclerosing cholangitis.

Pygo1 pygopus family PHD finger 1.

RIN RNA integrity number.

RNA ribonucleic acid.

RNAseq RNA sequencing.

Rnf43 ring finger protein 43.

RSC regenerative stem cell.

Rspo R-spondin.

Rspo2 R-spondin 2.

Rspo3 R-spondin 3.

S-CRC sporadic CRC.

SAM sequence alignment map.

SB small bowel.

SCORT stratification in colorectal cancer.

scRNAseq single cell RNA sequencing.

Sfrp1 secreted frizzled related protein 1.

Sfrp2 secreted frizzled related protein 2.

Smad2 SMAD family member 2.

Smad3 SMAD family member 3.

Smad4 SMAD family member 4.

Smad6 SMAD family member 6.

Smad7 SMAD family member 7.

SNA single nucleotide alteration.

SSL sessile serrated lesion.

SSL-like sessile serrated lesion-like.

Stat3 signal transducer and activator of transcription 3.

Stk3 serine/threonine kinase 3.

Stk4 serine/threonine kinase 4.

TA tubular adenoma.

TA-like tubular adenoma-like.

Taz transcriptional Co-activator with PDZ-binding Motif.

TBE tris-borate-EDTA.

Tcf transcription factor.

TCGA the cancer genome atlas.

TEAD TEA domain transcription factor.

Tgfa transforming growth factor alpha.

Tgfb transforming growth factor beta.

Tgfbr1 transforming growth factor beta receptor I.

Tgfbr2 transforming growth factor beta receptors II.

TP53 tumour protein p53.

Trp53 tumour protein p53.

TSA traditional serrated adenoma.

TSA-like traditional serrated adenoma-like.

TVA tubulovillous adenoma.

TVA-like tubulovillous adenoma-like.

UC ulcerative colitis.

UK United Kingdom.

UV ultraviolet.

VA villous adenoma.

VA-like villous adenoma-like.

Vegf vascular endothelial growth factor.

WNT wingless-related integration site.

Wnt3A Wnt family member 3A.

Wnt5A Wnt family member 5A.

Yap1 yes1 associated protein.

Znrf3 zinc and ring finger 3.

List of Tables

1.1 Main adenocarcinoma histological subtypes 25
1.2 Main precursor lesion histological subtypes 26
1.3 Key features of the CMS subtypes 27
1.4 Key features of the CRIS subtypes 28
1.5 Differences between S-CRC and CA-CRC 37

2.1 Mouse models details 41
2.2 Primer sequences 43
2.3 FACS staining mix 57
2.4 IBD patients recruitment criterion 67
2.5 Choi *et al.* morphological criteria of nonconventional dysplasia 70
2.6 Sonication programme 72

4.1 Fusions detected in the colitis-associated LGD 128

5.1 Clinical and endoscopic details of the cohort in strand A 161
5.2 Clinical and endoscopic details of the cohort in strand B 175
5.3 Univariate cox analysis for relevant clinical and endoscopic factors . . . 178
5.4 Multivariate cox analysis for relevant clinical and endoscopic factors . . 178
5.5 CNA differences between CA-CRC and S-CRC from Baker *et al.* 2019 . . 182
5.6 CNAs occurring at the greatest difference rate between progressors and nonprogressors 183
5.7 Genes within chromosomal regions affected by CNA 184
5.8 Clinical and endoscopic details of cohort in strand C 186

A.1 Details of Mouse models and derived organoids used in the study 213

B.1 RSC gene list 214
B.2 Statistical analyses results of stem cell index across human premalignant and malignant subtypes 221
B.3 Statistical analyses results of stem cell index across murine intestinal neoplasia models 222
B.4 Statistical analyses results of Apc^{Min};$Lgr5^{DTR}$ mice assessment 223
B.5 Statistical analyses results of FOxTROT analysis 223

C.1 Statistical analyses results of investigations carried in murine models of WNT disruption 225

List of Figures

1.1 Intestinal diagram 3
1.2 Stem cell niche diagram 6
1.3 WNT pathway diagram 8
1.4 BMP pathway diagram 10
1.5 Notch pathway diagram 11
1.6 TGFB pathway diagram 12
1.7 The Hippo-YAP/TAZ signalling pathway diagram 14
1.8 Main genetics differences between CA-CRC and S-CRC at different stages of the carcinogenic pathway 36

3.1 A historical timeline of the intestinal stem cell debate 80
3.2 The stem cell dynamics under homeostatic conditions and during regeneration 82
3.3 Molecular and morphological assessment of stem cell markers 88
3.4 Application of stem cell index and markers to human precursor lesions . 90
3.5 Stem cell index and expression in human malignant lesions 91
3.6 Stem cell index and marker expression across murine intestinal neoplasia models 93
3.7 Single cell clonogenicity experiment sorting strategy. 95
3.8 Sorting results and organoid growth of CBC and RSC cells 96
3.9 SNA and CNA alterations associated with CBC and RSC phenotypes . . 98
3.10 Driver pathways associated with CBC and RSC phenotypes 100

3.11 Assessment of *Lgr5*-positive cells loss in Apc^{Min};$Lgr5^{DTR}$ compared to Apc^{Min} mice . . . 102
3.12 Stem cell dynamics assessment in Apc^{Min};$Lgr5^{DTR}$ compared to Apc^{Min} mice . . . 103
3.13 Signalling pathways, proliferation, apoptosis and polyp size . . . 104
3.14 Assessment of the clinical implication of the adaptive response to selective pressure of cancer stem cell in FoxTROT . . . 107
3.15 Fitness landscape model . . . 110

4.1 The WNT signalling pathway in CRC . . . 118
4.2 Molecular pathogenesis of different colorectal precursor lesions . . . 120
4.3 Breakdown of the colitis, sporadic and normal control samples interrogated in this study . . . 123
4.4 *AXIN2* expression in human colitis and sporadic precursor lesions, and normal controls. . . . 124
4.5 Workflow of the multi-omic analysis performed to confirm the findings from the *AXIN2* expression. . . . 125
4.6 Oncoprint of WNT driver genes SNA . . . 126
4.7 *RSPO2* and *RSPO3* expression in bulk colitis-associated LGD RNA sequencing data . . . 127
4.8 *RSPO2* and *RSPO3* ISH in normal controls and colitis-associated LGD with *RSPO* expression in the upper end of the expression continuum . . . 129
4.9 Breakdown of the multifaceted phenotyping of the colitis-associated lesions . . . 131
4.10 Distribution of WNT alterations in colitis-associated LGD subset . . . 132
4.11 *NOTUM* expression in LD and LI human lesions. . . . 133
4.12 *Notum* expression in LI and LD mouse lesions. . . . 134
4.13 Quantification of *Notum* expression in mouse lesions. . . . 135
4.14 Assessment of Paneth cell population across mouse models of WNT disruption . . . 137

4.15 Quantification of Paneth cell population across mouse models of WNT disruption . . . 138
4.16 Assessment of the effects Rspo3 overexpression on Paneth cells in a case-control multiple timepoint experiment . . . 139
4.17 Assessment of the effects Rspo3 overexpression on goblet cells in a case-control multiple timepoint experiment . . . 141
4.18 Assessment of autocrine/paracrine effect Rspo3 overexpression on cell fate determination through quantification of *Rspo3*-positive and -negative Paneth cells . . . 142
4.19 Paneth cells and high *RSPO* expression: from mice to human lesions. . . 144

5.1 Illustration of progressor and nonprogressor patient/lesion groups. . . . 157
5.2 Breakdown of number of patients, lesions and regions in strand A cohort. 159
5.3 Summary diagram of the histological assessment criteria. . . . 163
5.4 Diagnostic agreement in the assessment of sporadic regions . . . 165
5.5 Frequency of dysplasia classes and subclasses in the sporadic cohort. . . 166
5.6 Diagnostic agreement in the assessment of colitis regions . . . 167
5.7 Breakdown of analytical strategy to investigate association of nonconventional dysplasia with progression . . . 169
5.8 Association of nonconventional dysplasia with progression . . . 170
5.9 Breakdown of analytical strategy to investigate association of high-risk histological subtypes of dysplasia with progression . . . 172
5.10 Association of dysplasia with high-risk histologies with progression . . . 173
5.11 Breakdown of the strand B cohort. . . . 175
5.12 Copy number profile of a progressor LGD and a nonprogressor LGD. . . 176
5.13 CNA burden in nonprogressor and progressor LGD of the discovery cohort . . . 177
5.14 Kaplan-Meier analysis of the discovery cohort . . . 179
5.15 CNA burden in nonprogressor and progressor LGD and associated Kaplan-Meier analysis of the validation cohort . . . 180

5.16 Genomewide frequency of CNA in progressor and nonprogressor lesions 181
5.17 Breakdown of number of patients, lesions and regions in strand C cohort. 186
5.18 Outline of the data gathered from digital histopathology analysis 188
5.19 Exploratory analysis 189
5.20 Assessment of histological features on the epithelial compartment. . . . 191
5.21 Assessment of histological features in the most aberrant regions based on CNA burden. 192

6.1 Clinical translation of main findings. 202
6.2 Model outlining the cell dynamics between LI and LD cells and their neighbouring cells 207

C.1 Breakdown of the multifaceted phenotyping of the colitis-associated lesions. 224

Chapter 1

General introduction

1.1 The intestinal tract

The small and large intestines are a long, continuous tube extending from the stomach to the anus. The small intestine is the organ responsible for nutrient breakdown and absorption. Although longer than the large intestine, being 6-7 metres long, the small intestine owes its name to its smaller lumen diameter. The small intestine consists of three main structural and functional segments: the duodenum, jejunum and ileum [1]. The large intestine, also known as large bowel, is the terminal portion of the gastrointestinal (GI) tract and its main functions include absorption of remaining nutrients, electrolytes and water, synthesis of certain vitamins, production and elimination of faeces. Although, only about 1-1.5 metre long, this portion of the GI tract is known as large intestine, due to its large lumen, which is twice the diameter of the small intestine at $\sim$6 centimetre. The large intestine comprises the caecum, colon, rectum and anal canal [2].

The cross-sectional structure of the intestines consist of four layers: mucosa, submucosa, muscularis and serosa (Figure 1.1 A and B). The mucosa is further divided into epithelium, lamina propria and muscularis propria. The epithelial layer lining the intestines is comprised of a monolayer of epithelial cells, which functions as an effective semipermeable barrier. The intercellular junctions between the epithelial cells separate the inner and outer environment, thereby allowing the selective entrance of nu-

trients and electrolytes, while blocking the passage of potentially harmful substances [1]. As a consequence of its function, the intestinal epithelium is constantly exposed to injury from various endogenous irritants and aggressive factors, including ingested elements, such as ethanol and certain drugs, and bacteria. Accordingly, maintenance of the intestinal barrier requires a delicate and dynamic balance between epithelial loss and generation of new cells.

1.1.1 Intestinal architecture

The single layer of columnar epithelial cells lining the small intestine is organised into two main structural units: the crypts and the villi (Figure 1.1 B). The crypts, also known as crypts of Lieberkühn, are the basic morphological and functional units of the gut. They are continuous pit-like invaginations comprising intestinal stem cells, transit-amplifying cell and differentiating cells [3]. Within the crypts, cells maintain a strict spatial arrangement. The slender shaped intestinal stem cells (5-16 cells), also known as crypt base columnar (CBC) cells, are found at the base of the crypts wedged between larger, granule containing Paneth cells, within a discrete microenvironment known as the stem cell niche. They divide each day to rapidly produce immediate stem cell progeny, known as transit-amplifying cells. These are located at position +3 and +5 of the crypt and comprise 120-150 cells. Rapid cell proliferation of these progeny cells allows for the expansion of the progenitor cell populations, thereby enabling rapid epithelial cell turnover. Following proliferation, daughter cells differentiate into various mature epithelial cell types as they move away from the base into the upper part of the crypt and eventually to the villi. The villi are surface-extending projections into the intestinal lumen. They form functional secretory and absorptive units and are comprised of diverse fully differentiated cells capable of fulfilling the digestion and absorption-associated functions of the organ. Two major epithelial lineages of differentiated cells are: the enterocyte or absorptive lineage, and the secretory lineage. The secretory lineage comprises Paneth cells (which function as both niche cells for stem cells and antimicrobial molecules secreting cells), mucus-producing goblet cells, hormone-secreting enteroendocrine cells and tuft cells. Mature cells at the tip of the

villus undergo apoptosis every 4-5 days after birth, likely as a result of being exposed to the hostile luminal environment [4].

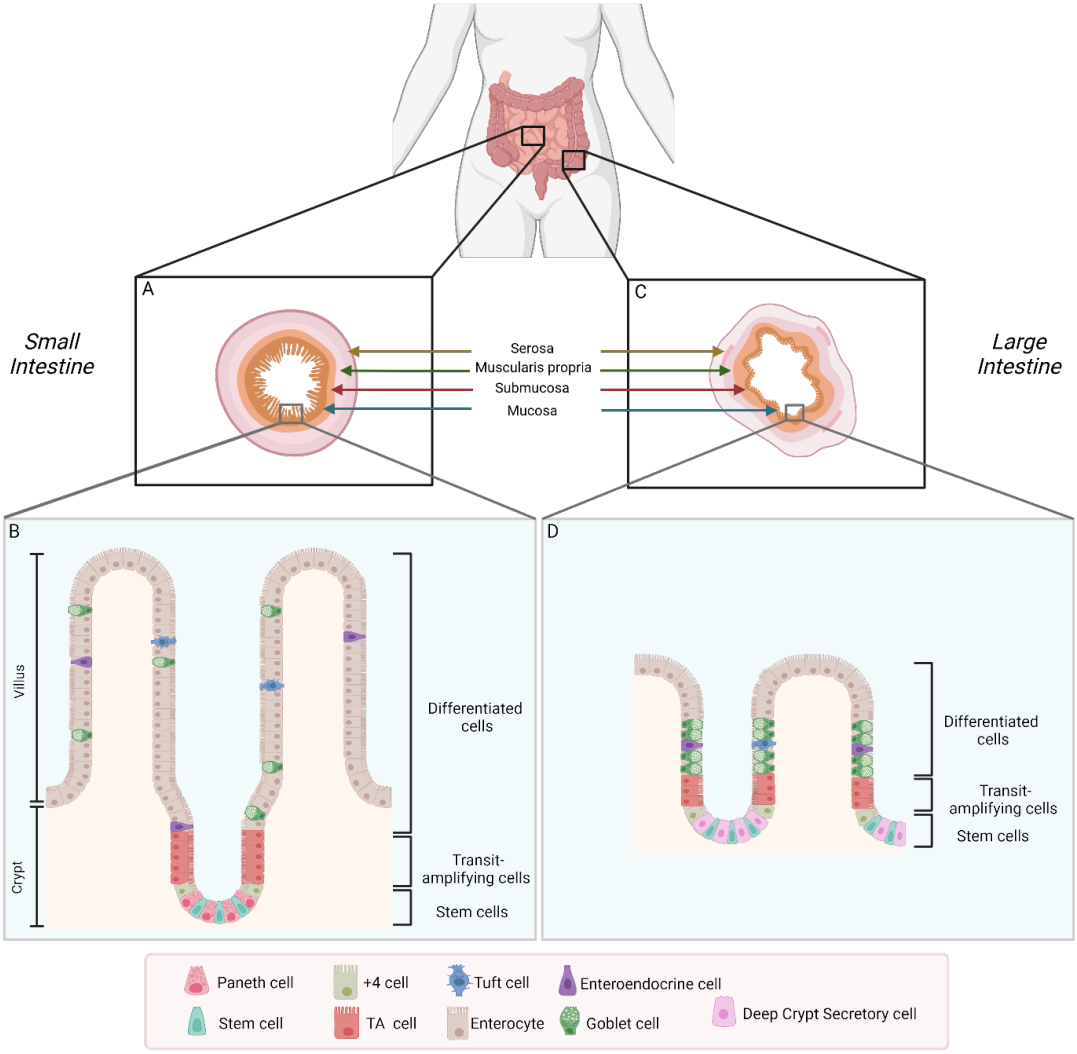


Figure 1.1: Diagram of the intestines: the gross anatomy, cross-sectional structure and the intestinal epithelium.

While very similar, the epithelia of the small and large bowel are different in several structural and organisational features [5]. The colonic lumen is lined by a single layer of tightly linked columnar intestinal epithelial cells. It is characterised by the absence of villi, and the presence of a thin brush border and numerous goblet cells. The crypts of the large bowel, which consist of invaginations of the monolayer of epithelial cells into the surrounding stroma and extracellular matrix, are unbranched and straight. Intestinal stem cells and endocrine cells are found at the base of the crypts. The intestinal

stem cells are responsible for generating colonocytes and supporting epithelial cells, including goblet cells, tuft cells and enteroendocrine cells. Paneth cells are usually not found in the colonic epithelium, although a few may be found under normal circumstances in the caecum and ascending colon [5, 6]. Instead, inter-mingled with intestinal stem cells are c-Kit-positive, deep crypt secretory cells. These are mucous-type cells located at the bottom of the colonic crypt, distinguishable from goblet cells and presumed to be Paneth cell equivalent [7]. The human colon self-renews every 4-7 days, with cells migrating from the bottom of the crypt to the top as they differentiate, to then eventually be lost by apoptosis [5, 6] (Figure 1.1 D).

1.1.2 Intestinal cellular dynamics

The intestinal epithelium is one of the fastest renewing tissues in mammals, a property evolved to be able to selectively absorb nutrients and water, while protecting the body from the toxic contents of the gut's lumen. It is a remarkable example of tissue that relies on adult stem cells to maintain itself and exert its functions. Adult intestinal stem cells are defined by two properties: self renewal, the ability to perpetuate the stem cell pool throughout life by undergoing cell division, and multipotency, the ability to give rise to all mature intestinal cell types [8]. Homeostasis of the intestinal epithelium is supported by proliferation of the intestinal stem cells located at the bottom of the crypt, within the stem cell niche. This allows continuous renewal of the epithelium, as the cells that have been exposed to the toxic luminal contents undergo apoptosis and shed into the lumen. However, during injury the intestinal epithelium demonstrates extraordinary stem cell plasticity. Cellular plasticity is the ability of adult fully differentiated cells to undergo diverse processes to adopt new phenotypic and functional identities in response to perturbations, to thereby facilitate tissue regeneration and restoration of homeostasis. This has been observed to be an important response mechanism against damage [4]. The intestinal stem cell dynamics in colorectal cancer (CRC) is the main focus of chapter 3 of this thesis, thus the topic will be outlined and discussed in further detail there.

1.1.3 The stem cell niche

The stem cell niche is a discrete microenvironment confined to the base of the crypt. It provides an optimal microenvironment for the maintenance of stem cells in their undifferentiated state and the production of differentiated progeny. The intestinal stem cell niche consists of two main components: the physical component, comprising the extracellular matrix, and the cellular component, comprising the stromal and epithelial niches [9] (Figure 1.2). The extracellular matrix is composed of an intricate network of structural proteins, such as fibronectins and collagens, that provide the scaffold required to maintain the three-dimensional structure of the intestine. The stromal niche comprises a collection of cells embedded within the extracellular matrix, including fibroblasts, myofibroblasts, endothelial cells, pericytes, neural cells, smooth muscle cells and immune cells. Making up the epithelial niche are Paneth cells, the nurse cells of the intestinal stem cells, and the intestinal stem cells themselves [10]. Crosstalk between the intestinal stem cells and their niche is regulated by several signalling pathways, including the wingless-related integration site (WNT), Notch, bone morphogenetic protein (BMP), transforming growth factor beta (TGFB) and Hedgehog (HH) pathways. Accordingly, cellular constituents of the niche produce ligands, cytokines and growth factors, to elaborate diverse paracrine signals that regulate stem cell behaviour, dictating the balance between stem cell renewal and differentiation [11, 12]. For example, pericytes and Paneth cells support the stem cell niche by secreting WNT ligands and R-spondin (RSPO), a WNT signalling potentiator [11, 13]. Within the stem cell niche WNT and BMP signalling pathways have antagonistic roles on the intestinal stem cell phenotype. It is the finely balanced intercompartmental cross talk between these pathways that maintains cell fate determination. High WNT/low BMP gradient at the crypt base drives stem cell and transit-amplifying cell proliferation, whereas daughter cell differentiation in the higher levels of the crypt is supported by low WNT/high BMP expression [14] (Figure 1.2).

The intestinal stem cell niche is critical for the maintenance of stem cells and the regenerative potential of the intestinal epithelium [10]. This is evidenced by the fact

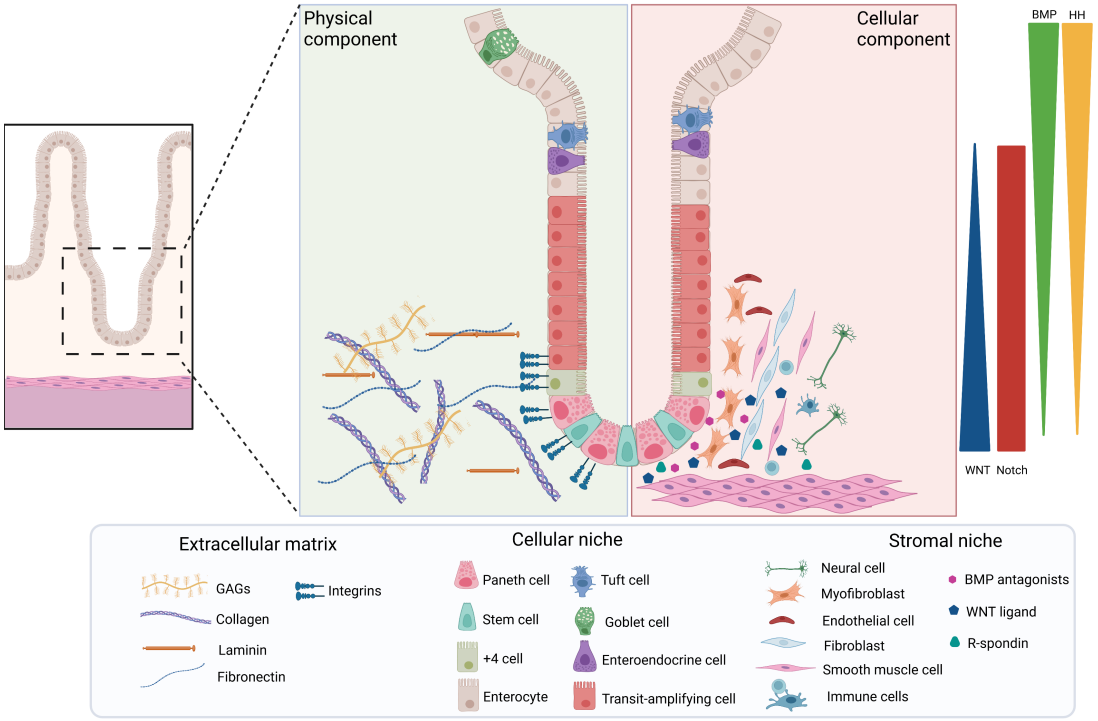


Figure 1.2: Stem cell niche diagram. Adapted from [10].

that genetic deletion of the WNT ligand producing Foxl1-positive fibroblasts leads to severe disruption of the intestinal epithelium [15]. Similarly, R-spondin 3 (*Rspo3*) deletion in PDGFRa-positive cells makes the intestinal tissue hypersensitive to stress [13].

1.2 Signalling pathways in intestinal homeostasis

Cellular processes required for the maintenance of the intestinal epithelium during homeostasis, including cell proliferation, differentiation, angiogenesis and apoptosis, are controlled by key signalling pathways. These include: the WNT, BMP, TGFB, Notch and the Hippo-yes1 associated protein (YAP1)/transcriptional Co-activator with PDZ-binding Motif (TAZ) signalling pathways. Tightly controlled gradients and cross-talk between these signalling pathways are responsible for cell fate determination.

1.2.1 The WNT signalling pathway

The WNT signalling pathway plays a crucial role in many developmental and biological processes, including cell fate determination, proliferation, migration and asym-

metric cell division. In the intestine, the pathway is important in the maintenance and self-renewal of the intestinal crypt stem cell compartment [16]. The WNT signalling pathway, which is generally activated by binding of WNT ligands to their frizzled family core receptors and co-receptors, has highly diverse and context dependent output. Consequently, two main pathways have been described as main transducers of WNT signalling: a) the canonical β-catenin dependent pathway, particularly significant in cell proliferation, and b) the non-canonical, β-catenin independent pathway, believed to be mainly involved in tissue morphogenesis during development and cell mobility [16]. This thesis will focus on the canonical β-catenin dependent pathway.

The WNT ligands are secreted glycoproteins that are heavily modified by glycosylation in the endoplasmic reticulum and palmitoylated by the membrane-bound O-acyltransferase porcupine prior to transport and release in the extracellular matrix. These modifications are necessary for their activity [16, 17]. The key element and signal transducer of canonical WNT signalling is β-catenin. β-catenin associates with deoxyribonucleic acid (DNA)-binding proteins of the lymphoid enhancer factor 1 (LEF1)/transcription factor (TCF) family to regulate the expression of context-specific target genes. Absence of WNT ligands leads to the phosphorylation of β-catenin by a destruction complex, comprising the scaffold protein adenomatous polyposis coli (APC), AXIN, and the serine/threonine kinases glycogen synthase kinase 3 beta (GSK3B) and casein kinase 1 alpha 1 (CSNK1A1). The phosphorylated β-catenin is then ubiquitinated by beta-transducin repeat containing E3 ubiquitin protein ligase pseudogene 1 (BTRCP1) and targeted for proteasomal degradation. In the absence of β-catenin, transcription of target genes is repressed through the formation of a repressive complex, comprising TCF/LEF1, the transducin-like enhancer protein (TLE/Groucho) and histone deacetylases (HDAC). The pathway is activated upon binding of WNT ligands to frizzled (FZD) receptors and LDL receptor related protein (LRP) co-receptors. This triggers a series of events, starting from membrane translocation of AXIN, which binds to a conversed sequence of the cytoplasmic tail of LRP5/6, followed by phosphorylation of LRP receptors by CSNK1A1 and GSK3B and recruitment of Dishevelled proteins to the plasma membrane, where they polymerise and

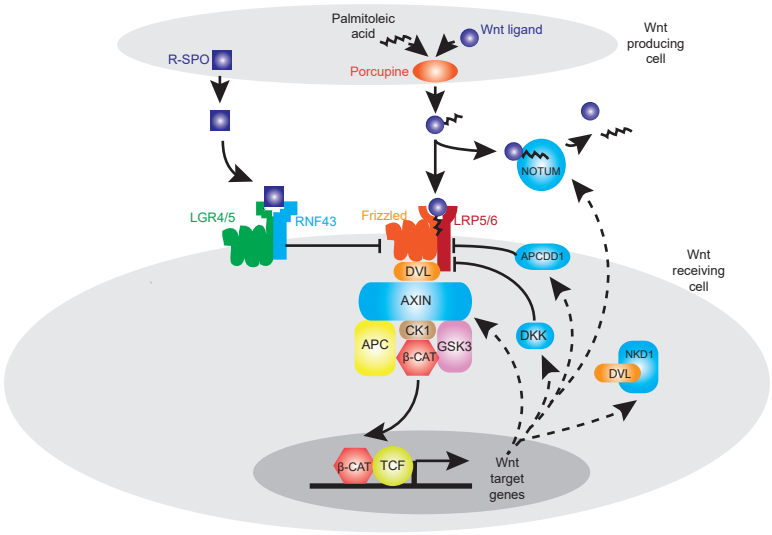


Figure 1.3: Diagram of activated canonical WNT pathway modified from [18].

are activated. They then disrupt the APC/AXIN/GSK3B destruction complex. This results in the cytoplasmic stabilisation and accumulation of β-catenin, which then translocates into the nucleus. Here, it forms a complex with the TCF/LEF1 family transcription factors, causing displacement of TLE/Groucho and recruitment of histone modifying co-activators, including Creb binding protein (CREBBP)/p300, BRG1, Bcl9 transcription coactivator (BCL9) and pygopus family PHD finger 1 (Pygo1). This initiates transcription of its various target genes, including MYC proto-oncogene, bHLH transcription factor (*MYC*), which are known to regulate stemness, cell cycle progression and proliferation [17] (Figure 1.3).

The RSPO signalling pathway has recently been described as key in WNT signalling amplification. In the absence of RSPO ligands, two E3 ubiquitin ligases, ring finger protein 43 (RNF43) and zinc and ring finger 3 (ZNRF3), degrade the cell surface WNT receptors FZD. However, binding of RSPO receptors to leucine-rich repeat-containing G-protein coupled receptors 4 (LGR4), 5 and 6 receptors inhibits the activity of the ligases and amplifies WNT signalling activity by increasing the availability of FZD receptors for WNT ligands [17].

Due to its pivotal role in stem cell control, the WNT signalling pathway is tightly regulated by several negative feedback loops, which control and fine-tune WNT sig-

nalling. These loops are activated upon WNT signalling stimulation and can be subdivided based on their mechanism of action as [18]:

- Secreted antagonists, such as FZD-related proteins;
- Ligand palmitoylated hydrolases, including Notum;
- Receptor inhibitors, e.g. dickkopf WNT signalling pathway inhibitor 1 (DKK1) and apc down-regulated 1 (APCDD1);
- Signal transduction inhibitors, e.g. AXIN1/2.

1.2.1.1 WNT signalling in physiological conditions

The WNT signalling pathway is a key regulator of the intestinal stem cells [16], which are dependent on canonical WNT signalling for survival, as shown by multiple *in vivo* and *ex vivo* studies. *In vivo* WNT inhibition prevents epithelial cell proliferation [19]. Similarly, *ex vivo* studies showed that two potent WNT agonists, Wnt family member 3A (WNT3A) and RSPO1, or an endogenous source of WNT, such as Paneth cells, are essential for intestinal stem cell expansion [20].

WNT activation is restricted to the bottom of the crypt and forms a signalling gradient along the crypt axis. Organoid work has shown that high WNT stimulates stemness and proliferation, while withdrawal of WNT prompts cell differentiation and spontaneous crypt-villus organisation, recapitulating what occurs *in vivo* and highlighting the role of WNT in stem cell organisation in the epithelium [20].

1.2.2 The BMP signalling pathway

BMP is a member of the TGFB superfamily of proteins. BMPs signal through a multimeric cell surface complex comprised of two type I receptors and two type II receptors (BMPR1 and BMPR2, respectively in the Figure 1.4). The BMP pathway is activated when the secreted glycoprotein BMP binds to type II receptors. The type II receptors then phosphorylate the type I receptors, which in turn recruit and phosphorylate the pathway specific R-SMADs (SMAD1/5/8). The activated SMADs can then form

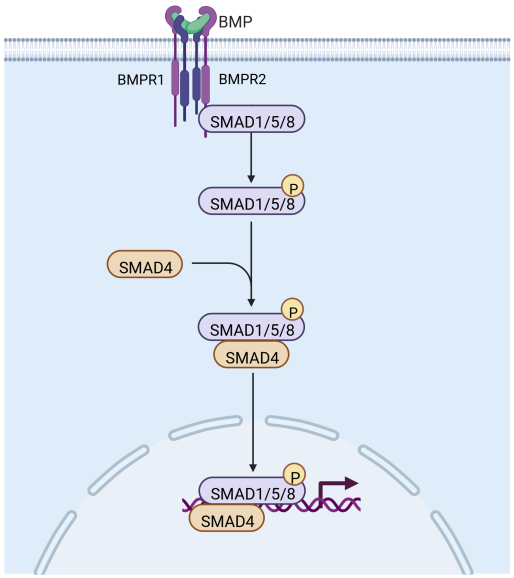


Figure 1.4: Diagram of the BMP pathway in its activated form.

trimers with SMAD family member 4 (SMAD4) and translocate to the nucleus, where they ultimately regulate transcription and expression of target genes, such as inhibitor of DNA binding (ID) [21] (Figure 1.4).

The BMP pathway is controlled by several regulatory mechanisms, including: antagonism by secreted receptors and ligands (e.g. ligand traps, such as Noggin and Gremlin), receptor trafficking by titrating receptors with non-activating ligands (such as BMP3) and negative intracellular feedback by SMAD family member 6 (SMAD6)/7 [21].

1.2.2.1 BMP signalling pathway in physiological conditions

The BMP pathway is involved in the induction of differentiation in the crypt. As previously outlined, the stem cell niche is a BMP-low environment maintained by the secretion of BMP inhibitors. Conversely, in the villi the expression of BMP ligands creates a BMP-high environment that promotes differentiation [4].

1.2.3 The Notch signalling pathway

The Notch signalling pathway is a highly conserved pathway, pivotal in normal cell development, differentiation, proliferation and apoptosis. The pathway relies on cell-

to-cell interaction, where a cell presents a Notch ligand to an adjacent cell expressing a Notch receptor. Following binding of a Notch ligand, the Notch receptor undergoes two proteolytic cleavages. First, ADAM10 metalloprotease catalyses the S2 cleavage, which results in the formation of a substrate for S3 cleavage by the γ-secretase protein complex. Next, γ-secretase protein complex cleavage causes the release of the active form of Notch, also known notch intracellular domain (NCID). Subsequently, NCID translocates to the nucleus, where it binds to the inactive CBF1, suppressor of hairless, lag-1 (CSL) transcription factor forming a complex. The formation of the complex causes displacement of CSL's co-repressors and binding of its co-activators mastermind-like proteins (MAML), p300 and histone acetyltransferases. This in turn leads to activation of the transcription complex and subsequent expression of target genes (Figure 1.5). The Notch signalling pathway has several downstream target genes, including hes family bHLH transcription factor 1 (*HES1*) and Deltex [22].

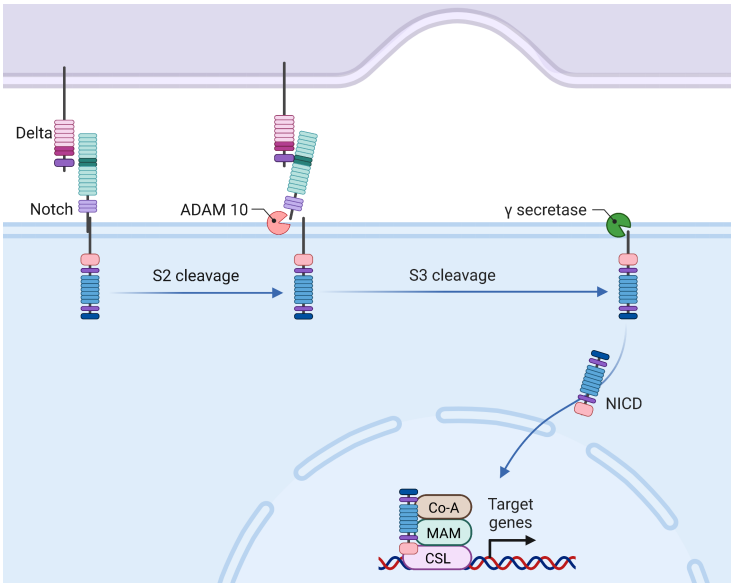


Figure 1.5: Diagram of the activated Notch pathway.

1.2.3.1 Notch signalling pathway in physiological conditions

Notch mediates intestinal development and homeostasis by regulating the differentiation of stem cell progenitors and colonic goblet cells. Specifically, Notch activation results in expression of HES1, which in turn represses atonal bHLH transcription factor

1 (ATOH1) expression. ATOH1 is important in differentiation of secretory cells and its activation promotes differentiation towards secretory fate, while its repression causes an absence of secretory cells. Thus, when inactivated, Notch promotes intestinal cells ATOH1-dependent secretory differentiation, acting as a binary switch through lateral inhibition [4].

1.2.4 The TGFB signalling pathway

The TGFB signalling pathway is activated when TGFB ligands bind to transforming growth factor beta receptors II (TGFBR2). This leads to the recruitment and phosphorylation of transforming growth factor beta receptor I (TGFBR1), which in turn phosphorylates and thereby activates downstream transcription factors: receptor-associated SMAD proteins (R-SMAD proteins) or SMAD family member 2 (SMAD2) and SMAD family member 3 (SMAD3). Activated SMAD proteins can then bind to SMAD4 and form a complex that can translocate into the nucleus where it regulates the transcription of target genes (Figure 1.6). Downstream targets of TGFB signalling include important cell cycle-checkpoint genes, such as cyclin dependent kinase inhibitor 1A (*CDKN1A*), cyclin dependent kinase inhibitor 1B (*CDKN1B*) and cyclin dependent kinase inhibitor 2B (*CDKN2B*), which when activated cause cell growth arrest [23].

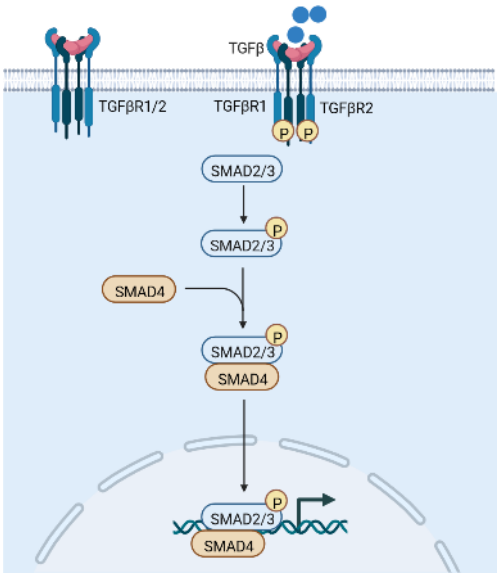


Figure 1.6: Diagram of activated TGFB signalling.

TGFB signalling is tightly regulated by several mechanisms. Negative regulation mechanisms include: regulation of TGFB-ligand access to receptors by ligand trap proteins and the inhibitory functions of several SMAD proteins, including SMAD6 and SMAD family member 7 (SMAD7). Acting as positive regulators are proteins that contribute to the recruitment of R-SMAD proteins to TGFBR1 and proteins that enhance SMAD activations, such as Endoglin and Axin [23].

1.2.4.1 TGFB signalling pathway in physiological conditions

The TGFB signalling pathway is important in the control of cell proliferation, differentiation, migration, apoptosis, adhesion and other biological processes. Under homeostatic conditions, TGFB functions as tumour suppressor by inhibiting cell proliferation and inducing apoptosis and differentiation [23].

1.2.5 The Hippo-YAP/TAZ signalling pathway

The Hippo signalling pathway consists of upstream signals, a kinase core and downstream target genes. It controls cellular proliferation and differentiation by regulating the cellular localization of two proteins: YAP1 and TAZ (formal name WWTR1) [24].

The Hippo pathway is stimulated by several upstream signals, including cell contact, mechanical and hormonal signals, extracellular matrix composition and growth factors via mechanisms that are yet to be fully understood. Upon activation of the pathway, serine/threonine kinase 4 (STK4)/s phosphorylates large tumour suppressor kinase (LATS)1/2. This activates LATS1/2, which interacts with MOB kinase activator 1a (MOB1A)/MOB kinase activator 1b (MOB1B) and phosphorylates YAP1 and TAZ. The phosphorylation prevents YAP1/TAZ from being transported into the nucleus. Moreover, the initial phosphorylation of YAP1 primes it for subsequent phosphorylation in two location. This causes recognition, ubiquitination and degradation of YAP1 by BTRCP1. Inactive Hippo signalling enables translocation of YAP1/TAZ into the nucleus where they form a complex with TEA domain transcription factor (TEAD). TEAD can enhance or repress expression of downstream target genes involved in proliferation, regeneration and reprogramming [25] (Figure 1.7).

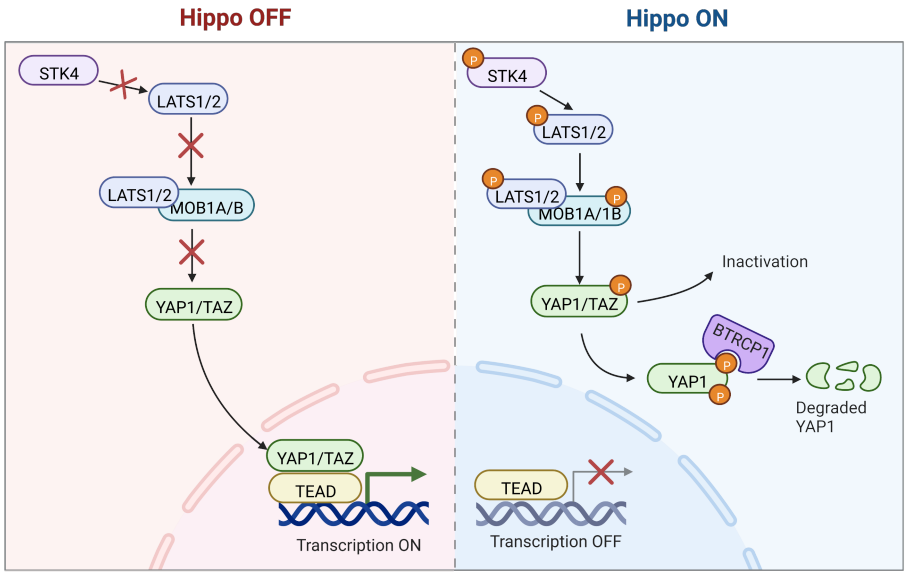


Figure 1.7: Hippo-YAP/TAZ signalling.

1.2.5.1 Hippo-YAP/TAZ signalling pathway in physiological conditions

During homeostasis, YAP1/TAZ play an important role in the proliferation and differentiation of intestinal stem cells. Studies have reported that while moderate levels of YAP1/TAZ activation enhances proliferation and differentiation of intestinal stem cells into mucus-secreting goblet cells, high levels of activation results in increased proliferation and inhibition of differentiation. YAP1/TAZ mechanosensing can serve as an intestinal stem cell switch between quiescence, proliferation and differentiation. To exemplify, fibronectin-based adhesion and high matrix stiffness was reported to significantly promote intestinal stem cell proliferation. Conversely, lamin-based adhesion on a soft matrix was shown to enhance differentiation [26].

1.3 Disruption of homeostasis

The intestinal epithelium functions as protective barrier against the harsh environment of the intestinal tract. Consequently, it is frequently subjected to chemical, microbiological and/or immunological mediated damage. Injury to the mucosa, characterised by epithelial ulceration and barrier breach, causes a skew in homeostatic epithelial stromal cross-talk and localised immune responses. These trigger a controlled tran-

sient dysregulation of the complex interacting pathways that control homeostatic cell fate. The rapid but temporary perturbation of cell-signalling creates a favourable microenvironment that promotes an adaptive cell reprogramming, characterised by stem cell plasticity, epithelial cell expansion, cell proliferation and migration. This allows for quick restoration of the epithelium. Following, epithelial regeneration homeostatic signalling is restored. The same signalling pathways that maintain homeostasis are involved in response to damage and regulate the regenerative capacity of the intestinal epithelium, including WNT, YAP1/TAZ and BMP signalling pathways.

Enhancement of the WNT signalling pathway has been found to stimulate regeneration following damage. Zhao *et al.* reported that administration of Rspo1 improved recovery in mice who had been subjected to chemotherapy [27]. Furthermore, several studies have shown that lack of WNT impedes regeneration, including models of *Myc* deletion, lack of Wnt family member 5A (WNT5A) [28] and lack of Rspo3 [29]. WNT ligands and Rspo3 have been associated with promotion of stem cell plasticity. Harnack *et al.* showed that mucosal injury causes an upregulation in stromal Rspo3, which induces WNT signalling in differentiated cells. This was shown to cause differentiated cells to re-acquire stem cell-like function and aid in crypt regeneration [29].

Recently, YAP1/TAZ signalling was reported as a key promoter of cellular plasticity following mucosal damage. Yui *et al.* showed that damage-induced changes in signalling between the extracellular matrix and epithelial cells causes the activation of YAP1/TAZ signalling. This in turn promotes reprogramming of the intestinal epithelium to a primitive highly proliferative state, characterised by novel expression of foetal markers and suppression of adult stem and differentiated cells, which enables rapid regeneration. This foetal-like state persists only until return to the normal homeostatic microenvironment [30].

Conversely, to the WNT and YAP1/TAZ signalling pathways, BMP, which as outlined above mediates cell differentiation in homeostasis, is inhibited during regeneration. This was shown by our group to be mediated by a rapid but transient upregulation of secreted BMP antagonist Gremlin 1 [31].

Failure to restore homeostatic controls results in chronic inflammation. Chronic or

recurrent inflammation can cause a persistence or hijacking of this adaptable cell reprogramming response, thereby increasing the risk for neoplasia. Inflammatory bowel disease (IBD) represents a perfect example of how chronic or recurrent inflammation can predispose to CRC.

1.3.1 Inflammatory Bowel Disease

IBD is the umbrella term used to describe chronic relapsing and remitting inflammatory disorders of the bowel. Two main forms of IBD are ulcerative colitis (UC) and Crohn's disease (CD). These idiopathic conditions are postulated to be the result of a complex interaction between environment, genetics and altered microbiota [32]. They both present with diarrhoea, abdominal pain, fatigue, weight loss and haematochezia [33].

CD is characterised by inflammation that can affect the entire GI tract, with lesions occurring most commonly in the small intestine and proximal colon. The inflammation presents itself as transmural, affecting multiple layers of the intestine, typically in a non-continuous, scattered manner. CD is commonly associated with complications, such as abscesses, fistulas and strictures [33].

UC is characterised by continuous, but localised inflammation. Inflammation usually affects the mucosa and submucosa (and sometimes, in cases of toxic megacolon, the muscularis propria). It usually arises in the rectum and extends proximally in a continuous fashion. The small bowel is rarely affected, except for backlash ileitis.

1.3.1.1 Epidemiology

IBD is a chronic, incurable disease with low mortality, but potentially high morbidity. The condition has emerged as a global health issue in the 21st century due to a substantial worldwide increase in incidence since the 1990s. Between 1990 and 2017, an 85% rise in global IBD incidence was registered, with an increase in individuals with IBD from 3.7 million to 6.8 million [34].

IBD incidence is highest in Western countries, particularly in North America and Europe, where 1 to 5 million and 2 million people are affected by the condition, respec-

tively [35]. In the United Kingdom (UK), overall IBD, CD and UC incidences were found to be 28.6, 10.2 and 15.7 per 100,000 persons, respectively. Although disease incidence seems to be stabilising in Western countries, it still remains high [36]. On the contrary, IBD was relatively rare in developing nations until recently [37]. Over the past decades, a rise in incidence has been observed in these regions, particularly in newly industrialised countries in Africa, Asia, South America and the Middle East. Although the incidence of IBD in newly developing countries is still considerably lower than that in Western countries, the rate of the rise is much higher. To exemplify, since 1990, Brazil has seen an annual percentage change of +11.1% for CD and +14.9% for UC [38]. The stabilisation of IBD incidence in high income countries and the rapid increase in prevalence in newly industrialised countries has turned IBD from being regarded as a disease of the Western world to a global phenomenon [38, 39].

Generally, IBD prevalence is similar in males and females [39]. Traditionally the age distribution in IBD was reported as following a bimodal distribution, with the highest peak age incidence between the ages of 15-25, and a second peak between the ages of 50-70 [40]. However, recent studies suggest regional differences in age distribution, with the bimodal distribution being typical of American cohorts, while European and Canadian cohorts showing peak incidences in the 15-29 and 20-29 age groups for CD and UC, respectively [41].

1.3.1.2 Aetiology

The aetiology of IBD still remains to be fully elucidated. Nevertheless, the condition has been speculated to be multifactorial and the result of an exaggerated immune response to altered gut microbiota in genetically predisposed individuals, in the presence of environmental triggers. A number of environmental risk factors have been identified. These include: cigarette smoking, use of non-steroid anti-inflammatory drugs and psychological stress [42].

1.3.1.3 Pathophysiology

IBD's remitting and relapsing nature is characterised by phases of injury and repair of the intestinal epithelium. Immune activation causes mucosal inflammation, characterised by marked infiltration of immune cells in the lamina propria, injury and ulceration. This is followed by regeneration and restoration [43], achieved via cellular plasticity and subsequent stem cell proliferative zone expansion. A common finding in IBD mucosa believed to be associated with expansion of the stem cell proliferative zone is increased crypt fission [44].

Current IBD therapies are mainly focused in attenuating the immune system to reduce mucosal inflammation. Unsuccessfully or inadequately treated chronic inflammation is associated with impaired mucosal healing. This can contribute to chronicity of the inflammation and mechanisms involved in mucosal healing, including cellular plasticity. This in turn is associated with a higher relative risk of CRC.

1.4 Inflammation, Wound healing and Cancer

The link between inflammation, wound healing and cancer has long been acknowledged [45]. Wound healing involves the transient activation of programmes involved in: immune and stromal cell activation, adaptive cell reprogramming, stem cell function, cellular migration, proliferation and tissue matrix remodelling. These programs, required for a rapid and effective response to wounding, are found to be highly activated in cancer cells. Likewise, the interplay of different cell types, including epithelial, mesenchymal and immune cells, is pivotal in both wound healing and tumorigenesis. Consequently, it is thought that the same signalling pathways and programmes that are transiently upregulated for wound healing are hijacked and constitutively activated in tumour formation and maintenance. These observations led Dvorak to define tumours as "wounds that do not heal" in 1986 [46]. The fact that pathways activated in response to tissue injury are co-opted and corrupted in cancer suggest an evolutionary trade-off between a robust regenerative response to wounding and cancer risk.

1.5 Colorectal Cancer

1.5.1 Epidemiology

CRC is the third most common cancer worldwide and the second leading cause of cancer death [47]. Marked geographic differences in CRC incidence and mortality rate have been reported. While Western countries have the highest CRC risk, newly industrialised countries adopting the Western lifestyle have the most significant increase in CRC incidence and mortality [47].

CRC is the second most common cancer in females and third most common in males globally, with over 1.9 million new cases in 2020 [47]. Australia, New Zealand, Europe and North America have the highest CRC rates, while the lowest rates are observed in Africa and South-Central Asia [48]. In the UK, CRC represents the 4th most common form of cancer. With 42,900 new cases every year, CRC accounts for 11% of all new cancer cases [49].

CRC is estimated to be responsible for 916,000 cancer deaths in 2020 [47]. In the UK, CRC is responsible for 16,600 deaths every year, i.e. 10% of all cancer death, leading it to be the second most common cause of cancer death [49].

CRC most commonly affects males (56% of cases) than females (44% of cases). Generally, both incidence and mortality are higher in white ethnicity compared to other ethnicities [49].

1.5.2 Aetiology

CRC is a multifactorial disease resulting from the interaction of multiple lifestyle, genetic and environmental factors. Although CRC's aetiology is complex, multiple risk factors have been identified, including:

- History of cancer or colon polyps;
- Genetics: a number of hereditary conditions such as familial adenomatous polyposis, Lynch syndrome and Juvenile polyps are associated with increased risk for CRC;

- Medical history of chronic GI inflammation, including IBD;
- Environmental factors, such as smoking, obesity, excessive alcohol consumption and a meat rich diet.

Additionally, CRC risk is known to be influenced by age, gender and race, and gut microbiome [50, 51].

1.5.3 Signalling pathways in intestinal carcinogenesis

CRC is driven by genetic and epigenetic alterations that cause cells to transform from normal to malignant. Many of these alterations map to key signalling pathways involved in development and homeostasis. This leads to the permanent dysregulation or co-option of signalling pathways involved in cell proliferation, differentiation, angiogenesis and apoptosis. Among the commonly disrupted pathways in CRC are the WNT, BMP, TGFB and Hippo-YAP1/TAZ signalling pathways.

Aberrant WNT activation is a well-recognised hallmark of CRC, with the vast majority of sporadic CRC (S-CRC) bearing WNT-activating mutations [16]. The role of the WNT signalling pathway in colitis-associated CRC (CA-CRC) is the focus of chapter 4 of this thesis, thus its role in CRC will be further outlined and discussed in that chapter.

The BMP pathway is down-regulated in the majority of CRCs. Loss of BMP expression through inactivating mutations in the genes involved in BMP signalling is observed in patients with hereditary mixed polyposis syndrome. Notably, inactivation of the BMP receptor bone morphogenetic protein type I receptor (Bmpr1) is found to elevate WNT signalling, causing expansion of the stem cell compartment [4, 21].

Aberrant Notch activation has been implicated in the initiation of CRC. It is thought to promote CRC through induction of cancer cell proliferation, survival, angiogenesis and epithelial-mesenchymal transition (EMT) [52]. Furthermore, an overexpression of Notch ligands, receptors and downstream targets has previously been reported in CRC cells [53].

Many CRCs lose TGFB expression in the initiation and further stages of carcino-

genesis. These cancers are resistant to TGFB-induced growth inhibition. However, TGFB seems to act as a promoter in late stages of CRC, where its high expression is linked with increased production of transforming growth factor alpha (TGFA), epidermal growth factor (EGF) and other mitogenic growth factors [54].

Dysregulated Hippo-YAP1/TAZ pathway has been implicated in CRC and many other cancer types. In CRC, Hippo signalling is mostly suppressed and YAP1/TAZ highly expressed. This is thought to promote migration, invasion, proliferation and EMT of colon cancer cells. In support of this, Camargo *et al.* found that Yap1 overexpression in mice promotes intestinal hyperplasia, characterised by absence of Paneth and goblet cells, and causes replacement of differentiated cells by novel Yap1-expressing progenitors [55]. Similarly, mice with conditional knockouts of the Hippo pathway components, *Stk4* and serine/threonine kinase 3 (*Stk3*), were reported to develop dysplastic lesions in the small and large intestines as well as colonic adenomas [56]. These studies support the postulation that Hippo is involved in tumorigenesis, with aberrant activation of YAP1/TAZ being oncogenic. In humans, increased YAP1 expression was associated with poor CRC prognosis [57], similarly to high TAZ expression, which was correlated with shorter survival [58]. Furthermore, *YAP1* and *TAZ* knockdowns in human colon cancer cell lines and xenograft mouse models resulted in decreased proliferation, in support of them being oncogenes of CRC [59].

1.5.4 Pathogenesis

The pathogenesis of CRC consists of a complex multistep process, comprising a sequential acquisition of genetic and epigenetic alterations that drive the initiation and progression of adenomas to carcinomas. Classically, CRC tumorigenesis begins with the transformation of normal epithelium to a benign adenoma, which then progresses to invasive carcinoma and eventually metastatic cancer. The acquisition or accruing of specific genetic and epigenetic alterations is responsible for a particular event within the sequence leading to CRC. The multiple stages pattern of CRC carcinogenesis was first theorised by Cuthbert Dukes, who described colorectal adenomas as potential precursors of CRC in the early part of the 20th century, and further elaborated by Basil

Morson, later in the 20th century. Subsequently, in the 1990s Fearon and Vogelstein proposed the Vogelgram, which assigned the stepwise progression to the accumulation of mutations in genetic drivers. Since then, the proposed multistep genetic model of colorectal carcinogenesis has been empirically confirmed through genetic studies.

Three main CRC carcinogenic pathways have been described: the chromosomal instability (CIN), the microsatellite instability (MSI) and the CpG island methylator phenotype (CIMP) pathways [60–63].

The CIN pathway, also known as the traditional pathway, is the most common pathway, occurring in 50-70% of S-CRC. It follows the original adenoma-carcinoma sequence outlined and is characterised by the stepwise accumulation in key tumour suppressor genes and oncogenes, starting with the inactivation of APC, followed by alteration in KRAS proto-oncogene (*KRAS*), *SMAD2/4* and loss of tumour protein p53 (*TP53*). Loss of APC is believed to trigger the cascade of events that eventually leads to the development of cancer [60–63].

The MSI pathway, also known as the alternative pathway, is followed by 10-30% of CRCs. This pathway is characterised by genetic hypermutability resulting from an impaired DNA mismatch repair (MMR) system. This occurs due to inactivation of the genes involved DNA MMR through point mutation or due to aberrant methylation of CpG of MutL homolog 1 (*MLH1*). Consequently, due to the lack of a functional MMR system, MSI cancers can more readily acquire mutations in cancer-associated genes, especially those located within microsatellite regions. Similarly to the CIN cancers, one of the earliest events in the MSI pathway is loss of APC [60–63].

The CIMP pathway was only recently outlined. Cancers progressing down the CIMP pathway are characterised by the presence of B-Raf proto-oncogene, serine/threonine kinase (*BRAF*) activating mutation and epigenetic silencing of genes involved in cell differentiation, DNA repair and cell cycle control. A key difference between this pathway and the other two is the absence of APC loss as an initiating event [60–63].

Conversely to the above classification of the molecular carcinogenic pathway of CRC, more recently single cell transcriptomic studies of precancerous [64] and cancerous [65] CRC lesions have suggested CRC carcinogenesis to follow two main path-

ways: a conventional pathway, similar to the CIN pathway, and a serrated pathway, comprising MSI and microsatellite stable (MSS) tumours. Specifically, single cell transcriptomic analysis of CRC cells led to the identification of two main epithelial cell type: iCMS2 and iCMS3, each characterised by distinct copy number genetics, transcriptional profiles and gene regulatory units, and giving rise to CRCs, with distinct biology and developmental origins and trajectories. iCMS2 cancers were MSS and characterised by *APC* and *TP53* mutations, copy number changes and high WNT activity, while the iCMS3 group comprising both MSS and MSI cancers, was characterised by *KRAS*, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (*PIK3CA*) and *BRAF* mutation, few copy number changes, high mitogen-activated protein kinase (MAPK) activity and upregulation of inflammatory pathways. While, iCMS2 cancers transcriptomically related to adenomas, iCMS3 related to serrated lesions [65]. Interestingly, a multi-omic human pre-cancer atlas, integrating single cell transcriptomics, genomics and immunohistopathology also identified two main cell types, one attributable to adenomas, showing similar properties to the iCMS2 cells, and the other to serrated lesions, similar to the iCMS3 cells [64]. Both studies suggested a conventional pathway of carcinogenesis, characterised by initial expansion from leucine-rich repeat-containing G-protein coupled receptors 5 (LGR5)-positive cells, early loss of function of APC providing promitent WNT pathway activation, and the acquisition of copy number changes and aneuploidy, accelerated by *TP53* mutation. This pathway was associated to iCMS2 cancers and conventional adenomas. While the serrated pathway, attributed to iCMS3 cancers and serrated precancerous lesions, was suggested to be characterised by MAPK activation and upregulation of inflammatory pathways, with initial expansion occuring through dedifferentiation of progenitor cells.

1.5.5 Colorectal cancer classification

CRC comprises a group of heterogeneous conditions, arising from different genomic, epigenetic, environmental and physiological factors. Consequently, multiple classification systems for CRC have been described.

1.5.5.1 Aetiology based classification

Based on aetiology of disease, CRC can be broadly divided into sporadic, familial, hereditary and colitis-associated. Sporadic CRC (S-CRC) occurs in the absence of contribution from germline CRC causing mutations, significant family history of cancer or IBD [66]. It accounts for 70% of CRC cases [67].

Hereditary CRC (H-CRC) accounts for 2-8% of CRC cases. They are CRCs that arise in the setting of Mendelian disorders, such as Lynch syndrome, familial adenomatous polyposis, MUTYH-associated polyposis, juvenile polyposis, hereditary mixed polyposis and Peutz-Jeghers syndrome [68].

In familial CRC (F-CRC), within affected families CRC occurs too frequently to be considered sporadic, yet not in a pattern consistent with inherited syndromes. Approximately 25% of all CRC cases are considered to be familial [67].

Colitis-associated CRC (CA-CRC) is a form of CRC that arises from longstanding colitis in patients with IBD. It constitutes 2% of CRC cases [69].

1.5.5.2 Histological classification

The vast majority of CRC, more than 90%, are adenocarcinomas, while the remaining minority comprises other types of carcinoma (squamous cell, adenosquamous, undifferentiated and carcinoma with sarcomatoid component), neuroendocrine tumours, mesenchymal tumours (gastrointestinal stromal tumour and leiomyosarcoma) and lymphoma.

Adenocarcinomas are defined as tumours originated from epithelial cells of the colorectal mucosa, characterised by gland formation. The grading of adenocarcinoma is based on the percentage of glandular differentiation: grade 1 well differentiated (>95% gland formation), grade 2 moderately differentiated (50-95% gland formation) and grade 3 poorly differentiated (<50% gland formation); with grade 3 being prognostically the worst grade. Histological subtypes of colorectal carcinoma are not as important as the pathological grade and stage. Consequently, the majority of adenocarcinomas are diagnosed as ”adenocarcinoma not otherwise specified”. Never-

theless, histological subtypes with specific clinical and molecular characteristics have been differentiated and listed in the WHO 2019 classification system [70, 71]. The main histopathological subtypes are reported in Table 1.1.

Table 1.1: Main adenocarcinoma histological subtypes. Adapted from: [70].

Histopathological type	Description
Mucinous adenocarcinoma	At least 50% of neoplasm composed of extracellular mucin lakes with presence of malignant epithelial structures.
Signet ring cell carcinoma	More than 50% of neoplasm with intracytoplasmic mucins, which often displaces the nucleus toward the periphery.
Medullary carcinoma	Neoplasm with sheets of large cells with vesicular nuclei, prominent nucleoli and abundant eosinophilic cytoplasm.
Serrated adenocarcinoma	Neoplasm with glandular serration, often with areas of mucinous component.
Micropapillary adenocarcinoma	Neoplasm with small aggregates of malignant cells in relation to lacunar spaces simulating vascular channels.

More important is the histological classification of precursor lesions of CRC. This is due to the association of histological subtypes with different carcinogenic pathways and molecular features. Most CRCs arise from adenomas and/or dysplasia. The main histological precursor lesion subtypes of S-CRC are conventional adenomas, comprising tubular adenoma (TA), tubulovillous adenoma (TVA) and villous adenoma (VA), and serrated lesions, comprising sessile serrated lesion (SSL) and traditional serrated adenoma (TSA) (features in Table 1.2). Conventional adenomas are the more common type of precursor lesions and follow the conventional 'adenoma-carcinoma' carcinogenic pathway; while on the other hand, malignant transformation of serrated lesions progresses via the MSI or CIMP pathways. Dysplasia in precursor lesions of the S-CRC is stratified into low grade dysplasia (LGD) and high grade dysplasia (HGD) based on the degree of cytological and/or architectural atypia. LGD is defined by dysplastic epithelium with polarized, pseudostratified, elongated and hyperchromatic nuclei in simple crypts. HGD shows more severe atypia and is defined by high nuclear:cytoplasmic ratio, nuclear pleomorphism, loss of nuclear polarity and intraluminal cribriforming in at least two or more glands [72].

Table 1.2: Main precursor lesion histological subtypes. Adapted from [72].

Subtype	Description
TA	Largely conserved normal crypt architecture, with variable elongation of the crypts and increase in the number of glands. Enlarged, hyperchromatic nuclei, with nuclear atypia and stratification, with loss of nuclear polarity.
TVA	>25% of the architecture is composed of structures resembling small intestinal villi, with cellular atypia similar to the tubular ones.
VA	>75% of adenoma with villous architecture.
SSL	Crypts with prominent serration; bland cytology with variable amount of goblet cells and cells with microvesicular mucin droplets.
TSA	Villous polyps with tall cells that contain prominent eosinophilic cytoplasm and pencillate nuclei.

1.5.5.3 Molecular classification

The development of CRC is marked by the occurrence of multiple genetic events. Consequently, several classification systems based on gene expression patterns have been described to stratify CRC into molecularly and clinically distinct groups. Two main ones are the consensus molecular subtypes (CMS) classification system and the colorectal cancer intrinsic subtypes (CRIS) classification system.

The CMS classification system is one of the most robust gene expression profile based CRC classifications [73]. Developed in 2015 by a large international consortium, the CMS classification system was compiled by analysing 4151 CRC samples from 18 CRC datasets. It is based on comprehensive gene expression profiles and classifies CRC into 4 molecularly and biologically distinct subtypes [74]:

- CMS1 (microsatellite instability immune): microsatellite unstable, hyper-mutated, immune-infiltrated and enriched for BRAFV600E tumours;
- CMS2 (canonical): epithelial and immune deprived, MSS, CIN tumours, driven by WNT and MYC activation, with marked epidermal growth factor receptor (EGFR) signalling independence;
- CMS3 (metabolic): epithelial tumours characterised by metabolic deregulation and MAPK pathway alterations, they can be either MSI or CIN, immune desert

or enriched;

- CMS4 (mesenchymal): CIN and MSS tumours, characterised by stromal infiltration and activation of the TGFB and vascular endothelial growth factor (VEGF) pathways.

Additional features of the subtypes are described in Table 1.3. Notably, in addition to the four CMS subtypes, a fifth group comprising lesions with mixed features (13%) was also described. These lesions have been proposed to represent a transition phenotype or intratumoral heterogeneity [74].

Table 1.3: Key features of the CMS subtypes. Adapted from [75].

CMS Class	CMS1	CMS2	CMS3	CMS4
Prevalence	14%	37%	13%	23%
Molecular features	MSI-High CIMP-High BRAFV600E	WNT-High MYC-High SRC-High	Metabolic dysregulation	EMT-High TGFB-High Matrix remodelling
Stroma	Immune Cells			Desmoplasia Fibroblasts
Immune phenotype	Immune infiltration	Immunological blunt	Immunological blunt	Inflamed infiltration
Outcome	Worse survival after relapse			Worse relapse-free and overall survival
Therapy Response	5-FU NR, Immunotherapy R	FOLFOX R Cetuximab R	FOLFOX R	FOLFOX NR FOLFIRI R Cetuximab NR

Key: NR = No Response, R= Response.

CMS2 and CMS3 are associated with favourable prognosis. Conversely, CMS4 is associated with the worst prognosis, showing worse relapse-free and overall survival and CMS1 is associated with worse survival after relapse. The associations of the different CMS subtypes with specific clinical outcomes have been validated by retrospective analyses of clinical trials [76].

The CRIS classification is a cancer cell-intrinsic classification system, developed in 2017 through the transcriptional profiling of 515 xenografts derived from 244 patients. This classification system focuses on the gene expression profiles originating only from

the epithelial cancer cells, removing the "molecular noise" contributed by stromal cells. Based on the CRIS classification CRCs can be discerned into 5 subtypes [77]:

- CRIS-A: mucinous and glycolytic tumours, enriched for either *KRAS* mutations or MSI;
- CRIS-B: tumours with TGFB pathway activation and characterised by EMT;
- CRIS-C: tumours with EGFR pathway activation and thereby sensitive to EGFR inhibitors;
- CRIS-D: tumours characterised by WNT-activation, insulin like growth factor 2 (*IGF2*) gene overexpression and amplification;
- CRIS-E: tumours harbouring *TP53* mutations and with Paneth cell-like phenotype.

Features of each subtype are outlined in the table below.

Table 1.4: Key features of the CRIS subtypes. Adapted from [75].

CRIS Class	CRIS-A	CRIS-B	CRIS-C	CRIS-D	CRIS-E
Molecular features	MSI-High CIMP-High BRAFV600E	No enriched features	CIN-High TP53mut MYC amplified	CIN-High IGF2 amplified	CIN-High KRAS-mut TP53mut
Pathway	Glycolysis Inflammation MSI Mucinous	EMT-High TGFB	EGFR-High ERBB3-High LGR5-positive like	IGFR-High FGFR1-high WNT-High	WNT-High Paneth cell-like
Outcome	+	-	+	++	+
Therapy Response	Cetuximab NR		Cetuximab R	Cetuximab NR	Cetuximab NR

Key: NR= No Response, R= Response, KRAS-mut= *KRAS* mutation, TP53mut= *TP53* mutation.

1.5.6 Disease models of colorectal cancer

The use of disease models in CRC research has aided in better understanding the pathogenesis of the disease and in the development of potential treatment strategies. Two main and commonly used disease models are animal models and organoids.

1.5.6.1 Mouse models

Mice are good animal models of carcinogenesis because they are genetically and physiologically close to humans, with 85% of the genomic sequences between mice and humans being still conserved. Moreover, extensive genetic information is available on individual mouse strains [78, 79]. An ideal model of CRC should replicate all features and stages of CRC evolution as they occur in humans. Nevertheless, although several mice models approximate some features of human CRC, none meet all these characteristics. This is due to both intrinsic differences between mice and human anatomy and physiology, and the heterogeneous nature of CRC.

The intestine of mice and humans are similar in terms of development, structure and functions, with these two species sharing major biological and histological intestinal features. However, a few differences have recently been highlighted. Organoid studies have shown a difference in niche factor dependence between mice and human organoids, which could be indicative of a difference in molecular mechanisms of self-renewal. Moreover, intestinal epithelial cells unique to the human epithelium and not found in the mice, including motilin-positive enteroendocrine cells and BEST4-positive/OTOP2-positive cells, have recently been described. Also, noteworthy are the environmental differences between humans and mice, such as differences in microbiota and diet, which translate into differences in environmental signals and their effect on the intestinal epithelium. In mice the majority of lesions occur in the small intestine, conversely to humans where the majority of the lesions are colonic. Moreover, the mouse small intestines seem to be more sensitive to WNT activation than the human small intestine [80].

Several mouse models have been developed overtime to investigate the development and characteristics of different types of CRCs [79]. Notably, four main types of mouse models are: spontaneous, chemically-induced, genetically-engineered and inoculated. Spontaneous and chemically-induced models are the simplest models of CRC, they replicate human S-CRC and are often employed in the investigation of the influence of diet and the microbiota on carcinogenesis. Inoculated models mimic some

features of CRC metastasis and are used to better characterise this stage of CRC and for the evaluation of anti-metastatic drugs. Genetically-engineered models are important tools in studying the contribution of genetic alterations in the development and progression of CRC and their sensitivity to treatment [78, 81, 82]. All models used in this thesis are genetically engineered.

1.5.6.2 Organoids

Organoids are self-renewing, self-organising multicellular 3-dimension *in vitro* tissue constructs exhibiting organ-specific features highly similar to that of their organ of origin *in vivo*. They are used to recreate the architecture, physiology and pathology of human organs with fine detail and thereby aid in the study of aspects of an organ and its associated disease *in vitro* [83]. Intestinal organoids were the first organoids to be established. They are also the most mature organoids produced and studied. The first organoid was established by Sato *et al.* from mouse intestinal tissue [84]. Two years later the same group established a method for successful and long-term culture of organoids from primary human colonic epithelial cells [20]. Currently, most CRC studies adopt Sato's protocols or slightly altered versions of them [85]. The use of organoids in CRC research comes with several advantages. The main one is that organoids are comprised of multiple cell types that more closely resemble the mileu of cells *in vivo*, with conditions, tissue structure and cell-to-cell interactions displaying many characteristics of the organ or tissue *in vivo* [85, 86].

1.6 Colitis-associated colorectal cancer

CA-CRC is the most feared complication in patients with IBD [87]. The first large meta-analysis estimated UC cancer risk to be 2% after 10 years, 8% after 20 years and 18% after 30 years of disease duration [88]. Similarly, the overall cancer risk in CD patients was reported as being 2.3% for 10 years, 5.6% for 20 years and 8.3% after 30 years of disease duration [89]. Since the first large meta-analysis, discrepant studies have reported a declined, continued high or increase in CA-CRC cancer risk [90]. A

meta-analysis from 2013, showed a reduction in CA-CRC incidence rate in UC patients from 4.96 per 1000 in studies published in the 1950s to 1.21 per 1000 patient-years in studies published between 2010 and 2013 [91]. Likewise, in the same year, Lutgens *et al.* [92] reported cumulative CRC risk in IBD patients to be 1%, 2% and 5% after 10, 20, >30 years, respectively. Conversely, other studies, including a study from St Mark's Hospital in London, which has the longest running surveillance programme in the world, and another from Mumbai, reported a continued high cancer risk [93, 94]. On the other end, a Japanese study found an increase in CA-CRC cases between 1995-1998 compared to 1978-1994 [95]. The discrepancy between studies in quantifying CRC risk in IBD patients is likely to be reflective of geographic differences, differences in study population and disease duration, but also advances in surveillance and treatment, bias arising from selection, surveillance and lead time [93]. Despite uncertainty on the exact CRC risk in IBD patients, several risk factors have been consistently associated with CA-CRC. These include younger age at diagnosis, and extent, severity and duration of inflammation. Highest risk is observed in IBD patients with longstanding, extensive and severe inflammation [96].

Current strategies to prevent CA-CRC development include: chemoprevention and early detection of dysplasia via surveillance colonoscopy. Endoscopic surveillance with concomitant histological testing of biopsied mucosal tissue is the current gold-standard for IBD patients' cancer risk management. However, this strategy is burdensome and expensive. Further discussion of this can be found in chapter 5 of this thesis. Increasing data on the role played by inflammation on CA-CRC carcinogenesis have led to the proposal of anti-inflammatory drugs as a chemoprevention strategy. This aims at controlling mucosal inflammation and promoting mucosal healing. However, the efficacy of currently available drugs for cancer prevention in IBD patients is yet to be determined, and most of the evidence available is weak and comes from observational studies [97].

Importantly, CA-CRC is associated with a poor prognosis, with a large meta-analysis reporting a poorer overall survival in patients with CA-CRC compared to patients with S-CRC [98].

1.6.1 Pathogenesis: chronic inflammation and colitis-associated colorectal cancer

The molecular events underlying the association between IBD and the increased risk of CRC are yet to be fully elucidated. The current understanding is that CA-CRC is the result of exposure to prolonged and recurrent inflammation. This is supported by the direct correlation between CA-CRC risk and the severity and duration of active disease, as well as the positive tumour-preventing effects of long-term administration of non-steroidal anti-inflammatory drugs (NSAIDS) [99].

Under the assumption that chronic inflammation is the main culprit responsible for the carcinogenesis of CA-CRC, multiple non-exclusive theories have been proposed to explain the causative relationship between inflammation and CA-CRC. Some of these explain the contribution of inflammation to carcinogenesis as being direct, whereby carcinogenesis is driven by oxidative stress and DNA damage caused by inflammation. Others propose indirect mechanisms, driven by cytokines produced by inflammatory cells and intestinal epithelial cells in response to inflammation.

One of the main theories postulates that inflammation is the driver of a different more aggressive oncogenic pathway. Evidence supporting this hypothesis comes from Baker *et al.*'s evolutionary genomic analysis on multiple regions of carcinoma, adjacent non-cancerous mucosa and blood from IBD patients with CA-CRC. The study led to the detection of a high single nucleotide alteration (SNA) burden in non-dysplastic epithelium prior to cancer formation, indicating that the evolutionary trajectory of CA-CRC begins in the highly mutated, field cancerised mucosa prior to cancer growth. Furthermore, the authors found that mutational processes in CA-CRCs were dominated by ageing-associated signatures. This led to the suggestion that mutational accumulation is the biological cost of inflammation and injury-induced cell turnover required to regenerate the epithelium. The identification of genetic differences between S-CRCs and CA-CRCs further suggested that the unique mutational composition of the neoplastic colitis-associated lesions was the result of the natural selection imposed by the microenvironment of the colitic bowel to the mutations generated by the accel-

erated ageing process. Based on their findings, Baker *et al.* proposed chronic inflammation to be the driving force that accelerates mutational accumulation and provides a distinctive selective pressure, rather than being mutagenic *per se* [96].

Another postulation is based on the view that cycles of inflammation and regeneration typical of chronic inflammation can co-opt and deregulate pro-tumorigenic signalling pathways. An example of this in IBD is provided by signalling pathways of the innate immune system, whose main function is to protect the host from pathogens and initiate tissue repair. These pathways can be corrupted and subsequently enable cancer development. Supporting the hypothesis above is a study by Kirchberger *et al.* on interleukin 22 (IL-22), an immune system cytokine which has a host protective role in acute intestinal infection and inflammation. IL-22 exerts its role by inducing the production of antimicrobial peptides, mucins and chemokines and epithelial regeneration [100, 101]. This cytokine is known to mainly act via ligation of interleukin 22 receptor (IL22R) on epithelial cells, which result in signal transducer and activator of transcription 3 (STAT3) activation [102]. STAT3 is a transcriptional regulator with key role in mature tissue function and its activation is associated with epithelial proliferation, pivotal for maintenance of tissue homeostasis in the gut [103]. Persistent IL-22, as in chronic inflammation, has been suggested to directly act on the regulator of epithelial regeneration, causing hyperproliferation and subsequent tumour maintenance. IL-22 is elevated in IBD patients and Kirchberger *et al.* found that colon tumours in mouse models require IL-22 for their maintenance. Based on this, the authors suggested IL-22 to be involved in tumour maintenance in CA-CRC carcinogenesis [104].

Nuclear factor kappa B (NFKB1), an important transcription factor in cell survival, inflammation and immunity, has been shown to play an important role as a link between inflammation and cancer. In a study investigating the activity of Nfkb1, Schwitalla *et al.* found that increased Nfkb1 caused an amplification in WNT signalling. This, in turn, caused the dedifferentiation of epithelial non-stem cells into tumour-initiating cells. Detection of a dramatically high number of epithelial cells with stem cell-like capacity in colitis patients led to the suggestion that chronic inflammation might increase the number of putative tumour-initiating cells by causing dedifferentiation.

This led the authors to propose that the increased cancer risk in IBD patient might be attributable to the ectopic and chronic release of pro-inflammatory cytokines in the adult stem cell niche, which in turn causes an increased number and ectopic presence of stem cells. In other words, the inflammatory stroma might induce a bidirectional conversion between tumour-initiating stem cells and non-stem cells, which culminates in the development of CA-CRC [105].

Although multiple hypotheses have been proposed to explain the increased risk of cancer in IBD patients, some of which have discussed above, noteworthy is that they are not mutually exclusive and in fact it is more probable that CA-CRC is the end result of multiple mechanisms acting concomitantly. Similar to the molecular carcinogenic pathway, the cancer cell of origin and the cancer stem cell of CA-CRC are yet to be identified, although a number of cell populations involved in regeneration have been proposed. The cancer cell of origin is the normal cell to receive the first carcinogenic mutation, whilst the cancer stem cells are the cell population within a tumour that uniquely sustains malignant growth.

1.6.2 Colitis-associated colorectal cancer versus sporadic colorectal cancer

CA-CRC is phenotypically, biologically and molecularly distinct from S-CRC (Table 1.5). Clinically, CA-CRCs often occur at younger age. They are frequently synchronous, have a more proximal distribution and present with mucinous or signet ring cell histology at a higher frequency. Moreover, they often originate from flat dysplasia, in a field of inflammation and scarring, in contrast with the commonly adenomatous origin of S-CRCs [96]. Colitis-associated tumorigenesis tends to histologically proceed through increasing grades of dysplasia. It comprises a transition from LGD to HGD, to culmination in adenocarcinoma [106]. This differs from the well-established adenoma-carcinoma carcinogenic pathway that most sporadic tumours follow [107], as well as from the most recently described and less common CIMP pathway. Furthermore, CA-CRC tends to be multifocal. The multifocal nature of CA-CRC might be explained by field cancerisation, a commonly occuring and recognised mechanism underlying CA-CRC carcinogenesis. Field cancerisation occurs when prior to the growth of a ma-

lignant lesion, a normal cell lineage acquires pro-tumorigenic alterations that are positively selected for in the microenvironment of an otherwise healthy organ. The mutant clone can grow to produce large patches or fields of cells that are primed to eventually progress to a neoplasm [108, 109].

Although the molecular events underlying CA-CRC tumorigenesis are yet to be fully elucidated, unique genetic and epigenetic alterations have been detected in CA-CRC compared to S-CRC. Genetically, mutational differences, including differences in type, frequency, timing and mutated genes have been identified between S-CRC and CA-CRC (Figure 1.8) [110]. *TP53* mutations, which are rarely detected in precursor adenomas of S-CRC, are a common early event in the carcinogenesis of CA-CRC being found in 50-85% of precancerous lesions and even in non-neoplastic mucosa. On the contrary, loss of APC, a common early event in S-CRC patients, is less frequent in CA-CRC and when present it occurs later in carcinogenesis. Similarly, *KRAS* alterations, which are common in S-CRC, occur at a lower frequency and later in CA-CRC. Conversely, *MYC* amplifications (26%) are more commonly found in CA-CRC than S-CRC (4%)[110–112]. Another important difference is the early occurrence of aneuploidy in CA-CRC, as opposed to S-CRC, where it occurs late in carcinogenesis. Baker *et al.* reported the presence of aneuploidy in non-dysplastic colitis epithelium, leading to the suggestion that CIN might be an early event in the carcinogenic pathways of CA-CRC. The group also reported the presence of unique and recurrent losses and gains (e.g. loss of 5q and 17q) in CA-CRC compared to S-CRC [96] (Figure 1.8).

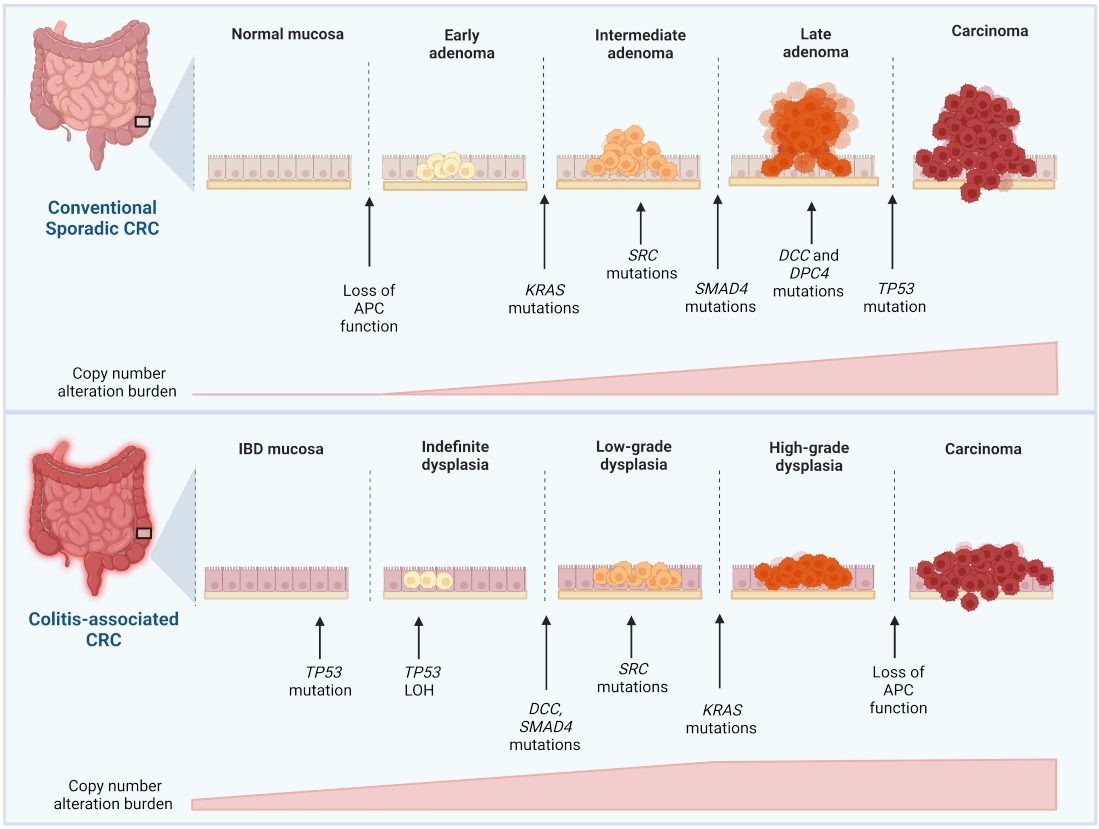


Figure 1.8: Molecular differences between CA-CRC and S-CRC at different stages of the carcinogenic pathway. Adapted from [113].

Epigenetically, aberrant promoter methylation was detected exclusively in CA-CRC in polymeric immunoglobulin receptor (*PIGR*) and oncostatin M receptor (*OSMR*), two mucosal immunity genes. Likewise, methylation of anaphase-promoting complex subunit 1 (*ANAPC1*), apc regulator of WNT signalling pathway 2 (APC2), secreted frizzled related protein 1 (*SFRP1*) and secreted frizzled related protein 2 (*SFRP2*) was observed in the progression from IBD to CA-CRC [114]. Lastly, immunoprofiling focusing on the expression level of various surface markers in immune cells, also found significant differences between sporadic and colitis-associated malignant lesions [115].

Table 1.5: Differences between S-CRC and CA-CRC. Adapted from [97, 116]

	S-CRC	CA-CRC
Epidemiology		
Age	Older age onset (>50 years old)	Younger age of IBD onset
Risk factors	Diet, smoking, obesity, CRC family history, *Helicobacter pylori*, alcohol, colonic polyps and others	Extensive colitis, increased IBD duration, CRC family history, primary sclerosing cholangitis (PSC)
Pathogenesis		
Pre-malignant lesion	Adenomatous polyp (polypoid/sessile)	Flat dysplasia mainly
Carcinogenic sequence	Adenoma-carcinoma	Inflammation-dysplasia- carcinoma
Early-stage genetic alterations	*APC* and *KRAS* mutation	Aneuploidy and *TP53* mutation
Late-stage genetic alterations	*TP53*	*APC* and *KRAS*
Clinical features		
Endoscopic characteristics	Commonly polypoid lesions; some sessile	Flat dysplasia. Synchronous and recurrent tumours
Histological characteristics	Majority adenocarcinoma. Fewer mucinous/signet ring cell morphology differentiation	Majority adenocarcinoma. Mucinous/signet ring cell differentiation more common.
Mortality and prognosis		Poor prognosis compared with S-CRC. Increase in recurrence.

1.7 Thesis objectives

In an effort to better characterise CA-CRC and improve patient management three main investigations were carried out in this thesis.

The stem cell dynamics in S-CRC and CA-CRC

Intestinal homeostasis is maintained by Lgr5-positive stem cells, though following injury, dedifferentiation results in the emergence of Lgr5-negative stem cells. This cellular plasticity of the gut's epithelium is pivotal in regeneration. However, it also plays a central role in tumour transformation, progression and adaptive response to therapy. In this regard, chapter 3 aimed at assessing the stem cell composition and dynamics in S-CRC and CA-CRC.

The role of WNT signalling pathway in CA-CRC

The role of the WNT signalling pathway is well-recognised and established in S-CRC. However, little is known about the contribution of WNT in CA-CRC. Thus, chapter 4 aimed at assessing the WNT signalling landscape in colitis-associated precursor lesions and the role of the WNT signalling pathway in CA-CRC carcinogenesis.

The early detection of CA-CRC

The current gold standard management strategy for IBD patients' increased cancer risk relies on endoscopic surveillance, often with concomitant biopsy and histological assessment, with the aim of detecting dysplasia as a precancerous lesion. However, this management strategy is ineffective and costly. In light of this, chapter 5 aimed at assessing three stratification strategies to determine cancer risk in LGD from IBD-patients.

Chapter 2

Materials and Methods

General methods utilised throughout the thesis will first be outlined in depth under the section general material and methods. This will be followed by a chapter specific outline of the experiments and analysis carried out for each results' chapter.

2.1 General Materials and Methods

2.1.1 Human

2.1.1.1 Human cohorts and ethical approval

Multiple human cohorts were used for the investigations carried out in this thesis. All human samples were obtained following ethical approval and individual informed consent from multiple centres.

Colitis cohort: obtained from London North West Healthcare NHS Trust (St Mark's Hospital), London (London Stanmore committee 10/LO/1613, London-Fulham Research Ethics committee 18/LO/2051), Oxford University Hospitals NHS Foundation Trust, Oxford (Oxford GI Illness ethics, 18/1L/JE/Early-detection, OCHRe:18/A11, London-Fulham Research Ethics committee 18/LO/2051), Belfast Health and Social Care Trust (16/NI, 0030) and the Newcastle Upon Tyne Hospitals NHS Foundation Trust (London-Fulham Research Ethics committee 18/LO/2051).

Normal GI cohort: obtained from Oxford University Hospitals NHS Foundation

Trust using Oxford GI Illness ethics.

Stratification in colorectal cancer (SCORT) WS4 Polyp Validation adenoma cohort: obtained from Belfast Health and Social Care Trust (16/NI,0030).

SCORT Oxford polyps: obtained Oxford University Hospitals NHS Foundation Trust (REC 15/A003).

Oxford Biomedical research centre (BRC) cancer cohort: obtained from Oxford University Hospitals NHS Foundation Trust under the REC: 17/NW/0252.

2.1.2 Mouse

2.1.2.1 Housing

Mouse work was performed in compliance with the Home Office UK regulations and Animal (Scientific Procedures) Act 1986, under PPL P0B63BC4D Leedham (Leedham) Epithelial-mesenchymal interaction in tissue regeneration and carcinogenesis and PIL I80E72907.

All mice were housed in the specific-pathogen-free Functional Genetics Facility at the Wellcome Centre for Human Genetics, University of Oxford. Mice were kept in individually ventilated cages, where they had unrestrained access to regular diet and water. Day-to-day care of the mice, including food and water changes, daily health monitoring, cage cleaning and weaning was carried out by the facility's technicians. Procedures were performed with the assistance and/or guidance of Dr Hayley Belnoue-Davis.

2.1.2.2 Mouse strains and models

Mice used in this thesis were maintained on a C57BL/6J background for ≥ 6 generations. Details on the models used are outlined in Table 2.1.

2.1.2.3 Breeding

Breeding colonies were established from sexual maturity (i.e. 6 weeks of age) onwards. Typically, breeding pairs suitable to obtain offspring of the genotype of interest were

Table 2.1: Mouse models details

Allele	*Description*	*Reference*
Apc$^{fl/fl}$	Allele with a *loxP* site in intron 13 and another with a PGK-neo casette into intro 14. Cre mediated recombination deletes a region of *Apc* encompassing exon 14 and induces a frameshift mutation at codon 580. This results in the production of a truncated protein with greatly impaired function.	Shibata1997 [117]
Apc^{Min}	Allele with a point mutation that changes nucleotide 2549 from a T to an A, resulting in a premature stop codon and truncation of the Apc protein, which causes an increased susceptibility for CRC.	Moser *et al.* 1990 [118]
Atoh1-Cre^{ERT2}	Tamoxifen-inducible Cre recombinase, whose expression is driven by the *Atoh1* gene promoter. Recombination is inducible in all Atoh1-positive cells.	Tomic *et al.* 2018 [119]
Gli1-Cre^{ERT2}	Tamoxifen-inducible Cre recombinase, whose expression is driven by the *Gli1* gene promoter. Recombination is inducible in all Gli1-positive cells.	Ahn *et al.* 2004 [120]
Kras^{G12D}	Allele with the G12D point mutation in exon 1 (commonly found in human cancer), whose expression is blocked by a loxP-flanked stop codon. Thus, expression of the mutant *Kras* is achieved only upon Cre recombination.	Jackstadt *et al.* 2019 [121]
Lgr5^{DTR}	A human diphtheria toxin receptor (DTR) cassette is knocked-in to the first exon of *Lgr5*. This allows for the *Lgr5*-positive cells to express DTRs and be ablated through the administration diphtheria toxin.	Tian *et al.* 2011 [122]
Rosa$^{26N1icd/+}$	A targeting vector containing a floxed neo-cassette followed by a mouse *Notch1* sequence fragment and GFP inserted at the endogenous locus. Excision of the floxed neo-cassette by recombination results in the expression of the *Notch1* sequence encoding an intracellular portion of Notch1, but lacking the c-terminal PEST domain	Jackstadt *et al.* 2019 [121].
Rspo3$^{inv/+}$	*Rspo3* coding sequence inserted in the inverse orientation, flanked by two pairs of oppositely oriented, homologous Lox sites and behind the ubiquitous CAGGS promoter. This enables conditional *Rspo3* expression upon Cre activity following tamoxifen induced recombination.	Hilkens *et al.* 2017 [123]
tdTomato	Tamoxifen-inducible Cre designed to have a *loxP*-flanked STOP cassette that prevents the transcription of *tdTomato*, a CAG promoter-driven red fluorescent protein variant. Cre recomination results in a robust tdTomato fluorescence expression.	Madisen *et al.* 2010 [124]
Trp53$^{fl/fl}$	Allele with *loxP* sites flanking exon 2 through 10. Upon tamoxifen-induced recombination the exons are excised.	Marino *et al.* 2000 [125]
Vil1-Cre^{ERT2}	Tamoxifen-inducible Cre recombinase whose expression is driven by a 9 kbp segment of the regulatory region of the epithelial *Vil* gene. Consequently, recombination is inducible in epithelial cells along the crypt-villus axis and in undifferentiated progenitor cells in the crypt region.	el Marjou *et al.* 2004 [126]

allowed to breed up to the obtainment of 5-6 litters. However, where an enhancement of breeding was required, breeding trios, consisting of two female mice and one male mouse were utilised. For mouse models of *Rspo3* overexpression, only females bearing the $Rspo3^{inv/+}$ allele and wild type males were used due to the reduced fertility of the males with the $Rspo3^{inv/+}$ allele.

2.1.2.4 Earclip

Following weaning at 3 weeks of age, ear punches, unique by number and/or location, were taken for identification purposes. The tissue from these punches was then used to provide material for DNA extraction to genotype the mice.

2.1.2.5 Genotyping

DNA was extracted following the protocol outlined in Truett's *et al.*, a simple method that enables rapid and cost-effective isolation of DNA of sufficient quality for genotyping purposes, but not for other downstream applications (e.g. next generation sequencing). The earclip was incubated in 50 µL of HotSHOT reagent (25 mM NaOH, 0.2mM disodium ethylenediaminetetraacetic acid (EDTA), pH 12) at 100°C for 30 minutes, followed by the addition of 50 µL of neutralising reagent (40 mM Tris-HCL, pH 5) to halt the reaction.

For the polymerase chain reaction (PCR) reaction, generally a master mix consisting of template DNA, PCR buffer, MgCl2, deoxynucleoside triphosphate (dNTP), primers and purified water was prepared.

A standard PCR reaction comprised the following steps:

- Initial denaturation at 95°C for 5 minutes
- Annealing (usually 35 cycles) of 94°C for 1 minute
- Extension at 55° for 1 minute
- Final extension at 72°C for 10 minutes

PCR master mixes and reactions were optimised based on the primers and target size. The concentration of the PCR master mix components, additives and annealing temperatures were modified accordingly.

PCR product were analysed by gel electrophoresis. For this, 2% agarose gels stained with ethidium bromide were used. PCR products were loaded onto the gels with loading buffer and run alongside a 1 Kb DNA ladder using tris-borate-EDTA (TBE) running buffer at 120V for 30 minutes. The gels were visualised on ChemiDoc MP Imaging System (Bio-rad).

Table 2.2: Primer sequences

Allele	Forward primer	Reverse primer
$Apc^{fl/fl}$	TTTGTTGTTGTTGTTCGGGG	TCTTTTACCTCAGCCCACTCTT
$Atoh1\text{-}Cre^{ERT2}$	TTTGTTGTTGTTGTTCGGGG	TCTTTTACCTCAGCCCACTCTT
Cre^{ERT2}	GGTCGATGCAACGAGTGATGAGG	CCAGAGACGAAATCCATCGCTCG
$Lgr5^{DTR}$	AGGCAGATCTGGACCTTTTGA	GCAGACAGACAGATGACAGC
$Rspo3^{inv/+}$	CGCCATTAATCGATCCCG	CCTATCTGCTTCATGCCAATCC
tdTomato (i)	CTACAAGGTGAAGATGCGCG	CTACAAGGTGAAGATGCGCG
tdTomato (ii)	CCGACATCCCCGATTACAAG	TAGTAGTAGCCGGGCAGTTG
$Trp53^{fl/fl}$	AAGGGGTATGAGGGACAAGG	TGTCTTTCCCCAGCACAGT
Apc^{Min} *(i)*	GCCATCCCTTCACGTTAG	GAAGACAGAAAAGGGGAGGG
Apc^{Min} *(i)*	GCCATCCCTTCACGTTAG	TTCCACTTTGGCATAAGGC
$Kras^{G12D}$	TGTCTTTCCCCAGCACAGT	CTGCATAGTACGCTATACCCTGT
$Kras^{G12D}$	TGTCTTTCCCCAGCACAGT	GCAGGTCGAGGGACCTAATA

2.1.2.6 Induction of recombination

Recombination in inducible models was activated by intraperitoneal injection with tamoxifen (T5648, Sigma-Aldrich) dissolved in corn oil and ethanol. Concentration and number of injections varied between models and are reported in apposite chapter specific sections.

2.1.2.7 Collection

Mice were sacrificed at pre-determined experimental timepoints or at humane endpoint, when displaying symptoms of intestinal phenotype or general illness (e.g. hunching, pale feet). All mice were euthanized by cervical dislocation. The intestinal tract

was removed and divided into four segments: three segments of approximately equal length comprising the small intestine (small bowel (SB)1, SB2, and SB3) and one segment comprising the large intestine. The intestines were washed by flushing phosphate-buffered saline (PBS) through the lumen with a syringe. They were opened longitudinally using a scalpel and gut rolling device [127]. They were fixed overnight in 10% neutralised formalin, before being transferred to 70% ethanol whilst awaiting to be processed for embedding.

2.1.2.8 Processing

Fixed gut sections were rolled, pinned and placed in an appropriately sized cassette. They were then processed for embedding in a Bavimed tissue processor using the following programme:

- 70% ethanol for 45 minutes
- 90% ethanol for 45 minutes
- 100% ethanol for 30 minutes
- 100% ethanol for 45 minutes
- 100% ethanol for 45 minutes
- Histo-Clear for 30 minutes
- Histo-Clear for 30 minutes
- Histo-Clear for 30 minutes
- molten wax for 30 minutes (60-65°C)
- molten wax for 45 minutes
- molten wax for 45 minutes

2.1.2.9 Embedding

Processed samples were embedded in paraffin wax in the Leica paraffin embedding station (EG1140H).

2.1.3 Sectioning

Formalin-fixed paraffin-embedded (FFPE) blocks sectioning was undertaken using a HistoCore Rotary Microtome (RM2255, Leica) and varied according to the purpose of the tissue.

- For histological assessment, FFPE blocks were sectioned at 4 μm thickness onto charged glass slides (Superfrost plus) using diethyl pyrocarbonate (DEPC)-treated water in the wash bath. The sections were allowed to dry at room temperature.
- For DNA extraction, six sequential 10 μm sections were cut using DEPC-treated water in the wash bath. They were mounted onto 1.0 PEN-membrane covered Membrane slides (Zeiss, 415190-9042-000) or charged glass slides (Superfrost plus). Membrane slides were ultraviolet (UV)-treated for 30 minutes prior to use.
- For ribonucleic acid (RNA) extraction, three 10 micron sections were cut in DEPC-treated water onto either charged slides (Superfrost plus) or directly placed in Eppendorf tubes for RNA extraction.

2.1.4 Histology

2.1.4.1 Haematoxylin and Eosin staining

Haematoxylin and eosin (H&E) staining was performed according to standard procedure following the steps outlined below.

- Deparaffinisation: Xylene for 5 minutes;
- Hydration: 100% ethanol for 5 minutes, 90% ethanol for 5 minutes, 70% ethanol for 5 minutes, tap water for 2 minutes;

- Haematoxylin staining: Harris Haematoxylin for 30 seconds, running water until water is clear, brief dip in 1% acid alcohol, water for 5 minutes;
- Eosin staining: Eosin for 3 minutes, quick dip in water;
- Dehydration: 70% ethanol for 5 minutes, 90% ethanol for 5 minutes, 100% ethanol for 5 minutes;
- Clearing: Xylene for 1 minute;
- Mount on coverslip with DPX.

2.1.4.2 Haematoxylin staining

Four micron sections were haematoxylin stained using a Leica ST5010 Autostainer XL, following the programme below:

- Deparaffinisation: Xylene for 3 minutes, three consecutive times;
- Hydration: 100% ethanol for 3 minutes, 90% ethanol for 3 minutes, 70% ethanol for 3 minutes, water for 2 minutes;
- Staining: Haem Gill 3 for 1:30 minutes, water 2 minutes, 1% acid alcohol for 45 seconds, water for 2 minutes;
- Dehydration: 70% ethanol for 2 minutes, 90% ethanol for 2 minutes, 100% ethanol for 2 minutes twice;
- Clearing: Xylene for 2 minutes;
- Exit and Coverslip.

2.1.4.3 *In situ* hybridisation

In situ hybridisation (ISH) was performed on 4 µm sections mounted on charged glass slides. The sections were first baked at 60°C for 1 hour, dewaxed with a two-times incubation of 5 minutes in xylene and rehydrated through a two-times wash in 100% ethanol.

For chromogenic ISH, the RNAscope 2.5 HD Reagent Kit (ACDBio, 322300) was used. The manufacturer's protocol was followed. Slides were counterstained in haematoxylin, dehydrated, cleared and then mounted with a coverslip using DPX.

For fluorescent ISH, the RNAscope Fluorescent Multiplex Reagent Kit V2 (ACDBio, 323110) was used, following the protocol outlined by the manufacturer. The stained slides were mounted with a coverslip using the mounting agent ProLong Gold Antifade Mountant (ThermoFisher, P36930) and stored at 4°C in the dark.

2.1.4.4 Immunohistochemistry

Immunohistochemistry (IHC) was performed on 4 μm sections mounted on charged glass slides. The sections were first baked overnight at 60°C, dewaxed in xylene for 5 minutes, and rehydrated through graded alcohols to water for 2 minutes each.

For chromogenic IHC, the following protocol was followed: antigen retrieval was performed by pressure-cooking the sections in 10 mM citrate buffer for 5 minutes. Endogenous peroxidase activity was blocked by incubating in 3% hydrogen peroxidase (in methanol) for 20 minutes. Next, sections were blocked with 1.5% serum for 30 minutes, after which they were incubated with primary antibodies for 2 hours. The sections were incubated with appropriate secondary antibodies for 30 minutes at room temperature. For chromogenic visualization, sections were incubated with ABC (VectorLabs) for 30 minutes and stained using DAB solution (VectorLabs). After, they were counterstained with haematoxylin, dehydrated and mounted.

For fluorescent IHC, the following protocol was followed: antigen retrieval was performed by pressure-cooking the sections in 10 mM citrate buffer for 5 minutes. Next, non-specific binding sites for antibodies were blocked by incubating the slides for 1 hour at room temperature with 5% normal goat serum (NGS). The sections were then incubated with the primary antibody diluted in 1% NGS at 4°C overnight. Sections were washed three times for 10 minutes each in PBS prior to applying the secondary antibody. The secondary antibody was diluted in 1% NGS and allowed to incubate for 40 minutes at 40°C. Finally, the sections were washed three times for 10 minutes each in PBS and mounted with 4',6-diamidino-2-phenylindole (DAPI) con-

taining mounting agent (VectorLabs, H-1800). The mounted slides were stored at 4°C in the dark.

2.1.4.5 Imaging and analysis

H&E, chromogenic ISH and IHC slides were imaged and scanned using the slide scanner Aperio CS2, at x20 or x40 magnification.

Fluorescent ISH slide, fluorescent IHC and fluorescent ISH-IHC slides were imaged using either the Leica DMi8 SP8 FALCON Confocal microscope or the Olympus SpinSR Sora microscope. Fluorescent images were processed using ImageJ [128], a Java-based image processing program.

Images were quantified using the digital pathology software QuPath [129] and HALO. Details of the quantifications are provided in the chapter specific sections.

2.1.5 Nucleic acids

2.1.5.1 Epithelial enrichment for DNA extraction

The tissue sections were first dewaxed in xylene for 5 minutes, and rehydrated using 5 minutes incubations in serial dilutions of ethanol i.e. 100%, 90%, 70% and water. To allow crypt visualisation and identification of epithelium-rich tissue, sections were stained with a 1:6 dilution of methyl green, followed by two brief water washes and a 5 minutes incubation in 100% ethanol. A sequential annotated H&E was used as guide for epithelial enrichment. The latter was undertaken either by laser-capture microdissection (LCM) using a Zeiss PALM Micro Beam microdissection system or manual scraping with a pipette tip. The choice of epithelial enrichment method was based on the size and cellularity of the tissue. In both instances, enriched sections were stored at 4°C until extraction.

2.1.5.2 DNA extraction from FFPE samples

DNA extraction from FFPE samples was undertaken using the High Pure FFPET DNA Isolation Kit (Roche, 06650767001). The manufacturer's protocol was followed with a

few alterations to enhance DNA yield. Briefly, the enriched epithelium was digested in 60 µL of tissue lysis buffer and proteinase overnight (14-16 hours) at 56°C, followed by enzyme denaturation at 90°C for 1 hour. After cooling the lysate, 67 µL of DNA binding buffer were added, vigorously mixed and allowed to incubate at room temperature for 10 minutes. Next, 33 µL of isopropanol were added and thoroughly mixed to precipitate the RNA. The lysate was passed twice through the DNA-binding spin column by centrifugation. This was followed by three washes: one with 500 µL of wash buffer 1 and two with 500 µL of wash buffer 2. The sample was centrifuged at 8000 RCF for 60 seconds and the collection tubes were emptied each time. The DNA-binding spin column was then dried by centrifugation at 1600 RCF for 1 minute and transferred into a fresh 1.5 mL reaction tube. This was followed by two sequential applications of preheated (56 °C) elution buffer. The elution buffer was incubated for 5 minutes at room temperature and passed by centrifugation to release the DNA bound to the spin columns each time. The volume of elution buffer used for the two application varied between 30-45 µL based on tissue size. The samples were quality checked and stored at -20 °C.

2.1.5.3 Assessment of DNA quantity and quality

DNA quantity was determined with either a Qubit Fluorometer using the Qubit double-stranded DNA (dsDNA) High Sensitivity Assay (ThermoFisher Scientific, Q32854) or the Qubit dsDNA Broad Range Assay (ThermoFisher Scientific, Q32853). DNA quality was assessed by Agilent Tapestation, using the High Sensitivity D1000 DNA ScreenTape (Agilent, 5067-5584) and reagents (Agilent, 5067-5585).

2.1.5.4 Tissue enrichment for RNA extraction

Where required, tissue enrichment was undertaken by manually scraping the tissue with a pipette tip. A sequential annotated H&E was used as guide for the process.

2.1.5.5 RNA extraction from FFPE samples

RNA extraction from FFPE samples was undertaken using the High Pure FFPET RNA Isolation Kit (Roche, 06650775001). The manufacturer's protocol was followed with adaptations to enhance RNA yield. In brief, tissue was first digested by adding 8 µL of 10% SDS and 20 µL of proteinase K to the 50 µL of RNA tissue lysis buffer and incubating at 85°C for 30 minutes. This was followed by a second digestion, consisting in the addition of 40 µL of proteinase K and incubation for 30 more minutes at 55°C. Both incubations were undertaken in a shaking heat block set at 600 RPM. The lysate was cooled at room temperature. Equal volumes, i.e. 162.5 µL, of binding buffer and absolute ethanol were added to it and mixed thoroughly. The lysate was then passed through the filter via centrifugation at 6000g for 30 seconds. The filter was dried by centrifugation at 16000g for 2 minutes, before being placed in a new collection tube. DNase treatment was performed to remove DNA bound to the filter column by incubating DNase working solution (i.e. 90 µL of DNase Incubation Buffer and 10 µL of DNase) for 15 minutes at room temperature. This was followed by three washes of the spin columns, one with 500 µL of wash buffer I and two with 500 µL of wash buffer II. The sample was centrifuged at 6000g for 20 seconds and the collection tubes were emptied each time. The DNA-binding spin column was dried by centrifugation at 16000g for 2 minutes and transferred into a fresh 1.5 mL reaction tube. This was followed by two sequential applications of the elution buffer heated at 56°C. The elution buffer was incubated for 5 minutes at room temperature and passed by centrifugation to release the RNA bound to the spin columns each time. The volume of elution buffer used for the two applications varied between 30-35 µL based on tissue size. The samples were quality checked and stored at -80 °C.

2.1.5.6 RNA extraction from fresh frozen tissue

The tissue was initially disrupted using an electrical homogeniser (Polytron). RNA was extracted using the Qiagen RNeasy Micro Kit (Qiagen, 74004), following the manufacturer's protocol for the purification of total RNA from animal and human tissue.

Two alterations to the protocol were made to improve RNA yield:

- Pre-heating the RNase-free water to 56°C for elution;
- Performing two sequential applications of the RNase-free water, allowing the water to incubate in the filter for 5 minutes each time before centrifugation.

The optional on-column DNase digestion was undertaken.

2.1.5.7 Assessment of RNA quality and quantity

RNA quantity was determined with a Qubit Fluorometer (ThermoFisher) using the Qubit RNA High Sensitivity Assay (ThermoFisher Scientific, Q32855) or the Qubit RNA Broad Range Assay (ThermoFisher, Q10211). RNA quality was assessed by Nanodrop (ThermoFisher). The absorption measures at 260 nm and 280 nm were taken to determine the 260/280 ratio, an established measure of nucleic acid purity. RNA samples with a 260/280 ratio of $\sim$2 were considered pure. In a representative subset of samples, RNA quality was also assessed by Agilent Tapestation, using the RNA TapeStation ScreenTape (Agilent, 5067-5579) and sample buffer (5067-5580). For FFPE samples, DV200 values i.e. the proportion of RNA fragments within the sample that are greater than 200 bp in length, were used as a measure of quality instead of RINs, which are known to be universally poor for RNA derived from FFPE samples.

2.1.6 3′ RNA sequencing

2.1.6.1 3′RNA sequencing library preparation for FFPE samples

Library preparation was undertaken with the Lexogen QuantSeq 3′ mRNA-seq Library Prep Kit FWD HT for Illumina (Lexogen, 015.96), following the manufacturer's modified protocol for FFPE samples. Further optimisations were made to the first step of the protocol, due to the low-input low-quantity nature of the samples. Specifically, the samples were preheated at 42° before the addition of the master mix and kept at 42° prior to being subjected to higher temperatures as required by the next step of the protocol. RNA input varying from 53-10000 ng and 14-21 PCR cycles were used.

The final library quantity was assessed by Qubit Fluorometer using the Qubit DNA High sensitivity kit. Library quality was assessed by Agilent Tapestation using the High sensitivity D1000 DNA system. Libraries with the following characteristics were considered to have passed quality check:

- Size greater than 150 bp, with unbound adaptor comprising less than 10% of the final library and without a second peak between 1000-9000 bp;
- Concentration greater than 0.2 ng/µL in 15 µL.

2.1.6.2 3'RNA sequencing library preparation for fresh frozen samples

Library preparation was undertaken with the Lexogen QuantSeq 3' mRNA-seq Library Prep Kit FWD HT for Illumina (Lexogen, 015.96), following the manufacturer's protocol. RNA input varying from 10-10000 ng and 14-17 PCR cycles were used. Quality and quantity of the libraries was assessed as outlined above.

2.1.6.3 3'RNA sequencing library pooling and sequencing

Uniquely indexed libraries were merged in pools of 16-20 libraries via equimolar pooling to a final volume of 100 µL using the following formula:

$$Estimated\ molarity = \frac{library\ concentration\ (\text{ng}/\mu\text{L})}{dsDNA\ molecular\ weight\ x\ peak\ fragment\ size}$$

Libraries were sequenced by:

- Barts Cancer Institute Genome Centre on Illumina's NextSeq500 (75 bp, paired end)
- Oxford Genomic Centre on Illumina's HiSeq4000 first and subsequently on Illumina's NextSeq500 (75 bp, single read) (due to decommissioning of the HiSeq4000)

aiming for an average mean depth of 1x per sample after initial processing.

2.1.6.4 3'RNA sequencing raw data processing

3'RNA sequencing (3'RNAseq) rawdata processing and quality check was performed on the Oxford University Biomedical Research Computing using a pipeline established for internal use by Dr Gabriel Valbuena.

Quality check was performed with fastQC. Adapters, polyA read through and low quality tails were trimmed. Trimmed fastq files were aligned to the reference genome (Grch38) with STAR. The R function FeatureCounts was used to quantify mapped reads to each gene. Individual sample feature counts were collated into a single dataframe for downstream analysis using R.

Raw feature counts were further analysed using the R packages EdgeR and limma. Only samples with 1 million minimum mapped read were taken forward. Samples with <1 million mapped reads were resequenced to achieve such cut off. This cut-off was set to ensure confidence when reporting quantitation in count per million (CPM).

2.2 Chapter 3 specific materials and methods

A study specific outline of the experiments and analysis carried out in chapter 3 is provided here. General methods used are briefly mentioned, with references to the appropriate sections above where the protocols are described in detail. Methods and analyses specific to solely this study are described in detail here.

2.2.1 Sample selection

The study was conducted on samples from three human cohorts: Oxford BRC cancer cohort, SCORT WS4 Polyp Validation adenoma cohort and Colitis cohort. Furthermore, human genetic and/or transcriptomic data was obtained from the SCORT WS4 Spinal Cancer cohort, the Fluoropyrimidine, Oxaliplatin and Targeted-Receptor pre-Operative Therapy (FOxTROT) trial, the the cancer genome atlas (TCGA) colorectal cancer cohort and the biopsies of surgical specimens (BOSS) study [130]. Human colon single cell RNA sequencing (scRNAseq) data was accessed from Lee *at al.* 2020 [131].

Two main mouse cohorts were used (details in the appendix Table A.1):

- The Oxford mouse cohort: internally established and molecularly profiled by members of the Leedham laboratory;
- The ACRCelerate cohort: comprising primary murine tumours and organoids, established and molecularly profiled by collaborators at the Beatson Institute for Cancer Research. RNA sequencing data and tissue block for histology were obtained from the collaborators.

2.2.2 Derivation of molecular signatures

Molecular signatures specific to the CBC cells and regenerative stem cell (RSC) were identified in the literature. For the CBC stem cells gene signature, the "Lgr5 intestinal stem cell signature" reported in Muñoz *et al.* [132] was used. The RSCs gene signature was derived by combining and curating the "Repairing epithelium genelist" from Yui *et al.* [30] and "Spheroid upregulated gene list" from Mustata *et al.* [133].

2.2.3 RSC signature curation

The genelist obtained from the combination of the Yui *et al.* and Mustata *et al.* gene signatures was curated by Dr Gabriel Valbuena. The signature genes were characterised for cell type-associated expression using hierarchical clustering in published scRNAseq data from colorectal tumours and normal colon tissue [131]. Genes that showed the following expression patterns were excluded:

- Genes from clusters that represented epithelial-expressed genes with high stromal expression;
- Stromal-expressed genes with no distinctive expression in other cell types;
- Genes with no clear cell-type specific expression.

2.2.4 Identification of signature markers and assessment of expression of markers

The CBC and RSC signatures were interrogated to identify molecular markers that could be used for on-slide assessment. Markers for each signature were chosen based on the literature. Using human scRNAseq and fluorescent ISH the expression of CBC and RSC genes in the normal cell compartment was assessed.

2.2.5 Stem cell index calculation

To calculate the stem cell index a sample's enrichment score for the RSC and CBC signatures was first measured using the gene enrichment analysis method gene set variation analysis (GSVA) [134]. This was performed on either normalised CPM in RNA sequencing (RNAseq) data or from mean per gene values in microarray data. The CBC score was then subtracted from the RSC score.

2.2.6 Assessment of stem cell marker expression

Fluorescent ISH on both human and mouse FFPE samples was performed as per section 2.1.4.3. Probes were purchased from ACDbio: Hs-ANXA1 (465319), Mm-Anxa1 (509291), Hs-PLAUR (542701), Mm-Ly6a-C2 (427571-C2), Hs-Lgr5-C2 (311021-C2), Mm-Lgr5-C2 (312171-C2) Mm-Lgr5-C1 (312171).

2.2.7 Transcriptomic profiling

Samples from the Colitis, SCORT WS4 Polyp and Oxford mouse cohort were transcriptomically profiled in-house. The samples from Colitis and SCORT WS4 Polyp were FFPE, consequently RNA extraction and sequencing was performed as per sections 2.1.5.5 and 2.1.6.1. The Oxford mouse cohort comprised fresh frozen samples, thus the methods 2.1.5.6 and 2.1.6.2 were followed.

The ACRCelerate cohort was established and transcriptomically profiled by collaborators in the Sansom laboratory at the Beatson Institute for Cancer Research. RNA extraction was performed with the RNeasy Micro Kit from Qiagen (74004), followed by

DNase treatment using the DNA-free kit from Life Technologies (AM1906). RNA sequencing libraries were prepared using the TruSeq RNA sample prep kit v2 (Illumina) and sequenced on an Illumina NextSeq (High Output run 75 cycles paired end reads). The raw data was preprocessed and aligned to mouse genome build GRCm38.98. The counts were normalised.

RNAseq data was used to calculate the stem cell index.

2.2.8 Stem cell sorting by FACS and clonogenicity analysis

Stem cells were sorted from *KPN* tumours. *KPN* mice were injected with 2 mg of tamoxifen to induce recombination. Mice were aged to allow development of intestinal phenotype. Tumours were removed with a scalpel and cut into smaller fragments before being washed with 1 mL PBS for five times. The tissue was subjected to manual dissociation followed by enzymatic digestion through an approximately 1 hour incubation in 5 mL 10x Trypsin (5 mgmilliL, Gibco), 1x DNase buffer (500 µL) and 200U recombinant DNase I (20 µL, Roche, 04716728001) at 37°C in a shaker set at 100 RPM. The incubation time was altered according to tumour size. The tumour fragments were then further dissociated by manual shaking in 5 mL ADF five times. After this, the resulting suspension was spun down, the supernatant was aspirated and resuspended in 1X PBS and 1% fetal bovine serum (FBS) (also referred to as fluorescence-activated cell sorting (FACS) buffer). The cell suspension was passed through a 7 mm cell strainer.

The cell suspension was first incubated in Mouse BD Fc Block™ purified anti-mouse CD16/CD32 (BD Biosciences, 553141) at 4°C for 5 minutes (1 µL of block per million of cell), before staining with the antibody mix. The antibody mix included antibodies for the following markers: cluster of differentiation 31 (CD31) (Thermo Fisher Scientific, 25-0311-82), cluster of differentiation 45 (CD45) (BD Biosciences, 561868), epithelial cellular adhesion molecule (EpCAM) (BD Biosciences, 740281), lymphocyte antigen 6 complex (Ly6a) (BioLegend, 108119) and EPH receptor B2 (Ephb2) (primary antibody: R&D Systems, AF467; secondary antibody: R&D Systems, F0108). Moreover, a live/dead stain, LD Fixable Green Dead Cell Stain Kit (ThermoFisher, L34969) was used to discriminate between dead and alive cells. The experiment included an

unstained control, FMOs of Ly6a and Ephb2 and the sample to sort. Staining mixes were the prepared as per Table 2.3.

The apposite antibody mix was added to each sample and mixed thoroughly, except for the unstained control, which was stored at 4°C until sorting. The antibody mixes with the samples (50 µL of sample to sort in 50 µL of antibody mix) were then allowed to incubate for 30 minutes at 4°C in the dark. At the end of the incubation, cells were washed with 1 mL of FACS buffer and centrifuged. The supernatant was removed and cells were resuspended in FACS buffer. The Ephb2 antibody was unconjugated, therefore a second incubation with the secondary antibody was undertaken for the samples stained with Ephb2. Following a second wash of the cells, each cell suspension was resuspended in 1 mL of FACS buffer and incubated for 30 minutes protected from light with the live/dead stain (1 µL per million of cells). Finally, all cell suspensions were spun down, resuspended in 350 µL FACS buffer and transferred to FACS polystyrene tubes.

Table 2.3: FACS staining mix

	EpCAM	CD31	CD45	Ly6a	Ephb2-primary	Ephb2-secondary	FACS buffer
Unstained control	0	0	0	0	0	0	50
Ly6a FMO	1	1	0.25	0	0.5 µg/million cells	10 µL/million cells	200
Ephb2 FMO	1	1	0.25	4	0	0	200
Cells to sort	1	1	0.25	4	0.5 µg/million cells	10 µL/million cells	200

FMO = fluorescence minus one (FMO), Unit of volumes = (µL)

In the meantime, 1 mL of media was added to Eppendorf tubes coated with FBS, to be used for cell collection from sorting. The stained cells were sorted for the following populations on a BD FACS Aria Fusion Flow Cytometer: (a) Ly6a-positive;Ephb2-

positive, (b) Ly6a-positive;Ephb2-negative, (c) Ephb2-positive;Ly6a-negative, and (d) Ly6a-negative;Ephb2-negative. The gating strategy used is reported in Figure 3.7.

Sorted cells were plated (12-well plate) in matrigel supplemented with Jagged-1 (Anaspec, 1 μM) and overlaid with advanced DMEM/F12 culture media (Thermo Fisher Scientific). The culture media was supplemented with: GlutaMAX (Thermo Fisher Scientific, 1%) and Penicillin/Streptomycin (Thermo Fisher Scientific, 0.5%) human EGF (Peprotech; 50 ng/mL), murine Noggin (Peprotech; 100ng/mL), human R-spondin1 (R&D Systems; 500 ng/mL), murine Wnt3a (Cell guidance systems; 100 ng/mL), Chir99021 (Stemgent; 3 μM), Prostaglandin E2 (PGE2; Sigma; 2.5 μM), Nicotinamide (Sigma; 10 mM), N-2 supplement (Thermo Fisher Scientific; 1%) and B-27 supplement (Thermo Fisher Scientific; 2%). In the first 3 days, cells were provided with Y-27632 as additional supplement. An approximately equal number of cells for each cell type was plated.

At day 7, the plates were imaged and the number of organoids counted manually. Results from the count were assessed to determine organoid growth efficiency and plotted on R.

2.2.9 Analysis of driver genes associated with variable stem cell molecular phenotype

Mutational and copy number alteration (CNA) analysis was performed on data obtained from the TCGA Colorectal adenocarcinoma project (TCGA-COADREAD). Mutation Annotation format files were generated by *SomaticSniper*. Mean segment copy number calls and RSEM-normalised RNA sequencing files were downloaded from the genomic data commons (GDC) portal. First, the stem cell index score were calculated using GSVA and lesions were segregated into polarised deciles for stem cell marker expression. Mutational and CNA analysis was then carried out on those within CBC and RSC predominant deciles.

For SNA analysis, mutations were filtered for pathogenicity using the *SIFT* and *Polyphen* predictors. An extensive list of driver genes was assembled using a combination of common CRC drivers, genes in the Intogen Colorectal Cancer (COREAD) driver

gene list and a list of genes of specific interest for our laboratory. A pairwise comparison of SNA mutations between predominantly CBC and RSC deciles was performed and SNA frequencies were plotted on R.

For the CNA analysis, values smaller than -0.3 were categorised as a loss (-1), whereas values above +0.3 were categorised as gains (+1). Gains and losses were plotted using the plotFreq function from the R package copynumber.

Common SNAs and CNAs were stratified based on signalling pathway.

This analysis was carried out by Dr Gabriel Valbuena and Dr Ester Gil Vazquez.

2.2.10 Analysis of pathways associated with variable stem cell molecular phenotype

Pathway analysis was performed on both mouse and human transcriptome data, using the previously published signatures: Yap1 [30, 135], Wnt [136, 137], Kras [138], interferon gamma (Infg) [139] and Tgfb [140].

For the analysis, single sample enrichment scores were calculated using GSVA to segregate mouse and human tumours by their stem cell index, and lesions in CBC and RSC predominant deciles were taken forward for analysis. Correlation scores were calculated using the *cor.test* function in R. Gene set enrichment analysis (GSEA) was performed using the *fgsea* package.

For the pathway analysis on the human data, mouse genes were mapped to the most confident human orthologs, based on the Ensembl annotation (March 2020).

2.2.11 Determination of stem cell phenotype dynamicity in *in vivo* murine model

In vivo response to selective pressure was assessed using $Apc^{Min};Lgr5^{DTR}$ mice. Lgr5-positive cells were ablated through a single intraperitoneal injection of one dose (50 $\mu g/kg^{-1}$) of diphtheria toxin (Merck, 322326-1 mg) in saline. Mice were then collected at timepoints 24 hours, day 3 and day 5 following injection. $Apc^{Min};Lgr5^{DTR}$ mice not injected with diphtheria toxin and Apc^{Min} were used as controls for the experiment.

Sample collection was performed as per 2.1.2.7. Following collection, polyps were processed for RNA sequencing and the rest of the bowel was processed, embedded and sectioned as per 2.1.2.8, 2.1.2.9 and 2.1.3. Sections were H&E stained as per 2.1.4.1.

Chromogenic ISH was performed for Lgr5 in normal tissue and polyps from Apc^{Min} and unrecombined Apc^{Min};$Lgr5^{DTR}$ mice. This was followed by quantification of CBC cells by QuPath and plots and statistical testing on R.

Fluorescent ISH for stem cell markers, i.e. Lgr5, annexin 1 (Anxa1) and Ly6a, was performed as per 2.1.4.3.

Ki67 and Caspase 3 IHC was performed as per 2.1.4.4, using the following antibodies: anti-Ki67 (Cell Signaling Technologies, CS12102S, 1:500) and anti-Caspase 3 (R&D Systems, AF835, 1:800). Caspase 3-positive and Ki67-positive cells were quantified using QuPath. Annotations of polyp area were first created for each tissue sample. Cells were then identified using a custom algorithm established via stain separation using colour recostruction. Positive cell detection analysis was run to identify all cells, Caspase 3-positive and Ki67-positive cells. Positive cells were reported as percentage. Staining and quantification was performed by Dr Hayley Belnoue-Davis and Dr Eoghan Mulholland. Results were plotted and statistically tested on R.

GSEA comparing diphtheria toxin untreated Apc^{Min};$Lgr5^{DTR}$ mice against day 1 treated Apc^{Min};$Lgr5^{DTR}$ mice was performed for key signalling pathways.

2.2.12 Analysis of associations of stem cell phenotype to therapy response

Associations between stem cell phenotype and therapy response were investigated in data from Trial Arm A of the FOxTROT Trial.

The analysis was undertaken on 75 patients, each of whom had pre-treatment biopsies and post-treatment resection samples collected and molecularly profiled. Patients' response to treatment was originally determined by histological assessment of the pre- and post-treatment samples, whereby patients were reported as patients with:

- No regression, where no change was detected;
- Mild regression, where residual cancer cells were higher than fibrosis;

- Moderate regression, where increased presence of residual cancer, but still high fibrosis was detected;
- Marked regression, where presence of fibrosis with rare residual cancer cells was found;
- Complete regression, where no residual cancer cells were found.

Based on the criteria above, I classified patients with no or mild regression as non-responders and patients with marked or moderate regression as responders. No patients were reported as having complete regression. The stem cell index was calculated for all samples. Samples were classified as RSC-positive, where their index score was positive, or CBC-positive, where their index score was negative. Subsequently, patients were grouped based on the difference between their pre- and post-treatment samples' index score into:

- Patients with a static stem cell phenotype, where their pre- and post-treatment samples were either both RSC-positive or both CBC-positive;
- Patients with a plastic stem cell phenotype, where the tumour shifted either from being RSC-positive pre-treatment to CBC-positive post-treatment, or from CBC-positive pre-treatment to RSC-positive post-treatment.

Statistical analyses were performed in R. Changes in stem cell index and proliferative score were assessed with Fisher's exact test and ANOVA. Differences in treatment responses between static and plastic stem cell phenotypes were evaluated with Fisher's exact test.

2.2.13 Assessment of intratumor heterogeneity in the BOSS study

To assess intratumor heterogeneity, the stem cell index in the transcriptome data derived from non-adjacent, multiregional biopsies in the BOSS study was assessed. Using R, the intra-class correlation coefficient, within-subject coefficient of variation and between-subject coefficient of variation were calculated. The analysis was performed by Dr Gabriel Valbuena.

2.3 Chapter 4 specific materials and methods

A study specific outline of the experiments and analyses carried out in chapter 4 is provided here, with general methods being briefly mentioned and referenced to the appropriate sections above and methods, and analyses specific to this chapter being described in detail below.

2.3.1 Phenotyping of the WNT signalling landscape in colitis-associated human samples

2.3.1.1 Sample selection

The study was undertaken on the following samples:

- Colitis-associated LGD samples from the Colitis cohort;
- Sporadic polyps from the SCORT WS4 Polyp Validation adenoma cohort;
- Normal colon samples from the Normal GI cohort.

2.3.1.2 Sample processing

All samples in this study were FFPE. FFPE blocks were sectioned for H&E staining, DNA and RNA extraction and ISH (as per 2.1.3). H&E were assessed by skilled pathologists for the identification of dysplasia in the colitis lesions, identification of sporadic lesions, and to confirm no pathological features in the normal control samples.

Following assessment, nucleic acids were extracted guided by the histopathological assessment as per sections 2.1.5.2 and 2.1.5.5. For transcriptomic profiling, 3'RNAseq was performed as per sections 2.1.6.1 and 2.1.6.3.

The SNA profiling of the sporadic polyps was undertaken at the Sanger Institute in Cambridge as part of a programme of work for the SCORT consortium.

2.3.1.3 Assessment of the expression of the biomarker *AXIN2*

3'RNAseq sequencing raw data was preprocessed, as per 2.1.6.4, to obtain normalised CPM. Normalised CPM data was interrogated for axin 2 (*AXIN2*) expression using

R. The expression of *AXIN2* in the colitis LGD was assessed against that of ligand-independent (LI) sporadic lesions, ligand-dependent (LD) sporadic lesions and the normal samples. Based on the *AXIN2* expression of the sporadic LI and LD lesions, an exploratory cut-off at the bottom quartile of the LI lesions was used to predict ligand dependency in the colitis-associated LGD. Expression was plotted using the R package *ggplot2*.

2.3.1.4 Mutational phenotyping for WNT genes of colitis-associated LGD

For the identification of SNA, a targeted panel was performed by collaborators in Trevor Graham's laboratory at the Institute of Cancer Research on the colitis samples with sufficient DNA available. The Target: Pan-Cancer (524) TMB/MSI Panel (Nonacus) kit with the "non-fragmentation" option for DNA processing was used for library generation, following the manufacturer's instructions. Libraries were sequenced on a NextSeq500 or NovaSeq 6000 system (Illumina).

2.3.1.5 Assessment of *RSPO* expression

The expression of R-spondin 2 (*RSPO2*) and *RSPO3* was assessed on the normalised CPM data. Colitis-associated LGD samples with an *RSPO2* expression within the upper fifteenth percentile (top 15%) and those with *RSPO3* expression within the upper twentieth percentile (top 20%) were considered as having an *RSPO* expression within the upper end of the expression continuum.

2.3.1.6 Validation of *RSPO* expression by fusion panel and *in situ* hybridisation

The *RSPO* fusion panel was performed for samples with a *RSPO2* expression above the fifteenth percentile and samples with a *RSPO3* expression in the upper twentieth percentile, where sufficient RNA was available (details in Figure 4.9 and Figure C.1). This aimed at determining the presence of fusions within this subset of samples. The fusion panel was performed by the Genomics Birmingham core facility. The custom designed Qiagen RNAscan 50 target panel (Qiagen, CFHS-10094Z-50) and QIaseq 96-Index I Set A index (Qiagen, 333727) were used. The Qiagen protocol HB-2170-003-HB-

QIAseq RNAscan Panels-0220-WW was followed and 26 PCR cycles were used. Raw data processing and fusion calling was also performed by Genomics Birmingham.

Following, chromogenic ISH was performed in the same group of samples, where sufficient tissue sections were available. The method outlined in 2.1.4.3 was followed. ACDbio probes Hs-RSPO2 (423991-C1) and Hs-RSPO3-03 (491461-C1) were used.

2.3.2 Assessment of WNT alterations in mouse models of WNT disruption

2.3.2.1 Mouse models

The following models of WNT disruption were used:

- ***ApcVil*** - *Apc$^{fl/fl}$;Vil1-Cre^{ERT2}*

  A tamoxifen-inducible model of *Apc*-mutant colorectal lesions. Upon Cre recombination induced by tamoxifen, all epithelial cells along the crypt-villus axis and the undifferentiated progenitor cells in the crypt region express an Apc protein with greatly impaired function.

- ***Rspo3Vil*** - *Rspo3$^{inv/+}$;Vil1-Cre^{ERT2}*

  A tamoxifen-inducible model of *Rspo3* fusion. Upon Cre recombination induced by tamoxifen, all epithelial cells along the crypt-villus axis and the undifferentiated progenitor cells in the crypt region express Rspo3.

- ***Rspo3Atoh*** - *Rspo3$^{inv/+}$;Atoh1$^{-/-}$-Cre^{ERT2}*

  A tamoxifen-inducible model of *Rspo3* fusion. The expression of the Cre recombinase in this model is driven from the *Atoh1* locus. Upon Cre recombination induced by tamoxifen, all epithelial Atoh1-positive cells express Rspo3.

- ***Rspo3Gli*** - *Rspo3$^{inv/+}$;Gli1-Cre^{ERT2}*

  A tamoxifen-inducible model of Rspo3 stromal overexpression. Tamoxifen-induced Cre recombination causes all stromal Gli1-positive cells to express Rspo3.

- ***tdTomAtoh*** - *$^{td/+}$Tom;Atoh1$^{-/-}$-Cre^{ERT2}*

  The expression of the Cre recombinase in this model is driven from the *Atoh1* lo-

cus. In this model all Atoh1-positive cells express robust tdTomato fluorescence following Cre-mediated recombination.

ApcVil, *Rspo3Vil* and *Rspo3Gli* were previously established and phenotyped models. Further details on the models can be found in Table 2.1.

2.3.2.2 *In situ* hybridisation for *NOTUM*

Sections of *ApcVil*, *Rspo3Vil*, *Rspo3Gli* and *Rspo3Atoh* samples were stained for notum, palmitoleoyl-protein carboxylesterase (*Notum*) by chromogenic ISH, using the ACDbio Mm-Notum probe (428981-C1). The protocol outlined in 2.1.4.3 was followed. Stained slides were imaged with the slide scanner Aperio CS4. Assessment and quantification of *Notum* staining in murine lesions was performed using QuPath. Number of positive cells in 1000 x 1000 μm^2 area were quantified.

Similarly, fluorescent ISH for *NOTUM* was also performed on human precursor lesions with *APC* alteration and *RSPO* fusion, using the ACDbio probe Hs-NOTUM (430311-C1). Slides were imaged on the confocal microscope Leica DMi8 SP8 FALCON.

2.3.2.3 Assessment of the effects of Rspo3 on cell fate determination

A four timepoint experiment was undertaken using *Rspo3Atoh* mice, with *tdTomAtoh* mice as controls to investigate the effects of Rspo3 overexpression on cell fate. Mice >6 weeks in age were injected once with 0.0025 mg of tamoxifen to induce recombination. They were then monitored and euthanized at 30 hours, day 4 and day 30 following injection.

Upon culling, the intestines were collected and processed as per sections 2.1.2.7, 2.1.2.8 and 2.1.2.9. Gut rolls were sectioned for H&E and ISH.

2.3.2.4 Paneth cells and goblet cells characterisation by *in situ* hybridisation and fluorescent immunohistochemistry

ISH for *Rspo3* followed by IHC for lysozymes was undertaken on *Rspo3Atoh* samples using a combined ISH-IHC protocol. Fluorescent ISH was undertaken as per 2.1.4.3, using the mm-Rspo3 (ACDbio, 402011-C1) probe. After the last washing step in the ISH, the sections were rinsed in PBS and then washed in fresh PBS for 5 minutes. Next, the blocking and antibody incubation steps were carried out as outlined in sections 2.1.4.4 of the method for fluorescent IHC. For the lysozyme IHC, the primary antibody Lysozyme (Dako, Rabbit anti-human EC 3.2.1.17 A099) and the secondary antibody Alexa fluorophore 680 (Invitrogen, Goat anti-Rabbit A32734) were used.

Dual-colour lysozyme-tomato fluorescent IHC was undertaken as per 2.1.4.4 on the *tdTomAtoh* samples. For lysozymes, the primary antibody Lysozyme and the secondary antibody Alexa fluorophore 680 were used. For tdTomato, the primary antibody mCherry (NovoBio, Mouse anti-Mouse NBP1-96752) and the secondary antibody Alexa fluorophore 488 (Invitrogen, Goat anti-Mouse A32723) were used.

All primary antibodies were used at 1:500 concentration and all secondary antibodies were used at 1:1000 concentration.

Stained slides were imaged on the Olympus SpinSR Sora microscope. For quantification purposes Z-stack images of 15 randomly chosen quadrants were taken. Images were assessed and quantified on QuPath, specifically 3 quantifications were carried out:

- *Quantification of Paneth cells - defined as lysozyme positive cells.* To determine the number of Paneth cells across the different mice genotypes and timepoints, one villus and its two associated crypts were annotated. Following visual quality control, the number of lysozyme-positive cells and the total number of cells were quantified using QuPath's "Positive cell detection" function.

- *Quantification of goblet cells - defined as number of lysozyme-positive cells above the ischimus in the villi.* To determine the number of goblet cells, one villus was annotated. Following visual quality control, the number of *Rspo3*-positive cells in

Rspo3Atoh and tomato-positive cells in *tdTomAtoh*, and the number of total cells were counted using the "Positive cell detection" function.

- *Quantification of Rspo3-expressing lysozyme-positive Paneth cells and Rspo3-negative lysozyme-positive Paneth cells*. One villus and its two associated crypt were highlighted. Following visual quality control, the number of *Rspo3*-expressing lysozyme-positive cells, *Rspo3*-negative lysozyme-positive cells, and the total number of cells were counted using the "Positive cell detection" function.

Each quantification was performed on five different regions per mouse assessed. Three mice were assessed per timepoint and genotype.

2.4 Chapter 5 specific materials and methods

A study specific outline of the experiments and analyses carried out in chapter 5 is provided here. General methods used are briefly mentioned, with references to the appropriate sections above where the protocols are described in detail. Methods and analyses specific to solely this study are described in detail here.

2.4.1 Sample Selection

IBD patients fitting the criterion outlined below (Table 2.4) were recruited into the study from St Mark's Hospital, London, Oxford University Hospitals national health service (NHS) Foundation Trust, Oxford, Belfast Health and Social Care Trust and Newcastle Upon Tyne Hospitals NHS Foundation Trust.

Table 2.4: IBD patients recruitment criterion

Recruitment Criterion	
Inclusion	Exclusion
Patients with confirmed IBD diagnosis	Patients unwilling or unable to give informed consent
Patients with 5 years follow-up information	
Progressors: patients who progressed to cancer within 5 years from the 1st LGD	Healthy volunteers
Non-progressor: patients who did not progress to cancer within 5 years from the 1st LGD	Patients under the age of 18

All samples were FFPE archival tissue from endoscopic procedures. In addition to tissue, relevant clinical and endoscopic information were collected, including:

- Clinical details: patient age, gender, IBD-duration and PSC-presence;
- Endoscopic details: lesion size, shape and location.

Censor time was defined as the number of days from the procedure at which the LGD tissue samples was obtained until the date of the procedure at which HGD or CRC was first detected for progressors or the date of the last available colonoscopy confirming no progression to HGD or CRC in nonprogressors.

Sporadic precursor lesions were obtained under ethical approval from Belfast Health and Social Care Trust and Newcastle Upon Tyne Hospitals NHS Foundation Trust. All sporadic lesions were collected from individuals without pre-existing bowel conditions.

All samples underwent histological review by expert pathologists prior to enrolment. Specifically, for the colitis lesions, only lesions with confirmed LGD upon review were enrolled into the study.

2.4.2 Histopathological assessment of colitis-associated progressor and non-progressor dysplasia

2.4.2.1 Slides staining and scanning

Tissue sections were cut and H&E stained as per sections 2.1.3 and 2.1.4.1 of materials and methods. They were scanned using the Aperio CS2 scanner at X40 magnification. Slide scans were annotated using ImageScope. Sufficiently large lesions were divided into multiple regions to allow for multiregional phenotyping. Where lesions were divided into multiple regions, each region was annotated on a separate file. The annotated files were uploaded into the web image viewer SlideScore for assessors to access. All regions were anonymised, renamed and randomised, with each pathologist viewing the regions in a different order.

2.4.2.2 Assessors and assessments

The study was undertaken with a panel of six expert and dedicated pathologists. They had been actively practising and dedicated to the field of colorectal dysplasia for a minimum of ~ 10 years. The panel was multicentred and international: three pathologist practising in England, one in Northern Ireland, one in Switzerland and one in Singapore. Each pathologist assessed all regions independently and scored the regions using the web questionnaire provided by the viewer; no discussions were held between pathologists. Each pathologist was asked to independently score a region's class, subclass and grade, following the Choi *et al.* [141] criteria (reported in Table 2.5). The option to score as no dysplasia, where no dysplasia was seen was also provided. Both scoring and grading was solely based on the assessment of the annotated area (i.e. region). The outcome measures were:

1. Inter-rater agreement in scoring a region's class, subclass and grade, measured using percentage agreement and kappa value;

2. The percentage of progressor regions with diagnosis of nonconventional dysplasia;

3. The percentage of progressor regions with high-risk histology diagnosis.

Notably, the assessment and downstream analysis was performed on a per-region basis. At no point were the values denotated back to per-lesion or per-patient.

2.4.2.3 Assessment of diagnostic agreement

The inter-rater agreement was measured in percentage agreement and kappa values. For each region, the class, subclass and grade scoring of the pathologists was compared, and the following were determined:

- The agreement score of each pathologist, by calculating the mean of the percentage agreement of each pathologist across all assessed regions (referred to as agreement score);

Table 2.5: Choi *et al.* morphological criteria of nonconventional dysplasia

	Hypermucinous type	*Intestinal type*		*Crypt cell type*	Serrated type		
	Hypermucinous	*DPD*	*GCD*	*CCA/D*	*TSA-like*	*SSL-like*	*Serrated NOS*
Architecture	Tubulovillous/villous	Tubular	Tubular	Flat	Tubullovillous/villous with serrattion	Tubular with serration	Tubular with serration
Defining features	Tall mucinous cells with typically mildly elongated, hyperchromatic nuclei Hypermucinous component should represent >50% of the lesion	Intestinal type cells with mostly elongated, hyperchromatic nuclei Increased Paneth cell differentiation involving at least two contiguous crypts in two different foci (beyond what is present in background mucosa)	Intestinal type cells with mostly elongated, hyperchromatic nuclei Complete or near-complete absence of goblet cells	Mostly round-to-oval, non-stratified nuclei Atypia can be limited to the crypt base without obvious surface involvement	Columnar cells with mostly elongated nuclei, intensely eosinophilic cytoplasm, and ectopic crypts TSA-like component should represent >50% of the lesion	Prominent serration and dilation at the crypt base and surface, including dilated L- or inverted T-shaped crypts at the interface with muscularis mucosa SSL-like component should represent >50% of the lesion	Often complex serration but without definite features of TSA or SSL Serrated lesion NOS component should represent >50% of the lesion
Other Features	Degree of atypia tends to decrease from the crypts to the surface of the villi	Some loss of goblet cells allowed, but no complete or near-complete absence of goblet cells	Scattered Paneth cells allowed, but not in multiple clusters of dysplastic crypts as seen in DPD Similar degree of Paneth cell differentiation is present in the lesion and in adjacent mucosa	Some loss of goblet cells allowed, but no complete or near-complete absence of goblet cells Occasional Paneth cells may be seen, but similar degree of Paneth cell differentiation is present in the lesion and in adjacent mucosa		May show dysplasia, which may be confined to the lower portion or involve the entire thickness of the mucosa	May show dysplasia, which may be confined to the lower portion or involve the entire thickness of the mucosa 'HP-like' changes may be present in adjacent mucosa or represent the entire lesion

- Study percentage agreement, determined by first calculating the percentage agreement across pathologists for each region and then taking the mean of such scores (referred to as study inter-rater agreement);
- Cohen's kappa: Cohen's kappa between two pathologists at each time;
- Fisher's kappa.

All calculations were performed on R.

2.4.2.4 Assessment of association between nonconventional dysplasia and progression

To assess the correlation between nonconventional dysplasia and progression, two approaches were taken:

1. Regions with a majority agreement (i.e. >4 pathologists agreed on a particular diagnosis) were selected and the number and percentage of conventional, nonconventional and mixed regions which were progressor/nonprogressor were determined. Fisher's exact test between conventional and nonconventional dysplasia was then performed on R.
2. Scores of mixed dysplasia (i.e. lesions with mixed conventional and nonconventional) were first denoted as nonconventional to focus on the worst histology visible, as per common practice in histology, which according to the Choi WT *et al.* [141] paper is nonconventional dysplasia. Regions with majority agreement were then selected and the number and percentage of conventional/nonconventional regions which were progressor/nonprogressor were determined and Fisher's exact test was performed.

To examine the correlation between high-risk histological subclass as defined by Lee *et al.* [142] and progression, two analyses were performed on regions with a majority diagnosis >4. In the first analysis, the scores were assessed without prior denotation of the mixed dysplasia, whilst in the second analysis scores of mixed dysplasia were denoted to the worst histology amongst the two reported. For example, for a

region where hypermucinous and TA were reported, the region was denoted to being hypermucinous, for the same rationale as above. For both analyses, the number and percentage of progressor and nonprogressor regions with a high-risk/low-risk histology were calculated and t-tests were performed.

2.4.3 Molecular assessment of colitis-associated progressor and nonprogressor dysplasia

2.4.3.1 DNA isolation

FFPE blocks were sectioned for DNA extraction as per section 2.1.3. Sufficiently large lesions were divided into multiple regions prior to tissue processing to allow for multiregional sequencing. Epithelial enrichment and DNA isolation were performed as 2.1.5.1 and 2.1.5.2.

2.4.3.2 Sonication of lowly fragmented DNA for lpWGS library preparation

Prior to library preparation, sonication was performed for samples with low DNA fragmentation with the Covaris SonoLab 7.2 machine and microTUBE-15 AFA Beads Screw-Cap (Covaris, 520145). The programme below was used to obtain a mean fragment size of 200-300 bp, ideal for library preparation for Illumina sequencing.

Table 2.6: Sonication programme

Temperature		
Min	Setpoint	Max
18°C	20°C	22°C
Treatment time: 28 seconds		
Peak Power	Duty Factor	Cycles/Burst
30	20	50

2.4.3.3 lpWGS library preparation, pooling and sequencing

Library preparation for low pass whole genome sequencing (lpWGS) was undertaken using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs, E7103). The protocol provided by the manufacturer was followed with minor

modifications. These modifications were:

- Undertaking a DNA repair step with the NEBNext FFPE DNA Repair kit (New England BioLabs, M6630L);
- Halving the volume of all recommended reagents for cost-effectiveness;
- Skipping the DNA fragment size selection step prior to adaptor addition;
- Optimising adaptor concentration and number of PCR cycles to use to quantity of input DNA.

Between 4-100 ng of input DNA was used. Shortly, the input DNA was initially repaired for damage caused by fixation and storage typical of FFPE samples. Following, end repair, 5′ phosphorylation and dA-Tailing was performed, before adaptor ligation. The adaptor ligated libraries were cleaned through a bead-clean up using Agencourt AMPure XP (Beckman Coulter, A63881) and barcoded using the NEBNext Multiplex Oligo for Illumina kit - Dual index primer set 1 (New England BioLabs, E7600S). The libraries were amplified by PCR; 7-12 cycles were used based on the amount of input DNA, to generate libraries of at least 50 ng. Lastly, the PCR products were subjected to a bead clean-up, before being quality checked.

The final library quantity was assessed by Qubit Fluorometer (ThermoFisher Scientific, Q32854). Library quality was assessed by Agilent Tapestation using the High sensitivity D1000 DNA system. Libraries greater than 184 bp in size, with unbound adaptor comprising less than 10% of the final library and of at least 50 ng passed quality check. Libraries with unique and compatible indexes (as outlines by the manufacturer) were combined by equimolar pooling to a final volume of 100 µL. Pools of 28-32 samples were prepared. Libraries were sequenced either by the Oxford Genome Centre or the Barts Cancer Institure Genome Centre on Illumina's NextSeq 500 (Mid-Output Run, 75 bp, paired end reads), aiming for an average mean depth of 0.1x per sample after initial processing.

2.4.3.4 lpWGS raw data processing

LpWGS data processing and copy number (CN) profile generation was performed by Dr Kathleen Curtius on the QMUL High Performance Computing Cluster. Raw FastQ files (four per end read per sample) were merged and assigned new read groups. Reads in the generated files were subjected to adaptor trimming using the package Skewer. Trimmed reads less than 35 bases pairs in size were filtered out. FastQC was used to generate a quality control report, which was reviewed. Reads passing initial quality check were aligned against the reference genome hg39 using the package Burrows-Wheeler Aligner. The resulting sequence alignment map (SAM) files were sorted to produce binary alignment map (BAM) files, following removal of poorly mapped reads using the package SAMtools. This was followed by the removal of duplicate reads using Picard Tools. For quality check purposes, BAMQC was used to generate a quality control report for each BAM file and GATK is used to estimate the mean genome coverage.

2.4.3.5 Analysis of copy number alterations

Following preprocessing of lpWGS raw data, BAM files were processed on R with the software package QDNASeq using a pre-established pipeline. Briefly, BAM reads were arranged into bins of 500 kbp using QDNAseq. This was followed by correction for GC content and filtering of backlisted regions and loess residuals. Sex chromosomes were excluded from downstream analysis. The remaining bins, i.e. 4401 bins of 500kbp, were normalised against the median, log transformed and segmented using DNAcopy. Each bin was assigned a call of gain (value ≥ 1), loss (value ≤ 1) or no change (value 0) relative to median copy number of the sample.

Downstream analysis consisted in analysing and comparing the following between samples and specifically between patient groups:

- Genomic aneuploidy: proportion of bins with a CN call;
- Number of chromosomal CNA events: number of individual segments with a CNA call;

- Chromosomal arm loss of gain of interest: a chromosomal arm gain or loss of interest was considered to be present if a segment(s) within that arm have been annotated to have a gain, loss, deletion or amplification by CGHcall;
- Frequency of chromosomal gains and/or losses.

Statistical comparison of these measurements were made on R using a two-sided unpaired t-test (for genomic aneuploidy and number of chromosomal CNA) or Fisher's exact test (for chromosomal arm loss of gain of interest), with a p-value of 0.05 used as a cut-off for significance. On analyses with multiple comparisons, p-values were adjusted with Benjamini-Hochberg correction. Plots were generated using the packages vioplot and *ggplot2* on R.

CNA analysis was carried out on a discovery and a validation cohort. The discovery cohort comprised samples collected from London North West Healthcare NHS Trust. The second independent validation cohort, comprised samples collected from Oxford (Oxford Hospitals NHS Foundation Trust), Newcastle (Newcastle Upon Tyne Hospitals NHS Foundation Trust) and Belfast (Belfast Health and Social Care Trust).

CNA burden comparison between nonprogressor and progressor patients and survival analysis, was performed on per patient basis, i.e. one sample per patient. For patients where samples from multiple regions and/or multiple timepoints were profiled, the region with maximal number of chromosomal CNA at the first available timepoint was taken forward.

A univariate cox regression analysis of clinical and endoscopic features, and genomic findings was performed to determine which features are predictive of progression. This was followed by a multivariate cox regression analysis. In both cases, the R package *survival* was used. Both analyses were performed on samples from the discovery cohort.

To determine whether CNA can predict cancer risk and therefore assess its utility as a cancer risk stratification tool, a Kaplan-Meier survival analysis was performed. For this, all patients were first ranked by number of abnormal copy profile segments, and patients in the upper quartile were then compared to the rest of the patients. Sur-

vival differences from the Kaplan-Meier analysis were evaluated using a log-rank test. Patients with LGD were censored when HGD or CRC was first detected or at 5 years from LGD detection.

The distribution of individual CN gains and losses was reviewed on a per-patient basis to determine whether specific CNAs were associated with the LGD phenotype and with progression. First, the CNAs commonly occurring in both progressor and nonprogressor LGD were assessed. Then the CNAs more commonly occurring in progressor LGD than nonprogressor LGD were determined. Lastly, an investigation of common CRC genes and genes commonly mutated in CA-CRC located within the chromosomal regions significantly more likely to be gained/lost in progressor LGD was carried out. The list of commonly CA-CRC genes was obtained from Baker *et al.* [96]. The list of driver colorectal cancer genes included common CRC driver genes, genes in the COREAD driver gene list and genes of interest to our laboratory.

The univariate cox regression analysis was performed by Dr Ibrahim Al-Bakir, whilst the CNA burden analysis and survival analysis were performed by Dr Kathleen Curtius.

2.4.4 Digital cytohistological assessment of colitis-associated progressor and nonprogressor dysplasia

2.4.4.1 Haematoxylin staining and scanning

Tissue sections were haematoxylin stained following the method outlined in section 2.1.4.2. Stained slides were scanned using a NanoZoomer S210 Digital slide scanner at X40 magnification (Hamamatsu). Slide scans were digitally annotated using ScanScope as per the annotations used in strand A and strand B of the study. As per strand A, the assessment was undertaken on a per-region basis.

2.4.4.2 HALO analysis

Digital analysis of the slide scans was undertaken by Dr Viktor Koelzer using the platform HALO. Each annotated area (i.e. region) was first segmented according to tissue

type into epithelial and stromal compartments. Visual control of the tissue segmentation was undertaken. Staining intensity for haematoxylin, nucleus size and nucleus roundness were recorded using a Cytonuclear v1.4 algorithm.

2.4.4.3 HALO data analysis

R was used to analyse the data generated by HALO and determine whether a progressor and nonprogressor LGD can be discerned on the basis of one or more nuclear features. Specifically, the following analyses were performed:

- Exploratory analysis

    1. PCAs of all tissue components: performed using average tissue values (i.e. summary data) and single cell values (i.e. single cell data) for all three tissue components, epithelium, lymphocyte-rich stroma and lymphocyte-poor stroma;
    2. PCAs of the epithelial component: performed on average tissue values and single cell values specific to only epithelial cells.

- Hypothesis driven analysis

    1. Using average tissue values: average nucleus area, nucleus roundness and nuclear staining intensity of the epithelial compartment were compared between nonprogressor and progressor regions. Violin plots were generated and T-test performed on R. P-values were adjusted using the Bonferroni method;
    2. Using single cell values: epithelial cells in the top quartile based on nucleus area, nucleus roundness and nuclear staining intensity were compared between nonprogressor and progressor regions. Focusing on one feature at a time, all epithelial cells within each region were ranked and those within the top quartile for that feature were subset. The mean value of the top quartile cells was calculated for each region. The means for the nonprogressor regions and the progressor regions were compared using violin plots

and t-statistics. P-values from the t-tests were adjusted using the Bonferroni method;

3. Using the molecular data and single cell values: epithelial cells in the top quartile based on nucleus area, nucleus roundness and nuclear staining intensity were compared in the worse nonprogressor and progressor regions based on CNA. For this analysis, the region with the highest CNA burden per-patient were selected. Focusing on one feature at a time, all epithelial cells of each selected region were ranked and the cells within the top quartile for that feature were subset. The mean value of the top quartile cells was calculated for each region. The means for the nonprogressor region and the progressor region were compared using violin plots and t-statistics. P-values from the t-tests were adjusted using the Bonferroni method.

Chapter 3

The stem cell population dynamics in sporadic and colitis-associated colorectal cancer

This chapter investigates the stem cell dynamics in S-CRC and CA-CRC in human and murine colorectal cancer lesions.

3.1 Introduction

During homeostasis, perpetual self-renewal of the intestinal epithelium is maintained by the CBC cells located within the stem cell niche. CBC are also known as LGR5-positive cells due to their high expression of LGR5, a specific marker of this stem cell population. CBC are WNT/β-catenin signalling dependent long-lived stem cells that maintain the stem cell pool. They are able to produce all mature lineages of intestinal epithelia to replace the mature cells found at the tip of the villous. Notably, only the stem cells in the crypt retain long-term self-renewing ability [3]. Upon cell intrinsic or extrinsic perturbation of homeostasis, for instance following cytotoxic damage or surgical resection, the epithelium shows remarkable regenerative capacity. However, despite their pivotal role in the maintenance of homeostasis, CBC cells play a limited role in regeneration. The identity of the cells and the biological processes underlying

regeneration have been an area of extensive investigation and discussion (Figure 3.1). Early work by the Potten group showed that the gut could recover from radiation-induced damage that eliminated CBC cells almost completely. Originally, it was hypothesised that quiescent cells at the +4 region of the crypt survived damage through their non-proliferative nature and later contributed to regenerating the epithelium by starting to cycle as a response to that same damage. However, the view of this cell population as quiescent has been challenged by recent data.

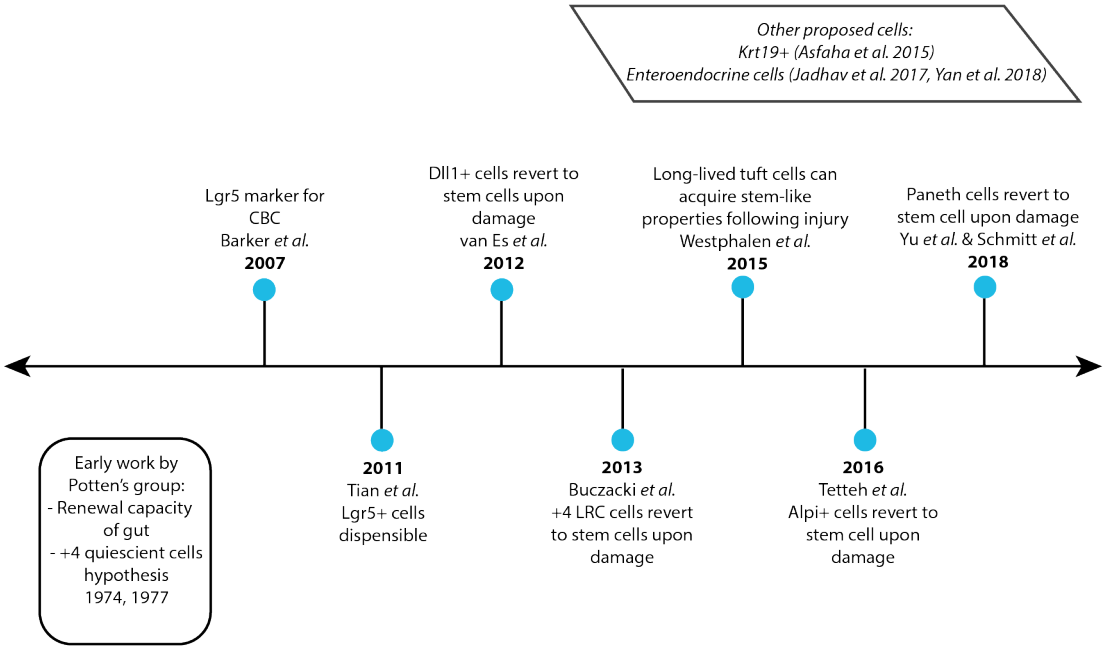


Figure 3.1: A historical timeline of the intestinal stem cell debate

In 2011 Tian *et al.* demonstrated that Lgr5-positive cells can be efficiently replenished following genetic ablation by cells higher up in the crypt, proving that Lgr5-positive cells are dispensable in regeneration/radiation response. This was the first description of cell fate plasticity in the intestine following damage to the stem cell compartment [122]. Since then, numerous studies have shown that upon ablation of Lgr5-positive cells a range of progenitor and even fully differentiated cell types are capable of undergoing dedifferentiation to re-acquire stem cell potential and regenerate the intestinal epithelium [12, 30, 143]. These include: delta-like1 (Dll1)-positive secretory cells [144], +4 label-retaining cells (LRC) [145], doublecortin like kinase 1 protein (Dlck1)-positive cells [146], alkaline-phosphate intestine (Alpi)-positive cells [147] and

Paneth cells [148, 149] (Figure 3.2).

Functional studies of cells expressing Dll1, also known as Dll1-positive secretory precursors, have revealed that these cells can contribute to regeneration. Lineage-tracing in Dll1-Cre mice has shown that under homeostatic conditions Dll1-positive secretory cells give rise to short-lived secretory clones. However, following damage many of these cells were found to re-acquire stem cell potential and generate long-lived clones of both absorptive and secretory lineages. This reacquisition of stemness of the Dll1-positive cells, which are immediate descendants of Lgr5-positive stem cells, was found to be associated with their renewed proximity to Paneth cells, known source of WNT, Notch and EGF signals [144].

Lineage-tracing studies have allowed for the identification and further study of a label-retaining stem cell population, which could putatively function as reserve stem cells. These cells, historically believed to be quiescent, reside at the +4 position of the crypt, express polycomb complex protein 1 (Bmi1), mTert, leucine-rich repeats and immunoglobulin-like domains 1 (Lrig1) and Hopx markers and display higher tolerance to radiation and WNT signalling deprivation. During homeostasis, they are short-lived precursors of Paneth and enteroendocrine cells. However, upon regenerative stimuli these alternative stem cells can re-enter the cell cycle and repopulate the epithelium [145].

A further population of cells implicated in regeneration following intestinal and colonic injury are Dlck1-positive tuft cells. This subpopulation of Dlck1-positive cells is a long-lived, post-mitotic population, with stem cell capacities. Dlck1-positive tuft cells, which are enriched in inflammation-induced carcinogenesis, were found to be activated during epithelial response to injury [146].

High regenerative capacity was also demonstrated for Alpi-expressing enterocyte progenitors. These are the most abundant and proliferative cells in a crypt, comprising the bulk of the transit-amplifying zone. Genetic fate-mapping of Alpi-positive cells before and upon loss of Lgr5-positive stem cells showed their dedifferentiation into cells exhibiting features of self-renewal and multipotency. The resulting dedifferentiated cells were observed to give rise to proliferative Paneth-like cells. Furthermore, their

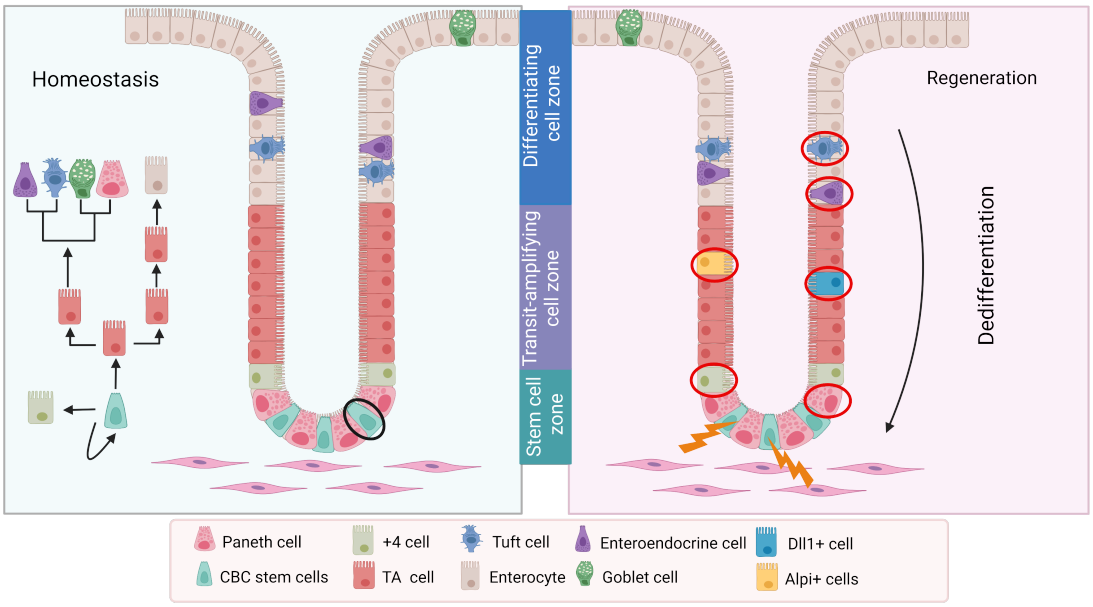


Figure 3.2: The stem cell dynamics under homeostatic conditions and during regeneration. During homeostasis, perpetual self-renewal is maintained by the CBC cells (circled in black). However, upon injury, a number of differentiated cells have been shown to undergo dedifferentiation to enable regeneration of the intestinal epithelium (circled in red).

dedifferentiation was associated with a downregulation of enterocyte specific genes expressed during homeostasis, and upregulation of Lgr5-positive stem cell specific genes [147].

Similarly, Paneth cells were reported to participate in intestinal regeneration following stem cell depletion. They are induced to mature by canonical WNT signalling and presumed to lack self-renewal capacity during homeostasis. Lineage studies demonstrated that irradiation induces a subset of Paneth cells to dedifferentiate, acquire stemness, undergo transient proliferation and re-differentiate. This was achieved through activation of Notch signalling [148].

Altogether, the studies mentioned above and others in the literature provide compelling evidence in support of the plasticity of various committed progenitor cells being an important means by which the intestinal epithelium copes with damage. However, the molecular mechanisms regulating cell plasticity are yet to be fully elucidated [12]. As of now, the ability of mature cells to dedifferentiate or switch lineages and the capacity of stem cells to maintain their stemness is known to be strongly dependent on the niche within which they are located and its associated signals [150]. Yui *et al.*'s

molecular investigation of colonic epithelial repair in both human and mouse revealed that tissue regeneration following severe injury is mediated by dedifferentiation and adaptive reversion of the remnant repairing epithelium to a primitive molecular phenotype characterised by the expression of foetal intestinal genes. Transient activation of epithelial YAP1 signalling was identified as the driver of tissue reprogramming into a highly proliferative epithelium. Specifically, in the model proposed by Yui *et al.*, damage-induced extracellular matrix remodelling initiates downstream signalling of the mechanosensory pathway, which entails activation of the focal-adhension pathway, cytoskeletal re-arrangement and YAP1/TAZ activation [30].

Neoplasia is thought to be driven by the activity of stem cells. The cancer stem cell theory suggests that a small number of dedicated stem cells drives tumour growth. These stem cells, also referred to as cancer stem cells, are a subpopulation of cancer cells capable of undergoing self-renewal, infinite cell division and multi-directional differentiation. One of the key features of cancer stem cells is the abnormal regulation of major pathways involved in stemness and proliferation and, unlike normal stem cells, cancer stem cells lack the necessary control system to prevent uncontrolled proliferation [151]. Cancer stem cells were first described in the context of myeloid leukaemia, but have since been identified in different cancer types [152–155]. In CRC, cancer stem cells are major players in tumorigenesis, metastasis and relapse. They are generally thought to originate from either adult intestinal stem cells or differentiated cells through gains of oncogenic genetic alterations [156, 157].

The first cancer stem cell population to be described in the context of CRC was the long-lived Lgr5-positive cells. Lgr5-positive cells were noted to be upregulated in some but not all CRC. To directly test CBC cells as the cell of origin, Barker *et al.* inactivated *Apc* via biallelic deletion in Lgr5-positive cells. This resulted in a bottom up tumorigenesis, characterised by the formation of microadenomas in mice, which rapidly progress to large tumours, suggesting a role of CBC as cancer-initiating cells [158]. Importantly, LGR5-positive cells are thought to be the cancer stem cells of most S-CRC.

More recently, Lgr5-negative cells have been proposed as cancer stem cells in CRC.

Specifically, in murine based studies, Schwitalla *et al.* found that inflammation and the resulting disrupted homeostatic morphogen gradient promote neoplastic growth from Lgr5-negative cells. Further to these observations, Westphalen found that Dlck1-positive cells became potent tumour initiating cells upon loss of *Apc* in an inflammatory setting [146]. Similarly, Gremlin 1 disruption was found to initiate cancer in ectopic crypts in Lgr5-negative cells away from the crypt base [159]. Furthermore, de Susa e Melo *et al.* observed that depletion of Lgr5-positive cells did not affect tumour growth, while Fumagalli *et al.* found metastatic dissemination and growth to be mostly driven by Lgr5-negative cells [105, 160, 161]. Comprehensively, these data demonstrate that Lgr5-positive cells are not present or apparently obligatory in all cases of CRC, supporting a role for an alternative, Lgr5-negative stem cell population. This is further supported by human data reporting minimal LGR5 expression in a subgroup of CRCs and more recently that serrated polyps arise from differentiated cells through a dedifferentiation process denominated as "gastric metaplasia" by the authors [64].

In the context of CA-CRC, the observation that inflammation and disrupted morphogen gradients can lead to neoplastic growth from Lgr5-negative cells and that CA-CRC arises in a field of inflammation as a consequence of chronic inflammation suggests Lgr5-negative cells as the potential cancer stem cells of colitis-associated neoplasia.

3.2 Study rationale, design study and aims

The cellular plasticity of the gut's epithelium is pivotal in regeneration and malignancy. Enhanced cellular plasticity is an important response mechanism against damage from intrinsic and extrinsic factors, including regeneration following active inflammation induced damage in IBD patients. However, intestinal cell plasticity plays an important role in CRC transformation, progression and adaptive response to therapy. Identification and characterisation of a lesion's stem cell phenotype can enable elucidation of the cellular processes that become disrupted during cancer development. This in turn can aid in the development of preventive and therapeutic measures.

The stem cell phenotype of tumours can also be used to track tumour cell adaptive changes. Starting from two main stem cell populations, the Lgr5-positive (referred to as crypt base columnar (CBC)) and the Lgr5-negative (referred to as regenerative stem cell (RSC)):

- A molecular and morphological analysis of the stem cell dynamics of human CRC lesions and mouse models of CRC was undertaken, with aims to:
    1. Assess the stem cell molecular phenotype of human pre-malignant and malignant lesions, with particular focus on colitis-associated lesions;
    2. Assess the stem cell molecular phenotype of a range of mouse models and derived organoids.
- An assessment of the factors that might influence phenotypic plasticity was undertaken to:
    1. Determine the factors that influence phenotypic plasticity;
    2. Determine the stem cell dynamics of CRCs, including CA-CRC.
- The stem cell response of tumours to selective pressure was assessed, with aims to:
    1. Determine whether the stem cell phenotype is fixed or dynamic;
    2. Determine the implications that stem cell plasticity has on response to treatment.

3.3 Results

3.3.1 Sample selection

The study was conducted on multiple human and murine cohorts. The former comprised human pre-malignant and malignant lesions, as well as human colon scRNAseq, data, data from the FOxTROT clinical trial and data from the BOSS study. The murine cohort included multiple murine CRC models established in the Leedham laboratory

in Oxford and in the Sansom laboratory at the Beatson Institute for Cancer Research. Details on both can be found in the materials and methods chapter (2.2.1).

3.3.2 Derivation of stem cell molecular signatures and markers

To assess the stem cell phenotype of the human and murine lesions, molecular signatures of the CBC cells and RSC were first identified in the literature. For the CBC stem cell gene signature, the "Lgr5 intestinal stem cell signature" reported in Muñoz *et al.* 2012 was used. This signature was originally derived by transcriptomically profiling FACS-sorted Lgr5-positive stem cells and their daughters [132]. The RSC signature was generated by aggregating and curating the following two molecular signatures:

- "Repairing epithelium gene list" [30], derived by transcriptomically profiling FACS-sorted cells from repairing epithelium;
- "Spheroid upregulated gene list" [133], derived by profiling a newly identified progenitor cell population with foetal features.

The curation of the combined signatures aimed at removing genes not contributing to the signature (details in materials and methods 2.2.2) and resulted in a RSC gene signature comprising 265 genes. This is reported in appendix Table B.1.

Using the molecular signatures, the stem cell index was established as a value to summarise stem cell behaviour within a particular sample along a spectrum from conventional CBC on one end to RSC on the other end. The stem cell index was calculated on transcriptomic data by first determining a sample's enrichment score for the RSC and CBC signatures, using the gene enrichment analysis method GSVA [134], followed by the subtraction of the CBC score from the RSC score (detailed in materials and methods 2.2.5).

The CBC and RSC signatures were interrogated to identify molecular markers that could be used for on-slide assessment. For the CBC, *LGR5* was selected, as it is a well-established and vastly used marker for CBC [158]. For the RSC, *ANXA1*, a widely used regenerative stem cell marker, was selected as primary marker for assessment in both human and mice [30, 143]. Moreover, two secondary markers: *Ly6a* and plasminogen

activator, urokinase receptor (*PLAUR*) were also chosen. *Ly6a* has previously been described as marker for RSC [30], but can only be used for mice as there is no human ortholog of the gene, while *PLAUR* was identified from the curated RSC gene list.

Subsequently, the expression of the CBC and RSC markers was assessed in the normal cell compartment by ISH and analysis of human scRNAseq data. Under homeostatic conditions, *LGR5* expression was found to be specific to epithelial cells located at the base of the crypt in both human and murine tissue by ISH. No epithelial homeostatic expression of *ANXA1* was seen in normal human colon and murine small intestine, although low expression was detected in the distal colon of the mice. However, non-epithelial expression of *ANXA1* could be seen on-slide in both human and murine normal intestinal tissue. Similar expression patterns were seen for *Ly6a* in mice, whereas no expression of *PLAUR* was detected in normal human colon (Figure 3.3 A).

Findings from the ISH for the stem cell markers *LGR5* and *ANXA1* were confirmed by the assessment of the scRNAseq data of normal cells, which showed *LGR5* expression to be epithelial-specific, while *ANXA1* expression was found to be absent in normal epithelial cells, but present in stromal, myeloid and T-cells (Figure 3.3 B). Furthermore, cell-specific assessment of *LGR5* and *ANXA1* expression in the scRNAseq data of human cancers showed an inversely proportional enrichment of these individual cell markers in the cancer epithelial cell compartment. This suggests that individual tumours may be variably skewed towards predominantly CBC or RSC stem cell phenotypes (Figure 3.3 B). Interestingly, the application of the full CBC and RSC molecular signatures in human normal tissue and cancer scRNAseq data also led to the separation of discrete stem cell populations, with only a small number of cells expressing both stem cell signatures (Figure 3.3 C).

Together these data showed that the stem cell molecular signatures and morphological markers selected can be used for the assessment of the epithelial stem cell population dynamics in both human and murine intestinal epithelium.

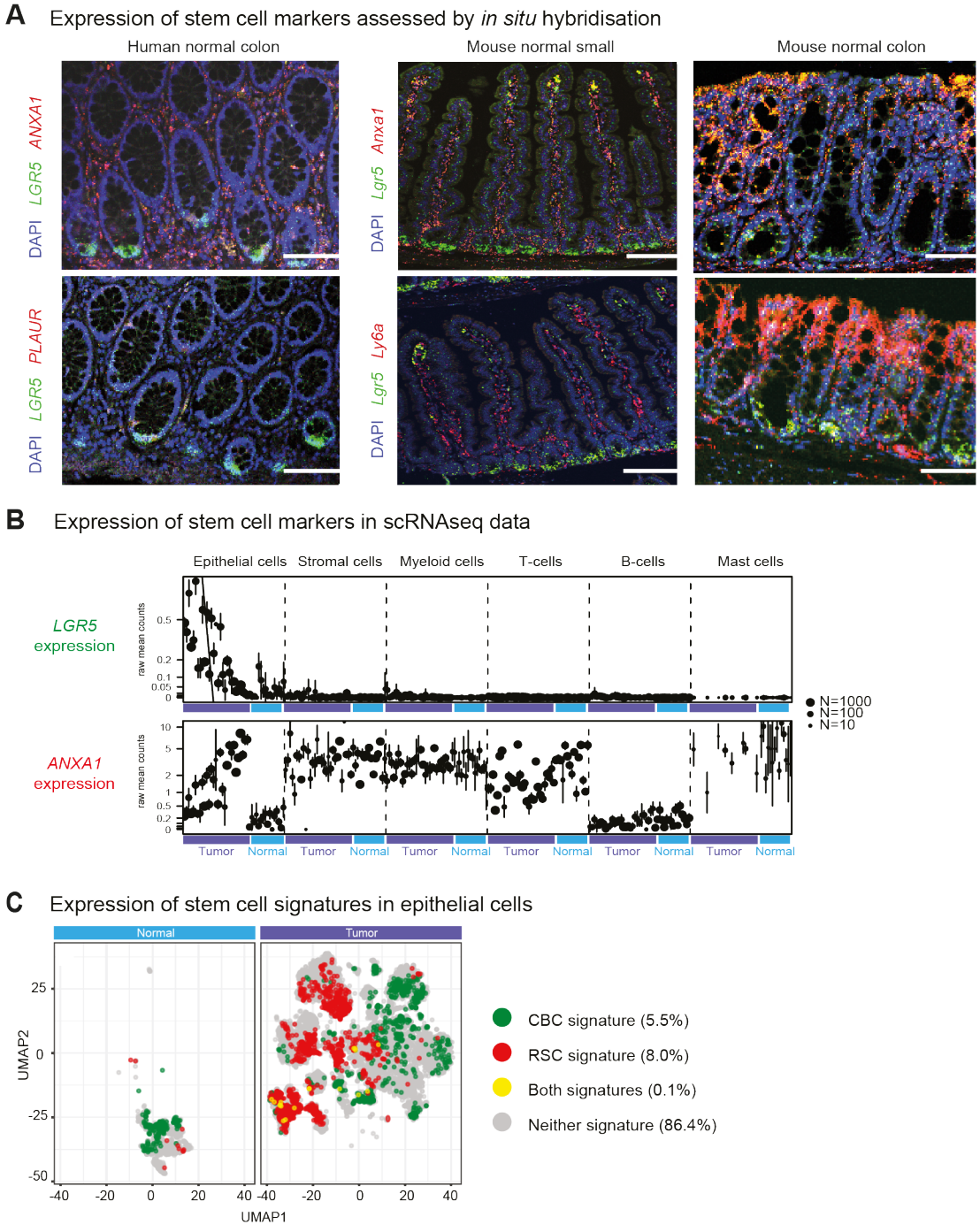


Figure 3.3: Molecular and morphological assessment of stem cell markers. **A.** Fluorescent ISH investigating the expression pattern of CBC and RSC cell markers in normal human and murine intestinal tissue. Scale bars, 100 µm. **B.** Distribution of human multicompartment scRNAseq expression of *LGR5* and *ANXA1*, CBC and RSC markers respectively, in normal (cyan) and cancer (purple) cell compartments. **C.** UMAP plot of normal and CRC single epithelial cells showing enriched CBC (green), RSC (red) and mixed CBC/RSC (yellow) gene expression. Cells with no enriched stem cell signature expression are shown in grey.

3.3.3 Stem cell marker expression across human premalignant and malignant subtypes

To determine the stem cell phenotypic landscape of CRC, the stem cell index and morphological markers were applied to human pre-malignant and malignant lesions. Firstly, the stem cell index was investigated in histological subtypes of sporadic precursor lesions (i.e. TVA, TSA and SSL) and colitis-associated LGD. Application of the stem cell index to bulk transcriptomic data from pre-malignant lesions showed differential enrichment of the two stem cell populations across the subtypes (Figure 3.4 A). Specifically, TVAs, which are described as the conventional type of adenomas arising and progressing following the adenoma-carcinoma sequence [162], were found to be enriched for CBC stem cells. On the contrary, TSAs and SSLs, which follow the serrated pathway [162], were enriched for RSC. Similarly to the serrated lesions, colitis-associated LGD were also enriched for RSC. The stem cell expression pattern observed in the bulk transcriptome data was confirmed by ISH, with differential epithelial marker *LGR5* and *ANXA1/PLAUR* expression being observed across representative lesions (Figure 3.4 B).

Next, the stem cell dynamics in CMS and CRIS CRC molecular subtypes and CA-CRC was assessed. Differential enrichment of stem cell populations was also demonstrated by assessment of both bulk transcriptome and ISH for CRC subtypes. CMS2 and CRIS-C,D and E cancer subtypes showed a predominantly CBC expression, while CMS4, CRIS-A,B and CA-CRC showed a predominantly RSC expression (Figure 3.5).

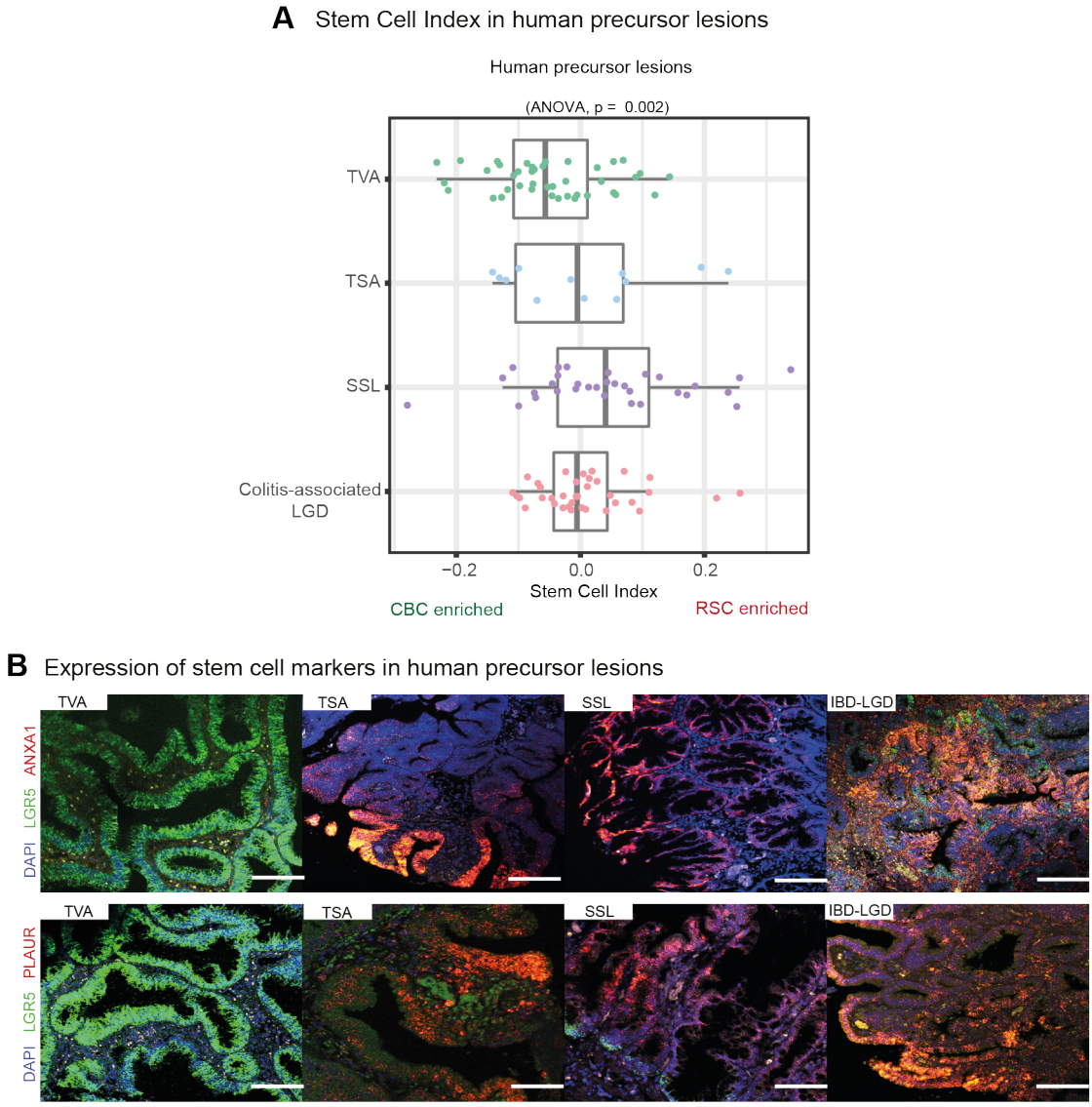


Figure 3.4: Application of stem cell index and markers to human precursor lesions. **A.** Stem cell index in human sporadic TVA, TSA, SSL and colitis-associated LGD. The stem cell index is laid out along the x-axis, with the left side (negative stem cell index) representing CBC enrichment and the right side (positive stem cell index) representing RSC enrichment. **B.** Dual colour fluorescent ISH for *LGR5* and *ANXA1* (Top) and for *LGR5* and *PLAUR* (Bottom) in representative human precursor lesion subtypes. Both assessments show a skew of TVA towards the CBC phenotype, and TSAs, SSLs and IBD-LGDs towards the RSC phenotype. Scale bars, 100 µm. Complete statistics in appendix Table B.2.

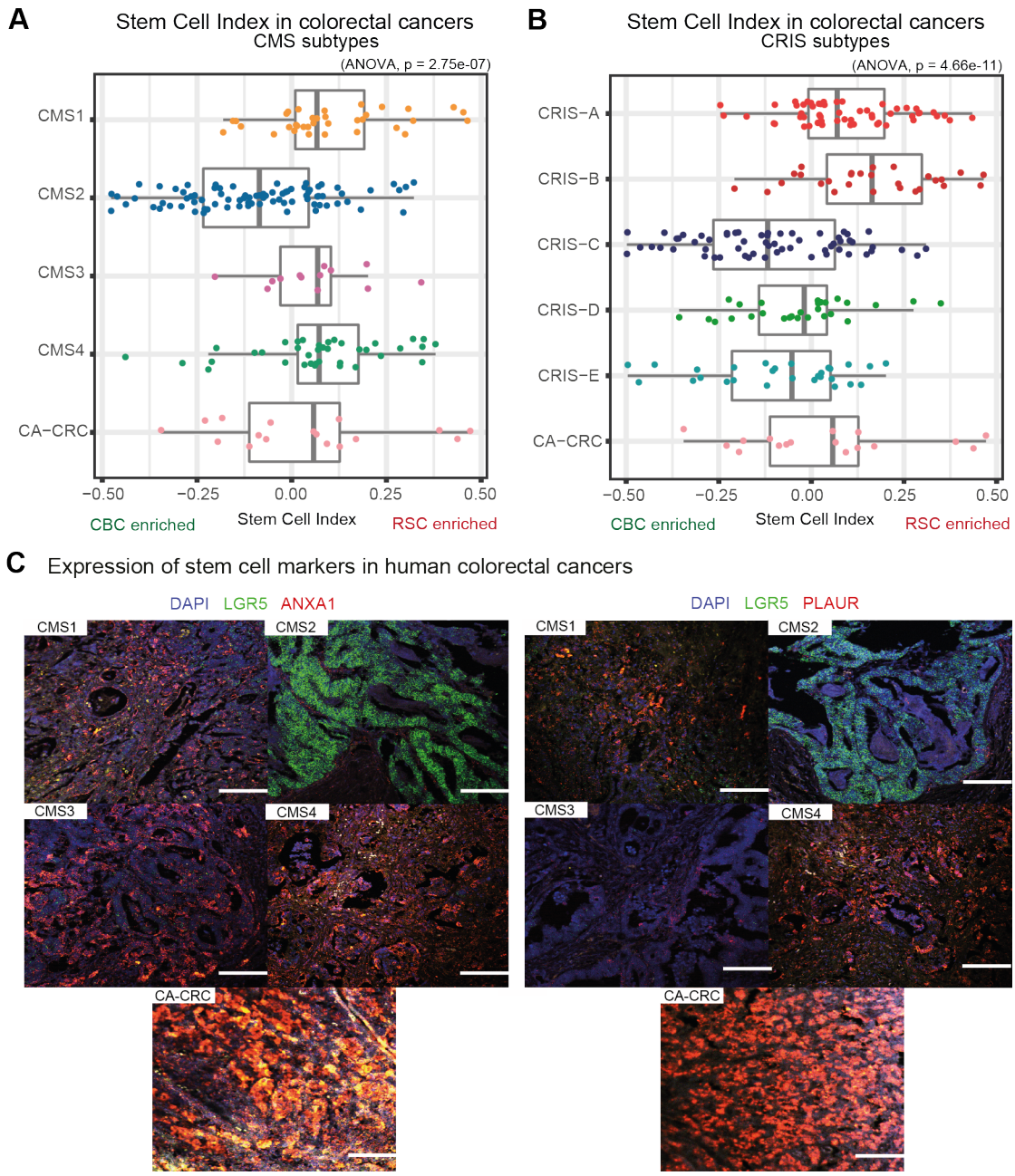


Figure 3.5: Stem cell index and expression in human malignant lesion subtypes. **A.** Stem cell index in CRC CMS subtypes and CA-CRC. **B.** Stem cell index in CRC CRIS subtypes and CA-CRC. **C.** Dual colour fluorescent ISH for *LGR5* and *ANXA1* (Left) and for *LGR5* and *PLAUR* (Right) in representative human CMS subtypes and CA-CRC cancers. Scale bars, 50 µm. Complete statistics in appendix Table B.2.

The data from the assessment of both pre-malignant and malignant human lesions revealed an enrichment of CBC cells in lesions following the conventional pathway and enrichment of RSCs in lesions progressing down the serrated and alternative pathways, and in colitis-associated carcinogenesis.

3.3.4 Stem cell expression across murine intestinal neoplasia models

Next, the stem cell composition of different murine models of neoplasia was assessed by applying the murine equivalents of the stem cell index and stem cell markers used for ISH. The murine models assessed differed in genetic alteration/s carried and were selected to represent conventional and serrated molecular carcinogenic pathways [163, 164]. Consequently, their assessment and a comparison of the stem cell composition of their human counterparts enabled determination of their quality as models, at least with regard to the stem cell dynamics in the tumours. Moreover, the control over the epithelial genotype of the models allowed for an assessment of the cumulative effect of alterations in driver genes on molecular phenotype in both autochthonous tumours and derived organoids.

Assessment of autochtonous mouse tumours showed that the stem cell index varied across the different models, with a noticeable skew in stem cell admixture being observed in association with alteration in some key drivers (Figure 3.6 A). Specifically, loss of Apc was found to skew towards a CBC enriched phenotype, while Mapk activation, through *Kras* or *Braf* mutation, or Tgfb disruption, *via Tgfbr1* knockout, skewed towards a RSC phenotype. In multi-allele models, e.g. *AKPT*, the cumulative accrual of driver genes correlated with variable degrees of admixture of stem cell marker expression. Observations made in bulk transcriptome were confirmed by ISH (Figure 3.6 C).

Further to the assessment of autochthonous tumours of murine models, tumour derived organoids, cultured in the absence of other niche cellular constituents, were also assessed with the aim of determining the cellular contributors to the stem cell phenotype (Figure 3.6 B). This was undertaken to specifically ensure that the phenomenon observed was indeed epithelial, and therefore rightfully attributed to alterations in stem cell population, and not a surrogate of differences in the stromal compartment between conventional and serrated lesions. In the mouse cancer organoids, each genotype had a distinct baseline stem cell index. However, the pattern of stem cell composition across different genotypes was conserved between the *ex vivo* autochthonous

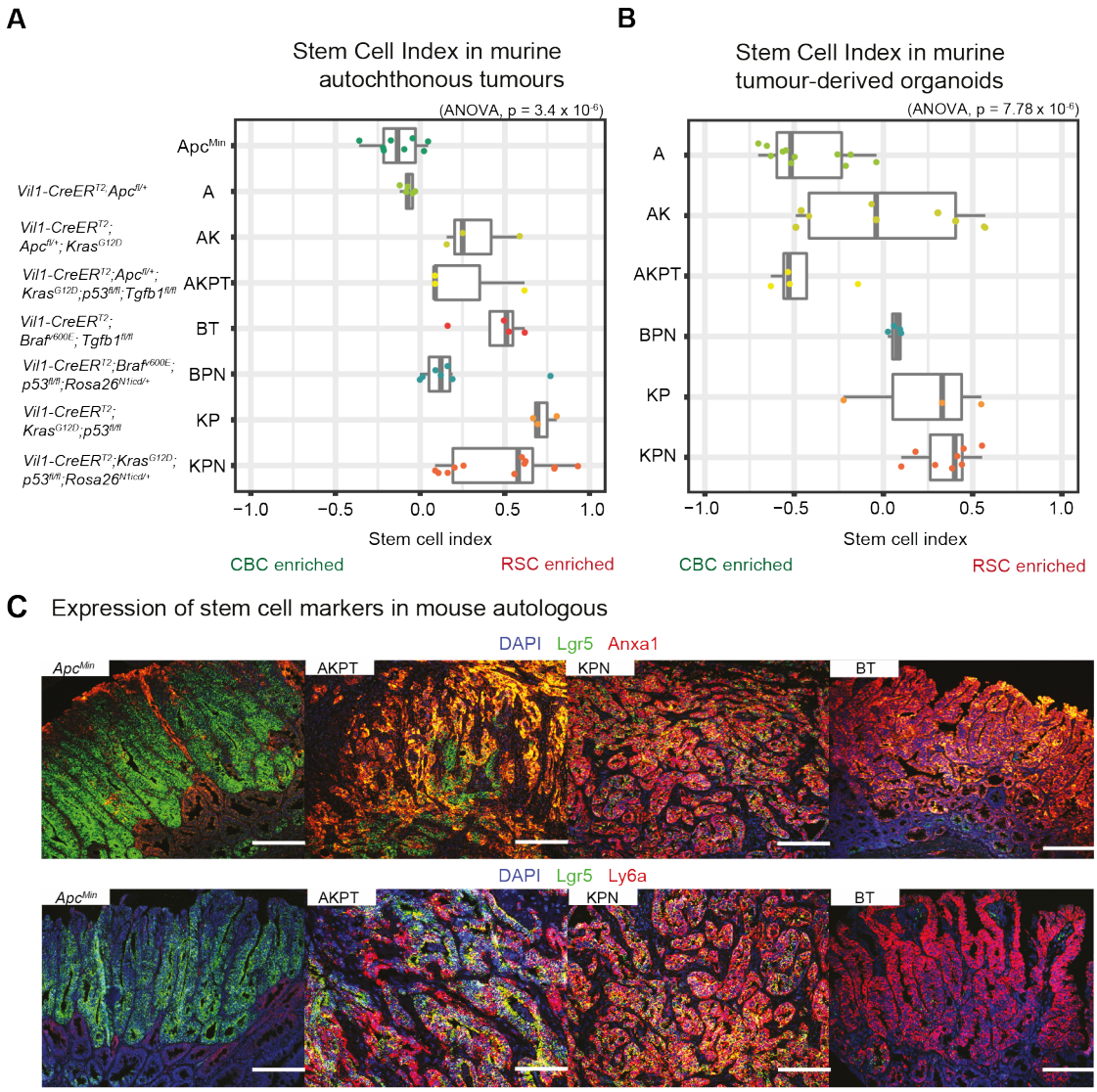


Figure 3.6: Mouse autochthonous tumours (**A**) and matched derived organoids (**B**) mapped across the CBC to RSC spectrum using the stem cell index. **C.** Dual colour fluorescent ISH in representative genotype tumours across the CBC to RSC spectrum. Variable stem cell index across RSC and CBC spectrum can be observed, with a clustering of individual lesions by mouse genotype. Noteworthy, the grade of the tumours varied from adenoma to adenocarcinoma across the different models, with the Apc^{Min} and *AK* lesions being adenomas, while lesions from the other models being adenocarcinomas. Scale bars, 100 µm. Complete statistics in appendix Table B.3.

tumours' bulk transcriptome and the derived *in vitro* organoids cultured in the absence of other niche cellular constituents (Figure 3.6 A and B). This suggests changes in epithelial gene expression as main contributors to the variable stem cell phenotype. Together with the results from the assessment of the human lesions, these data demonstrate that the stem cell index can be used to assess variable admixture of stem cell phenotype across murine and human lesions.

3.3.5 Stem cell potential of murine CBC and RSC cell populations

A single cell clonogenicity experiment was undertaken using a *KPN* mouse to demonstrate the retained stem cell potential within tumours of cells expressing CBC and RSC cell markers and assess whether CBC and RSC have comparable fitness peaks. The *KPN* model is a multi-allele model that phenocopies human serrated lesions and as observed in the previous section has a stem cell phenotype skewed towards the RSC.

KPN derived autochthonous adenocarcinomas were manually and enzymatically dissociated and the resulting single-cell suspensions were FACS-sorted and allowed to grow into organoids, following protocols previously used to enrich for CBC and RSC (details in material and methods 2.2.8). Cells were sorted following a controlled gating strategy (Figure 3.7). Variable but comparable numbers of CBC and RSC cell populations were obtained across biological repeats. However, cell populations expressing both stem cell markers were consistently found to be numerically fewer (Figure 3.8 A). This observation is consistent with the findings from the human scRNAseq data (Figure 3.3 C), where cells expressing both markers were seen to make up a smaller subpopulation.

Culturing of the isolated cell populations in organoid media demonstrated that CBCs, RSCs and cells enriched in both markers retained stem cell capacity as they were able to grow into organoids. Conversely, cells expressing neither CBC or RSC markers were not able to grow into organoids, although they stayed alive, thereby demonstrating little clonogenic capacity *ex vivo* (Figure 3.8 B and C).

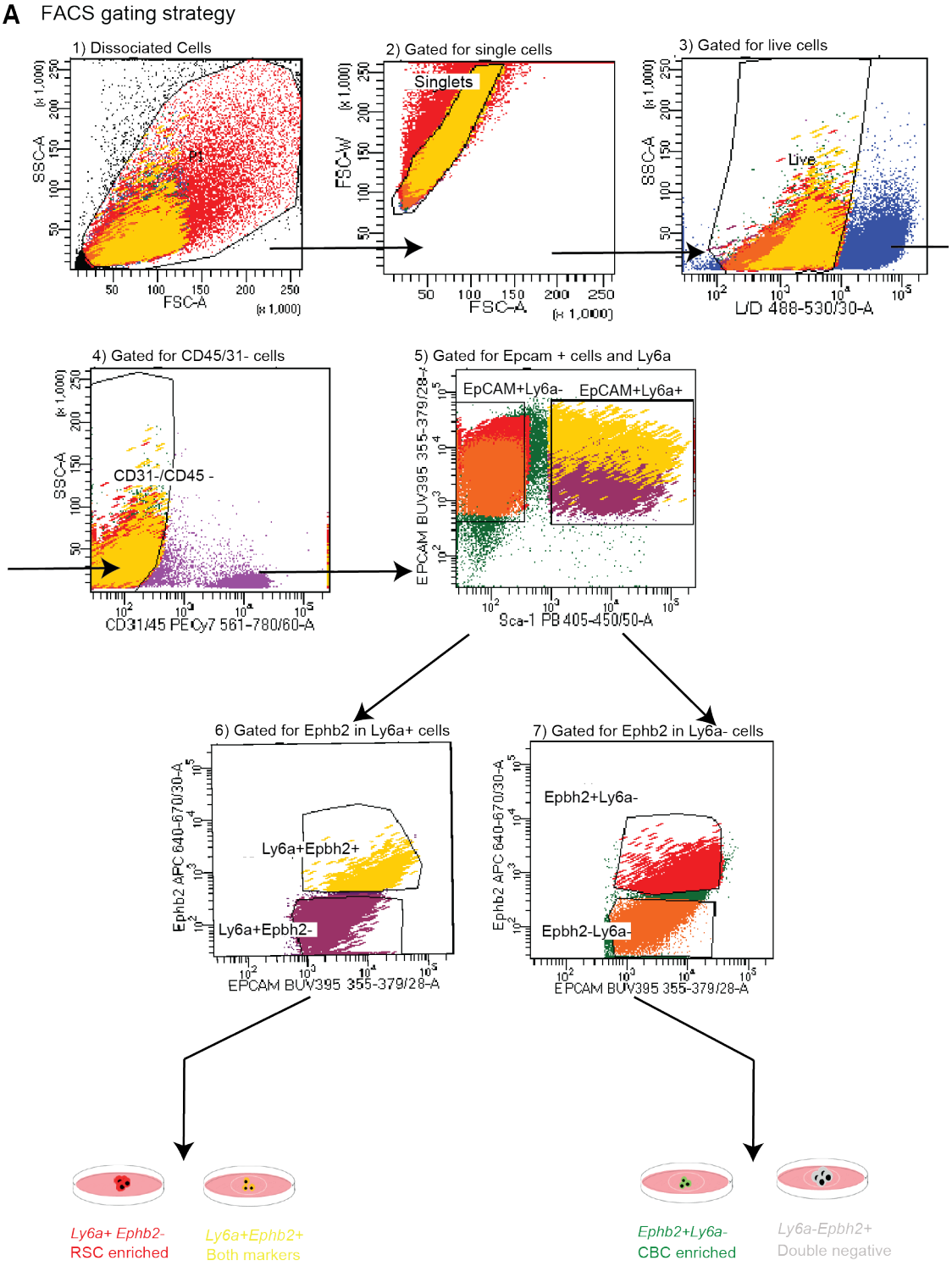


Figure 3.7: Single cell clonogenicity experiment sorting strategy. Cell suspensions dissociated from *KPN* mouse primary tumours (n = 5) were stained and FACS-sorted. CD31 and CD45 were used as markers of immune cells, EpCAM as marker of epithelial cells, Ly6a for Ly6a-positive cells (RSC), Ephb2 as a surrogate marker of Lgr5-positive, thus CBCs. Ephb2 was chosen as marker for Lgr5-positive cells because of a lack of a good antibody against Lgr5 and given that Ephb2 has previously been used as a suitable marker for CBCs [165]. Moreover, a live/dead stain was used to discriminate between dead and alive cells. Plots of the gating strategy employed are reported here. First only dissociated cells were selected, followed by singles and alive cells. From the alive cells the CD45/CD31 negative cells and EpCAM positive cells were selected. These gates aimed at removing immune cells and selecting for only epithelial cells. Lastly, from the EpCAM-positive cells, using Ephb2 and Ly6a the following populations were isolated: Ly6a+positive;Ephb2-negative (RSC), Ephb2-positive;Ly6a-negative (CBC), Ly6a-positive;Ephb2-positive (RSC and CBC positive cells/double positives cells) and Ly6a-negative;Ephb2-negative (RSC and CBC negative/double negative cells).

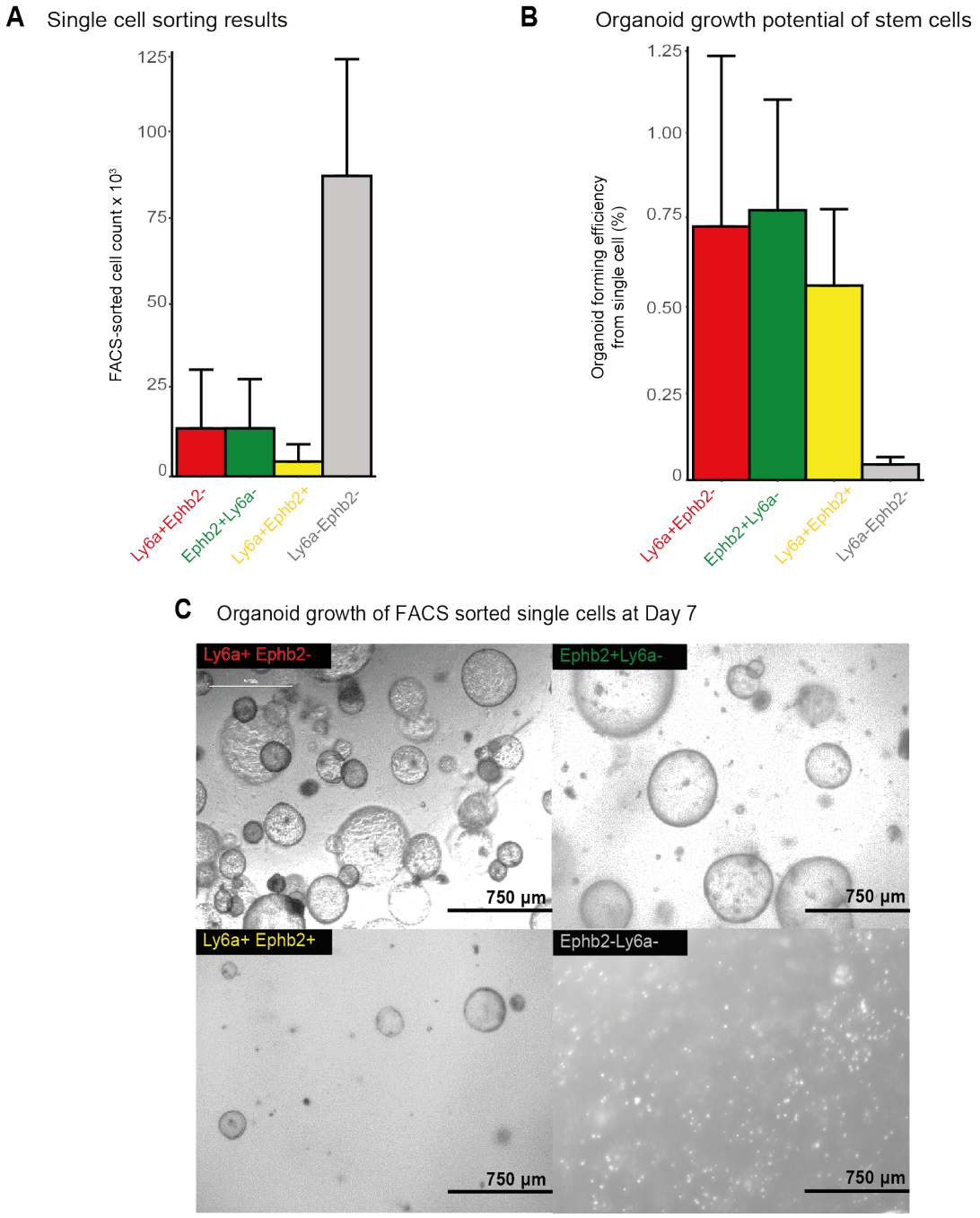


Figure 3.8: Sorting results and organoid growth of CBC and RSC cells. The isolated Ly6a-positive;Ephb2-negative (RSC), Ephb2-positive;Ly6a-negative (CBC), Ly6a-positive;Ephb2-positive (RSC and CBC positive cells/double positives cells) and Ly6a-negative;Ephb2-negative (RSC and CBC negative/double negative cells) single cell populations were allowed to grow into organoids in Matrigel. Cloning efficiency was then assessed 7 days following seeding by counting the number of viable cells. **A.** Number of FACS-sorted cells per cell population isolated. **B.** Organoid forming efficiency from single cells for each cell population reported as percentage. This was determined by dividing the number of organoids formed by the number of cells plated. **C.** Images of organoids/cells at day 7 for each cell population isolated.

3.3.6 Driver genes and pathways associated with CBC and RSC stem cell molecular phenotypes

Mouse models of CRC allow control over the epithelial genotype, which in turn enables assessment of the cumulative impact of individual driver genes on a tumour's molecular phenotype. In this study, this allowed for an assessment of the impact of driver genes on the stem cell phenotype of both autochthonous tumours and derived organoids. The assessment revealed that manipulation of some key drivers results in a notable skew in cell admixture, with epithelial *Apc* mutation being enriched in CBC cell marker expression, while Mapk activation and Tgfb disruption skewed the stem phenotype towards the RSC markers. Based on these findings, the distribution of driver genes in human tumours was assessed with the aim of identifying associations between the stem cell phenotypes and driver mutations in key signalling pathways to begin to assess the mechanistic differences between CBC and RSC phenotypes (Figure 3.9 and 3.10). TCGA lesions were segregated into polarised deciles for CBC and RSC expression and a pairwise comparison of SNA and CNA was undertaken between the top decile CBC (referred to as predominantly CBC) or RSC (predominantly RSC) expression (Figure 3.9 and 3.10 A). The analysis revealed differences in the frequency of the optimally selected SNA and CNA between predominant CBC and RSC tumours, with the most common alteration being *APC* mutation in CBC enriched lesions and *BRAF* mutation in those enriched for RSC. Interestingly, the variable selection of alterations between stem cell phenotypes was consistent with that observed in the murine models.

With regard to the alterations in important CRC signalling pathway genes, *APC* or catenin beta 1 (*CTNNB1*) alterations (referred to as ligand-independent (LI) alterations - outlined in chapter 4), were found in 84% of predominantly CBC cancers, as opposed to the predominantly RSC cancers, where they were found in only 35%. However, *RNF43* mutations and *RSPO3* alterations (also known as ligand-dependent alterations (LD) - chapter 4), which occur in a co-exclusive manner with LI alterations, were detected in the predominantly RSC cancers. Variable mutational patterns be-

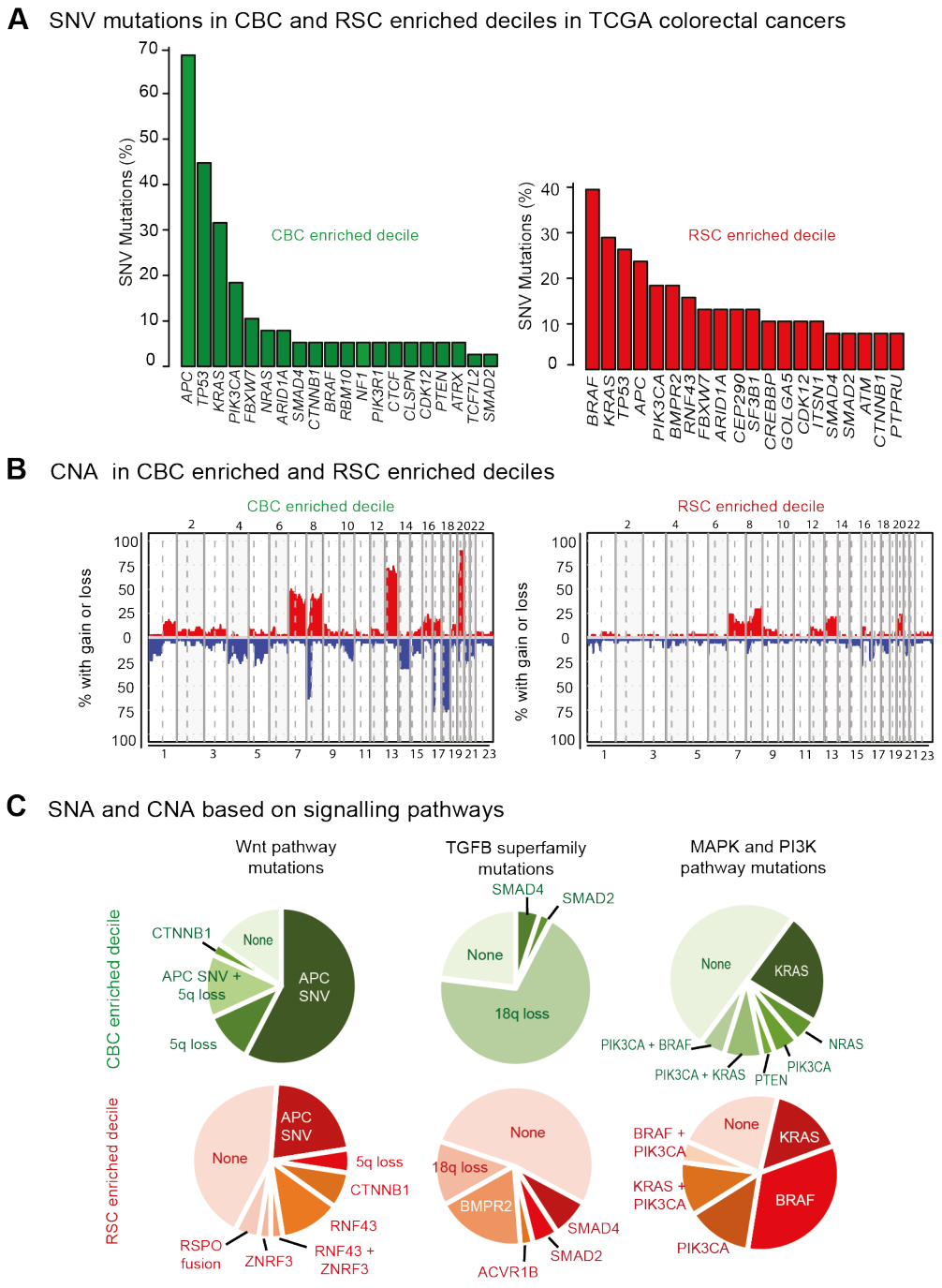


Figure 3.9: SNA and CNA alterations associated with CBC and RSC phenotypes. **A.** SNA comparison of mutation type and prevalence in CBC and RSC predominant TCGA tumours. **B.** CNA comparison in CBC and RSC predominant deciles within TCGA tumours. **C.** Comparison of prevalence and type of SNA and CNA disrupting the WNT, TGFB superfamily, and MAPK and PI3KCA signalling pathways in TCGA tumours divided in CBC and RSC predominant deciles. A differential selection for SNA and CNA can be seen between CBC and RSC predominant TCGA tumours.

tween predominantly RSC and CBC tumours were also seen in MAPK/PI3K and the TGFB signalling pathways, including a higher frequency of *BRAF* in predominantly RSC tumours.

These data indicated an association between stem cell phenotype and specific driver gene alterations in key signalling pathways. Consequently, a cross-species comparison of the signalling pathway disrupted in predominantly CBC and RSC tumours was undertaken. Using transcriptome data from TCGA and the mouse cohorts, GSEA for hallmark and selected pathways was performed on the top decile CBC and RSC human and murine tumours based on stem cell marker expression. Notably, a remarkable cross-species correlation between stem cell phenotype and disruption of specific signalling pathways was seen. WNT activation was correlated with CBC predominant lesions, consistent with the effect of *APC* mutation, while a correlation between RSC predominant lesions and KRAS, YAP1 and TGFB and other inflammatory pathways was found in both human and murine lesions (Figure 3.10 A and B). Correlation between inflammatory pathways and predominantly RSC lesions is particularly interesting given the regenerative nature of RSCs. It seems to suggest that the inflammatory pathways involved in regeneration might be involved in sustaining RSC tumour growth. Furthermore, using the mouse cohorts, the expression of the key pathways was mapped with CBC and RSC gene signatures across the various mouse models. The analysis revealed a statistically significant correlation between the CBC signature and the WNT signalling signature, while the Kras, Yap1 and Tgfb signatures were correlated with the RSC signature. This is consistent with the other findings in both human and murine lesions (Figure 3.10 C).

Together these data demonstrated that variance in stem cell phenotype is correlated with differences in epithelial cell-intrinsic driver gene mutations and conserved cross-species disruption of key microenvironmental signalling pathways.

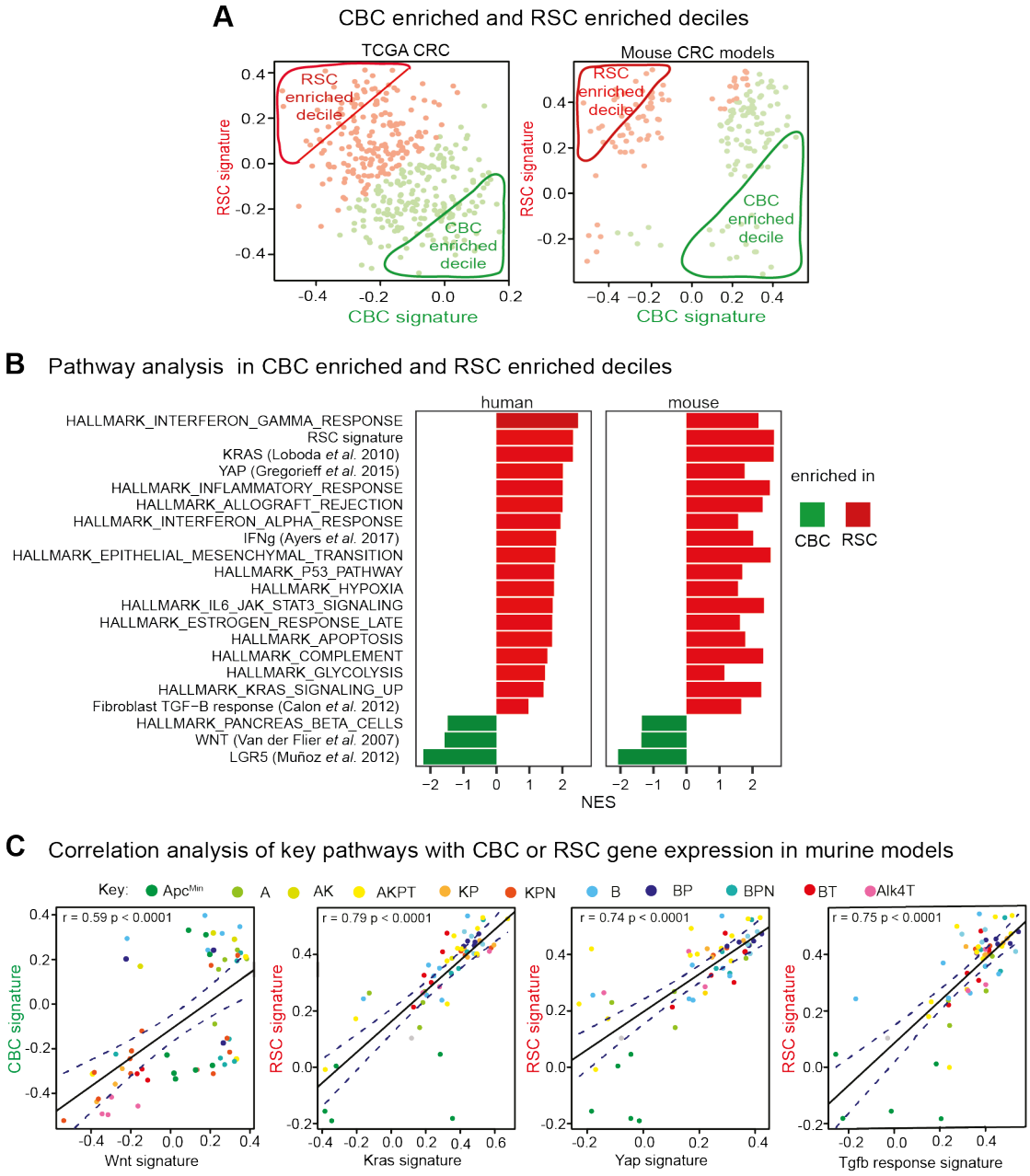


Figure 3.10: Driver pathways associated with CBC and RSC phenotypes. **A.** Segregation of mouse and human colorectal lesions by CBC and RSC signature expressions. Tumours are coloured in green (CBC) or red (RSC) based on their predominant, defined as above median, expression of signature. The 10% most polarised CBC and RSC expressing tumours (highlighted) were segregated in CBC and RSC and used for comparative analysis. **B.** GSEA of hallmark and selected pathways in bulk transcriptome from the CBC and RSC predominant human and murine tumours. Significance threshold was set at an adjusted p-value < 0.05. **C.** Correlation of key pathway expression signatures with CBC or RSC gene expression across a range of murine models of CRC. Driver alleles initialisation: A - $Apc^{fl/+}$, Apc^{Min} - Apc^{Min}, B - $Braf^{V600E}$, K - $Kras^{G12D}$, P - $P53^{fl/fl}$, T - $Tgfbr1^{fl/fl}$, N - $Rosa26^{N1icd/+}$, Alk4 - $Alk4^{fl/fl}$.

3.3.7 Stem cell phenotype dynamics in *in vivo* murine model

The molecular assessment of the stem cell phenotype of human and murine neoplasia revealed that their stem cell phenotypes ranged across a CBC to RSC phenotypic axis. All lesions demonstrated the co-occurrence of CBC, RSC and CBC-RSC cell populations, although at varying proportions. Next, an assessment of whether the stem cell phenotype of lesions remains fixed or can dynamically adapt to changes was carried out. The working hypothesis was that a tumour's stem cell phenotype is dynamic and can allow for interconversion between cells situated on CBC and RSC stem phenotype peaks as an adaptive response to selective pressure. This was based on previous knowledge that injury to CBC causes *Lgr5*-negative progenitor and differentiated cells to dedifferentiate, re-acquire stem cell potential and regenerate the epithelium [122]. Such proposition was tested *in vivo* in Apc^{Min};$Lgr5^{DTR}$ by ablating Lgr5-positive cells and assessing the stem cell phenotype of their lesions at different timepoints post-ablation.

Apc^{Min};$Lgr5^{DTR}$ mice were generated by crossing Apc^{Min} mice, which have a nonsense mutation affecting codon 850 in the *Apc* gene resulting in a truncated Apc protein that makes them prone to develop intestinal tumours, with the $Lgr5^{DTR}$ allele, generated by knocking-in a human diphtheria toxin receptor (DTR) cassette to the first exon of *Lgr5*. This allows for the Lgr5-positive cells to express DTRs and be ablated through the administration of diphtheria toxin. Notably, targeted introduction of the DTR cassette into Lgr5-positive cells had previously been reported to cause a reduction in Lgr5-positive CBC cells [166]. Consequently, prior to assessing the stem cell phenotype pre- and post- ablation, a comparison of the Apc^{Min};$Lgr5^{DTR}$ mice and Apc^{Min} at steady state was undertaken. In congruence with previous observations [166], a shift in stem cell index was noted in the Apc^{Min};$Lgr5^{DTR}$ mice in comparison to the Apc^{Min} mice in both normal tissue and polyps (Figure 3.11).

To assess the stem cell dynamics pre- and post- CBC ablation, Apc^{Min};$Lgr5^{DTR}$ mice were injected with diphtheria toxin and sacrificed after 24 hours, 3 days and 5 days and their stem cell phenotype was compared against that of Apc^{Min} and non-recombined

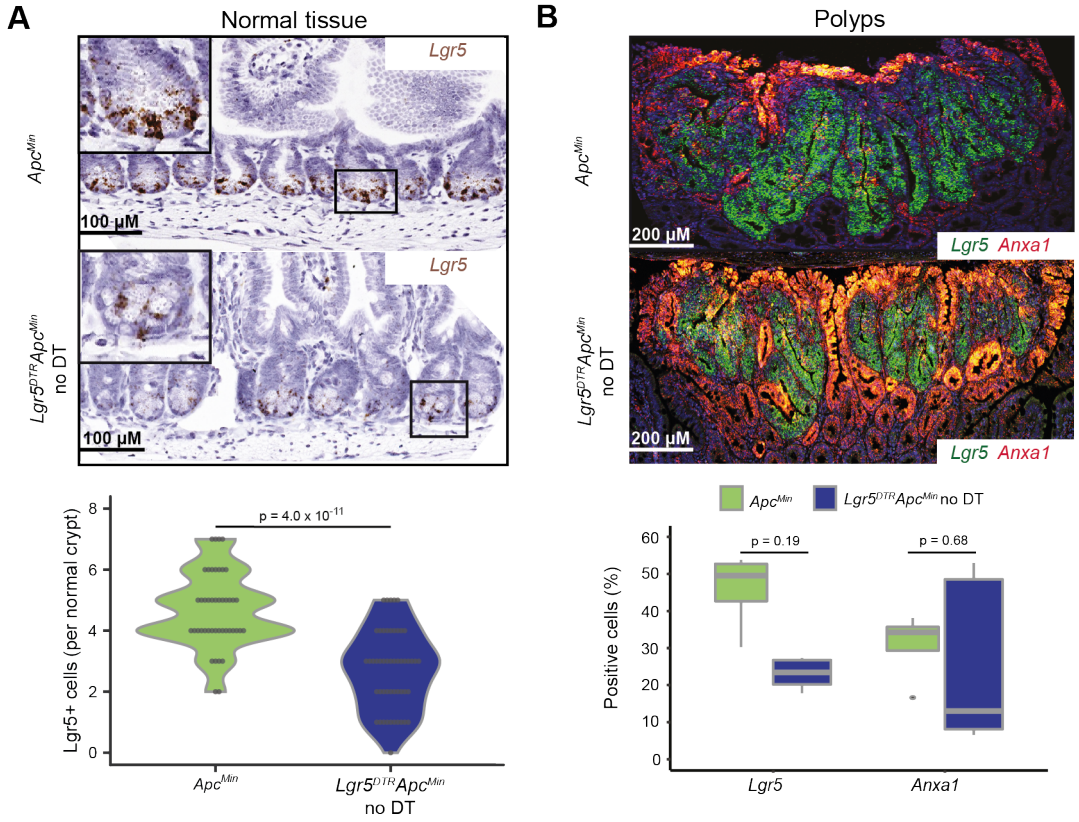


Figure 3.11: Determination of the impact of *Lgr5* hemizygosity on CBC count in normal and polyp tissue. **A.** Chromogenic *Lgr5* ISH with quantification results comparing CBC count in normal tissue between Apc^{Min};$Lgr5^{DTR}$ and Apc^{Min} mice in normal tissue. **B.** Dual colour fluorescent ISH for *Lgr5* and *Anxa1* and quantification results in Apc^{Min};$Lgr5^{DTR}$ and Apc^{Min} mice. In both assessments a reduction in CBC count is noticed in the Apc^{Min};$Lgr5^{DTR}$ mice.

Apc^{Min};$Lgr5^{DTR}$ mice. In addition to the stem cell phenotype, polyp size, cell proliferation and apoptosis were also assessed. Cell proliferation and apoptosis were measured by Ki67 and Caspase 3 IHC staining, respectively. Twenty-four hours following injection, in addition to the previously known ablation of Lgr5-positive cells attributed to the introduction of the DTR cassette, a profound upregulation of RSCs was observed, skewing the molecular stem cell phenotype of the model to be predominantly RSC (Figure 3.12 B and C). The upregulation of RSCs was correlated with a dramatic upregulation of Infg, Yap1 and Kras signalling pathways (Figure 3.13 A). However, 5 days following injection, a partial recovery of Lgr5-positive cells was recorded, resulting in lesions with an admixed stem cell population (Figure 3.12 B and C). Importantly, unlike the shifts in stem cell phenotype recorded across the pre- and post- ablation polyps, as well as the different timepoints post-ablation, no significant change in polyp size,

proliferation or apoptosis was detected (Figure 3.13 B and C).

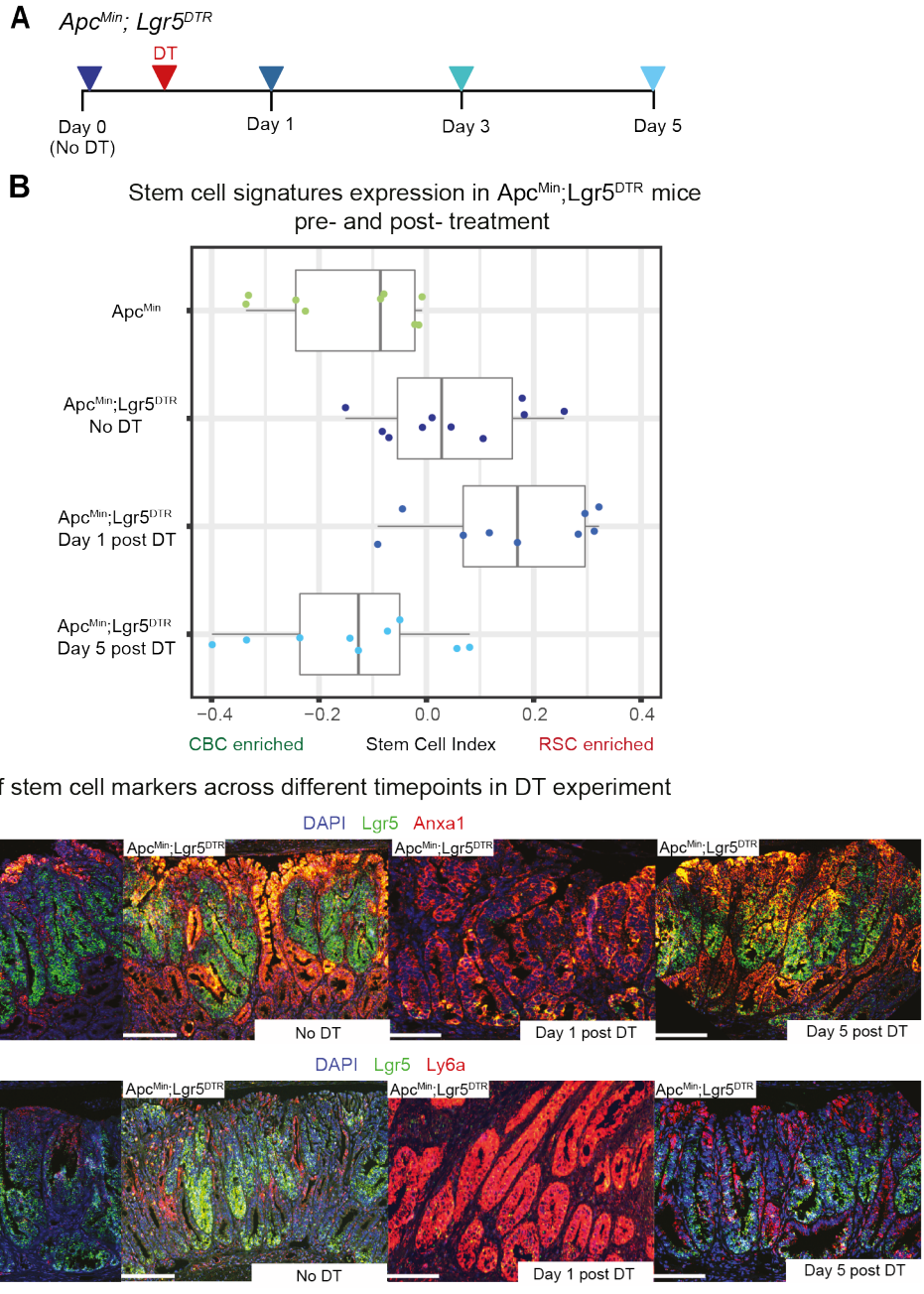


Figure 3.12: Stem cell dynamics assessment in $Apc^{Min};Lgr5^{DTR}$ compared to Apc^{Min} mice. **A.** Timeline displaying timing of recombination, diphtheria toxin injection and tissue harvest. **B.** Stem cell index mapping of polyps from Apc^{Min} and $Apc^{Min};Lgr5^{DTR}$ mice displaying a dynamic change in stem cell phenotype before and after diphtheria toxin injection and the resulting CBC ablation. **C.** Dual colour ISH for stem cell markers in representative Apc^{Min} and $Apc^{Min};Lgr5^{DTR}$ lesions confirming the shift in stem cell phenotype following CBC ablation seen on RNAseq data. Scale bars, 100 µm. Statistics in appendix Table B.4.

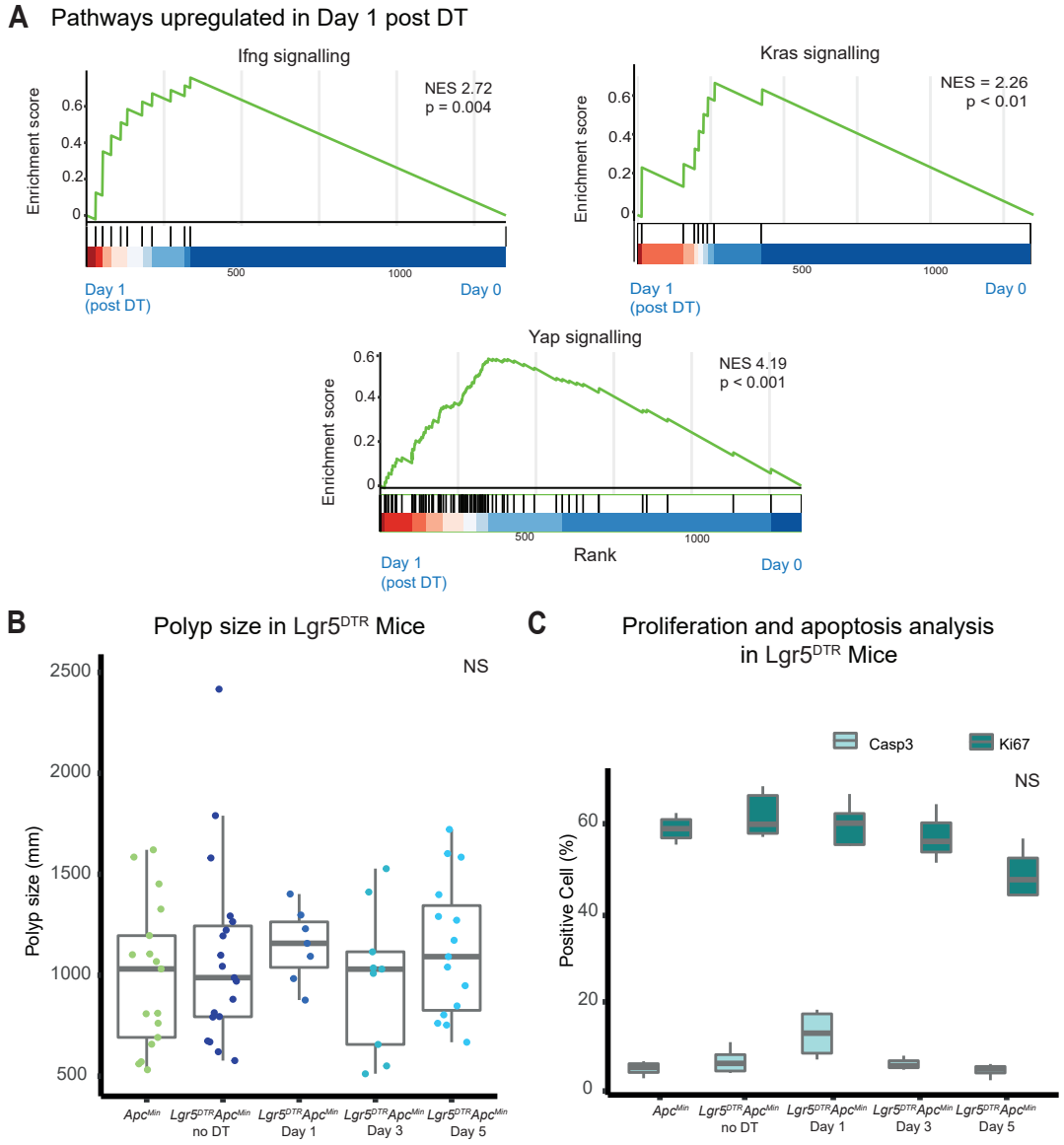


Figure 3.13: Signalling pathways, proliferation, apoptosis and polyp size. **A.** GSEA assessing Infg, Yap1 and Kras signalling pathways in polyps from day 0 (before diphtheria toxin mediated stem cell ablation) and day 1 (following diphtheria toxin mediated stem cell ablation). An enrichment of the three pathways in day 1 post ablation can bee seen. **B.** Assessment of polyp size (i.e. polyp diameter in µm) in Apc^{Min} and Apc^{Min};$Lgr5^{DTR}$ mice. No difference between all genotype/timepoints can be seen. **C.** Assessment of cell proliferation (by Ki67 staining quantification) and apoptosis (by Caspase 3 staining quantification) in Apc^{Min} and Apc^{Min};$Lgr5^{DTR}$ mice. No significant difference between genotype/timepoints was found. Statistics in appendix Table B.4.

Considering the ablation of Lgr5-positive cells following injection with diphtheria toxin, the non-significant rise in apoptosis at day 1 post-treatment, together with a shift towards a RSC phenotype, is indicative of the fact that by day 1, the RSC have already repopulated the ablated epithelium and that perhaps an earlier timepoint is required

to capture the initial apoptotic phase.

The data indicated that selective cell ablation prompts an adaptive shift in signalling pathways and stem cell composition to rapidly restore the stem cell equilibrium. Notably, the absence of change in polyp size, proliferation and apoptosis rates suggests that conventional tumour clinical response measures, such as change in lesion size, would not be sufficient to detect a change in a tumour's molecular phenotype.

Comprehensively, the data showed that the stem cell phenotype of tumours lies within a dynamic equilibrium and that selective pressure can cause an interconversion between CBC and RSC phenotypic peaks as an adaptive response.

3.3.8 Clinical implication of the adaptive response to selective pressure of cancer stem cells

Previous investigations highlighted that the stem cell phenotype of tumours, which ranges across a CBC to RSC phenotypic axis, lies within a dynamic equilibrium, where cells can interconvert between CBC and RSC phenotypic peaks as an adaptive response to selective pressure. Next, a comparable assessment of the molecular phenotypic shifts in human CRC was undertaken by interrogating data from the FOxTROT clinical trial Arm A.

The FOxTROT clinical trial is a multicentre, phase II/III study developed by the NCRI colorectal cancer clinical studies group and aimed at determining the probability of cure in patients with high-risk operable colon cancer who are given combination chemotherapy prior to surgery. The trial also aimed at establishing whether the added administration of the anti-EGFR monoclonal antibody panitumumab could improve results in *KRAS*/*NRAS*-WT patients randomised to neoadjuvant chemotherapy. In Trial Arm A, patients were given 6 weeks of pre-operative fluorouracil and oxaliplatin (OxFP) chemotherapy followed by surgery and then 18 or 6 weeks of post-operative OxFP chemotherapy. This precious resource is one of the extremely rare studies where tumour samples were taken before and after chemotherapy, enabling a temporal analysis of the tumour response to treatment, and thereby the stem cell adaptive response to the selective pressure imposed by treatment. First, the impact of chemotherapy on

the change in stem cell index was assessed between pre (specimen 1) and post (specimen 2) treatment samples. Transcriptomic analysis of the pre- and post- treatment samples showed a range of response to therapy, within which patients could be distiguished in two main categories:

- Patients with plastic tumours, i.e. tumours that had a very dynamic response between sampling, with shift in stem cell index from positive to negative or *vice versa*;
- Patients with static tumours, i.e. tumours that showed comparatively little difference in stem cell index between sampling, with both pre- and post- treatment lesions having an either positive or negative stem cell index.

As shown in Figure 3.14 B.

Next, the relationship between clinical response and stem cell phenotype were assessed. The analysis revealed that patients with a plastic tumour were significantly less likely to have a histologically detectable response to treatment, suggesting an association between a tumour's ability to adapt to change and clinical response (Figure 3.14 D). Moreover, plastic patients with their stem cell index shifting from predominantly CBC to predominantly RSC were found to be more commonly non-responders. However, this was not sufficiently significant to make adequate conclusions. Notably, while the stem cell index could detect a difference between plastic and static tumours, no difference between these tumour groups could be detected using cell proliferation. This highlights the fact that the adaptive response to treatment recorded in the change in stem cell index is intrinsic to the stem cell phenotype and not a surrogate measure of change in proliferation.

The sampling difference between the pre-treatment specimens, which were biopsies, and the post-treatment specimens, which were resections, led to the consideration that intra-tumour heterogeneity could be a potential confounder when interpreting results from the analyses. In fact, sample heterogeneity has previously been reported to contribute to highly variable CMS assignment in multiregional biopsy samples. This was attributed to differences in stroma/immune cells proportion between biopsies.

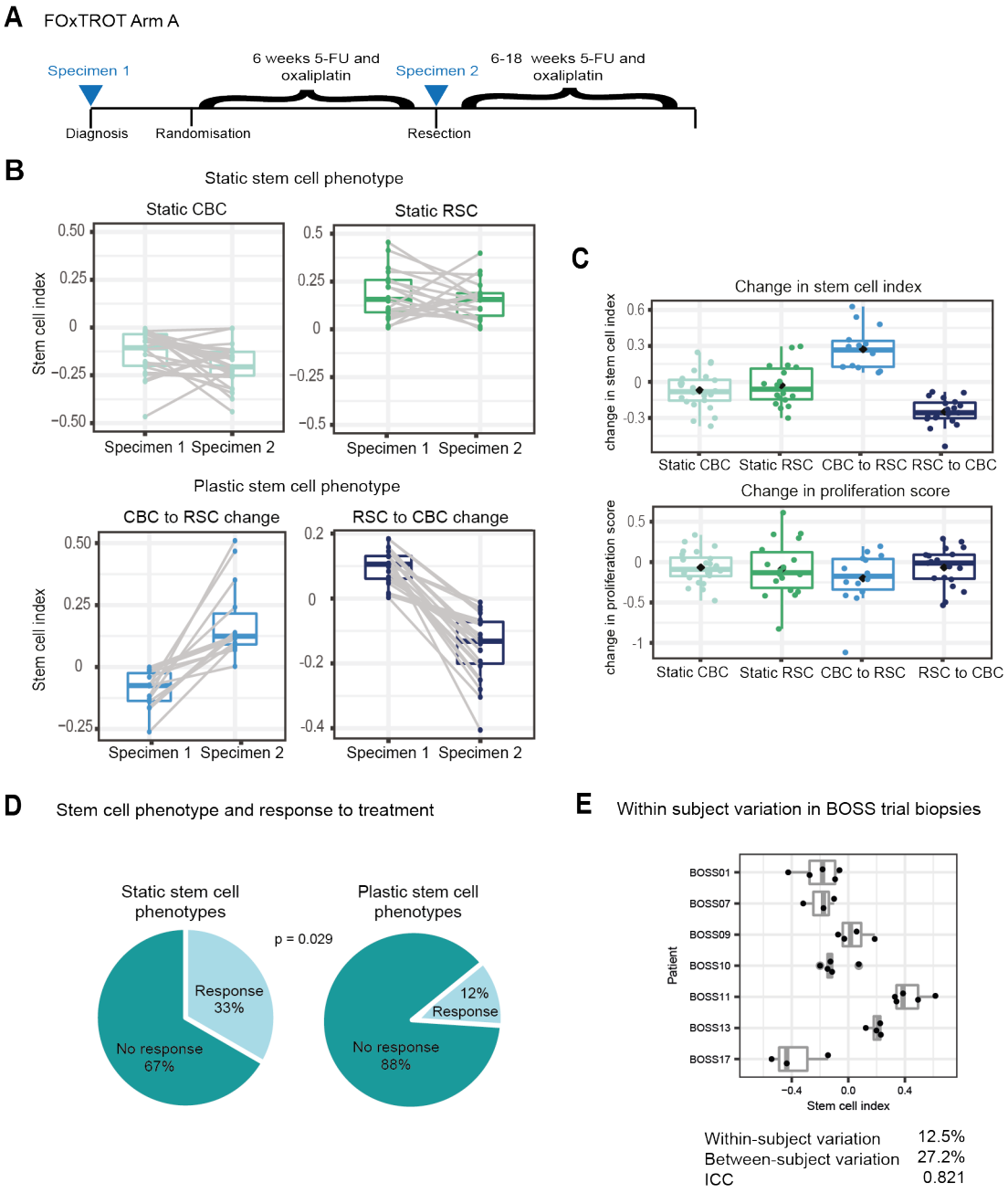


Figure 3.14: **A.** FOxTROT Arm A trial timeline showing specimen acquisition before and after 6 weeks of 5-FU and oxaliplatin chemotherapy. **B.** Ladder plots highlighting the stem cell index in specimen 1 and specimen 2 in static and plastic patients. Patients were grouped into static or plastic based on whether a shift in stem cell phenotype was noted between their pre- and post- treatment specimen. **C.** Assessment of stem cell index change in parallel with change in proliferation score within the same patient groups. **D.** Proportion of patients with and without response grouped by static and plastic stem cell phenotype. Patients with plastic lesions were statistically more likely to be non-responders. **E.** Within-subject variation of stem cell index from random non-adjacent biopsies from the BOSS trial. Complete statistics in appendix Table B.5.

To investigate the impact of intra-tumour heterogeneity on stem cell index, transcriptome data from non-adjacent multiregional biopsies derived from the BOSS study were interrogated. Assessment of the stem cell index showed a high intra-class correlation coefficient. Moreover, the within-subject coefficient variation, which measures the variation between random biopsies from the same tumour, was lower than the between-subject, i.e. between different tumours, coefficient of variation, demonstrating that stem cell index measurement is not susceptible to sample heterogeneity arising from contamination with non-epithelial tissue (Figure 3.14 E). Consequently, it could be concluded that measurement of stem cell index from biopsy samples produces a reasonably robust representation of the stem cell phenotype of the entire tumour.

3.4 Key findings

- CRCs are supported by an admixture of CBC and RSC populations, with distinct molecular phenotypes.
- CA-CRCs have a predominantly RSC phenotype.
- Stem cell admixtures can be measured from transcriptomic data using the newly established stem cell index.
- Stem cell phenotypes vary by genotype and subtype, shifting with cell signalling disruption.
- Stem cell phenotypes are adaptive, characterised by a dynamic response to selective pressure.
- Phenotype shift from selective pressures are associated with chemoresistance.

3.5 Discussion

During homeostasis, the self-renewal capacity of the intestinal epithelium is supported by CBC/LGR5-positive cells. However, upon homeostatic distress, the epithelium displays remarkable stem cell plasticity, characterised by dedifferentiation of the intesti-

nal epithelial cells through a transient disruption of morphogen signalling and easing of stringent homeostatic control over cell fate. This enables a rapid regeneration of the epithelium. The pathways evolved to enable the regenerative response can be co-opted and corrupted by cancer [150]. This co-option and corruption in cancer of pathways activated in response to tissue injury is believed to be the evolutionary trade-off between a robust regenerative response to wounding and cancer risk. Despite our understanding of this and the role of LGR5-positive cells in CRC, the specific contribution of stem cell plasticity to CRC development is yet to be fully elucidated.

In this chapter, a combined molecular and morphological assessment of the stem cell phenotypic heterogeneity across a wide range of murine and human intestinal tumours was carried out, with the development of a tool to measure such phenotypic heterogeneity, the stem cell index. The stem cell index is a simple and accessible tool, applicable to bulk transcriptomic data (both RNAseq and microarray) that can be used to determine a lesion's stem cell phenotype and track dynamic molecular phenotypic change.

A variable enrichment and admixture of CBC and RSC was found in human and murine lesions using the molecular-morphological analysis. The degree of CBC-RSC admixture was associated with epithelial mutation and microenvironmental signalling disruption. Moreover, the CBC-RSC was adaptive, showing a dynamic response to acute selective pressure. Based on these findings, a modified fitness landscape model of stem cell phenotype can be proposed to explain the stem cell dynamics in CRC. Within this fitness landscape model of stem cell phenotype, CBC and RSC stem cells are distinct, but interlinked, stem cell population peaks in the phenotypic axis. Skewing of the stem cell phenotype towards one of the peaks is dictated by differences in epithelial driver gene mutations in conjunction with conserved cross-species disruption of key microenvironmental signalling pathways. However, selective pressure can induce a transient shift towards an alternative stem cell phenotype through an alteration of the morphogenic signalling landscape (outlined in Figure 3.15).

Further to the proposal of the modified fitness model, the data gathered allow for several considerations to be made. The multifaceted assessment of both human and

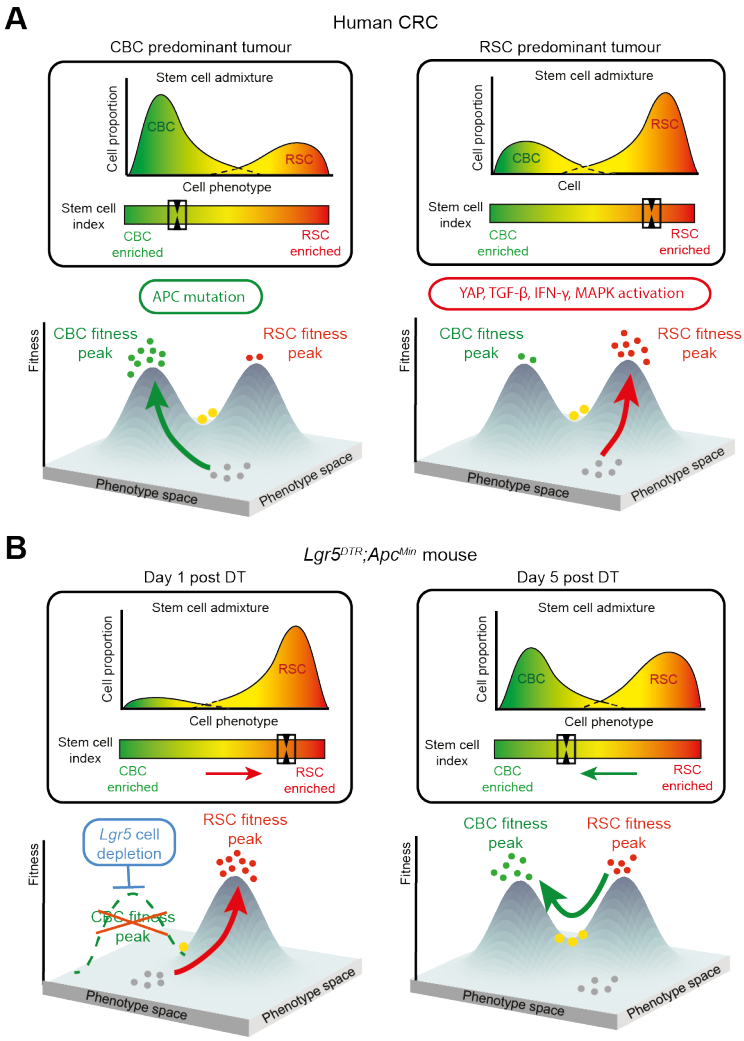


Figure 3.15: Fitness landscape model. A representation of the tumour stem cell phenotype as a fitness landscape model, characterised by two distinct but interlinked fitness peaks located along a phenotypic axis, Lgr5-positive CBC and Lgr5-negative RSC. **A.** Epithelial driver gene mutations in conjunction with conversed cross-species disruption of key microenvironmental signalling pathway dictate the skewing of the stem cell phenotype towards a peak. **B.** Application of selective pressure to a fitness peak will cause the morphogenic signalling landscape to change and shift the stem cell equilibrium towards the alternative phenotype. Key: green dots - CBC cells, red dots - RSC cells, yellow dots - cells expressing both CBC and RSC markers, grey dots - epithelial cells expressing neither stem cell marker.

murine lesions demonstrated the role of RSCs as cancer stem cells with similar *ex vivo* clonogenic capacity to that of CBC. Although the role of RSCs in regeneration has been previously described and their role in cancer hypothesised, this is believed to be the first cross-species characterisation of RSC as cancer stem cells.

An important aspect of the stem cell phenotype peaks described are the genetic signatures and signalling pathways associated with each peak. Focusing on the WNT signalling pathway, predominantly CBC lesions were enriched for LI alterations, while

predominantly RSC were enriched for LD alterations. Recently, LI alterations, specifically *Apc* mutations, in CBC cells were found to provide a competitive advantage over their neighbouring wild-type cells, thereby allowing for adenomatous growth [167, 168]. On the other hand, cellular dedifferentiation through a reversion of foetal development (such as that of RSCs) was found to be a WNT-indepedent process. Instead, it was found that it could be initiated by MAPK activation through either *BRAF*-activating mutations, epithelial damage response or stress triggered by loss of mismatch repair mechanism. These are features common to serrated lesions (i.e. *BRAF* alteration), colitis-associated lesions (i.e. epithelial damage response) and loss of mismatch repair mechanism (i.e. CMS1 cancers or cancers arising from serrated lesions). Han *et al.* found that *KRAS* or *BRAF* mutation, together with disruption of *TP53* and *SMAD4*, and transient TGFB exposure drove YAP1/TAZ-dependent epithelial cells' transcriptional reprogramming and lineage reversion [169]. This highlights the need for microenvironmental triggers for cellular lineage reversion and is consistent with the findings in this chapter where the RSC phenotype was associated with microenvironmental pro-regenerative pathways. Collectively, based on the data in this chapter and the literature, it can be deduced that while the CBC stem phenotype is primarily contributed by cell-intrinsic WNT ligand-indepent mutation, the RSC phenotype is supported by both cell-intrinsic (e.g. *BRAF* and *KRAS* mutations) and tumour microenvironment sources, with key pro-regenerative stem cell pathways being INFG, TGFB and YAP1.

Investigation of the stem cell response to selective pressure revealed the adaptive nature of the intestinal stem cell phenotype. Specifically, loss of *Lgr5*-positive cells in Apc^{Min};$Lgr5^{DTR}$ mice polyps was shown to induce a rapid adaptive response mediated by upregulation of Infg, Yap1 and Kras signalling pathways and an acute switch in stem cell phenotype. These are the same signalling pathways activated during epithelial regeneration and restoration of homeostasis following injury [30]. The adaptive phenotypic switch observed in the model is of interest in the clinical setting, particularly with regard to chemosensitivity and response to treatment. That is because the ability to undergo a phenotypic switch in response to selective pressure could trans-

late in a tumour's ability to undergo molecular changes to persist following selective pressure exerted by some form of therapy. This was partially shown by the analysis on the FOxTROT data, where the ability to undergo a shift in stem cell phenotype in response to selective pressure (seen in the plastic tumours) was found to be associated with poor response to treatment. Furthermore, just this year, at least two papers have reported an association between cellular reprogramming and chemoresistance. Solè *et al.* showed that sublethal doses of chemotherapy lead *TP53* wild-type CRC cells to acquire a YAP1-dependent foetal phenotype. This was found to be associated with high tumour-initiating and metastatic activity [170]. Similar observations were made by Alvarez-Varela *et al.* on Mex3a-positive cells, a newly identified latent Lgr5-positive cell population arising as a result of a lack of optimal growth conditions. The group found that upon chemotherapy persister Mex3a-positive cells downregulate the WNT/stem cell gene program and adopt a transient state reminiscent to that of Yap1 foetal intestinal progenitors, which protects them from chemotherapy [171]. Furthermore, prior to these two studies, Yap1 had been recognised to induce resistance in cancer [172]. Collectively, these findings suggest the mechanism behind chemoresistance and relapse after treatment to be adaptation to selective pressure through cancer stem cell phenotypic conversion. Further to its association with chemoresistance, switch in stem cell phenotype has been previously reported in the context of metastasis. Fumagalli *et al.* found circulating metastatic cells to be predominantly Lgr5-negative, but to require switching to Lgr5-positive for secondary outgrowth in the liver [161]. Importantly, both in this chapter and the studies from Solè *et al.* and Alvarez-Varela *et al.*, the adaptive response to pressure and stem cell phenotype switch was associated with an upregulation of the Yap1 signalling pathway and the co-option of foetal, revival and Yap1 expression programs. The identification and characterisation of the signalling pathways and mechanisms involved in cancer cell adaptivity is pivotal for drug development and improved patient management. Development of drugs targetting these adaptive signalling pathways would be beneficial for the prevention of stem cell plasticity and thereby prevention of emergence of therapy resistant cell clones.

Another translational potential highlighted by the findings in this chapter is the

use of the stem cell index in treatment decision-making and monitoring. The stem cell index has the potential to be used to decide on the best course of treatment based on a lesion's stem cell phenotype. Furthermore, its ability to effectively detect a phenotypic shift following selective pressure, where no changes are detectable by traditional response to treatment measures, such as change in polyp size and/or cell proliferation, highlights its potential use to monitor treatment response. Following administration of a first line of treatment, the stem cell index could be used to determine how the tumour is responding to it, and where a shift in stem cell phenotype is detected, it could warrant the need for a second line of treatment. Thus, it could be used to both monitor and time treatment.

Lastly, the assessment of the stem cell index across different histological and molecular subtypes of neoplasia revealed a differential clustering of the lesions subtypes across the RSC and CBC spectrum, with conventional lesions having a predominantly CBC phenotype, while serrated and colitis-associated lesions showing a predominantly RSC phenotype. This differential clustering was observed in precursor lesions already, suggesting that preferential stem cell expansion occurs early in the carcinogenic pathway, and was found to be conserved between precursor lesion and its equivalent cancer e.g. RSC predominance in colitis-associated LGD and CA-CRC. Findings on the preferential clustering of precancerous polyps along the RSC and CBC spectrum are of particular interest with regard to both cancer stem cells and cancer-initiating cells. Consistent with their cancer-initiating cell, TVAs were found to have a predominantly CBC stem cell phenotype. Conversely, serrated and colitis-associated lesions had a predominantly RSC phenotype. Interestingly, the cancer cell of origin of serrated lesions has recently been demonstrated, following previous speculations, to be different from CBC by Chen *et al.* [64]. The authors suggested serrated lesions to arise from "metaplasia" or dedifferentiation of differentiated cells. Transcriptomic profiling of these cells revealed similar expression patterns as the RSC described in this chapter, with expression of foetal markers, including *ANXA1*, *ANXA10*, as well as lack of expression of mature stem cell signatures [64]. This, together with the fact that both the cell populations described by Chen *et al.* and RSC arise from dedifferentiation of ma-

ture cells, would suggest that these could be the same cell population, in which case serrated lesions would follow TVAs in having the same cancer initiating cell and cancer stem cell. However, further work is needed to confirm this hypothesis. Similar to serrated lesions, the cancer cell of origin of colitis-associated lesions is yet to be fully elucidated, although it is known to be different from CBC (to be further discussed in chapter 6).

In conclusion, this chapter has demonstrated that the stem cell phenotype in CRC is variable, with tumours having variable degrees of CBC-RSC admixture, and adaptive, being able to dynamically respond to selective pressure. This highlights the importance of measuring the stem cell phenotype in deciding, monitoring and scheduling patients' treatment. Notably, this can be done using the the stem cell index. Lastly, future efforts should focus on the development of treatment directly targetting both stem cell populations and drugs that can inactivate the adaptive signalling pathways that drive stem cell plasticity.

3.6 Limitations and future work

The study had the following limitations:

- The SNA and CNA analysis was performed on sporadic cancers from TCGA, hence more reflective of sporadic lesions. Future work should be undertaken to determine whether the same genomic patterns are seen in CA-CRC.

- Whole exome somatic mutation screening of all the lesions from the autochthonous animal models was not performed. Therefore, it is possible for some of the stem cell index intra-genotype variation seen (particularly in *KPN* mice) to be the consequence of undetermined additional somatic mutation.

- The multi-allele mouse models used in this chapter are the result of simultaneous genetic recombination, thereby changes seen in the genotype-specific tumour landscape remodelling are the result of concomitant combined driver gene mutations. This is different from the cumulative, stepwise progression seen in human

disease.

Findings in this chapter have led to several lines of enquiry that should be further investigated, including the following:

- Development of a murine model for RSC: Unlike the Lgr5 models, which represent a good tool for the study of CBC cells, there are currently no RSC models that can be used to study RSCs. This could be developed by assessing the RSC gene list for a specific marker, expressed well and solely by epithelial cells, that can then be used to construct a murine model.

- Further investigation on the clinical implications that the stem cell phenotype might have in CA-CRC early detection and treatment. Specifically, I would be interested in assessing whether within the stem cell phenotypic spectrum there is a difference in progression risk between colitis-associated LGD sitting on either stem cell population end/peak. Similarly, it would also be interesting to assess whether there is a difference in response to treatment between CA-CRC sitting on either end of the spectrum.

- One exciting translational potential of this chapter is to develop treatment that can directly target the different stem cell populations. Recently, several *ANXA1* antibodies and blockers have been described [173, 174], with some being tested for the treatment of cancer. An example of this is the use of MDX-124 anti-annexin-A1 antibody for the treatment of pancreatic cancer [175]. It would be interesting to test these antibodies and blockers in CRC to determine if they can effectively target RSCs.

Chapter 4

The WNT signalling pathway in colitis-associated colorectal cancer

This chapter investigates the role of the WNT signalling pathway in CA-CRC. The chapter comprises two strands:

- Strand A: Genetic and transcriptomic profiling of the WNT signalling landscape in human colitis-associated lesions;
- Strand B: Assessment of the effects of ligand-dependent alterations on cell dynamics in mouse models of WNT disruption.

4.1 Introduction

Aberrant WNT signalling pathway activation is a hallmark of CRC. Colorectal tumours rely on abnormal WNT signalling to maintain stemness and a dedifferentiated state. Moreover, WNT has also been implicated in metastasis and modulation of anti-tumour immunity [4]. The principal mechanism of WNT aberrant activation is mutation of key WNT genes, with loss-of-function mutations in *APC* and *RNF43* and gain-of-function mutations in *CTNNB1* and *RSPO* being recurrent alterations. *APC* alterations account for 80% of the WNT alterations, while *RNF43*, *CTNNB1* and *RSPO* alterations account for only 20% of all known WNT alterations in S-CRC [176].

APC alterations are generally frameshift mutations affecting predominantly the mutation cluster regions between amino acids 1200 and 1600, and commonly gain their second hit from loss of heterozygosity. *CTNNB1* mutations are characteristically activating missense mutations that disrupt the phosphorylation sites for the components of destruction complex, thus preventing degradation of β-catenin. The most commonly occurring *RNF43* alterations are frameshift mutations at amino acids 117 and 659, which result in a truncated protein. Similarly to *APC* alterations, loss of heterozygosity of the wild-type *RNF43* allele is frequently seen. *RSPO* alterations have only been reported in *RSPO2* and *RSPO3*. They are most commonly gain-of-function gene fusions, arising from the replacement of an *RSPO2*/3 exon and promoter with that of a gene with higher basal expression, thereby inducing the overexpression of functional RSPO from epithelial cells. In the absence of *RSPO* fusions, *RSPO* excessive/perturbed physiological expression from the stromal cells has also been reported [176, 177]. Although the mechanism behind high stromal *RSPO* expression in CRC is yet to be determined, it has been previously linked to demethylation of *RSPO3*'s promoter region and deficiency in Keap1 in lung cancer [178].

APC, *CTNNB1*, *RSPO* and *RNF43* alterations can be divided into ligand-dependent (LD) or ligand-independent (LI). LI alterations affect the intracellular proteins *APC* and *CTNNB1* and drive a LI downstream activation of the WNT signalling pathway that results in constitutive activation. Conversely, LD alterations in *RSPO* and *RNF43* disrupt the synergistic *RSPO*-axis and can only amplify endogenous otherwise intact WNT ligand signalling (Figure 4.1). Recently, LI and LD lesions have been shown to differentially express WNT negative regulators that act at multiple levels of the pathway [18]. Activated upon WNT signalling stimulation, negative feedback loops act to constrain activation of the pathway. In LI lesions, with the mutations occurring downstream of the WNT-ligands and the negative regulation loops acting upstream of the destruction complex, the constitutive activation of the pathway is uncoupled from the action of negative regulators. Consequently, although WNT hyperactivation will cause the stimulation of the negative feedback loops, LI lesions are not sensitive to them (Figure 4.1).

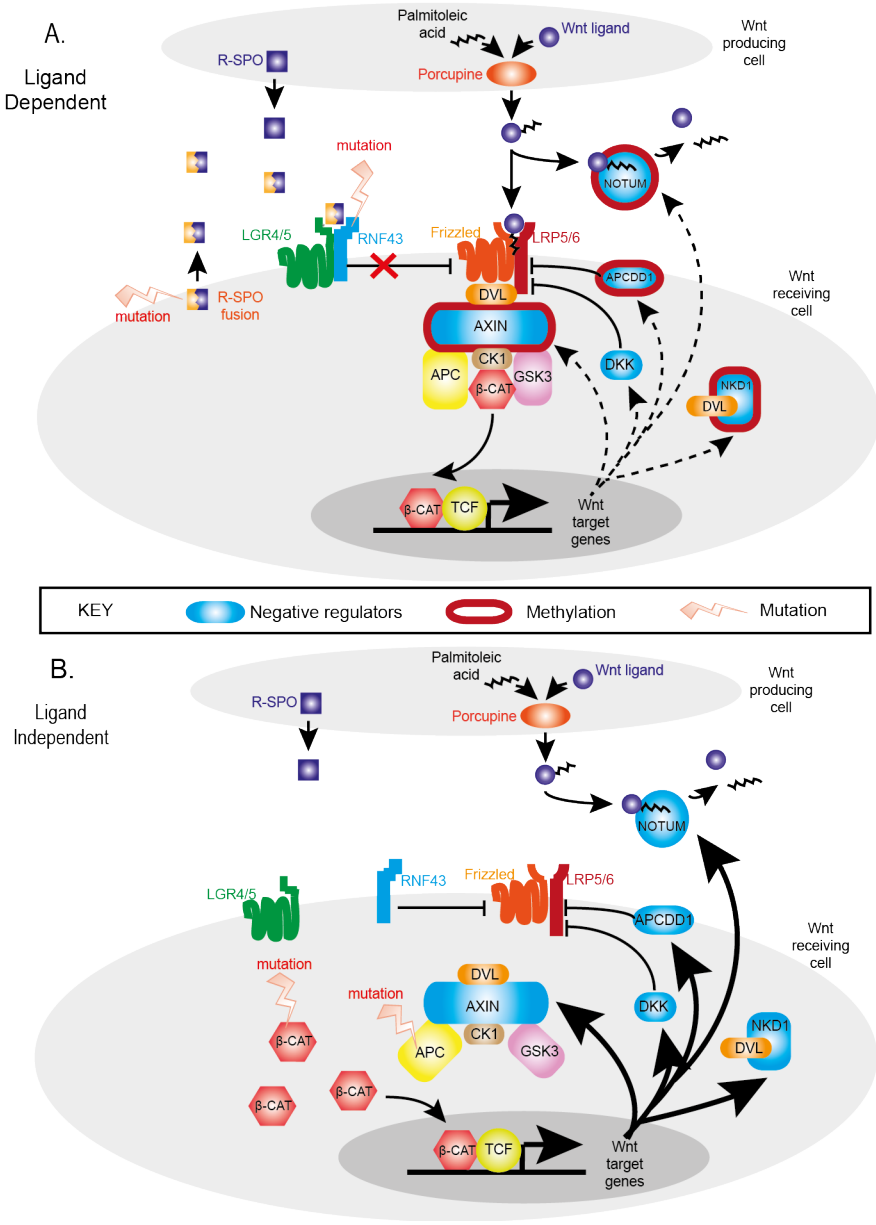


Figure 4.1: The WNT signalling pathway in CRC. From [18].

Conversely, LD mutations occur upstream of the WNT-ligands, thus LD retain their sensitivity to the negative feedback loops. In lesions with LD mutations, this physiological negative regulation constraint needs to be actively repressed, resulting in low expression of key WNT regulators. This has been reported to be partially associated with their epigenetic downregulation, particularly of AXIN2, leading to the postulation that LD WNT hyperactivation relies on two hits: a first mutational alteration in *RNF43* or *RSPO*, followed by epigenetic downregulation of physiological negative reg-

ulator mechanisms. The difference in negative feedback effectiveness and repression between LD and LI lesions results in differential expression of some WNT negative regulator genes between LI/LD lesions, with *AXIN2* being the most differentially expressed. This led to the emergence of *AXIN2* expression as a discriminatory molecular biomarker to establish ligand dependency in lesions [18]. Other regulators hypermethylated in LD lesions include *NKD1*, *WIF1*, *DKK1*/3 and *NOTUM*.

APC, *CTNNB1*, *RSPO* and *RNF43* occur mostly in a mutually exclusive manner. They are also selectively and preferentially acquired in different precursor and malignant CRC subtypes (as well as different forms of cancer). In terms of CRC precursor histological subtypes, TVAs and TAs are predominantly LI, selecting for mainly *APC* and less commonly *CTNNB1* mutations. On the contrary, serrated lesions are predominantly LD, selecting mostly for *RNF43* mutations and *RSPO* fusions, although a small minority of serrated lesions are *APC*-mutant. Thus, while LD alterations are exclusively found in serrated lesions, LI mutations are seen across all histological subtypes of sporadic precursor lesions. Furthermore, an assessment of the clinicopathological features of LD and LI CRCs found that LI alterations were significantly enriched for left-sided tumours, *KRAS* mutation and CMS2. Tumours bearing *RNF43* mutation were significantly more right-sided, *BRAF*-mutant, MSI and CMS1, while 50% of tumours with *RSPO* fusions were left-side with lower CMS1 frequency. These findings are consistent with LI having a conventional lesion aetiology and LD *RNF43*-mutant lesions originating predominantly from SSLs. They also suggest an enrichment of tumours arising from TSAs within the LD *RSPO*-fusion tumours [18]. This highlights the fact that different lesion subtypes follow different molecular carcinogenesis pathways (Figure 4.2).

The mutual exclusivity of different WNT alterations and preferential selection of specific mutations has been linked to the relative level of WNT/CTNNB1 activation and the ”just-right” theory. The ”just-right” theory poses that an intermediate level (optimal and not excessive; not too hot or too cold) of WNT hyperactivation is ideal for cell transformation. This is supported by several lines of experimental evidence, including the observation that *APC* mutations within the mutation cluster region pro-

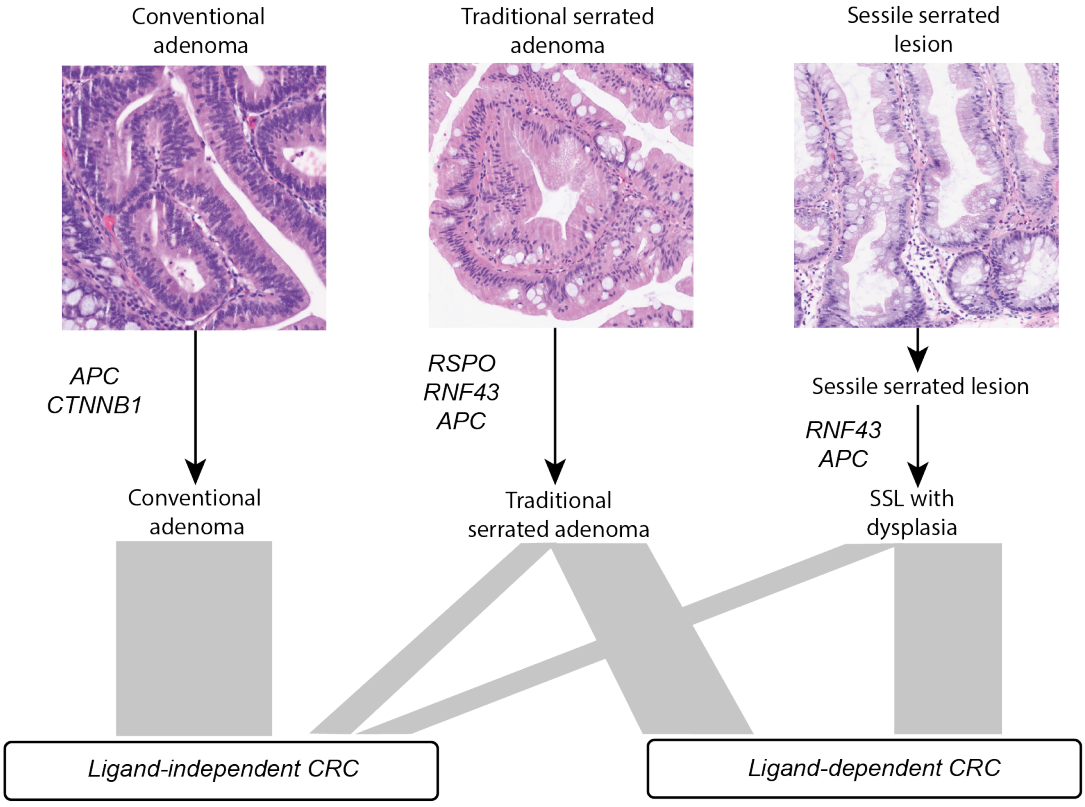


Figure 4.2: Molecular pathogenesis of different colorectal precursor lesions. Adapted from [176].

vide a hyperactivated, but not maximal WNT response [177]. On the contrary, ectopic expression of Rspo in *Apc*-mutant mice has been shown to result in reduction of proliferation and increase in apoptosis, highlighting that excessive WNT signalling has a negative impact on cell fitness [176]. The "just right" level of WNT seems to be variably set across different lesion types. This could be the result of different precursor and malignant CRC lesions following distinct evolutionary pathways, which in turn appear to select for either LD or LI mutations. Although the driving force of such selection is unclear, it may be explained by the variable cell-of-origin in different lesion subtypes [176]. Thus, strong constitutive WNT signalling might be required to induce conventional adenoma dysplasia in CBC, whilst a more contained WNT signalling might provide selective advantage in lesions that have been initiated by other drivers, e.g. epigenetic alterations in serrated lesions [179].

In the past decades extensive research has been undertaken to elucidate the specific contribution of the above described WNT genes and their associated mutations

to CRC development. Cell and animal models of aberrant WNT activation have aided in unveiling that WNT hyperactivation is sufficient to trigger increased proliferation, prevent cell death and halt differentiation in mutant cells. These features are thought to contribute to the general fitness of the mutant cells, their evolution to malignant cells and ultimately to the characteristics of a tumour. However, with increasing recognition of the role of the tissue environment on the natural selection of these mutation-driven characteristics, the elucidation of the interactions between malignant cells and their neighbouring wild-type cells becomes important. Recently, Yum *et al.*, Van Neerven *et al.* and Flanagan *et al.* provided insight into the dynamics of cancer cells and the role of the WNT pathway in this. Comprehensively, the authors demonstrated that malignant cells secrete factors that promote differentiation of neighbouring stem cells, thereby advancing the replacement of normal stem cells by those with cancer promoting mutations [167, 168, 180]. Using cell and animal models, Van Neerven *et al.* and Flanagan *et al.* revealed that *Apc*-mutant cells secrete the WNT negative regulator Notum. Because *Apc* mutation activates the WNT-signalling pathway downstream of Notum's site of action, the secreted Notum inhibits WNT signalling in neighbouring wild-type cells, without affecting the *Apc*-mutant cells. This aberrant WNT bestowed cell endogenous growth advantage coupled with suppression of neighbouring wild-type cells makes *Apc*-mutant cells super competitors [167, 168]. The same dynamics is expected for *CTNNB1*-mutant cells and their neighbouring cells, given the upstream nature of WNT hyperactivation as well as the increased NOTUM expression found in *CTNNB1*-mutant lesions. Conversely, the dynamics between LD-mutant cells, which have an upstream activation of the WNT, repressed expression of the negative regulators (including NOTUM) and secrete potent morphogens (in the case of *RSPO* fusions), and their neighbouring cells is yet to be described.

Although impressive progress has been made in our understanding of the role of the WNT signalling pathway in S-CRC, its role in CA-CRC is yet to be fully elucidated. Unlike S-CRC, where *APC* mutations are the most common WNT alteration, as well as an early and important event in carcinogenesis, in CA-CRC *APC* alterations are found in only 13% of the lesions. Moreover, in those cases where *APC* is mutated, the mu-

tations have either been reported to occur late in cancer, thus to not be important in the early stages of tumorigenesis, or to mainly occur in S-CRCs arising in IBD patients rather than true CA-CRCs. Similarly, *CTNNB1* mutations tend to also be very rare in CA-CRC. Nonetheless, despite the much lower prevalence of *APC* and *CTNNB1* mutations, nuclear β-catenin accumulation, a hallmark of canonical WNT signalling activation is a common finding in CA-CRCs. This suggests a dysregulation of the pathway. Several mechanisms of dysregulation in CA-CRC have been suggested including that: a) WNT could be epigenetically activated, thus bypassing the need for genetic alterations; b) WNT signalling could be stimulated by inflammation, possibly through NFKB1 and STAT3, whose elevated expression was shown to enhance β-driven transcription [105, 181]. More recently, the recognition of LD alterations as WNT-activating mutations and their association with an alternative carcinogenic pathway has opened the possibility that colitis-associated lesions might selecting for LD alterations. This is supported by a recent finding of *RNF43* mutations in 11% of the CA-CRC cases assessed [110], slightly higher than its prevalence in S-CRC (<10%) [18].

4.2 Study rationale, design study and aims

The aberrant activation of the WNT signalling pathway is a hallmark of CRC. Nevertheless, the role of the pathway in CA-CRC is unclear, with the most common alteration in S-CRC being present in only 13% of CA-CRC cases. Under the hypothesis that colitis-associated lesions might be selecting for LD alterations, the CA-CRC WNT landscape was investigated in two main lines work:

- Strand A: A comprehensive phenotyping of colitis-associated LGD comprising transcriptomic, genomic and histological profiling, with the aims to:

    1. Explore the WNT genetic and transcriptomic landscape in colitis-associated LGD;
    2. Assess the frequency and time of LD WNT mutations in CA-CRC precursor lesions.

- Strand B: An investigation of the cell dynamics in murine models of WNT disruption, based on the findings from the molecular phenotyping of the human precursor lesions, with the aims to:

    1. Determine why colitis-associated lesions might be preferentially selecting for LD alterations;
    2. Assess the role of RSPO3 on cell fate determination;
    3. Determine if RSPO3 promotes Paneth cell expansion.

4.3 Results

4.3.1 Phenotyping of colitis-associated precursor lesions

To assess the role of the WNT pathway in CA-CRC, the WNT landscape of colitis-associated LGD was investigated.

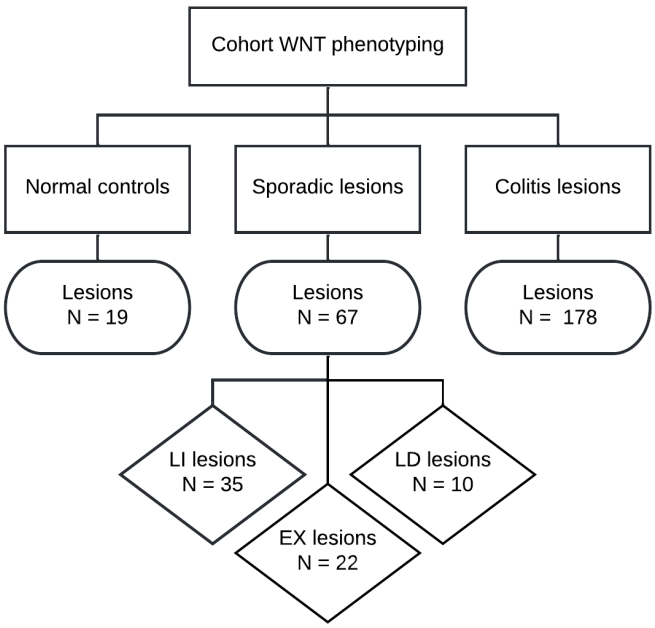


Figure 4.3: Breakdown of the colitis, sporadic and normal controls interrogated in this study. Key: EX - lesions without any known WNT alteration.

The investigation was carried out on FFPE colitis-associated LGD, sporadic precancerous lesions (divided into LD, LI and with no known WNT alterations based on

pre-established molecular ground truth) and normal controls (from patients without any known colorectal pathology) (Figure 4.3).

4.3.1.1 Assessment of *AXIN2* expression to predict ligand dependency in colitis-associated precursor lesions

To test the hypothesis that CA-CRC might be selecting for LD alterations, the expression of the previously established discriminatory biomarker *AXIN2* was assessed in colitis-associated LGD, with the aim to predict ligand dependency in these lesions. For this, transcriptomic profiling was performed on colitis-associated LGD, sporadic lesions and normal controls by 3'RNAseq (Figure 4.4).

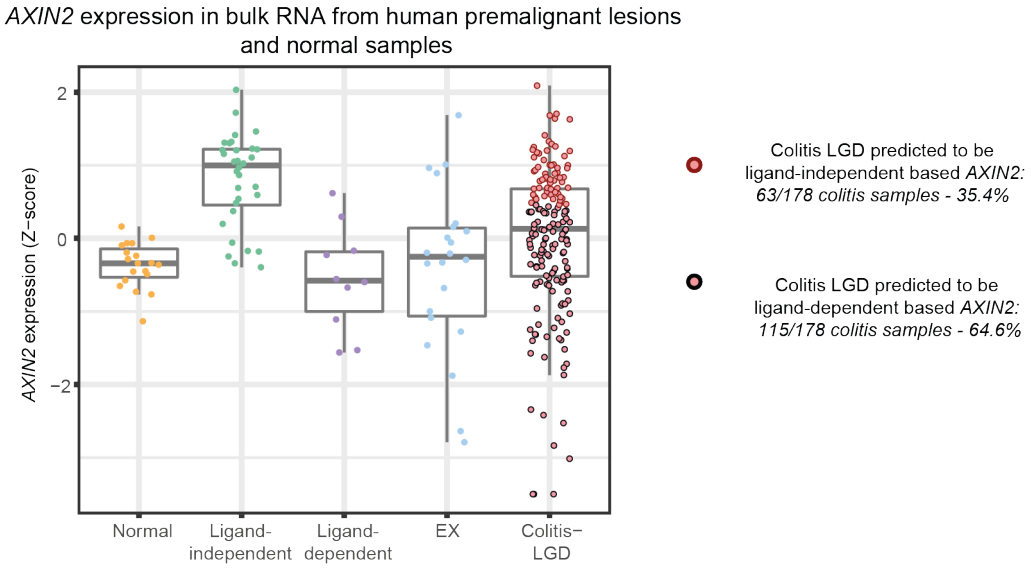


Figure 4.4: The expression of *AXIN2* was assessed in colitis-associated LGD against that of LI and LD sporadic lesions and normal controls. Furthermore, the sporadic cohort also included lesions without a known driver WNT alteration (EX). Variation in *AXIN2* expression can be seen among the colitis lesions. Using the lower quartile of the LI *AXIN2* expression as cut-off, colitis-associated lesions were predicted to be likely-LI (highlighted in red) or likely-LD (highlighted in black).

As per Kleeman *et al.*, *AXIN2* expression was low in LD and high in LI sporadic precursor lesions. In the colitis-associated LGD, a gradient in *AXIN2* expression was noticed. Using the *AXIN2* expression of the sporadic LI/LD precursor lesions, the lower quartile (i.e. bottom 25%) *AXIN2* expression of the sporadic LI polyps was used as cut-off to predict ligand-dependency. Colitis-associated lesions with *AXIN2* expres-

sion above the cut-off were predicted as likely-LI (63/178) and those below the cut-off as likely-LD (115/178). The majority of the colitis-associated LGD were predicted to be LD based on their *AXIN2* expression. To confirm this finding a multi-omic analysis was performed (outlined below).

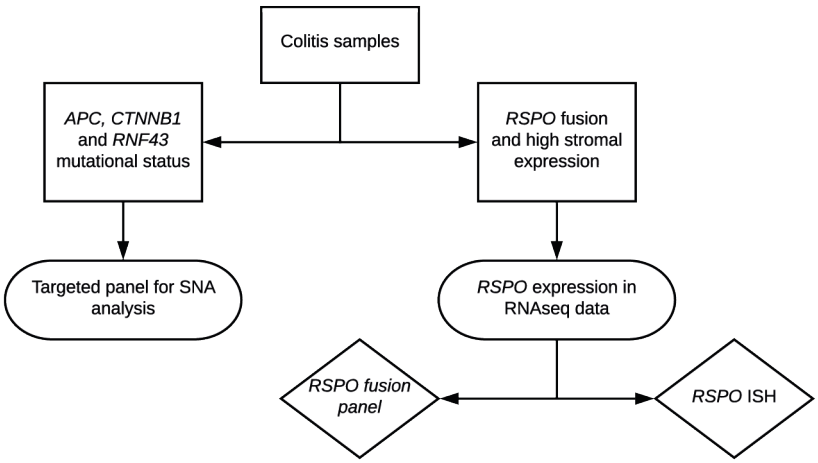


Figure 4.5: Workflow of the multi-omic analysis performed to confirm the findings from the *AXIN2* expression.

4.3.1.2 Mutational phenotyping for WNT genes of colitis-associated LGD

Mutational analysis of the colitis samples was undertaken to identify activating WNT alterations. Targeted DNA sequencing of key WNT driver genes (i.e. *APC*, *CTNNB1*, *RNF43*) was performed on the colitis-associated LGD with sufficient DNA i.e. 90 samples out of 178 samples (materials and methods 2.3.1.4). Due to technical issues results were only obtained for 32 samples. Of the 32 samples, two had *APC* mutations (6.25%) and seven had *RNF43* mutations (21.9%). The remaining 23 samples did not have SNAs in known WNT driver genes. No *CTNNB1* alternations were found in the colitis-associated LGD genotyped (Figure 4.6).

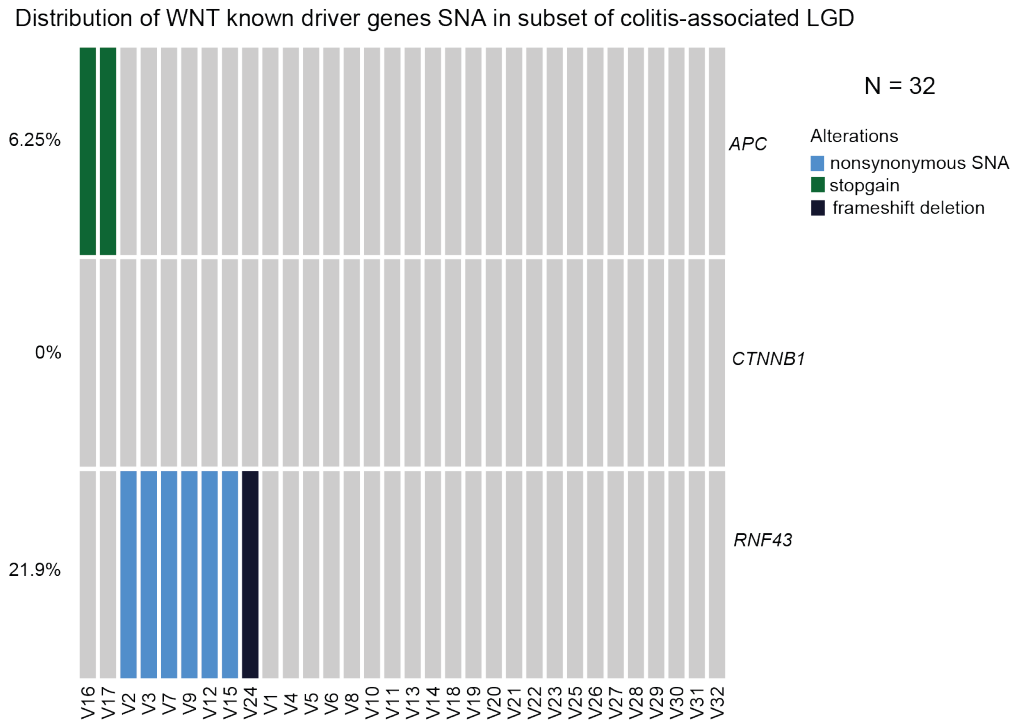


Figure 4.6: Oncoprint of WNT driver genes SNA.

4.3.1.3 Assessment of *RSPO* alterations

Next, the presence and frequency of *RSPO* alterations in colitis-associated LGD was assessed. Within the different RSPO subtypes, *RSPO2* and *RSPO3* were queried because they are the two most potent and highly expressed *RSPO*s in the intestine [182]. Moreover, activating translocations in these two *RSPO* subtypes have previously been shown to drive CRC development [18, 183–185].

First, the expression of *RSPO2* and *RSPO3* in colitis-associated LGD was interrogated, together with that of sporadic lesions and normal tissue in 3'RNAseq data. For both *RSPO2* and *RSPO3* varying expression were seen among the colitis-associated lesions, with several colitis-associated LGD showing high expression. Using the expression of the normal samples, sporadic lesions and sporadic LD lesions with *RSPO2* fusions, the upper fifteenth percentile of *RSPO2* expression within the entire cohort was used as cut-off for the upper end of the *RSPO2* expression continuum. Colitis samples with an *RSPO2* expression above this cut-off where categorised as having

"high" *RSPO2* (n = 30/178; highlighted in black in Figure 4.7 A). Similarly, based on the expression of the normal samples, sporadic lesions and sporadic LD lesions with known *RSPO3* fusion, the upper twentieth percentile of *RSPO3* expression within the entire cohort was used as cut-off for the upper end of the *RSPO3* expression continuum (n = 46/178; highlighted in red in Figure 4.7 B).

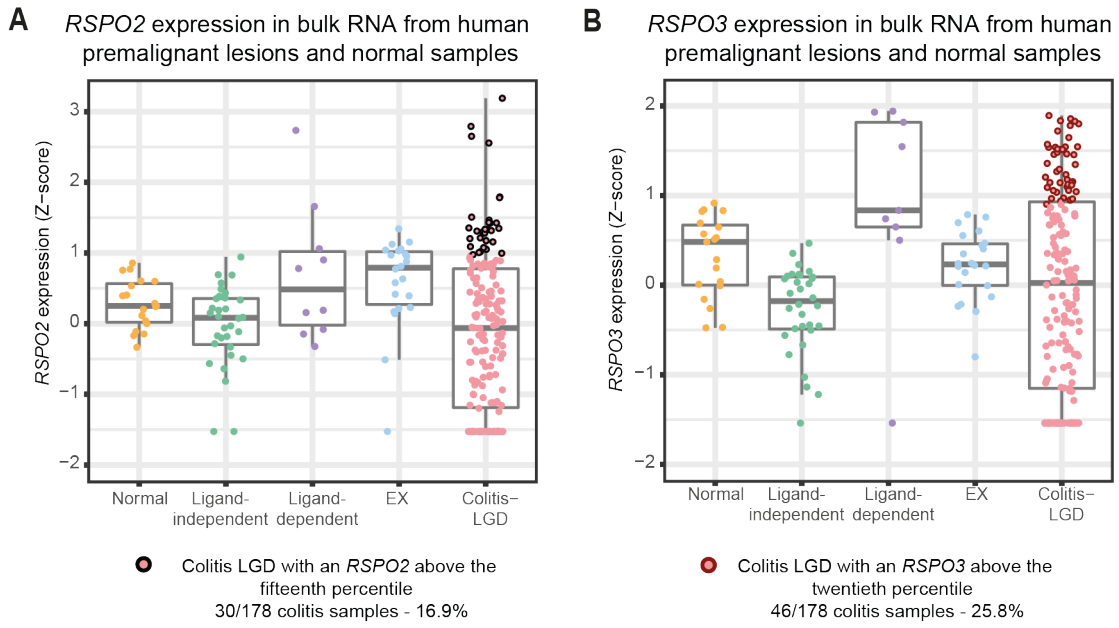


Figure 4.7: *RSPO2* and *RSPO3* expression in bulk colitis-associated LGD RNA sequencing data. **A.** Expression of *RSPO2* in colitis-associated LGD, sporadic precursor lesions and normal controls. Highlighted in black are the lesions with an *RSPO2* expression above the upper fifteenth percentile. **B.** Expression of *RSPO3* in the colitis-associated LGD, sporadic precursor lesions and normal controls. Lesions with an *RSPO3* expression above the upper twentieth percentile are highlighted in red.

High *RSPO* expression is suggestive of the presence of either *RSPO* fusions or high stromal RSPO, as previously shown by Kleeman *et al.* [18]. This in turn suggests that the colitis-associated LGD categorised as having high *RSPO2*/3 might have either one of these *RSPO* alterations. To confirm this, *RSPO2*/3 fusion panel and ISH was performed on these samples.

To identify *RSPO* fusions, an *RSPO* fusion panel was undertaken on samples with sufficient RNA. For *RSPO2*, out of the 30 samples with high *RSPO2* expression, the panel could be run only on 15 samples, whereas for *RSPO3*, it could only be run on 28 samples. Nonetheless, the fusion panel led to the identification of 11 fusions, of which

three were *RSPO2* fusions and eight *RSPO3* fusions (Table 4.1).

Table 4.1: Fusions detected in the colitis-associated LGD

Fusion	Samples with fusion	Novelty
CELF2-RSPO2	1	Previously described
EIF3E-RSPO2	1	Previously described
RSPO2-RBPMS	1	Previously described
PTPRK-RSPO3	1	Previously described
RSPO3-ARF1	1	Novel
RSPO3-CD53	1	Previously described
RSPO3-MGAT5	1	Previously described
RSPO3-OS9	1	Previously described
RSPO3-PIGR	1	Previously described
RSPO3-RNF146	1	Previously described
RSPO3-TACC1	1	Previously described

To confirm the results from the fusion panel and assess the samples for which the panel could not be run, ISH was undertaken looking for any profound epithelial expression of *RSPO2* and or *RSPO3*, which is indicative of *RSPO* fusion, or high stromal *RSPO* expression.

From the lesions with the upper fifteenth percentile *RSPO2* expression, two showed profound epithelial expression (Figure 4.8 C Left and Right). However, 26 showed inconclusive results and two could not be run due to the lack of tissue availability. From the samples with the upper twentieth percentile *RSPO3* expression, two showed profound epithelial expression (Figure 4.8 D Left), one showed high stromal expression (Figure 4.8 D Right) and the rest either showed inconclusive results or could not be assessed due to the lack of tissue availability. In both *RSPO2*/3 ISH, the inconclusive results were due to the poor quality of the FFPE sections. Collectively, four lesions were confirmed to have aberrant epithelial expression of *RSPO* consistent with a fusion (one of which previously identified with the fusion panel), and one to have high stromal *RSPO*.

From the assessment of *RSPO2*/3 in colitis-associated LGD, 76/178 were found to have *RSPO* expression within the upper end of the expression continuum. Of these, 15 had a confirmed fusion and one had confirmed high stromal expression. Notably, for the samples that had high *RSPO* expression, but were found to not carry *RSPO*

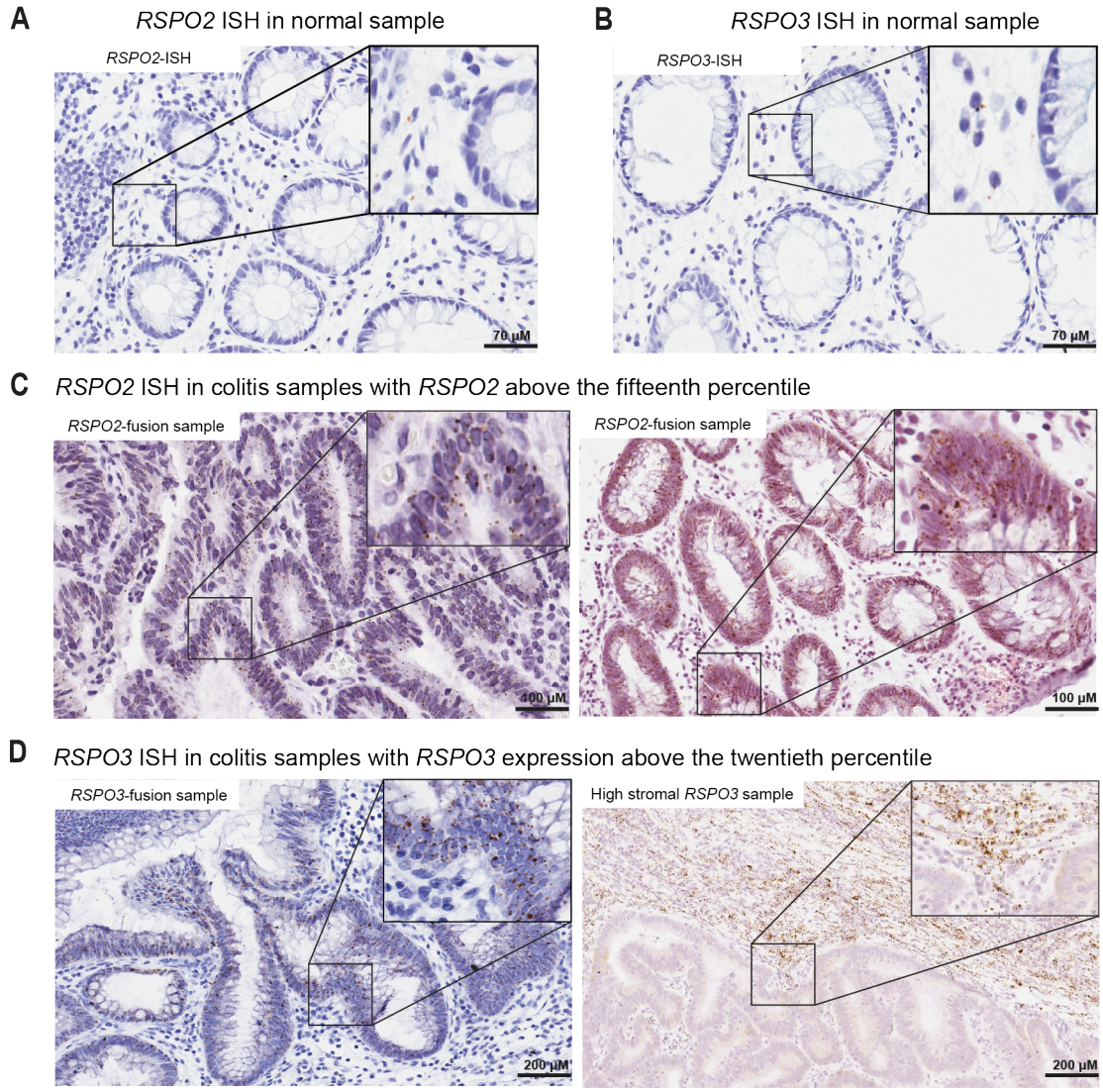


Figure 4.8: *RSPO2* and *RSPO3* ISH in colitis-associated LGD with *RSPO* expression above the set cut-offs and normal controls. **A.** Image of chromogenic ISH for *RSPO2* in a normal control sample to demonstrate non-pathogenic *RSPO2* expression. **B.** Image of chromogenic ISH for *RSPO3* in a normal control sample to demonstrate non-pathogenic *RSPO3* expression. **C.** Images of chromogenic ISH for *RSPO2* in two colitis-associated LGD. In both left and right images strong *RSPO2* expression can be seen in the epithelium (brown dots). The staining pattern is representative of *RSPO* fusion. **D.** Images of chromogenic ISH for *RSPO3* in two colitis-associated LGD. **Left.** A strong *RSPO3* hybridisation can be seen in the epithelium, representative of *RSPO* fusion. **Right.** A strong *RSPO3* expression in the stroma can be seen. This is typical of high stromal RSPO. X40 magnification.

fusions, high stromal RSPO cannot be excluded given the inconclusiveness of the ISH.

4.3.1.4 Comprehensive assessment of the multifaceted profiling of the colitis-associated LGD

A schematic outlining the results from the multifaceted analysis is provided in Figure 4.9 (an in depth version is provided in the appendix C.1).

Next, in an attempt to compare the frequency of the different WNT alterations, I focused on the small subset of lesions with complete characterisation, where molecular ground truth was established (n = 39). Within this subset *APC* mutations made up 5.1% of the alterations, *RNF43* represented 12.8% and *RSPO* alterations were present in 38.5% of the lesions. In 38.5% (n = 16) of the cases, no known WNT driver alteration was found (i.e. no SNA in WNT driver genes and no high *RSPO* expression or confirmed *RSPO* alteration). This was also seen for the sporadic lesions (referred to as EX). Whilst the possibility that these lesions might not carry any WNT alteration is valid, it is also likely that they carry an alteration in a gene yet to be described as a WNT driver gene. For instance, in the past years several genes have been implicated in WNT aberrant activation in CRC including *BCL9* and Bcl9-like (*BCL9L*) [186], aryl hydrocarbon receptor (*AHR*) [187], *AXIN2* [188]. Another possibility, is that they could be acquiring their WNT alteration through different means, including epigenetic silencing and CNA. Specifically with regard to CNA, colitis-associated LGD have been found to be enriched for loss of chromosomal arms 5q and 17q, location of *APC* and *RNF43* (shown in chapter 5). More simply, given the difficulty in detecting fusions from FFPE samples and that of sequencing *RNF43*, it is possible that some fusions and alterations are passing undetected. Interestingly, two lesions had both an *RSPO* and an *RNF43* alteration, although this is likely to be a sequencing artefact.

Collectively, although in a limited subset of samples and therefore to be taken with caution, these results show a higher frequency of LD alterations in colitis-associated LGD than LI mutations, with LD accounting for 56.4% of the WNT alterations detected in this cohort against the 5.1% of LI alterations. This suggests that colitis-associated lesions might be preferentially selecting for LD alterations (Figure 4.10).

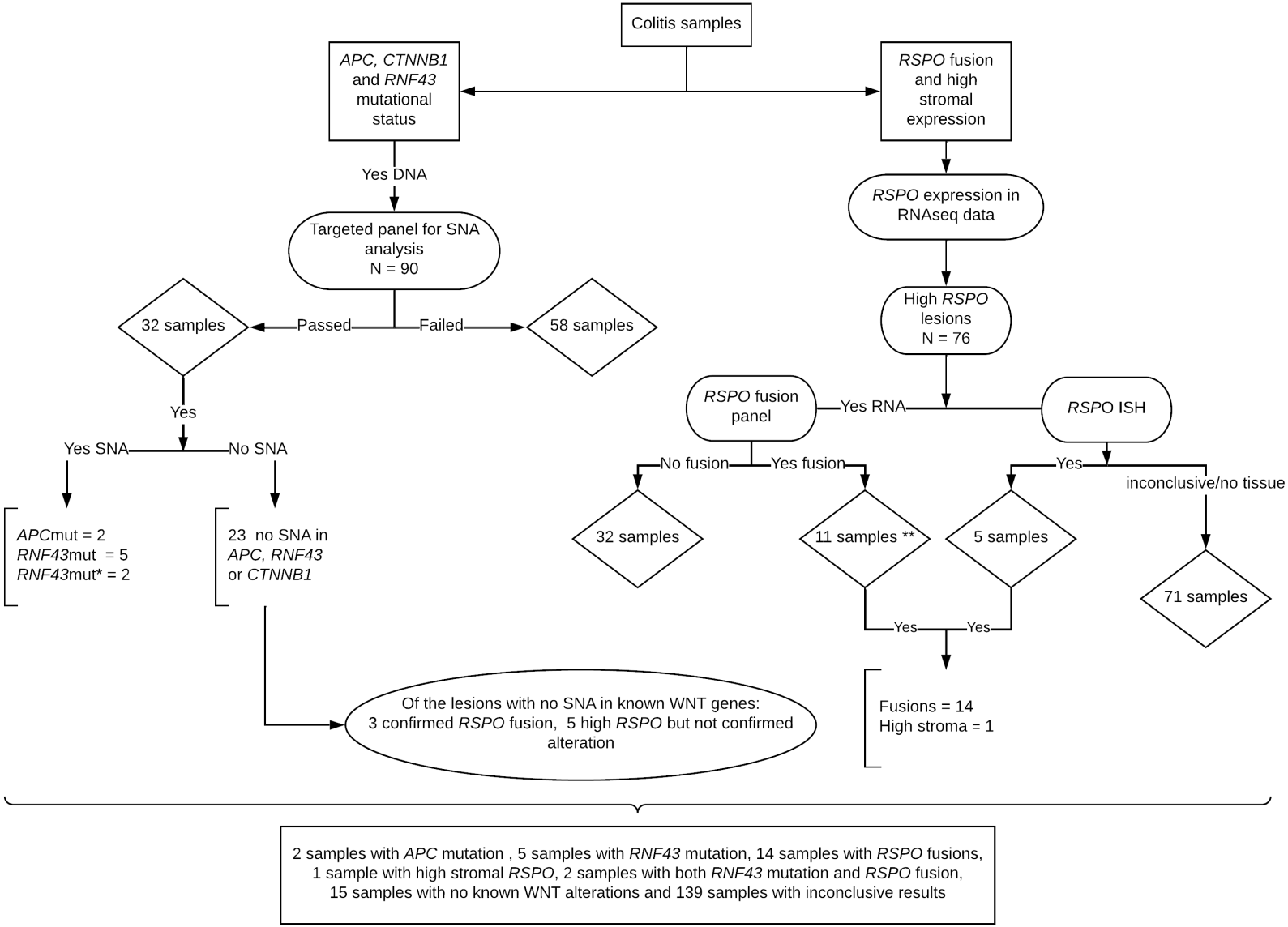


Figure 4.9: Breakdown of the multifaceted phenotyping of the colitis-associated lesions. The multifaceted analysis of the 178 colitis-associated LGD led to the identification of: two lesions with *APC* alteration, five with *RNF43* mutation, 15 with *RSPO* alterations, two with *RSPO* fusion and *RNF43* mutation. Moreover, 15 lesions were confirmed as not having any known WNT alterations. For the rest 139 lesions conclusive results could not be obtained due to the high drop-off of samples. Key: * - *RNF43* mutation with *RSPO* fusion, ** - 1 out of 11 fusion was confirmed by ISH.

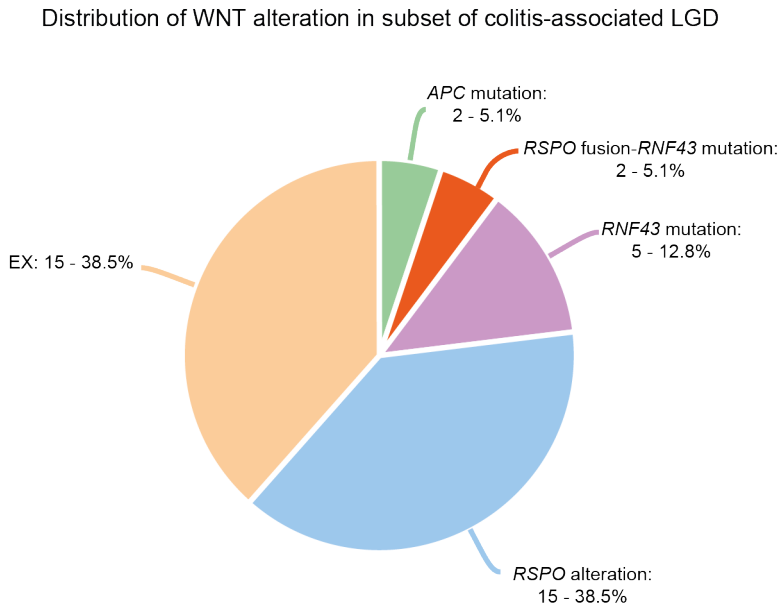


Figure 4.10: Distribution of WNT alterations in colitis-associated LGD subset. Within *RSPO* alterations, *RSPO* fusions and high stromal *RSPO* are included. Key: EX - lesions with no known WNT driver gene alteration

4.3.2 Assessment of WNT alterations in mouse models of WNT disruption

The molecular profiling of the human colitis-associated LGD suggested that colitis-associated lesions might preferentially select for LD alterations. The question naturally following this finding is why would they select for LD over LI alterations? LI alterations in intestinal stem cells were recently shown to give rise to super-competitor cells, which maintain their stemness through the constitutive activation of the WNT pathway, whilst prompting their neighbouring cells to differentiate by inhibiting WNT signalling. However, the effect of LD alterations on the cells carrying the mutation and their neighbouring cells are yet to be described. Consequently, to determine the reason behind the potential preferential selection of LD alterations in colitis-associated lesions, the cell dynamics in LD lesions and how this could be favoured in the colitic setting was investigated in human lesions and predominantly mouse models.

4.3.2.1 *NOTUM* expression pattern across human and mouse lesions bearing WNT alterations

Notum, a WNT deacylase that disrupts WNT ligand binding to frizzled receptors, is a well-recognised secreted WNT negative regulator. Notum is highly expressed in LI lesions. Recently, expression of Notum by *Apc*-mutant cells was found to have a WNT suppressive effect on wild-type neighbouring cells, thereby inhibiting proliferation and promoting their differentiation [167, 168]. To investigate the cell dynamics in LD lesions and assess LD cells' behaviour, the expression of *NOTUM* was compared between *APC*-mutant (n=3) and *RSPO*-fusion (n=3) human precursor lesions. As expected a profound *NOTUM* expression was seen in the *APC*-mutant lesions. Conversely, in the lesions with *RSPO* fusions no *NOTUM* expression was seen (Fig 4.11). This is congruent with WNT negative regulators being downregulated in LD lesions.

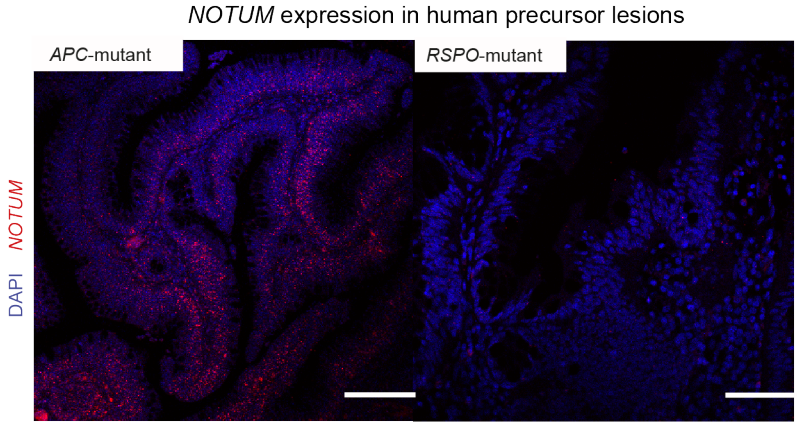


Figure 4.11: *NOTUM* expression in LD and LI human lesions. Images of fluorescent ISH for *NOTUM* in sporadic precursor lesions. Profound *NOTUM* expression can be seen in the *APC*-mutant lesion (Left), conversely to the absence of *NOTUM* expression in the lesion with *RSPO3* fusion (Right). Scale bars 50 μm.

An analogous investigation was then performed in *Apc*-mutant and *Rspo3*-overexpressing murine models. Three models of Rspo3 overexpression were used: a) Two epithelial models - $Rspo3^{inv/+}$;$Vil1\text{-}Cre^{ERT2}$ (*Rspo3Vil*) and $Rspo3^{inv/+}$;$Atoh1^{-/-}\text{-}Cre^{ERT2}$ (*Rspo3Atoh*), characterised by inducible constitutive Rspo3 expression from epithelial cells and modelling *RSPO* fusions; b) A stromal model, $Rspo3^{inv/+}$;$Gli1\text{-}Cre^{ERT2}$ (*Rspo3Gli*), characterised by inducible constitutive Rspo3 expression from stromal Gli-positive cells and

modelling high stromal RSPO lesions (details in material and methods 2.3.2.1).

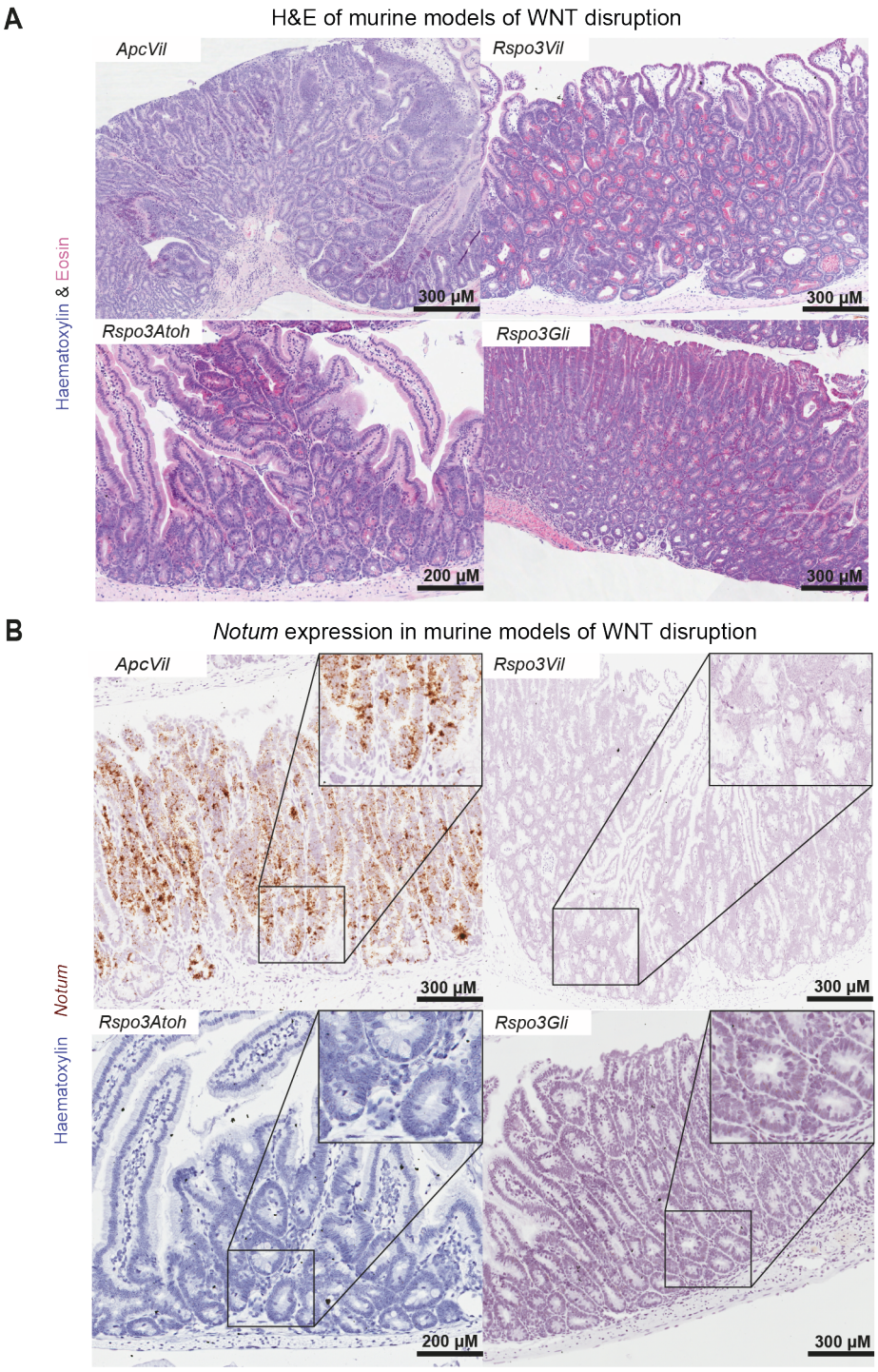


Figure 4.12: *Notum* expression in LI and LD lesions. **A.** H&E staining to show the phenotype of the mouse models of WNT disruption. **B.** Images of chromogenic ISH for *Notum*. High *Notum* expression is seen in the model of *Apc* disruption *ApcVil*, whilst no-to-very little expression of *Notum* can be noticed in the models of Rspo3 overexpression.

Chromogenic ISH for *NOTUM* was performed on the lesions from these models, followed by quantification of *Notum*-positive cells. In congruence with the human data, *Notum* expression was profound in the *Apc*-mutant model *ApcVil*, with an average of 4,618 positive cells in 1000x1000µm^2 area, whilst little-to-no *Notum* expression was found in all three models of Rspo3 overexpression (Figure 4.12 and 4.13).

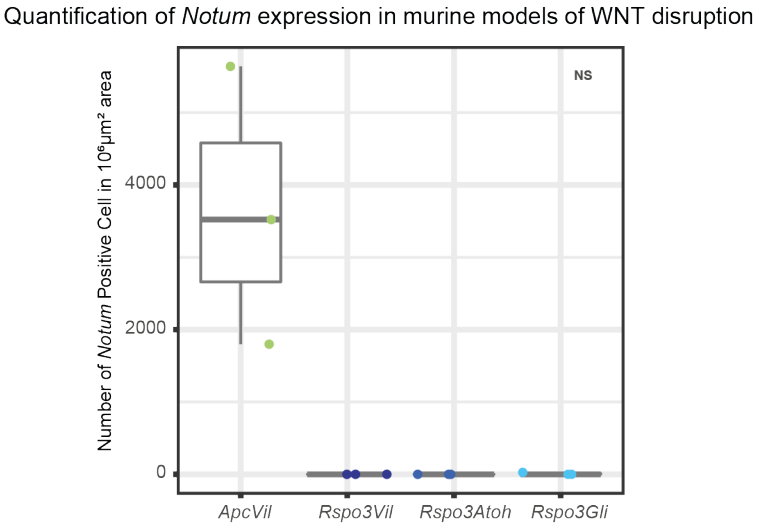


Figure 4.13: Quantification of *Notum* expression in mouse lesions. Between 1799-5639 *Notum*-positive cells in a 1000x1000µm^2 area were counted in the *ApcVil* model, while 0-27 *Notum*-positive cells in a 1000x1000µm^2 area were detected in the models of *Rspo3* overexpression.

The assessment of the expression of *NOTUM* in both human and mouse lesions showed that *RSPO*-overexpressing lesions do not express *NOTUM*. This is in agreement with the fact that WNT negative regulators, including *AXIN2* and *NOTUM*, are epigenetically silenced in LD lesions and suggests that the dynamics between cells with LD alterations and their wild-type neighbouring cells might be different from that of *APC*-mutant cells [18].

4.3.2.2 Characterisation of intestinal phenotype in murine models of WNT disruption

The comparison between the human and murine lesions carrying *Apc* mutations and *Rspo3* alterations revealed that cells with LD alterations do not express *NOTUM*, suggesting that they might not suppress their neighbouring cells' stemness. In the as-

sessment of the *Apc*-mutant and *Rspo3*-overexpressing models an important difference that was noticed in the intestinal phenotype induced by these different alterations was a dramatic increase in Paneth cells in the small intestine of the *Rspo3*-mutant models. Paneth cells, which are differentiated secretory cells located at the bottom of the crypt in between Lgr5-positive cells, produce WNT-ligands, as well as EGF and Notch, which in turn promote stem cell proliferation and maintenance. In fact, Paneth cells have been proposed as key nurse cells for Lgr5-positive cells. In the context of WNT signalling, an increase in Paneth cells would mean higher availability of secreted WNT-ligands, which would translate into increased WNT signalling.

To confirm the observation of an increased number of Paneth cells in the *Rspo3*-overexpressing models, combined staining for *Rspo3* and lysozyme (Paneth cell marker) was performed and the number of Paneth cells quantified. Proportion of Paneth cells, defined as number of Paneth cells over the total number of epithelial cells, was used as measure of number of Paneth cells to account for the difference in degree of hyperplasia across the models (materials and methods section 2.3.2.4). Both visually and quantitatively a significantly higher proportion of Paneth cells was found in the *Rspo3*-mutant models than the *Apc*-mutant model (Figure 4.14 and 4.15). Together with the Paneth cell hyperplasia seen in these mice, an extension in location of the Paneth cells from the base of the crypt towards the +4 region of the crypt and the isthmus was noticed (Figure 4.14).

The comparison between *Apc* and *Rspo3* models revealed a higher proportion of Paneth cells in the *Rspo3* models. This indicates that Rspo3 overexpression might cause an increase in Paneth cell number.

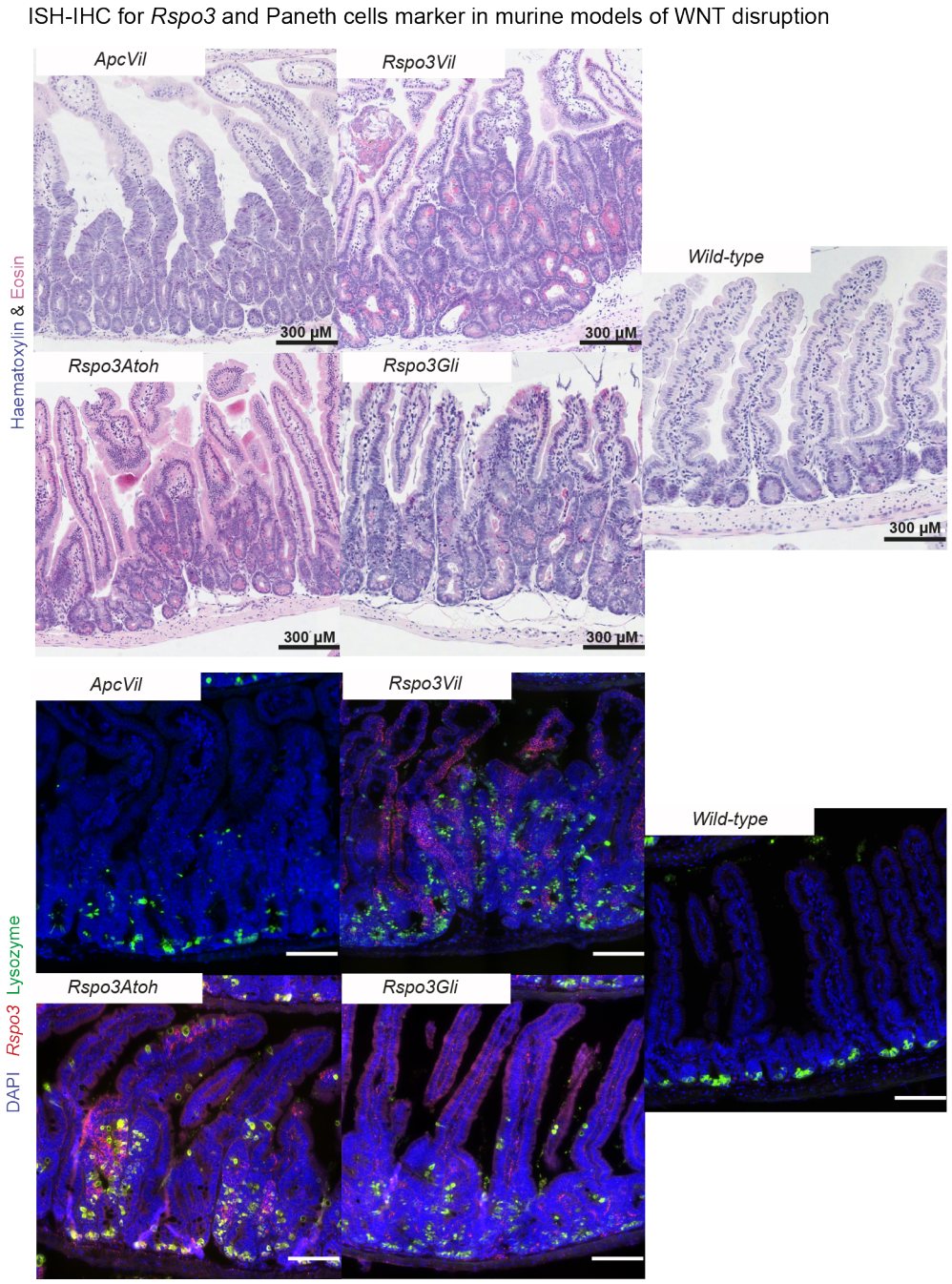


Figure 4.14: Paneth cells in LD and LI models of WNT disruption. **A.** H&E images of the mouse models of WNT disruption and a wild-type control to highlight the differences in phenotype. **B.** Images of combined fluorescent IHC-ISH for *Rspo3* and lysozyme of mouse models of WNT disruption and a wild-type control. In both H&E and IHC-ISH, a profound expansion of the Paneth cells can be noticed in the *Rspo3* models. Scale bars 300 µm.

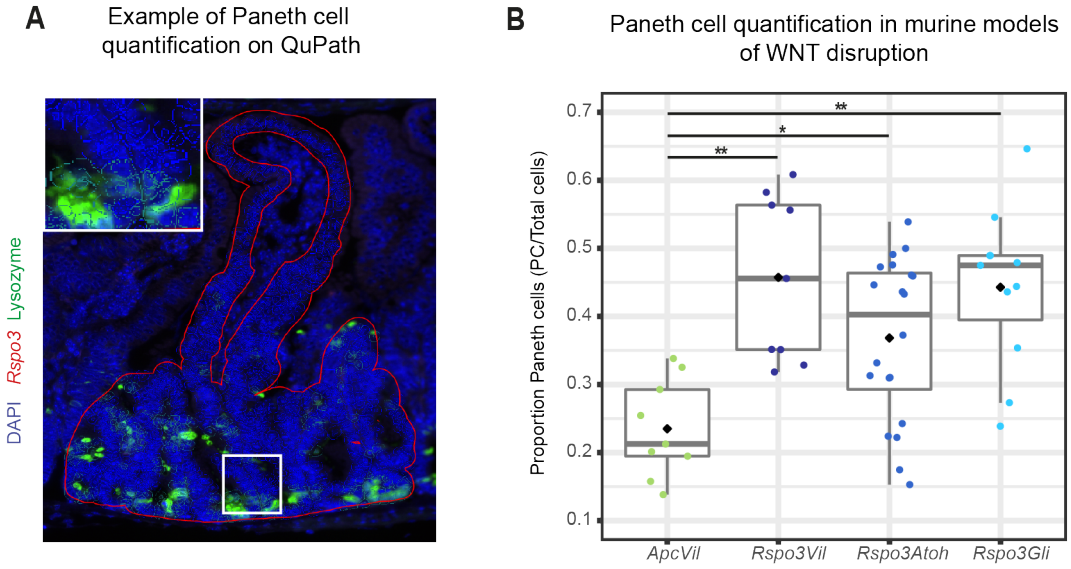


Figure 4.15: Comparison of the proportion of Paneth cells between the *Apc*-mutant model and the models of Rspo3 overexpression. **A.** An example of each area of quantification (marked in red), which comprised two crypts and one villous, and only epithelial cells. Five areas were quantified per mouse and three per genotype. The total number of cells and the number of lysozyme-positive cells (in green) were quantified using QuPath. These were then used to calculate the proportion of lysozyme-positive cells i.e. the proportion of Paneth cells. **B.** Plot of the proportion of Paneth cells (PC) across the different murine models. A higher proportion of Paneth cells was found in the *Rspo3* models (*Rspo3Vil, Rspo3Gli* and *Rspo3Atoh*), than the *ApcVil* model. Key: * adjusted p-value < 0.05; ** adjusted p-value < 0.01. Complete statistics in appendix Table C.1.

4.3.2.3 Assessment of the effects of Rspo3 on Paneth cells

The comparison of *Rspo3*-overexpressing models against the *Apc*-mutant lesions, showed a significantly higher proportion of Paneth cells in the *Rspo3* models, suggesting that Rspo3 might cause an expansion of these key nurse cells. To assess such postulation a case-control three timepoint experiment was undertaken in the *Rspo3Atoh* model.

The *Rspo3Atoh* model is characterised by tamoxifen-inducible expression of Rspo3 from progenitor Atoh1-positive cells. The latter reside at the +4 position in the crypt and serve as secretory progenitor cells, which give rise to either goblet cells or Paneth cells under homeostatic conditions. This makes the *Rspo3Atoh* model a good model to assess the effect of Rspo3 on Paneth cells and on cell fate determination in general. *Atoh1*$^{-/-}$-*Cre*^{ERT2} was crossed with either *Rspo3*$^{inv/+}$ LSL (cases) or $^{td/+}$*Tom* LSL (controls). The purpose was to track the fate of either Rspo3 or tdTomato expressing secretory progenitor cells in a LD aberrant WNT versus homeostatic tissue environ-

ment (further details on models in materials and methods 2.3.2.1).

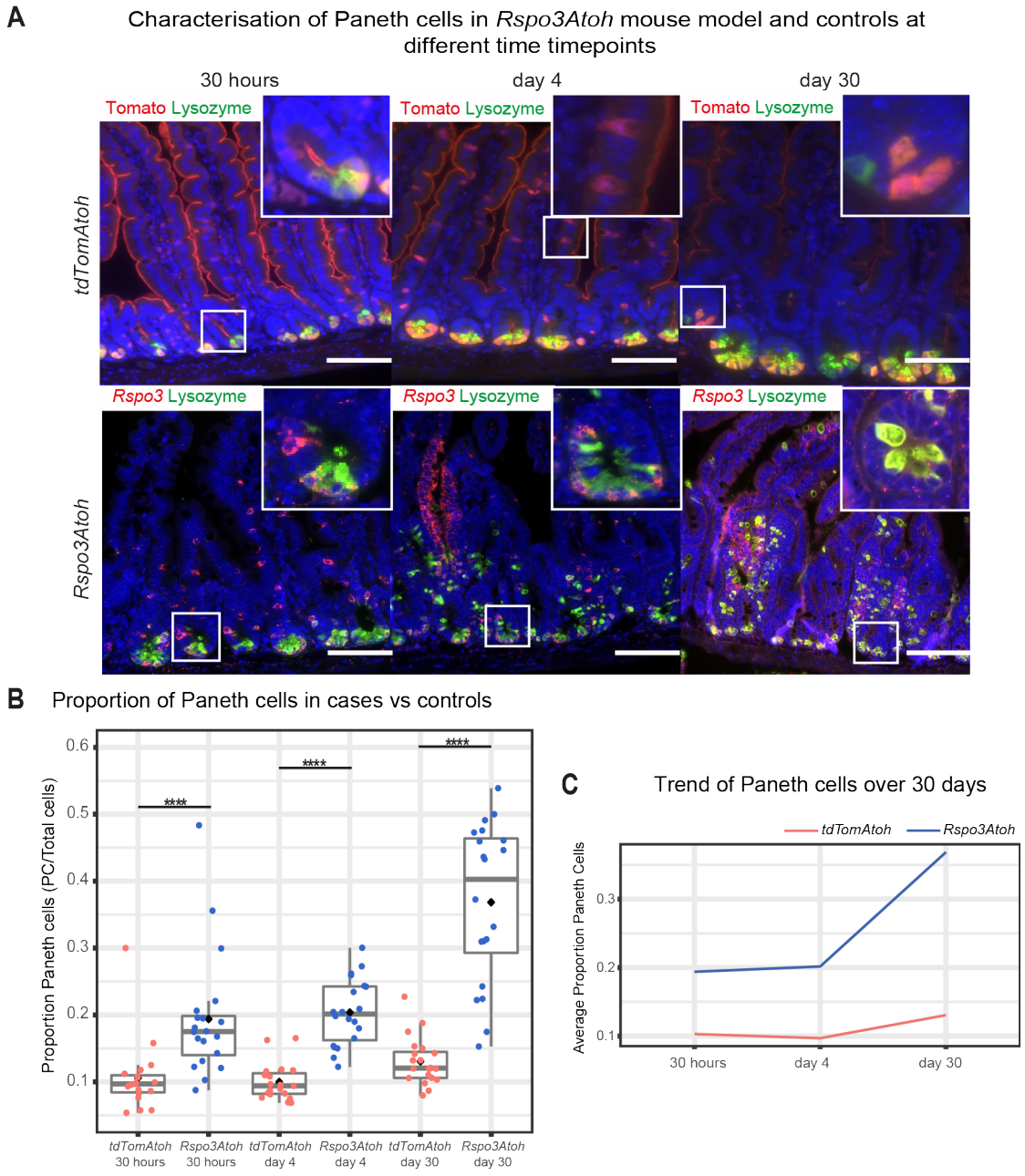


Figure 4.16: The effects of Rspo3 overexpression on Paneth cell number were assessed by comparing Paneth cell proportion in *Rspo3Atoh* mice at 30 hours, day 4 and day 30 from tamoxifen-induced Rspo3 expression in Atoh1-positive cells, and their associated *tdTomAtoh* controls. **A.** Images of combined *Rspo3* and lysozyme staining in the *Rspo3Atoh* samples, and combined Tomato and lysozyme staining in the *tdTomAtoh* mice. **B.** Plot of the proportion of Paneth cells (PC) in the *Rspo3Atoh* and *tdTomAtoh* models overtime, quantified as per Figure 4.15. **C.** Line plot to highlight the trend in proportion of Paneth cells overtime in the cases versus controls. The proportion of Paneth cells is significantly higher in the *Rspo3Atoh* mice compared to their controls. Scale bars 300 µm. Key: **** adjusted p-value < 0.0001. Complete statistics in appendix Table C.1.

Following tamoxifen induced recombination, *Rspo3Atoh* and *tdTomAtoh* mice were collected at 30 hours, day 4 and day 30. Their Paneth cells were assessed by staining (materials and methods 2.3.2.4). Quantification of Paneth cells in the case and controls showed a significant increase in proportion of Paneth cells in the *Rspo3Atoh* model at all timepoints compared to the *tdTomAtoh* controls (Figure 4.16 B). Notably, comparison between the different timepoints of *Rspo3Atoh* mice showed a progressive increase in Paneth cell number, with the highest proportion being observed at day 30 (Figure 4.16).

The data demonstrated that Rspo3 expression in secretory progenitor cells causes an expansion of Paneth cells. This is supported by previous studies reporting an increase in abundance of Paneth cells in Rspo3 induced lesions and loss of Paneth cells in *Rspo3* knockout mice [13, 123].

4.3.2.4 Assessment of the effects of Rspo3 on goblet cells

Assessment of the *Rspo3Atoh* and *tdTomAtoh* mice demonstrated that Rspo3 overexpression causes an expansion of Paneth cells. The finding raises questions on whether the effect of Rspo3 is specific to Paneth cells or whether Rspo3 causes an expansion of all differentiated secretory cells arising from Atoh1-positive cells. To determine this, goblet cells were quantified in the *Rspo3Atoh* and *tdTomAtoh* mice.

Goblet cells, which are located in the villi, arise from the differentiation of Atoh1-positive cells. Using *Rspo3* and tdTomato as markers of lineage tracing Atoh1-positive cells in the *Rspo3Atoh* and *tdTomAtoh* mice, respectively, *Rspo3*/tdTomato-positive cells in the villi were identified as mature goblets and quantified. At 30 hours and day 4, goblet cells were less abundant in the *Rspo3Atoh* mice than in *tdTomAtoh*. However, at day 30, by which lineage tracing cells are expected to have shed into the lumen based on the rapid intestinal cell turnover, a few goblet cells were still seen in *Rspo3Atoh* mice, while no lineage tracing goblet cells were found in the *tdTomAtoh* mice. This suggests a longer persistence of *Rspo3*-positive goblet cells than their wild-type counterpart. The findings suggest that Rspo3 overexpression skewing of Atoh1-positive cells towards Paneth cells may be at the expense of goblet cells. However, it does appear that excess WNT also prolongs survival of differentiated goblet cells.

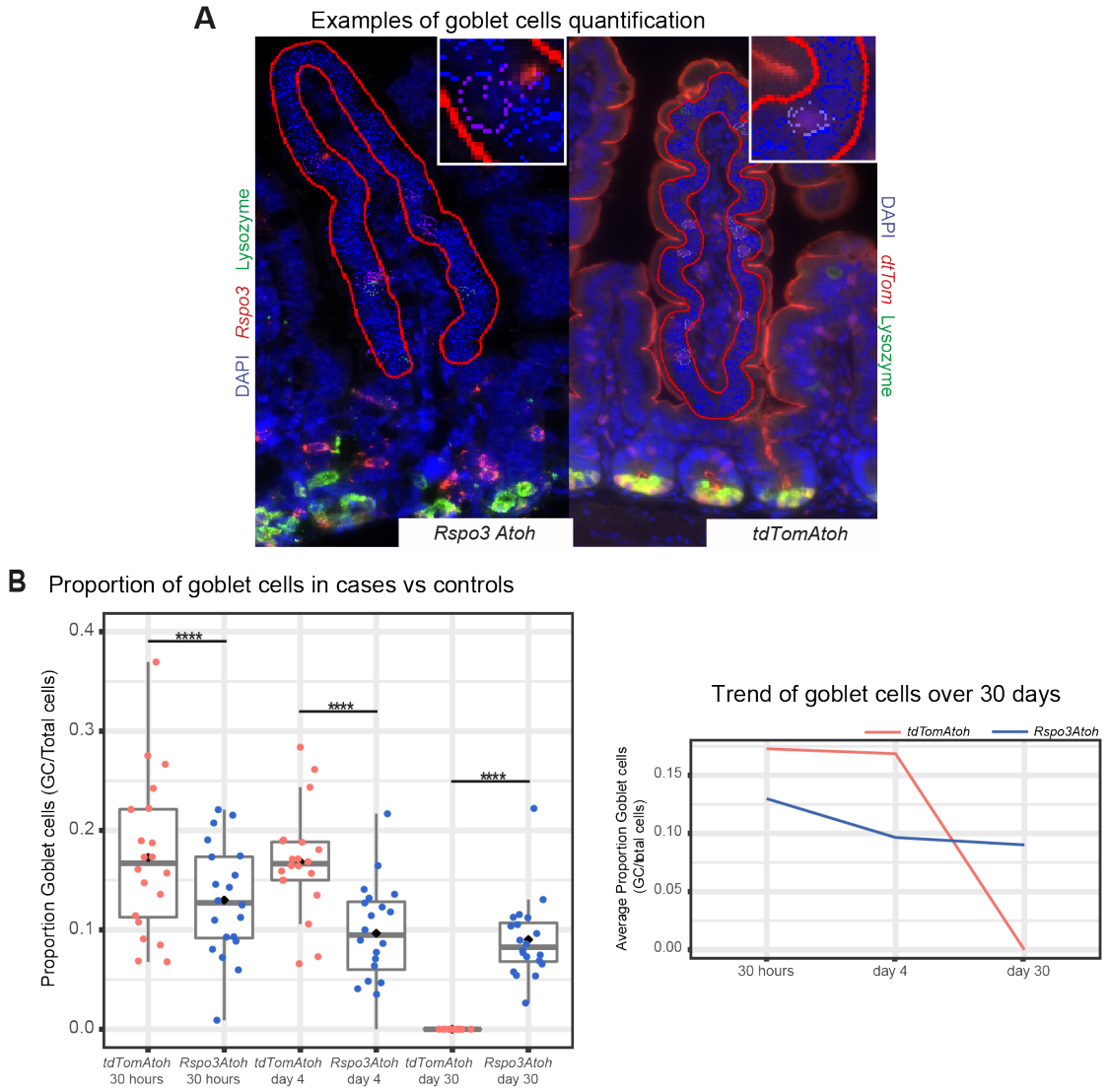


Figure 4.17: The effects of Rspo3 overexpression on goblet cell number were assessed by comparing the proportion of goblet cells in *Rspo3Atoh* mice at 30 hours, day 4 and day 30 following induction of Rspo3 expression in Atoh1-positive cells, against their associated *tdTomAtoh* controls. **A.** Examples of each area of quantification (marked in red), which comprised only the villous (above the isthmus) and only epithelial cells. Goblet cells were measured as *Rspo3*-positive lysozyme-negative (red only) cells in the villous in the *Rspo3Atoh* mice and tdTomato-positive lysozyme-negative cells (red only) in the villous in the *tdTomAtoh* mice. The number of goblet cells and total cells were quantified by QuPath. These were used to calculate the proportion of goblet cells. **B.** Plot of the proportion of goblet cells (GC) in the *Rspo3Atoh* and *tdTomAtoh* models overtime. **C.** Line plot to highlight the trend in proportion of goblets cells overtime in the cases and controls. Key: **** adjusted p-value < 0.0001. Statistical analysis details in Table C.1.

Noteworthy, as goblet cells were quantified as the number of Rspo3/tdTomato-positive lysozyme-negative cells in the villi, the total number of goblet cells may include a few non-Paneth, non-goblet cell progenitors. For a more accurate quantifica-

tion of goblet cells mucin staining, which is the gold-standard for mucin identification, would be advantageous.

4.3.2.5 Assessment of the autocrine/paracrine nature of Rspo3's effect on cell fate determination

Having shown that Rspo3 skews cell fate towards Paneth cells, the next question to be tackled was whether Rspo3's effect was solely autocrine or paracrine as well. In

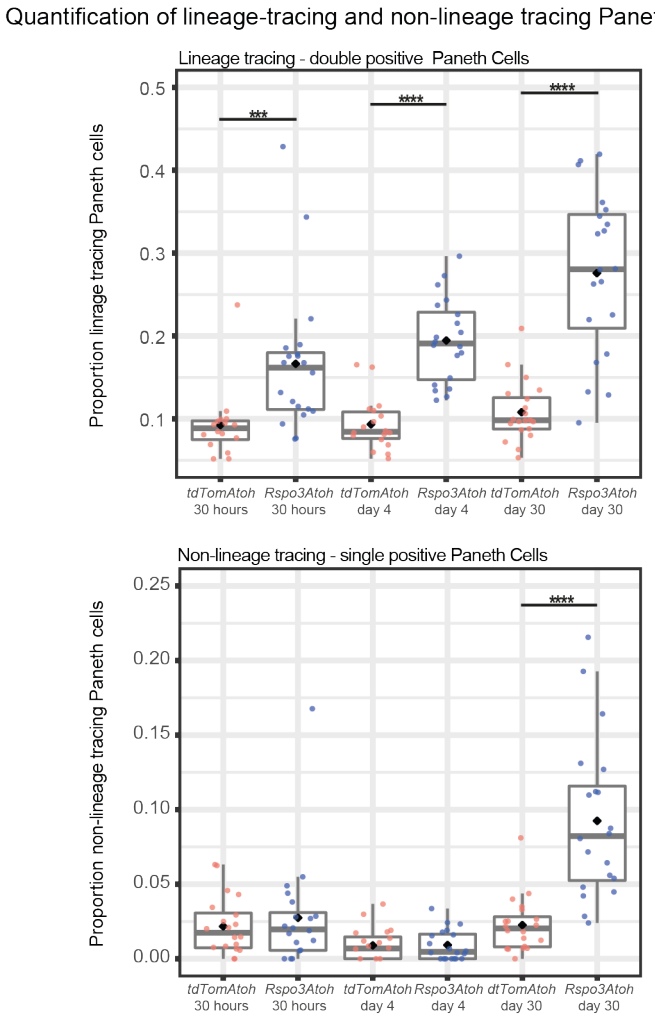


Figure 4.18: Assessment of whether Rspo3 has a cell endogenous or exogenous effect on cell fate determination, through quantification of *Rspo3*-positive and -negative Paneth cells. A higher proportion of *Rspo3*-positive Paneth cells can be seen at all timepoints of the *Rspo3Atoh* model, which confirms an autocrine effect of Rspo3. Although the number of lysozyme-only positive cells is similar to that of the controls at 30 hours and day 4, at day 30 a significantly higher proportion of the cells is observed in the *Rspo3Atoh* mice, suggesting that Rspo3 has also a paracrine effect. Key: **** adjusted p-value < 0.0001. Complete statistics in appendix Table C.1.

other words, was Paneth cell expansion the result of Rspo3-expressing Atoh1-positive cells proliferating and differentiating into Paneth cells or was Rspo3 expressed by the mutant Atoh1-positive cells having an exogenous effect on the neighbouring wild-type cells, causing them to also proliferate and differentiate into Paneth cells. To answer this question, the number of *Rspo3*-positive Paneth cells and *Rspo3*-negative Paneth cells was assessed in the different timepoint *Rspo3Atoh* mice, together with their respective controls. The quantification revealed the effect of Rspo3 on cell fate determination to be both cell endogenous and exogenous. In detail, an increase in *Rspo3*-positive Paneth cells could be seen from the first available timepoint, at 30 hours, with the highest proportion of Paneth cells being seen at day 30. The expansion of *Rspo3*-positive Paneth cells confirms the autocrine effect of Rspo3 on mutant-Atoh1-positive cells. The number of *Rspo3*-negative Paneth cells was similar to the controls at 30 hours and day 4, although a significant increase is noticed by day 30. This suggests that Rspo3 has also a cell exogenous effect on wild-type neighbouring cells, which is however only visible by day 30 (Figure 4.18).

4.3.2.6 Paneth cells in human colitis-associated lesions

In the *Rspo3* models, the Paneth cell expansion resulting from increased Rspo3 expression was histologically detectable as profound Paneth cell hyperplasia, with expansion of Paneth cells from the base of the crypt to the +4 region in the small intestine, and some Paneth cell metaplasia in the colon (Figure 4.19 A). Interestingly, Paneth cell metaplasia, the occurrence of Paneth cells in areas in which they are not normally found (for example the colon), is a common finding in IBD [189]. Moreover, dysplasia with increased Paneth cell differentiation has recently been described as a type of non-conventional dysplasia occurring in IBD patients [141]. Based on these observations, I wanted to investigate whether a correlation between the histological phenotype of increased Paneth cell differentiation and high *RSPO* expression, through *RSPO* fusion or aberrant high stromal expression, existed. In attempt to do so, a small subset (n = 28) of colitis-associated LGD, which had previously been histologically scored as dysplasia with increased Paneth cell differentiation by at least one expert pathologist and

transcriptomically profiled, was interrogated. Specifically, their *RSPO2* and *RSPO3* expressions were assessed (Figure 4.19).

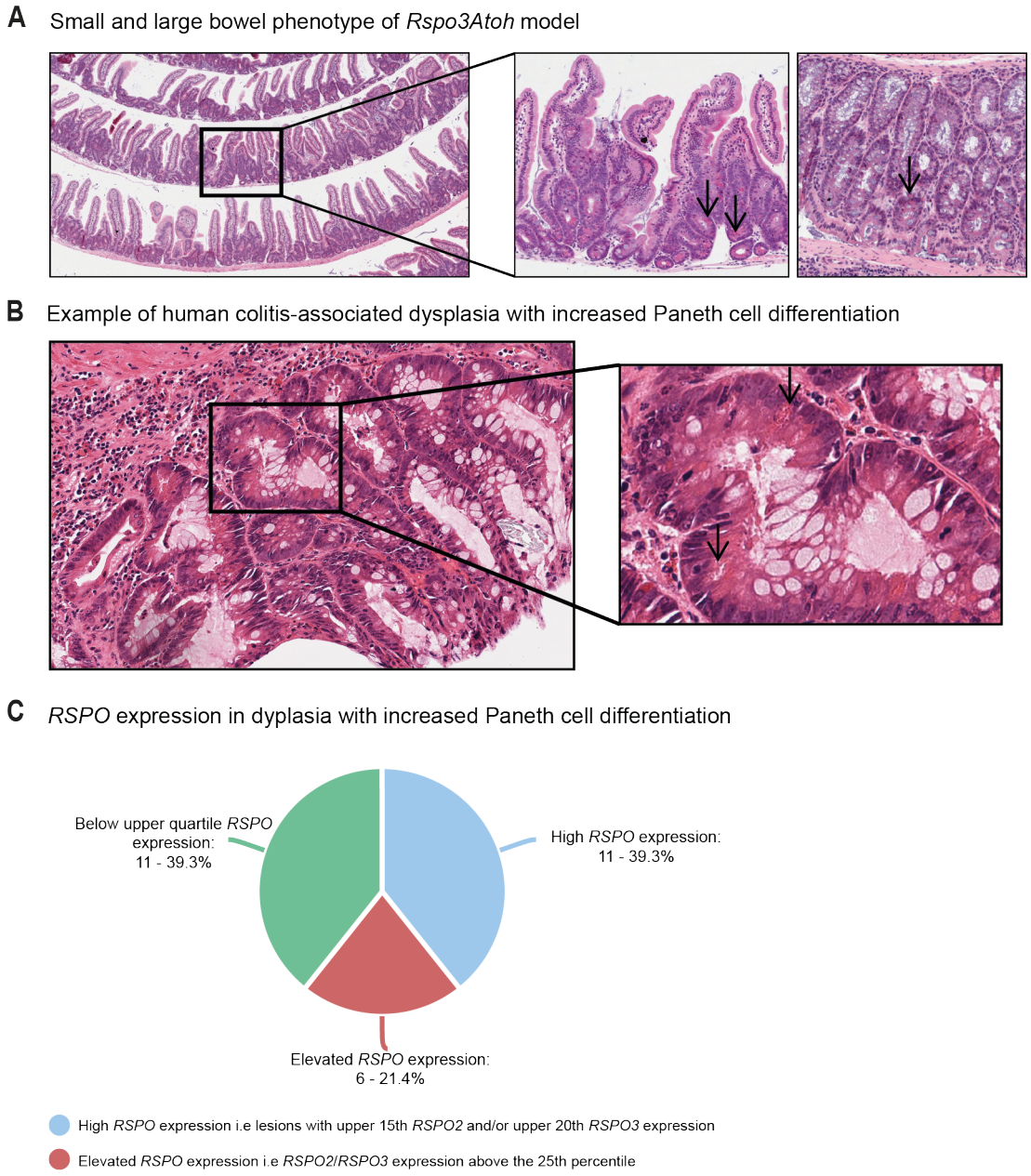


Figure 4.19: Association between Paneth cell hyperplasia, metaplasia and increased differentiation, and *RSPO* increased expression. **A.** The intestinal phenotype of *Rspo3Atoh* mice, with focus on the Paneth cell hyperplasia and metaplasia. Arrows pointing at Paneth cells. **B.** Example of human colitis-associated dysplasia with increased Paneth cell differentiation. Arrows pointing at Paneth cells. **C.** Distribution of *RSPO2*/3 expression in human colitis-associated dysplasia with increased Paneth cell differentiation.

Of the 28 lesions, 39.3% had a high *RSPO* expression, i.e. an *RSPO2* expression above the upper fifteenth percentile and/or *RSPO3* expression within the upper twentieth

percentile, 21.4% had an *RSPO2/3* expression above the upper quartile but below the cut-offs set for high *RSPO* expression and the rest 39.3% had below upper quartile *RSPO* expression. Whilst one of the lesions had a confirmed fusion, the others did not have a confirmed *RSPO* alterations. Collectively, 60% of the lesions had *RSPO* expression above the upper quartile. Although in a very small cohort and with the limitations of not having a confirmed *RSPO* alteration in the vast majority of the lesions, this result suggests that there might be an association between increased Paneth cell differentiation and high *RSPO* expression in human lesions. However, future work in a larger fully characterised cohort is required to determine this with confidence.

The data from this second half of the investigation showed that *Rspo3*-positive cells do not suppress the stemness of their neighbouring cells, instead they secrete Rspo, which generates a WNT milieu that favours both cell endogenous and exogenous skewing of secretory cell fate towards Paneth cells. The data also suggest that there might be an association between increased Paneth cell differentiation in human colitis-associated dysplasia and high *RSPO* expression.

4.4 Key findings

- In the limited set of colitis-associated lesions with complete data available, a greater proportion of colitis-associated LGD were LD than LI.
- Colitis-associated LGD seem to preferentially select for LD alterations.
- *Rspo3*-positive cells do not suppress the stemness of their neighbouring wild-type cells.
- Rspo3 causes an expansion of Paneth cells, source of WNT ligands.
- Secreted Rspo3 in the *Rspo3Atoh* model favours both cell endogenous and exogenous skewing of secretory cell fate towards Paneth cells.
- Dysplasia with increased Paneth cell differentiation might be associated with increased *RSPO2/3* expression.

4.5 Discussion

Aberrant WNT activation is a hallmark of CRC. However, little is known about the role of the WNT signalling pathway in CA-CRC. Here, comprehensive profiling of colitis-associated LGD allowed for the assessment of the WNT signalling landscape in CA-CRC. The multifaceted profiling of LGD suggested an enrichment of LD alterations, thereby hinting at a preferential selection for LD alterations in CA-CRC. Although technical issues meant that I could only establish the comprehensive molecular ground truth in a small subset of lesions, this is supported by two main lines of evidence. First, using *AXIN2* expression thresholding established in sporadic tumours, the majority of the colitis-associated LGD within the entire colitis cohort were predicted to be LD. *AXIN2* has been validated as discriminatory biomarker in both sporadic precursor lesions and cancer, with a sensitivity and specificity >90% [18]. Second, the frequencies of *AXIN2*, *CTNNB1* and *RNF43* alterations detected in the small subset of colitis-associated LGD with established molecular ground truth were similar to previous frequencies reported in CA-CRC [96, 110, 112]. To exemplify, similarly to the findings in this study, where *RNF43* was altered in 12.8% of the lesions, *RNF43* mutation has previously been reported to occur in 11% of CA-CRCs [110]. Noteworthy, the frequency of *APC* alteration was lower than in more recent studies [96, 110], but in congruence with older studies [190, 191].

Distinguishing between LD and LI alterations is important for therapeutic reasons. Unlike LI WNT activation, which has proven difficult to target, LD activation, being dependent on WNT ligand binding to their receptors to induce downstream activation of the signalling, retains sensitivity to WNT ligand inhibitors [176]. WNT ligand inhibition can be achieved through either deprivation of extracellular ligands or reduction of suppression of the negative feedback mechanisms using demethylating agents [18]. Based on this, three main types of agents have been developed: porcupine inhibitors, anti-RSPO and demethylating agents. Porcupine inhibitors, which induce inactivation of WNT ligands have been shown to repress WNT signalling, reduce tumour size and induce intestinal differentiation in murine models of LD CRC [192, 193].

In lesions with *RSPO* fusions, anti-RSPO was shown to cause inhibition of xenograft tumour growth and tumour regression by blocking the autocrine signalling loop [194, 195]. Both anti-RSPO and porcupine inhibitors have entered early phase clinical trials, where preliminary results in LD sporadic lesions have shown partial responses [196, 197]. However, they have also been associated with bone toxicity (which could be reduced with administration of drugs inhibiting bone degradation) [176, 197, 198]. Based on the observation that LD lesions depend on repression of negative regulators, demethylating agents can be used to reverse methylation of the genes involved in the negative feedback loops [176]. Treatment with azacitidine, a demethylating agent and approved treatment for myelodysplastic syndrome, in a CRC cell line with an *RNF43* mutation and comparatively low *AXIN2* expression, was shown to cause an increased in *AXIN2* expression [176, 199]. A different approach that circumnavigates the issue of bone toxicity associated with the administration of anti-RSPO and porcupine inhibitors is the use of asparaginase. Asparaginase is an enzyme that degrades the non-essential amino acid asparagine. Activation of WNT signalling upstream of glycogen synthase kinase (GSK3), such as in LD lesions, has been found to induce potent sensitisation to asparaginase. Sensitisation to asparaginase is mediated by the fact that GSK3 proteosomal degradation, which is a source of asparagine, is reduced by LD GSK3 inhibition. Reduced asparagine due to GSK3 inhibition in LD coupled with treatment with asparaginase causes a depletion of asparagine, which was found to be important for carcinogenic cell growth in leukaemia. Pre-clinical studies in CRC organoids and xenograft models showed asparaginase to be highly toxic to LD tumours, thereby suggesting it could be a viable and well-tolerated treatment for patients with LD CRC. Collectively, these studies show that several therapeutic strategies have been and are being developed targeting LD WNT activation. Thus, the finding that the majority of colitis-associated LGD might be LD opens the possibility for them to be directly targetable with these agents and increases the need to be able to molecularly identify these tumours.

The preferential selection of LD alterations in colitis-associated LGD is proposed to be the combined result of the cell dynamics between cells with LD alterations and their

neighbouring cells, and the requirement for increased regenerative capacity of the intestinal epithelium of IBD patients. In the context of chronic inflammation, such as in IBD, where the intestinal epithelium is subjected to cycles of inflammation-induced injury and regeneration, I propose cells carrying *Rspo3* fusions to provide a fitness advantage in a regenerative milieu through cell autonomous and non-cell autonomous skewing of residual cell behaviour. Skewing of secretory cells towards Paneth cells and thereby increasing Paneth cell proportion may encourage regenerative healing through an increase in Paneth cell-secreted antimicrobial peptides and trophic factors. Paneth cells are the main source of antimicrobials, such as α-defensins, c-type lectins, lysozymes and phospholipase A2, which protect against enteric bacterial infections. Furthermore, increased colonic Paneth cells in IBD has previously been proposed as an adaptive reaction in which α-defensins constitute a protective barrier that helps prevent secondary mucosal infection in already inflamed bowel [200]. Paneth cells are also an important source of WNT ligands, thus would encourage stem cell behaviour in residual cell through WNT ligand expression. Importantly, WNT-ligand secretion by Paneth cells, together with the lack of expression of WNT negative regulator Notum and the expression of Rspo3 *per se* would create an amplified and WNT ligand enriched milieu, which is believed to support increased cell proliferation. Hence, it can be hypothesised that alterations that promote healing and stemness would be favoured over alterations, such as *Apc* mutations, that give rise to super-competitors. Another factor that supports the hypothesis of LD being preferentially selected for in colitis-associated LGD is associated with the "just right theory". In an inflammatory environment, such as that of IBD, the intestinal epithelium is already subjected to upregulation of WNT signalling, usually from the stroma. In this context, LD mutations serve to amplify endogenous WNT signal, thus provide a more graduated and context dependent WNT, as opposed to the instantaneous, cell endogenous WNT excess generated by LI constitutive activation, which may provide an excessive WNT drive in cells that have exited the WNT-high crypt basal stem cell niche. Furthermore, based on previous observations that colitis-induced injury causes loss of Lgr5-positive cells [29] and that *APC* mutations generally occur in LGR5-positive cells, it can be further hypothesised that

APC mutations could also be negatively selected for during inflammation. Conversely, *RSPO* fusions in LGR5-negative cells would be able to fix within the intestinal epithelium as the LGR5-negative cells repopulate the crypt following inflammation.

Focusing on the Rspo3-induced skewing of secretory cell fate, Atoh1-positive cells are secretory progenitor cells that give rise to goblet and Paneth cells through yet to be fully elucidated differentiation signals [201]. This suggests that increased Atoh1-positive expansion would result in increased goblet and Paneth cells. However, results from the quantification of goblet and Paneth cells showed an increase in solely Paneth cell proportion, suggesting that WNT might be the differentiation signal that separates the development of Paneth cells and goblet cells from Atoh1-positive cells. This hypothesis is in line with the locations of the two secretory cell types: Paneth cells being located in the WNT-high environment of the crypt base, whereas goblet cell differentiation occurs as cells travel along the crypt-villus escalator, which is a WNT-low environment. It is also supported by a previous study showing that MAPK suppression in the crypts increased WNT signalling promoting Paneth and stem cells, whilst high-MAPK supported goblet cell properties [202].

Lastly, histologically Rspo3 overexpression was seen to induce Paneth cell hyperplasia and metaplasia in murine models. Moreover, assessment of human colitis-associated LGD led to the suggestion that colitis-associated dysplasia with increased Paneth cell differentiation could be associated with aberrant high *RSPO* expression, thus with *RSPO* alterations. On the other hand, *RNF43* mutation has previously been associated with mucinous morphology in biliary malignancies [203], as well as ovarian [204] and pancreatic cancer [205]. In CRC, *RNF43* mutation was linked to mucinous differentiation due to the association between MSI tumours and mucinous morphology, but whether mucinous differentiation is induced by *RNF43* mutation or reflective of the association with serrated pathway is yet to be determined [176]. However, recent association of *RNF43* mutation with signet-ring cell carcinoma [206], a rare subtype of colorectal cancer characterised by accumulation of mucins in the cytoplasm, and the association between *RNF43* mutation and mucinous histology in other cancers seem to support the hypothesis that *RNF43* mutation causes such phenotype. Nevertheless,

further work is needed to determine whether *RNF43* mutation causes mucin differentiation and whether this could be detected histologically. The observation that different histological patterns might be induced by *RSPO* and *RNF43* alteration suggests that different WNT alterations, might be associated with specific histological features. The association between WNT alterations and histological features is important because it has the potential to become a simple and inexpensive histological marker that could help predict a tumour's WNT alteration. This, as previously discussed, would be important in establishing treatment strategies. Future work should be undertaken using manual and automated digital pathological approaches to determine whether WNT alterations can be distinguished based on specific histological features.

4.6 Limitations and Future work

This study had two main limitation:

- The molecular profiling of human colitis-associated lesions was greatly impacted by the number of samples that could be fully characterised. This was due to the lack of sufficient material (i.e. DNA, RNA or tissue sections for ISH), the poor quality of FFPE samples, particularly for ISH and targeted RNAseq, and sample processing issues. Consequently, although a trend was observed, whereby more colitis-associated lesions seemed to select for LD alterations, validation in a second complete cohort is required.

- Although a general comparison between an *Apc*-mutant model and the *Rspo3*-overexpressing models was performed, a direct comparison between the effects of *Apc*-mutation and Rspo-overexpression on secretory cell fate in Atoh1-positive cells was not undertaken. For this investigation, an *Apc*-mutant allele (e.g. Apc^{tm2Cpe}, characterised by truncation of exon 14 resulting in an inactive *Apc* gene) should be crossed with the *Atoh1* allele. The resulting mice would then be used to recapitulate the multiple timepoint experiment undertaken on the *Rspo3Atoh* mice.

Further to the work to be undertaken in the future outlined above, two other lines

of enquiry should be further investigated:

- The cell dynamics between *Rnf43*-mutant cells and their neighbouring cells should be assessed to determine the effect of *Rnf43* mutation on cell fate determination.
- Insight into the histological features, suggested to be induced by different WNT alterations, warrants future work using digital pathological approaches and AI algorithms to determine whether WNT alterations can be distinguished based on specific histological features.

Chapter 5

The early detection of colitis-associated colorectal cancer

This chapter focuses on the early detection of CA-CRC, with the aim to identify a histological or molecular strategy that can enable early and objective identification of cancer risk in IBD patients. This study comprises three strands:

- Strand A: Early detection of CA-CRC by histopathological assessment;
- Strand B: Early detection of CA-CRC via CNA burden;
- Strand C: Early detection of CA-CRC by cytohistological assessment.

5.1 Introduction

CA-CRC remains the most feared complication of IBD. Consequently, to address their increased cancer risk, IBD patients are currently offered endoscopic surveillance, often with concomitant biopsy and histological assessment. The aim of this management strategy is to detect dysplasia as precursor lesion of CA-CRC [207]. In the UK, the IBD surveillance program was first introduced in 1971 at St Mark's Hospital, London, and since then has had several alterations guided by advances in the understanding of the disease and in endoscopic technologies [93]. The current UK guidelines, similar to the European and American, recommend surveillance to commence between 6-10 years

following onset of symptoms in adult IBD patients, except for patients with concomitant PSC, who should be offered annual surveillance from diagnosis [207]. Patients are stratified into low, intermediate and higher risk based on degree and extent of inflammation, presence of post-inflammatory polyps or strictures, dysplasia, PSC and family history of CRC under 50 years. Patients with dysplasia are categorised in the higher risk group and provided with the option to undergo endoscopic resection followed by surveillance colonoscopy or offered preventive colectomy, commonly in the cases of multifocal dysplasia [208].

The IBD surveillance program is thought to have played a key role in cancer prevention and early detection. A review of the surveillance program in St Mark's Hospital found a decrease in advanced cancer incidence rate over the four decades since its introduction [93]. Furthermore, endoscopic surveillance, together with other factors, including advances in endoscopic technology and management, better treatment and improved disease management is believed to have contributed to the reduction in CA-CRC incidence [93]. Nevertheless, the surveillance program is beset with difficulties and limitations. Colitis-associated dysplasia is challenging to detect at endoscopy as these lesions are often flat with indistinct margins and/or arise in a field of inflammation, scarring and post-inflammatory polyps. Furthermore, when detected their assessment is hindered by the poor inter-observer agreement among pathologists in their grading and the inability to confidently discriminate between regenerative epithelium, colitis-associated dysplasia and sporadic adenoma. Discriminating between sporadic, which have an estimated prevalence of 6.2% in IBD patients, and colitis associated lesions in IBD patients is particularly challenging, although important [209–211]. Although cancer risk in sporadic lesions is similar to that of sporadic adenomas occurring in disease-free segments (0-6%) and can be safely managed by endoscopic resection, their histological distinction from colitis-associated lesions is challenging [212]. Further to these diagnostic process-based difficulties, a number of disease-intrinsic factors have been shown to contribute additional difficulty. Firstly, the natural history of progression of CA-CRC from inflamed mucosa, to dysplasia, to cancer, has yet to be fully elucidated. Moreover, field cancerisation, the presence of widespread onco-

genic mutations in large patches of non-dysplastic epithelium, is a common finding in IBD patients. These render endoscopic identification of visible lesions and histological assessment of resected tissue an imperfect gold standard in early cancer detection.

Together these process-intrinsic and disease-intrinsic challenges impact the effectiveness of the current management strategy. These difficulties contribute to the high rate of interval cancers, i.e. cancers diagnosed within 3 years of having normal results on colonoscopy and before the next screening appointment, which can be as high as 30%. Furthermore, the chance of identifying previously undetected cancer upon histological assessment of surgical resections from patients undergoing prophylactic colectomy following dysplasia diagnosis is as high as 25% and 50% in those with LGD and HGD, respectively [93]. Conversely, other patients are subjected to frequent intrusive endoscopic surveillance regimes without ever developing CRC. In fact, the majority of LGD do not progress to HGD or cancers, further highlighting the lack of understanding of natural progression of the disease and how it negatively impacts the effectiveness of the surveillance program. The surveillance programme is also labour-intensive and costly, with an approximate cost to the UK NHS of £17,557/quality-adjusted life year [90]. Consequently, there is an imminent need to identify an adequate and effective risk stratification method that can allow stratification of patients whose dysplasia are at high-risk of progressing to cancers from those who are at low-risk. Several groups have attempted to address this need by trying to stratify patients based on:

- Lesion phenotype/macroscopic appearance;
- Lesion histological/microscopic appearance;
- Lesion molecular phenotype.

5.1.1 Stratification based on lesion macroscopic appearance

The Graham group investigated data collected from UC patients diagnosed with LGD between 1993-2012 as part of the UC surveillance program at St Mark's Hospital and identified a number of risk factors that could be adapted to stratify IBD patients with LGD into high or low risk of progression to HGD/CRC. Specifically, they found an

association between increased progression risk to HGD or CRC and lesions with the following macroscopic features: a) non-polypoid or endoscopically invisible, b) larger than 1 cm and c) preceded by indefinite dysplasia [211]. Recently, a retrospective multicentre validated cohort study by the same group confirmed and refined this stratification criteria. The study found significant association between progression and four features: a) larger than 1 cm dysplasia, b) unresectable or incomplete endoscopic resection, c) moderate to severe histological inflammation within 5 years of LGD diagnosis, and d) multifocality. These clinicopathological features were used to derive a risk prediction tool, which was validated and embedded in a publicly available web tool that can be used in the clinical setting, the UC-CaRE [208].

5.1.2 Stratification based on lesion microscopic appearance

Focusing on lesions' microscopic appearance for the identification of a risk stratification method, Choi *et al.* [141] and Lee *et al.* [142] investigated histological subtypes of colitis-associated colorectal dysplasia and their association with CRC risk. Literature on colorectal dysplasia is mostly based on well-recognised intestinal-type adenomatous dysplasia, also referred to as conventional dysplasia. However, more recently other histological patterns, collectively referred to as nonconventional dysplasia, have been identified. In a multicentre clinicopathological study investigating 106 dysplastic/serrated IBD-patient derived lesions, Choi *et al.* identified six morphologic patterns of nonconventional dysplasia and outlined morphological criteria for both conventional and nonconventional dysplasia subtypes. The authors also reported the commonality of nonconventional dysplasia in IBD, particularly among IBD patients with CRC [141]. In a follow-up study investigating the potential association between nonconventional dysplasia and CRC in dysplastic lesions from IBD patients, Lee *et al.* suggested nonconventional dysplasia, particularly crypt cell dysplasia (CCD), hypermucinous, goblet cell deficient (GCD) and SSL-like histological subtypes, to be associated with a higher risk of CA-CRC progression [142].

5.1.3 Stratification based on lesion molecular phenotype

CNAs are a common early event finding in CA-CRC. Aneuploidy has been shown to occur prior to or co-occur with dysplasia, with few CNAs occurring in most IBD patients. Several studies have reported a correlation between CNA, histological grade and progression risk in colitis, which could potentially be used to better stratify patients. Rubin *et al.* found a correlation between aneuploidy in mucosal biopsy specimens and the severity of histological abnormality in the specimens [213]. Furthermore, in patients without detected precursor lesions, the presence of aneuploidy in histologically non-dysplastic mucosa was associated with a greater risk of subsequent dysplasia development. The correlation between dysplasia and aneuploidy was later confirmed by Hartmann *et al.* [214] and Baker *et al.* [96]. Specifically, in a cross-sectional analysis of CNA in non-dysplastic mucosa, LGD, HGD and cancer, Baker *et al.* found that although CNA begin to accrue in non-dysplastic bowel, the progression from LGD to HGD often involved a punctuated CNA increase, where CNAs accrue and stabilise in the population in 'bursts' rather than continuously over time. Whilst, no significant difference in CNA burden was observed between HGD and cancer [96].

5.2 Study rationale, design study and aims

The current management strategy of IBD patients' increased CRC risk relies on endoscopic surveillance with concomitant histological assessment for the detection of dysplasia as precursor lesion of CA-CRC. However, the strategy is of limited efficacy, labour intensive and costly. This highlights the need for a stratification strategy that can provide an early and objective identification of cancer risk, which in turn can inform interventional decisions and guide endoscopic surveillance regimes, thereby enabling effective and cost-efficient management. To this end, an integrated, retrospective, comparative analysis was undertaken on dysplastic tissue from IBD patients. Patients and their associated lesions were divided into two cohorts (illustrated in Figure 5.1), defined by disease outcome over a period of 5 years:

- Progressors: IBD patients who developed HGD or CRC in the 5 years following

LGD detection at colonoscopy, where LGD was the maximal detected severity.

- Nonprogressors: IBD patients who did not develop HGD or CRC in the 5 years or more following a LGD detection at colonoscopy, where LGD was the maximal detected severity.

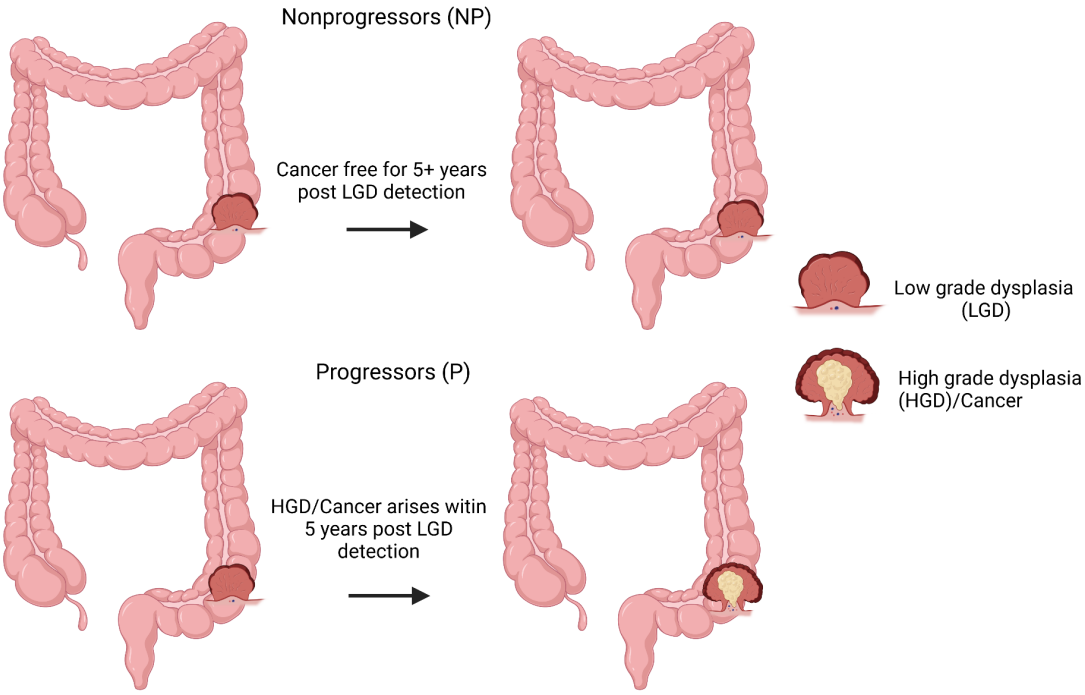


Figure 5.1: Illustration of progressor and nonprogressor patient/lesion groups.

The primary aim of this chapter was to identify and/or assess a histological or molecular strategy that can enable early and objective identification of cancer risk in IBD patients. The study comprised three main strands of work:

- Strand A: A histological classification of nonprogressor and progressor LGD using the classification criteria outlined by Choi *et al.*, followed by an assessment of the association between nonconventional dysplasia and progression, with the aims to:

  1. Determine whether histological assessment can aid in cancer risk stratification of LGD;

2. Test consensus agreement and practical application of the Choi *et al.* criteria in a real world multicentred setting;
3. Validation test for the association of nonconventional dysplasia with increased cancer risk.

This analysis was based on the Choi *et al.* and Lee *et al.* papers [141, 142].

- Strand B: A high-resolution CNA assessment of LGD using lpWGS, with the aims to:

  1. Determine whether progressor and nonprogressor lesions differ in CNA burden;
  2. Assess whether CNA burden can be used to discern between progressor and nonprogressor lesions.

  This investigation was based on Baker *et al.* finding that progression from LGD to HGD is marked by an increase in CNA [96].

- Strand C: A digital pathology analysis investigating cellular features of progressor and nonprogressor LGD, with the aims to:

  1. Determine whether molecular findings correlate with morphological cellular changes visible on slide;
  2. Determine if progression risk can be determined at the cellular level with the aid of digital pathology.

5.3 Results

5.3.1 Histopathological assessment of colitis-associated progressor and nonprogressor dysplasia

Choi *et al.* [141] and Lee *et al.* [142] proposed a novel classification criteria for colitis-associated dysplasia, based on specific histological features, and suggested an association between pathologically defined lesion categories and clinical risk for progression.

Here an assessment of the applicability of the Choi *et al.* classification system and the association between histopathological categories and cancer risk was undertaken.

5.3.1.1 Demographics and lesion details

Strand A of the study was undertaken on 263 histological regions from 149 precursor lesions, comprising 243 IBD regions and 20 regions of sporadic lesions (Figure 5.2).

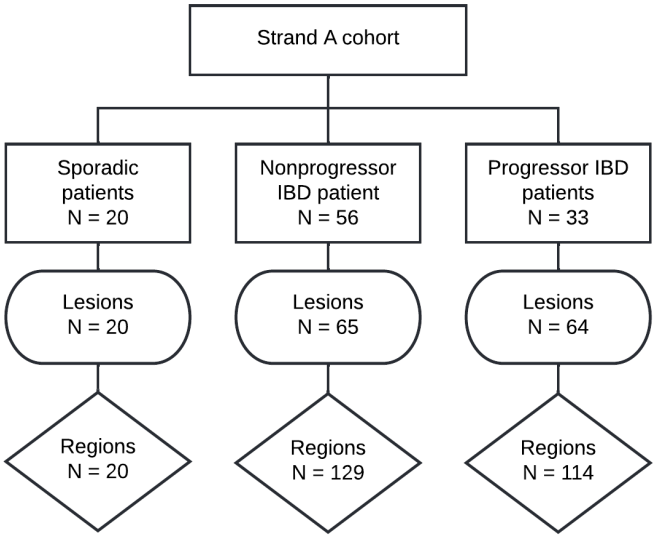


Figure 5.2: Breakdown of number of patients, lesions and regions in strand A cohort.

To better assess the applicability of the Choi *et al.* classification criteria, dysplastic sporadic lesions were used as controls. Sporadic lesions are generally considered to be easier to histologically assess than colitis-associated dysplasia, due to the fact that the latter arise in a field of inflammation, among regenerative epithelium, which share many features with dysplastic epithelium. The sporadic lesions (n = 20) were collected from 20 patients without underlying bowel conditions. Of the 20 patients, 10 were female, eight male and two of undetermined gender. No information on age or progression was available. Similarly, size and shape for most sporadic lesions was not specified. However, four lesions were reported as right-sided, two as left-sided and six as polypoid (Table 5.1).

The colitis-associated LGD specimens were collected from 89 IBD patients, with IBD duration ranging from 0-59 years. Of the 89 patients, 56 were progressors and

33 were nonprogressors. Notably, where available multiple lesions were collected per patient and sufficiently large lesions were divided into multiple regions to allow for multiregional profiling. Consequently, the progressor LGD samples consisted in 64 lesions, divided into 114 regions. The nonprogressor LGD samples consisted in 65 lesions divided into 129 regions (Figure 5.2). Most lesions were left sided. A skew in lesion size was observed between progressor and nonprogressor lesions. While the distribution of lesions of >10 mm and lesions of <10 mm was comparable for the nonprogressor lesions, the majority of the progressor lesions were >10 mm. The association between progression risk and endoscopically visible LGD >10 mm in size has been previous reported [208, 211]. It should be noted that lesion size was not specified for 15 nonprogressor and 32 progressor lesions and therefore, the observation in this specific dataset is to be taken with caution. With regard to lesion size, of the nonprogressor lesions 58.9% were polypoid, 34.1% were nonpolypoid, 3.9% were invisible and 3.1% had unspecified lesion shape. Conversely, of the progressor lesions 46.5% were polypoid, 27.2% were non-polypoid, 9.6% were invisible and 16.7% had unspecified lesion shape (Table 5.1).

Table 5.1: Clinical and endoscopic details of the cohort in strand A

Patient group		Nonprogressors	Progressors	Controls
Number of patients		56	33	20
Gender	Female	12	12	10
	Male	44	21	8
	Not available	-	-	2
Age	Median	60 yrs	58.12 yrs	Not available
	Range	28-78 yrs	32-82 yrs	
Disease duration	Median	27 yrs	21.59 yrs	Not applicable
	Range	0-59 yrs	0-50 yrs	
	Longstanding	-	4 yrs	
Number of lesions		65	64	20
Number of regions		129	114	20
Lesion location	Left	75	77	2
	Right	21	25	4
	Transverse	30	6	-
	Colon, NOS	3	6	14
Lesion size	<10 mm	57	29	-
	>10 mm	57	53	-
	Not specified	15	32	20
Lesion shape	Invisible	5	11	-
	Polypoid	76	53	6
	Non-polypoid	44	31	-
	Not specified	4	19	14

Location, size and shape reported on per-region basis. Key: yrs - years

5.3.1.2 Histological assessment and diagnostic agreement

The study was undertaken with a panel of six expert and dedicated pathologists (here referred to with Greek constellation names to maintain anonymity), who had been practising in the field of colorectal dysplasia for a minimum of $\sim$10 years. H&E stained slides were scanned, annotated and uploaded on the web image viewer SlideScore, where each pathologist assessed all histological regions independently and scored the regions using the web questionnaire provided by the viewer. Scoring was performed on a per-region basis (details in material and methods 2.4.2.2), following the criteria outlined by Choi *et al.* [141] (reported in Table 2.5 of materials and methods). The panel was asked to score the class, subclass and grade for each region as per below:

- Dysplasia class, with diagnostic possibilities of:

- Conventional (>80% of the tissue being conventional);
- Nonconventional (>80% of the tissue being nonconventional);
- Mixed (for regions with mixed conventional and nonconventional histological features, where no >80% majority tissue composition was found).

- Dysplasia subclass, with diagnostic possibilities of:
  - Tubular adenoma-like (TA-like), tubulovillous adenoma-like (TVA-like) and villous adenoma-like (VA-like) for conventional dysplasia;
  - Hypermucinous, dysplasia with increased Paneth cell differentiation (DPD), crypt cell atypia (CCA/D), goblet cell deficient (GCD), traditional serrated adenoma-like (TSA-like), sessile serrated lesion-like (SSL-like) and serrated lesion not otherwise specified (NOS) for nonconventional dysplasia;
  - For mixed regions i.e. regions with two or more histological components:
    * Major component (comprising >50% of the tissue); any of the conventional or nonconventional subclasses.
    * Minor component (comprising <50% of the tissue); any of the conventional or nonconventional subclasses.

- Dysplasia grade based on the worst grade identified within the region, with diagnostic possibilities of:
  - LGD;
  - HGD;
  - Adenocarcinoma;
  - No dysplasia.

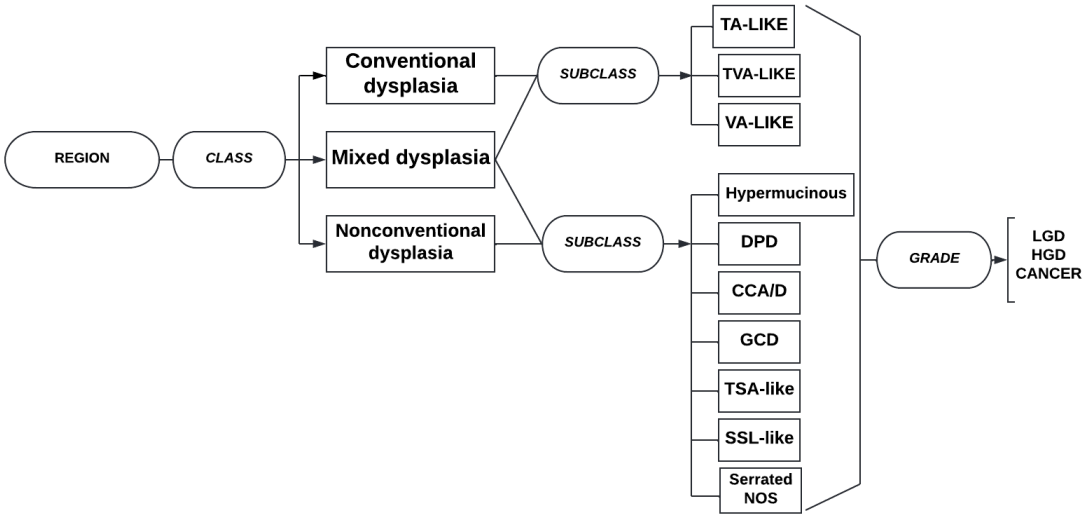


Figure 5.3: Summary diagram of the histological assessment criteria.

The scores from the pathologists were then analysed to determine the diagnostic agreement between the six pathologists. Two of the most commonly used metrics in pathology were adopted: percentage agreement and kappa coefficients. Percentage agreement is the simplest metrics to test for inter-rater agreement, measured by calculating the number of ratings in agreement divided by the total number of ratings. Using the percentage agreement, two main values were calculated for this specific study:

- Agreement score: the mean of the percentage agreement of each pathologist across all assessed regions;
- The study inter-rater agreement: determined by first calculating the percentage agreement across pathologists for each region and then taking the mean of such scores;

The minimum percentage agreement considered acceptable for most studies is 80% [215].

The kappa value (κ) is a measure of inter-rater agreement that accounts for chance. Here the Cohen's kappa and the Fleiss' kappa were used. Cohen's kappa measures the proportion of inter-rater agreement beyond chance in interpretation of observations for two observers. The Fleiss' kappa measures the proportion of inter-rater agreement beyond chance in interpretation of observations for three or more raters. When inter-

preting kappa coefficients, values ≤ 0 indicate no agreement, 0.01-0.20 none to slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial, and 0.81-1.00 almost perfect agreement [215].

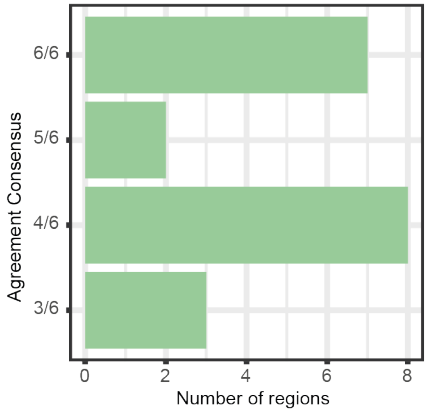
5.3.1.3 Diagnostic agreement in the assessment of sporadic regions

Based on previous expert assessment the sporadic regions consisted of: six TVAs, seven TSAs and seven SSLs. All six pathologists were aware of the presence of the sporadic samples in the cohort, although no other detail was provided to them. Two of the six pathologists commented on the questionnaire to report the regions as sporadic. However, information on whether the other four were able to recognise them is not available as it was outside the study's scope.

Assessment of the scoring of sporadic regions revealed suboptimal diagnostic agreement in determining region class and subclass. The lowest diagnostic agreement (percentage agreement of 61% and Fleiss' kappa equal to 0.266) was observed for region subclass, with all pathologists agreeing on the diagnosis of only one of 20 regions. Slightly better was the diagnostic agreement for region class, with a percentage agreement of 78% and a Fleiss' kappa of 0.37. Conversely, a good diagnostic agreement (i.e. inter-rater agreement of 91%) in grading regions was observed (Figure 5.4).

Next, the frequency of the different histological classes and subclasses among the sporadic cohort was assessed to determine whether they were in agreement with the previous diagnosis. Due to the suboptimal inter-rater agreement, majority agreement, i.e. agreement on a particular diagnosis by four or more pathologists, was used to determine the frequency of the histological class and subclasses. From the initial 20 regions, only 17 had a majority agreed histological class diagnosis. The majority of these lesions (10/17) were scored as nonconventional, whilst six were scored as conventional and one as mixed. Following, the region subclass scores from the pathologists were assessed in the regions with a majority agreed subclass diagnosis (12/20) (Figure 5.5). Despite the poor inter-rater agreement, assessment of the histological subclass scoring showed similar distribution and scoring to the original expert assessment, with only scores of TVA-like, TSA-like and SSL-like at an equal rate.

A Agreement among pathologists in sporadic regions class scoring



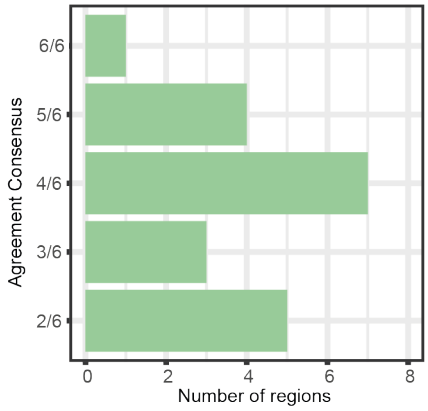
Percentage agreement

Assessors	Andromeda	Ara	Hydrus	Eridanus	Lyra	Tucana
Agreement score	73	76	70	65	64	58
Study inter-rater agreement			78			

Kappa values

Cohen's Kappa	Andromeda	Ara	Hydrus	Eridanus	Lyra
Andromeda					
Ara	0.808				
Hydrus	0.554	0.63			
Eridanus	0.386	0.325	0.3		
Lyra	0.375	0.444	0.449	0.414	
Tucana	0.216	0.263	0.0871	0.182	0.321
Fleiss' Kappa			0.37		

B Agreement among pathologists in sporadic regions subclass scoring



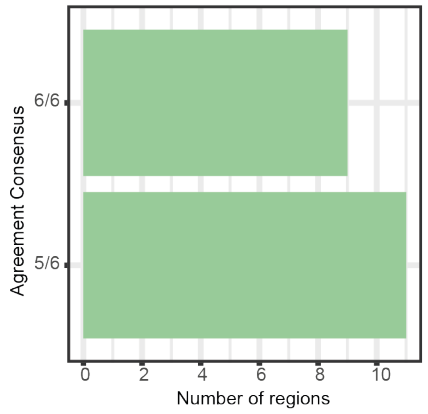
Percentage agreement

Assessors	Andromeda	Ara	Hydrus	Eridanus	Lyra	Tucana
Agreement score	49	58	40	52	53	40
Study inter-rater agreement			61			

Kappa Values

Cohen's Kappa	Andromeda	Ara	Hydrus	Eridanus	Lyra
Andromeda					
Ara	0.509				
Hydrus	0.128	0.2			
Eridanus	0.41	0.403	0.138		
Lyra	0.266	0.583	0.286	0.432	
Tucana	0.104	0.231	0.0556	0.185	0.222
Fleiss' Kappa			0.266		

C Agreement among pathologists in sporadic regions grading



Percentage agreement

Assessors	Andromeda	Ara	Hydrus	Eridanus	Lyra	Tucana
Agreement score	91	58	91	91	91	88
Study inter-rater agreement			91			

Kappa values

Cohen's Kappa	Andromeda	Ara	Hydrus	Eridanus	Lyra
Andromeda					
Ara	0.103				
Hydrus	1	0.103			
Eridanus	1	0.103	1		
Lyra	1	0.103	1	1	
Tucana	1	0.0393	0.487	0.487	0.487
Fleiss' Kappa			0.216		

Figure 5.4: Measures of diagnostic agreement among pathologists in assessing sporadic regions. **Left.** Breakdown of consensus agreement in the assessment of the regions. **Top Right.** Percentage agreement measured in agreement score for each pathologist and the study's inter-rater agreement. Both reported as percentage values. **Bottom Right.** Cohen's kappa measuring inter-rater agreement between two pathologists at a time (e.g. between Andromeda and Ara) and Fleiss' Kappa measuring agreement between all pathologists.

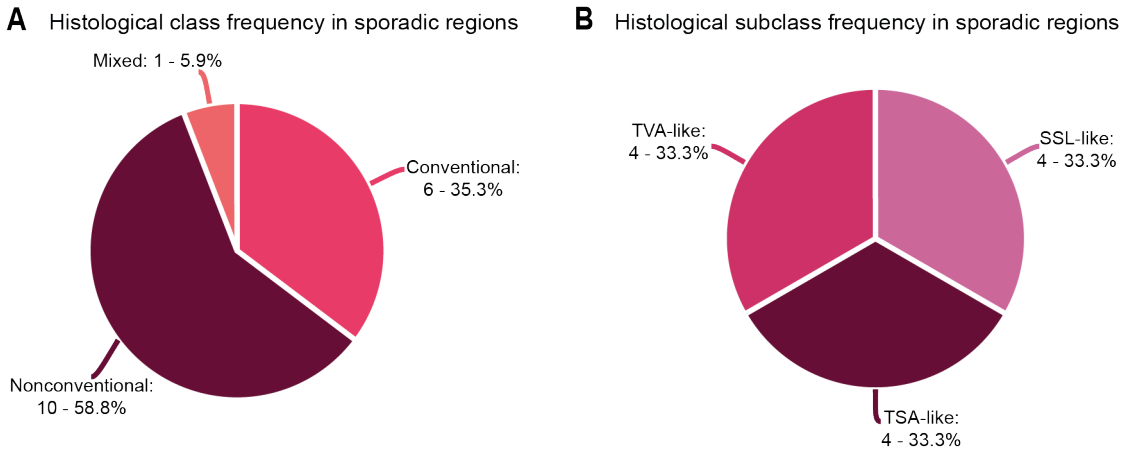


Figure 5.5: Frequency of dysplasia classes and subclasses in the sporadic cohort. **A.** Pie chart outlining the distribution of conventional, nonconventional and mixed dysplasia in the sporadic cohort. **B.** Pie chart outlining the distribution of histological subclasses in the sporadic cohort. In both figures only samples with a diagnosis agreed, but the majority of the pathologists were included.

Comprehensively, the suboptimal inter-rater agreement suggests a difficulty in classifying and subclassifying dysplasia (Figure 5.4). Nevertheless, assessment of the majority agreement scores and its similarity with the previous scoring of the regions evidences some degree of agreement between the pathologists.

5.3.1.4 Diagnostic agreement in the assessment of colitis regions

Next, the diagnostic agreement between pathologists in the assessment of the colitis-associated dysplasia was examined. Poor inter-rater agreement was found in the classification (69% percentage agreement) and subclassification (57% percentage agreement) of the colitis-associated dysplasia. On the contrary, acceptable inter-observer agreement, with a percentage agreement of 85%, was achieved in the grading of the regions. Notably, worse overall agreement was observed in the assessment of the colitis-associated dysplasia than in that of the sporadic regions for all three categories, with percentage agreement reduction from 78% to 69% for class, from 61% to 57% for subclass and from 91% to 85% for grade (Figure 5.6). This further confirms the already known increased difficulty in assessing colitis-associated dysplasia.

The poor inter-rater agreement found in the classification of the regions evidences the complexity of the classification criteria, particularly in the assessment of the colitis-associated LGD. This is further highlighted by the fact that higher agreement was

achieved between pathologists in the grading of both sporadic and colitis than in the classification of the dysplasia.

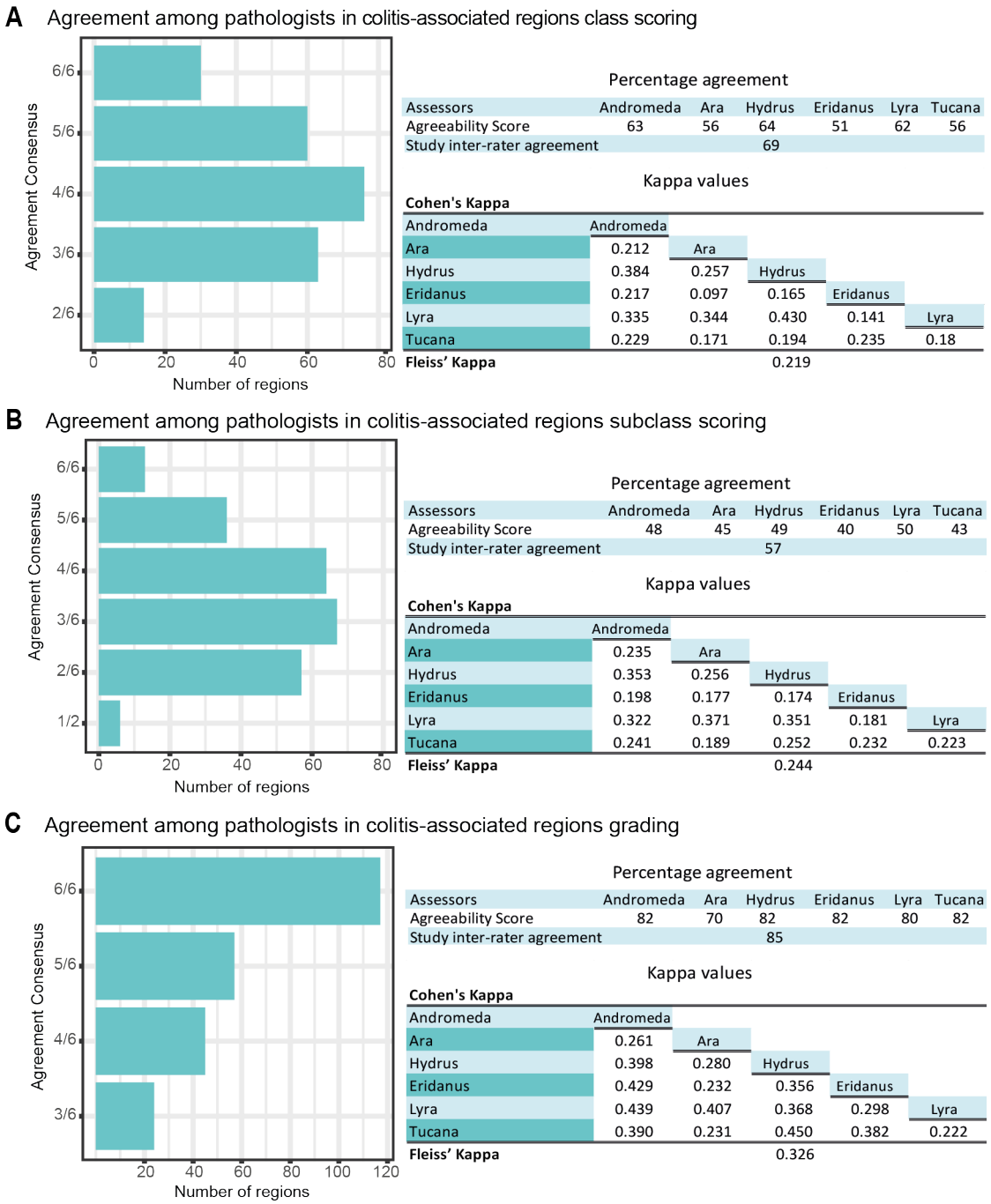


Figure 5.6: Measures of diagnostic agreement among pathologists in assessing colitis-associated LGD. **Left.** Breakdown of consensus agreement in the assessment of the regions. **Top Right.** Percentage agreement measured in agreement score for each pathologist and the study's inter-rater agreement. Both reported as percentage values. **Bottom Right.** Cohen's kappa measuring agreement between two pathologists at a time and Fleiss' Kappa measuring agreement between all pathologists.

5.3.1.5 Association between nonconventional dysplasia and progression

Despite the poor inter-rater agreement in the assessment of both colitis-associated and sporadic regions, an attempt at testing Choi *et al.*'s hypothesis that a nonconventional histological diagnosis is associated with increased progression risk was made. The poor diagnostic agreement between pathologists resulted in only 30 colitis regions with a diagnostic class agreed by all six pathologists. Consequently, to enable a numerically and scientifically meaningful assessment, the analysis was performed on regions with a majority agreement, where four or more pathologists agreed on a region's class. From the initial 243 histological regions, majority agreement was obtained only in 166. Of these, 73.5% were conventional, 21.1% were nonconventional and 1.8% were mixed. Interestingly, six regions that had a previous LGD diagnosis were scored as not having dysplasia. The analysis was first performed on regions with only a majority conventional or nonconventional dysplasia diagnosis (Figure 5.7 Analysis 1). Assessment of progression in nonconventional and conventional dysplasia revealed no association between progression and nonconventional dysplasia (p-value = 0.318, Figure 5.8 A).

Common histology practice is for diagnosis to be made based on the worse visible histology/grade. In the case of regions with mixed histology, nonconventional dysplasia is the worse visible histology based on Choi *et al.* findings. Consequently, in a second analysis scores of mixed conventional-nonconventional dysplasia were denoted as nonconventional prior to assessment of progression (Figure 5.7 Analysis 2). Denotation of the mixed dysplasia led to an increase in regions with a majority consensus to 198. With this second approach, the number of regions scored as conventional dysplasia remained unvaried. However, an increase in regions denoted as nonconventional was seen from 34 to 70. Whilst this could be due to the fact that as outlined above mixed conventional-nonconventional dysplasia were denoted as nonconventional, the doubling in regions with nonconventional dysplasia is also loosely suggestive of a tendency of nonconventional dysplasia to occur as mixed dysplasia. As per the findings from first analysis, no association between progression and nonconventional dysplasia was found (p-value = 0.368, Figure 5.8 B).

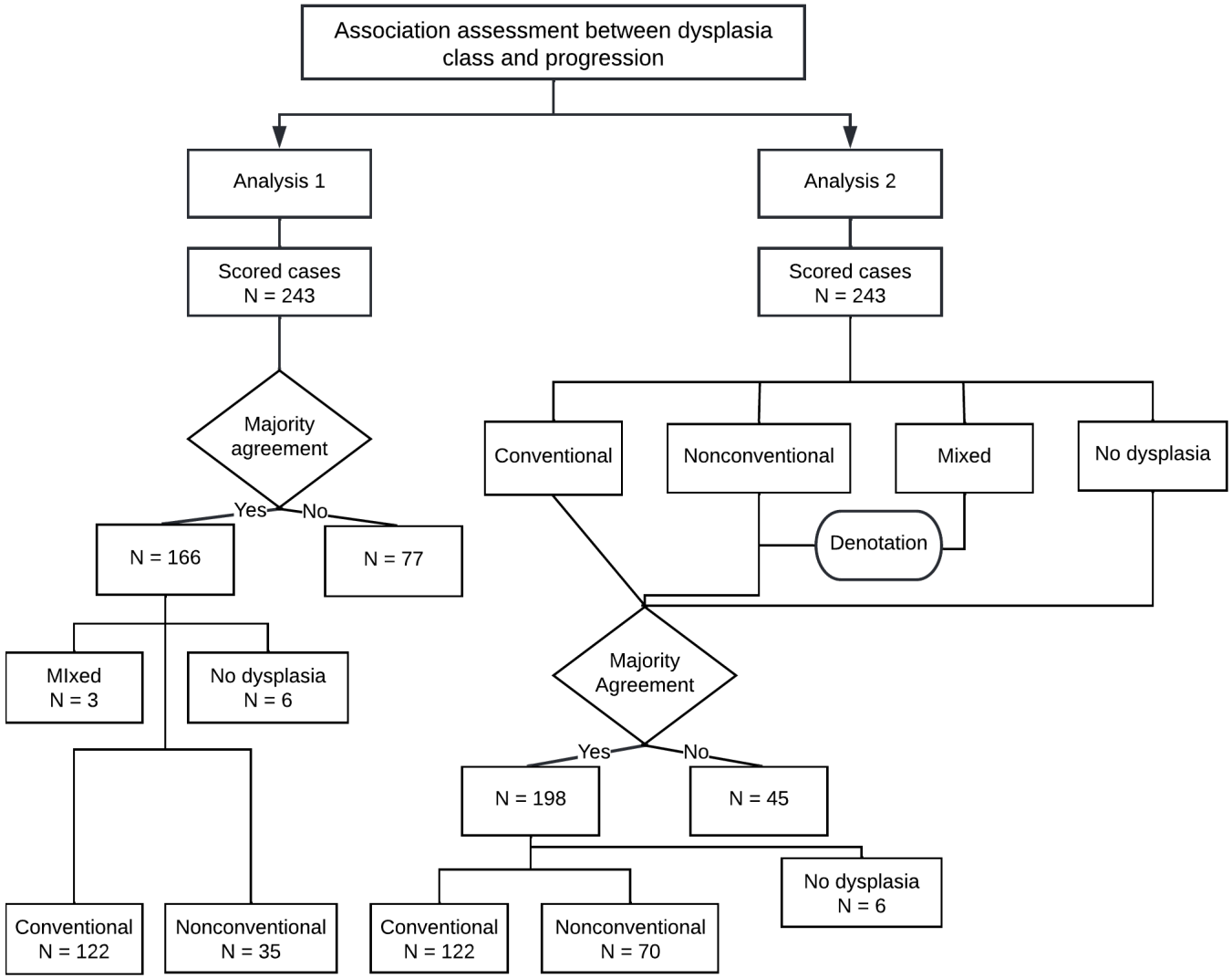
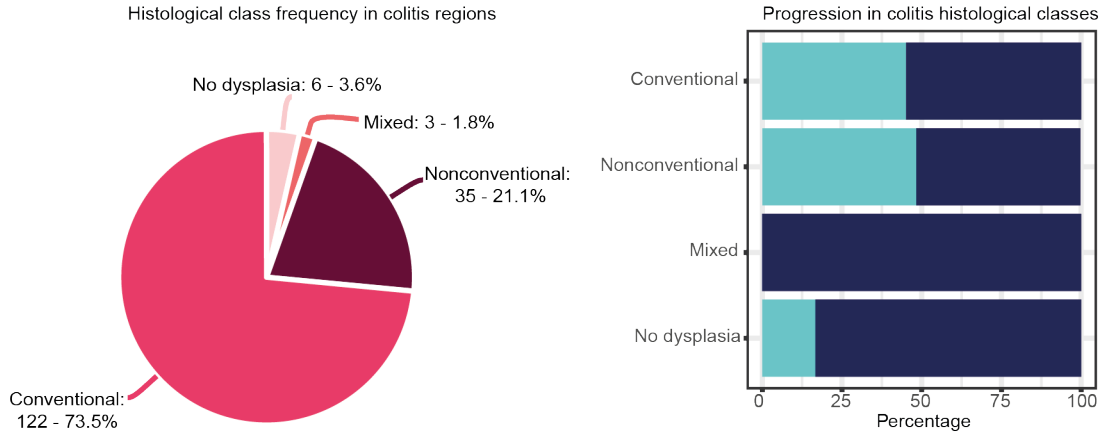


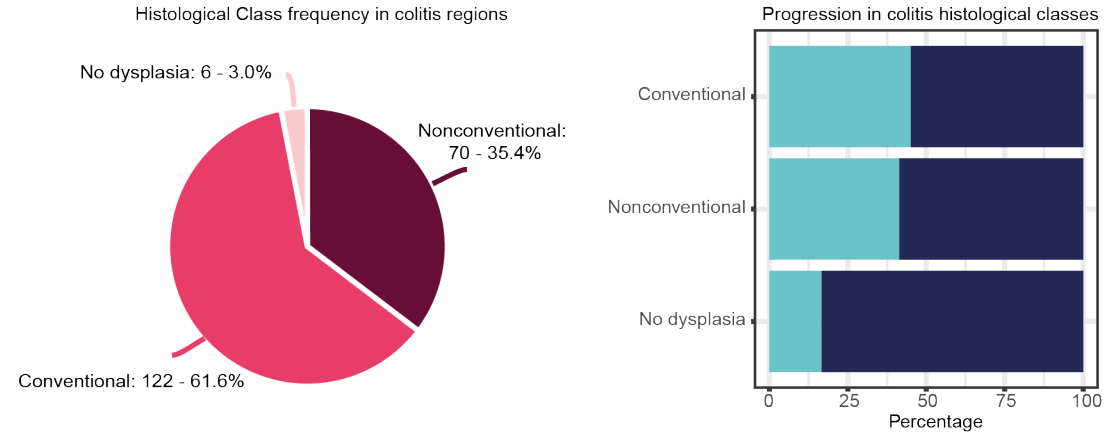
Figure 5.7: Breakdown of analytical strategy to investigate association of nonconventional dysplasia with progression. **Analysis 1.** Assessment without denotation of mixed scores. **Analysis 2.** Assessment where scores of mixed LGD were denoted to the worse visible class.

A Assessment of correlation between progression and nonconventional dysplasia



Histological class	Progressors	Nonprogressors	Total regions	p-value
Conventional	55	67	122	0.318
Nonconventional	17	18	35	
Mixed	0	3	3	-
No dysplasia	1	5	6	-

B Assessment of correlation between progression and nonconventional dysplasia following denomination of mixed cases



Histological class	Progressors	Nonprogressors	Total regions	p-value
Conventional	55	67	122	0.368
Nonconventional	29	41	70	
No dysplasia	1	5	6	-

Figure 5.8: Association of nonconventional dysplasia with progression. **Right.** A pie chart outlining the distribution of conventional, nonconventional and mixed dysplasia in the colitis cohort. **Left.** Distribution of progressor and nonprogressor LGD among the dysplasia regions. **Bottom Left**. Result of Fisher's t-test assessing whether nonconventional dysplasia are more likely to be progressors. **A**. Results assessment without denotation of mixed scores. **B**. Results assessment where scores of mixed LGD were denoted to the worse visible class.

All together, it can be concluded that the association between nonconventional dysplasia and progression reported by the Choi group could not be confirmed in this

dataset.

5.3.1.6 Association between progression and high-risk histologies

Having found no association between progression and nonconventional dysplasia, the next analysis focused on dysplasia subclasses, particularly on the high-risk histological subtypes. The Choi group, in their follow-up paper [142], identified 4 nonconventional histologies: CCD, hypermucinous, GCD and SSL-like, as more likely to be associated with HGD or CRC, more so than DPD, TSA-like and the conventional dysplasia subtypes.

Similarly to the region class, the poor diagnostic agreement between pathologists resulted in only 16 regions with a diagnostic subclass agreed by all pathologists. Consequently, the analysis was performed on regions with a majority subclass diagnosis. In a first analysis, regions with a majority agreed subclass diagnosis (113/243) were categorised into high-risk or low-risk based on their histology, without accounting scores of mixed histology (Figure 5.9 Analysis 1). The Choi group did not provide any information on cancer risk of serrated NOS lesions, thus regions with this histology could not be categorised based on cancer risk. The association between the high-risk histology and progression was then statistically assessed. In a second analysis, regions with two histological subclasses were denoted based on the worse histological class visible. Therefore, scores of mixed high- and low-risk histology were denoted to their high-risk histology (Figure 5.9 Analysis 2). This was followed by statistical assessment of the association between high-risk histology and progression. With both approaches, the most common histological subclass in this cohort was TA-like, accounting for 58.4% / 52.8% of the regions. This was followed by TVA-like regions (14.2% / 12.8%) and hypermucinous dysplasia (9.7% / 16%), with the latter being the most common nonconventional dysplasia in the cohort. The rest of the regions were TSA-like, VA-like, DPD or serrated NOS. No cases of GCD and CCA/D were found in this cohort (Figure 5.10 A).

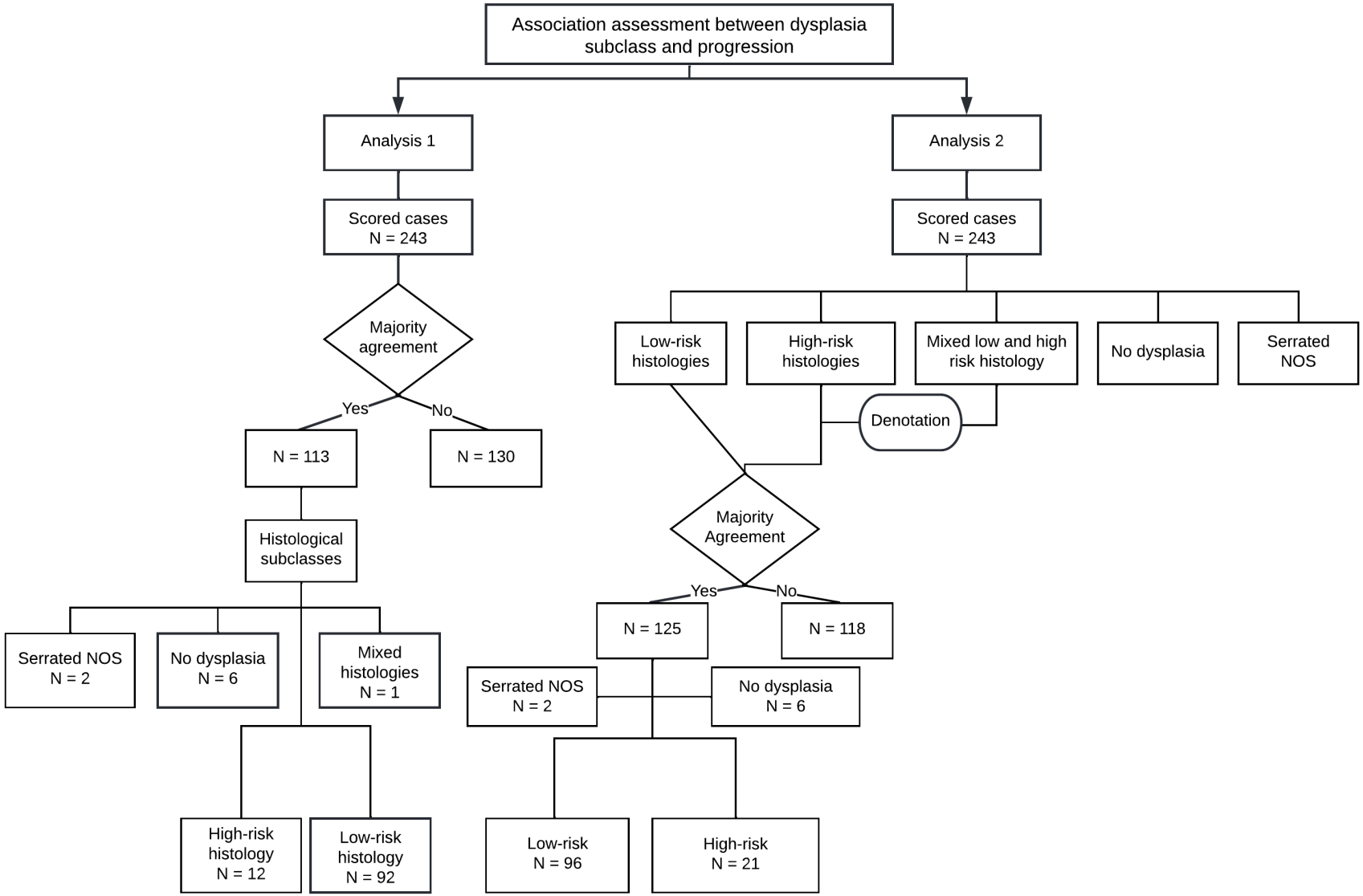
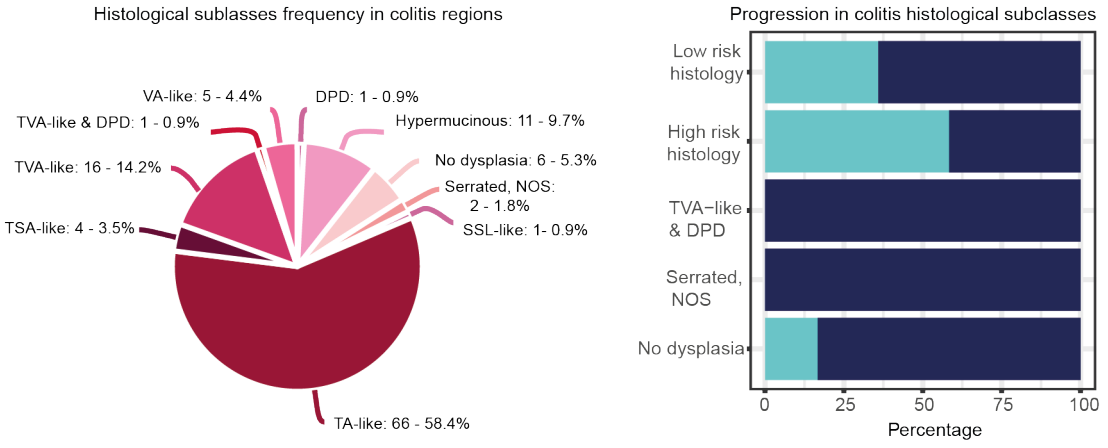


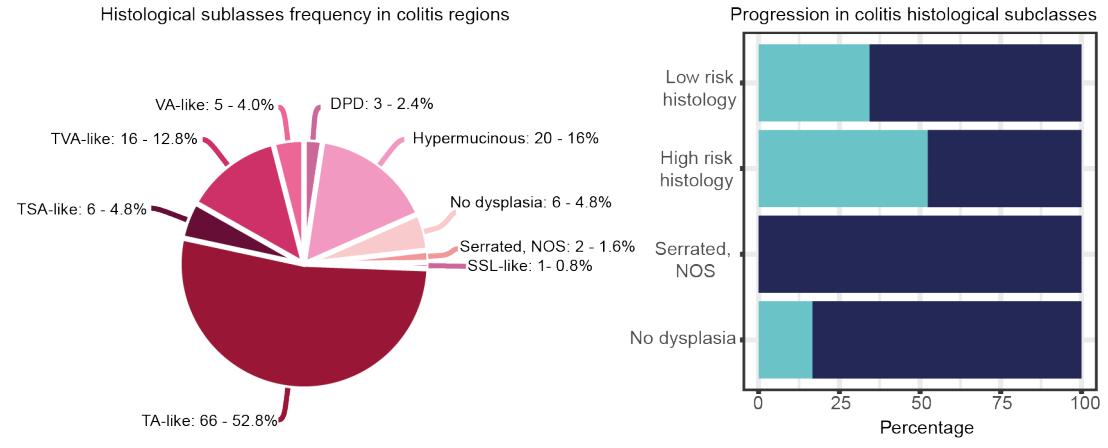
Figure 5.9: Breakdown of analytical strategy to investigate association of high-risk histological subtypes of dysplasia with progression. **Analysis 1.** Assessment without denotation of mixed scores. **Analysis 2.** Assessment where scores of mixed LGD were denoted to worse class visible.

A Assessment of correlation between progression and high-risk histological subclasses



Histological subclass	Progressors	Nonprogressors	Total regions	p-value
Low risk histology	33	59	92	0.376
High risk histology	7	5	12	
TVA-like & DPD	0	1	1	-
Serrated NOS	0	2	2	-
No dysplasia	1	5	6	-

B Assessment of correlation between progression and high-risk histological subclasses with denomination of mixed cases



Histological subclass	Progressors	Nonprogressors	Total regions	p-value
Low risk histology	33	63	96	0.099
High risk histology	11	10	21	
Serrated, NOS	0	2	2	-
No dysplasia	1	5	6	-

Figure 5.10: Association of dysplasia with high-risk histologies with progression. **Right.** Pie chart outlining the distribution of dysplasia subclasses in the colitis cohort. **Left.** Distribution of progressor and non-progressor LGD among the high-risk and low-risk histologies. The histological subtypes in the high-risk group are: hypermucinous and SSL-like. Those in the low-risk group are: TVA-like, VA-like, TSA-like, TA-like and DPD. **Bottom Left.** Results of Fisher's t-test assessing whether the high-risk histologies are more likely to be progressors. **A.** Assessment without denotation of mixed scores. **B.** Assessment where scores of mixed LGD were denoted to worse class visible.

The association between high-risk histology and progression could not be con-

firmed in this dataset using both analytical approaches. A weak trend was noticed when assessing high-risk histology and progression, whereby slightly more (7/12 and 11/21) high-risk regions were progressors than nonprogressors. However, this was not statistically significant and was further weakened by the poor inter-rater agreement found among the expert pathologists, which suggests these histological subclasses would be extremely difficult to confidently identify and base interventional decisions on in standard clinical practice.

Collectively, data from the histological assessment of the colitis-associated dysplasia using the Choi *et al.* criteria highlighted the difficulty and weakness of determining cancer risk prediction based on histological assessment and the need for a different diagnostic approach capable of objectively identifying high-risk profiles.

5.3.2 Molecular assessment of colitis-associated progressor and nonprogressor dysplasia

Having established that the Choi *et al.* histological classification of dysplasia cannot be used to reliably and confidently predict cancer risk, the next assessment focused on identifying a molecular strategy for patient stratification. Based on previous findings of a correlation between CNA, histological grade and progression risk, the CNA burden in nonprogressor and progressor lesions from IBD patients was compared to determine whether CNA burden could be used to predict cancer risk.

5.3.2.1 Demographics and lesion details

The study was undertaken on two multicentre cohorts of IBD patients: a discovery cohort and a validation cohort. The discovery cohort comprised 45 nonprogressor and 22 progressor lesions from 67 patients (using a 2:1 patient matching), whilst the validation cohort comprised 35 nonprogressor and 16 progressor lesions from 51 patients (Figure 5.11). The majority of the patients were male, with a median age around 60 and without PSC. A cohort specific fluctuation in disease duration was seen. Progressors and nonprogressor patients/lesions were well-matched for all criteria except lesion size and macroscopic morphology (Table 5.2).

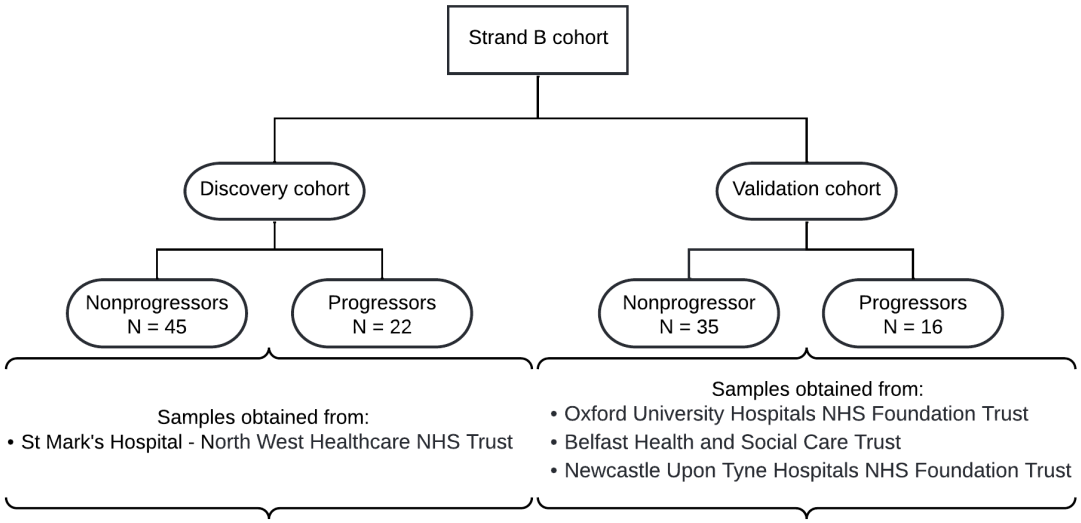


Figure 5.11: Breakdown of the strand B cohort.

Table 5.2: Clinical and endoscopic details of the cohort in strand B

Cohort		Discovery		Validation	
Patient group		Nonprogressors	Progressors	Nonprogressors	Progressors
Number of patients		45	22	35	16
Gender	Female	11	9	9	5
	Male	34	13	26	11
Age	Median	60 yrs	59.59 yrs	62 yrs	58 yrs
	Range	28-78 yrs	34-77 yrs	28-79 yrs	25-82 yrs
Disease Duration	Median	26.2 yrs	18	9 yrs	19
	Range	0-57 yrs	8-48 yrs	0-59 yrs	0-32 yrs
	NA	-	-	-	3
PSC	No	42	18	32	13
	Yes	3	4	3	3
Number of lesions		45	22	35	16
Lesion location	Left	31	15	20	12
	Right	5	3	9	1
	Transverse	9	4	4	1
	Unspecific	-	-	2	2
Lesion size	<10 mm	27	9	15	2
	>10 mm	1	13	7	4
	Unspecified	17	-	13	10
Lesion shape	Invisible	5	5	3	3
	Polypoid	26	11	27	6
	Non-polypoid	13	6	3	2
	Unspecified	1	-	2	5

5.3.2.2 Copy number profiles in progressor and nonprogressor patients

Several studies found that LGD with a higher rate of aneuploidy were more likely to progress to HGD/cancer [96, 142, 213, 214]. Here a high-resolution assessment of CNA in colitis-associated LGD from progressor and nonprogressor patients was undertaken using lpWGS to determine whether a difference between the two patient groups exists.

When broadly assessing CN profiles of nonprogressor and progressor, differences in the number of gains and/or losses were observed. CN profiles of progressor samples were characterised by the presence of gains and/or losses in several chromosomal regions, conversely to nonprogressor samples, which had very few CNAs, with at most 1 or 2 gains and/or losses (Figure 5.12).

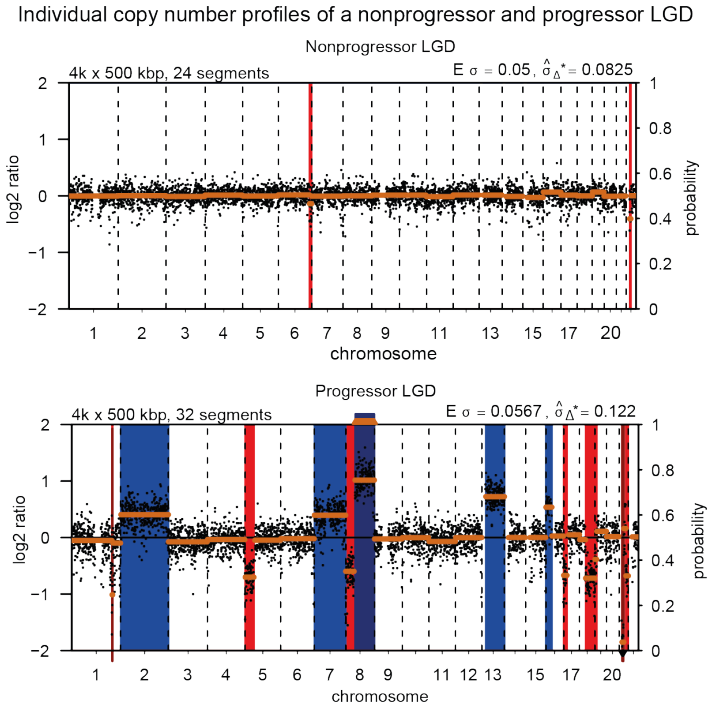


Figure 5.12: Representative CN profile of a progressor LGD and a nonprogressor LGD CN profiled were generated by segmenting the genome in 500kbp (black dots), removing sex chromosomes and problematic bins, and normalising, log-transforming and segmenting (visible by orange line) the remaining bins (details in 2.4.3.3, 2.4.3.4 and 2.4.3.5). As this tissue is diploid, there are 22 segments, one for each non-sex chromosome (x-axis). In blue are the CN gains, whilst in red are the CN losses. On the right is the probability of loss or gain, with 1 indicating a gain, 2 an amplification and -1 a loss and -2 a homozygous loss. **Top.** CN profile of a nonprogressor LGD, where only two small losses can be detected within chromosome 6 and chromosome 22. **Bottom.** CN profile of a progression LGD. Several gains and losses can be seen, including gains within chromosome 2, 7 and 13, and losses within chromosome 5 and 17.

Having observed a difference between the CN profiles of nonprogressor and progressor samples, the CNA burden in patients bearing nonprogressor LGD and patients bearing progressor LGD were quantified (2.4.3.5). On a per-patient basis, CNA burden events in progressor LGD was found to be statistically higher (p = 3.310e-05) than that of nonprogressor LGD, with a median number of abnormal segments equal to 35.7% in progressors LGD as opposed to the 2.2% in nonprogressor lesions (Figure 5.13).

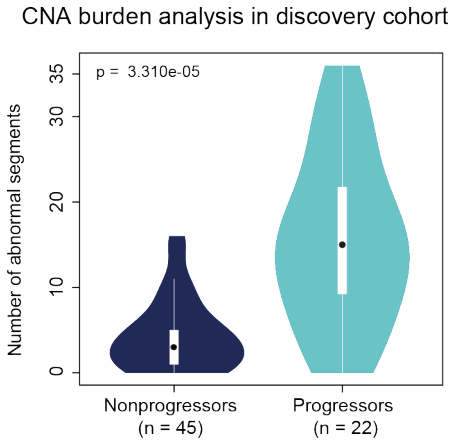


Figure 5.13: CNA burden in nonprogressor and progressor LGD of the discovery cohort. Violin plots comparing CNA burden between patients with nonprogressor LGD and patients with progressor LGD in the discovery cohort. The measure of CNA burden here is the number of abnormal segments. Highlighted by the solid white bars in the violin plots are the median and the interquartile range, whilst the lines represent the 95% confidence interval. Reported is the p-value from unpaired two-sided t-test comparing CNA burden in progressor and nonprogressor LGD.

Having found a statistically significant difference in CNA burden between progressor and nonprogressor LGD, we wanted to assess the ability of CNA burden and clinical and endoscopic features (including some well-recognised clinical risk factors) to predict cancer risk. Consequently, a univariate cox regression analysis of the relevant clinical variables and the genomic output was performed (Table 5.3). CNA events, together with only two well-recognised clinical risk factors i.e. multifocality and concomitant PSC diagnosis, were found to be significantly associated with an increased risk of progression to HGD and/or cancer. Number of abnormal CN segments was the strongest predictor for cancer risk (with a p-value < 0.001; Table 5.3). Next, to avoid overfitting in a small patient cohort, a multivariate cox regression analysis was performed on only the three variables significantly associated with progression. CNA

burden was found to be the only statistically significant risk factor for LGD progression (Table 5.4). Both results suggest high CNA burden to be a robust indicator of increased risk of progression.

Table 5.3: Univariate cox analysis for relevant clinical and endoscopic factors

Variable	Cox coefficient	HR (95% Cl)	p-value
Abnormal CN segments	0.109	1.15 (1.07 - 1.16)	< 0.001***
Age	-0.016	0.98 (0.95 - 1.01)	0.36
Disease duration	-0.022	0.97 (0.94 - 1.01)	0.20
PSC	1.22	3.38 (1.14-10.1)	0.028*
Gender male	-0.51	0.60 (0.26 - 1.40)	0.23
Multifocal LGD	1.24	3.45 (1.34 - 8.89)	0.010*
Size of visible lesions (mm)	0.018	1.02 (0.9 - 1.04)	0.13
Lesion shape			
a) Invisible	1		
b) Non-polypoid	-0.10	0.52 (0.30 - 2.77)	0.56
c) Polypoid	-0.10	0.52 (0.20 - 1.77)	0.56
Right-sided LGD	0.24	1.27 (0.53 - 3.02)	0.29

A univariate cox regression analysis of the relevant clinical and endoscopic factors and CNA burden, and their association with progression. Key: HR - Hazard ratio; Cl - confidence level.

Table 5.4: Multivariate cox analysis for relevant clinical and endoscopic factors

Variable	Cox coefficient	HR (95% Cl)	p-value
Number of abnormal CN segments	0.103	1.11 (1.06 - 1.16)	<0.001
PSC	0.61	1.84 (0.53-6.37)	0.338
Multifocal LGD	0.49	1.64 (0.53 - 5.08)	0.392

A multivariate cox regression analysis of only the three variables found to be statistically associated with progression by the univariate analysis. Key: HR - Hazard ratio; Cl - confidence level.

After having established that CNA burden can predict cancer risk, we assessed the performance of CNA burden as a clinical risk stratification tool. Patients were ranked based on number of abnormal segments and those in the upper quartile (i.e. patients with the highest number of abnormal CN segments) were compared to the rest of the patients using a Kaplan-Meier analysis. The upper quartile of IBD patients were estimated to be significantly more likely to develop HGD/cancer in the future than the lower three quartiles (Hazard ratio 7.5, p-value < 0.0001, negative predictive value = 91%, sensitivity = 89%, and specificity = 82%, Figure 5.14).

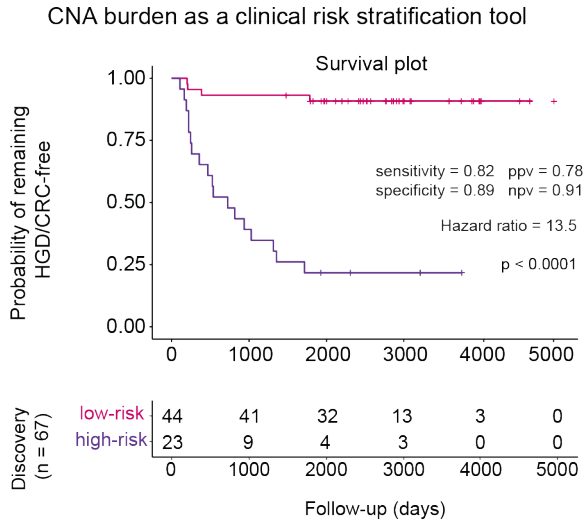


Figure 5.14: Assessment of the performance of CNA burden as a clinical risk stratification tool. Kaplan-Meier survival analysis comparing the upper quartile patients based on number of CNA events (in purple) and the rest of the patients (in pink). Key: PPV - 5 year positive predictive value, NPV - 5 year negative predictive value.

Lastly, the findings were interrogated in an independent validation cohort, comprising nonprogressor and progressor lesions collected from Oxford Hospitals NHS Foundation Trust, Newcastle Upon Tyne Hospitals NHS Foundation Trust and Belfast Health and Social Care Trust. In concordance with the discovery cohort, progressor LGD had significantly greater CNA events than non progressor LGD. Furthermore, the survival analysis comparing the upper quartile patients based on number of CNA events and the rest of the patients in the cohort confirmed that patients with higher CNA burden have a higher likelihood of progression (Figure 5.15). Collectively, the data showed that CNA burden in LGD can predict cancer risk in IBD patients with high specificity.

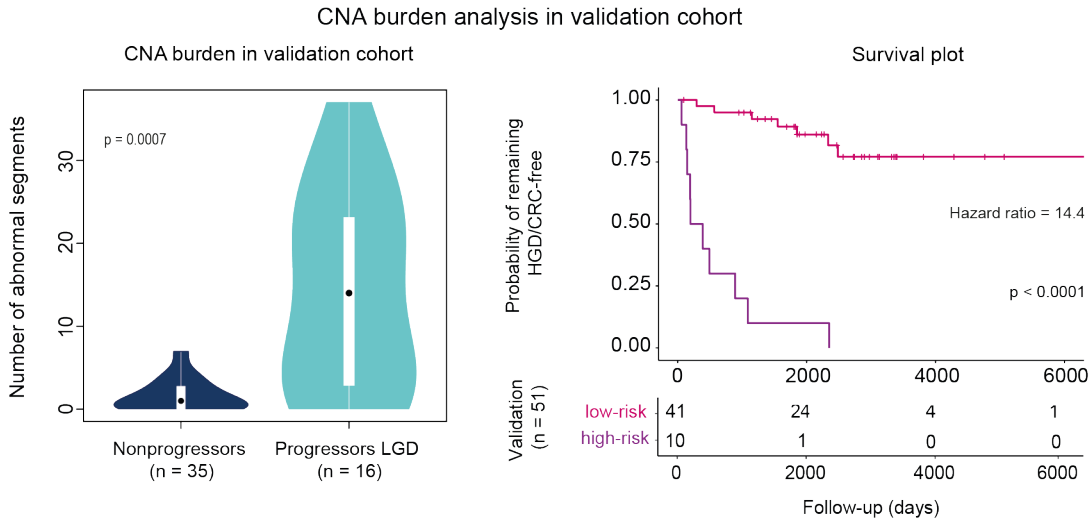


Figure 5.15: CNA burden in nonprogressor and progressor LGD and associated Kaplan-Meier analysis of the validation cohort. **Left.** Violin plots comparing CNA burden between patients with nonprogressor LGD and patients with progressor LGD in the validation cohort. The measure of CNA burden here is the number of abnormal segments. Highlighted by the solid white bars in the violin plots are the median and the interquartile range, whilst the lines represent the 95% confidence interval. Reported is the p-value from unpaired two-sided t-test comparing CNA burden in progressor and nonprogressor LGD. **Right.** Kaplan-Meier survival analysis comparing the top quartile patients based on number of CNA events (in purple; high-risk) and the rest of the patients (in pink) (low-risk). Key: PPV - 5 year positive predictive value, NPV - 5 year negative predictive value.

5.3.2.3 Copy number events

With lpWGS allowing for a high resolution CNA assessment, the next assessment focused on the distribution of specific CN gains and losses. The aim of this assessment was to identify CN changes specific to the colitis-associated LGD phenotype and to progression. First, CNAs commonly occurring in colitis-associated LGD were identified by assessing the alterations that occur in both progressor and nonprogressor patients (Figure 5.16). These comprised losses in sub-telomeric segments of chromosomes 4, 5, 7, 8 and 22. They are particularly important because of their potential role in discerning between true colitis-associated lesions and sporadic lesions.

To establish whether these CNA are specific to colitis-associated lesions a direct comparison between sporadic adenomas and colitis-associated LGD is needed, which however was out of this study's scope. Nevertheless, previous work from Baker *et al.* [96] can aid in making meaningful observations on the results obtained. Baker *et al.* in-

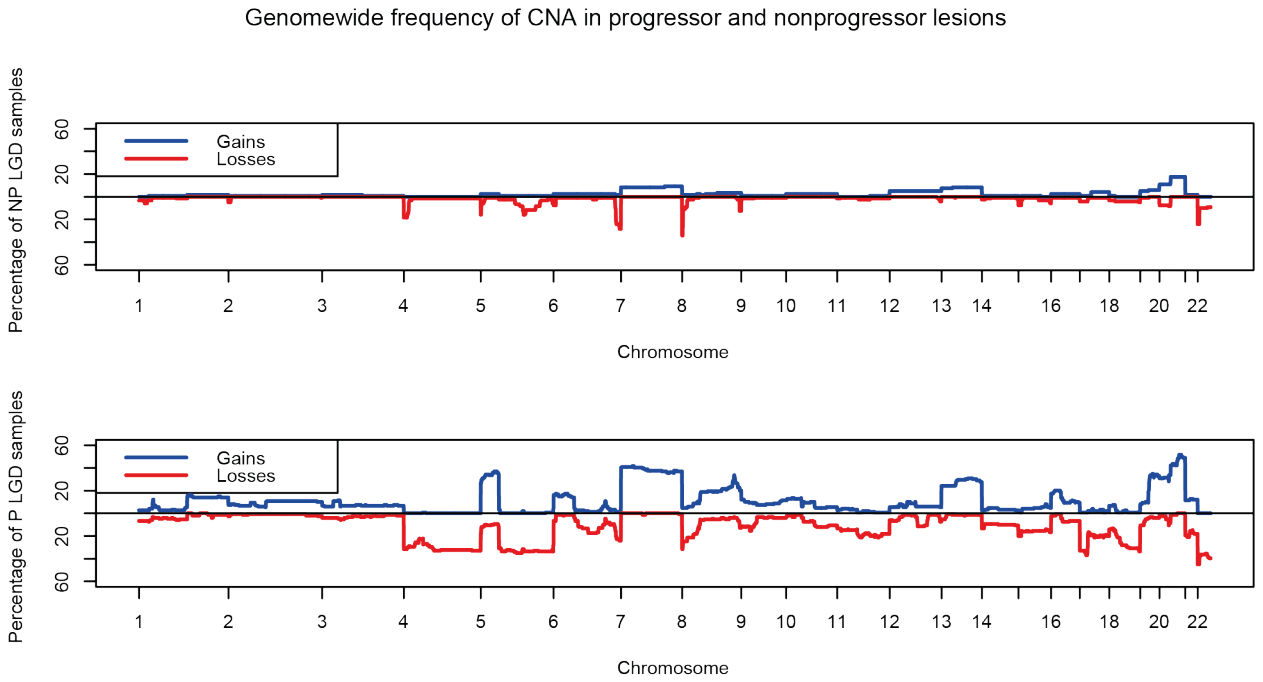


Figure 5.16: Frequency of gains and losses across all non-sex chromosomes in progressor and nonprogressor lesions

vestigated CNA differences between sporadic and colitis-associated CRC and found: six chromosomal arms to be more likely gained, 10 most likely to be lost and five less likely to be lost in CA-CRC compared to S-CRC (Table 5.5). This was despite CA-CRCs and S-CRCs having similar proportions of CNAs. Among the chromosomal alterations more significantly associated with CA-CRC, were loss of 4p and 5q. Here, these two CN losses were found to be common alterations in colitis-associated LGD, suggesting that they could be specific to colitis-associated lesions.

Table 5.5: CNA differences between CA-CRC and S-CRC from Baker *et al.* 2019

Chromosomal alteration	Percentage CA-CRC	Percentage S-CRC	Odds ratio	Adjusted* Fisher p-value
Gain on 1q	38.4%	17.0%	2.89	0.027
Loss on 2q	15.0%	3.0%	5.44	0.027
Loss on 4p	55.4%	25.0%	3.67	0.005
Loss on 4q	51.4%	24.0%	3.46	0.008
Gain on 5p	42.8%	14.0%	4.51	0.003
Loss on 5q	54.7%	17.0%	5.87	0.000
Gain on 6p	37.1%	14.0%	3.66	0.011
Loss on 8p	27.1%	50.0%	0.38	0.041
Loss on 8q	0.1%	12.0%	0.00	0.047
Gain on 10p	21.8%	5.0%	5.40	0.011
Gain on 10q	12.9%	2.0%	7.12	0.027
Loss on 11q	36.1%	11.0%	4.37	0.005
Loss on 12q	24.0%	7.0%	4.39	0.014
Loss on 16p	18.1%	5.0%	3.95	0.041
Loss on 17q	37.4%	15.0%	2.68	0.047
Loss on 18p	35.6%	61.0%	0.344	0.024
Loss on 19p	16.5%	4.0%	5.19	0.024
Gain on 19q	32.1%	15.0%	2.68	0.047
Loss on 20p	11.0%	32.0%	0.24	0.026
Loss on 20q	0.5%	15.0%	0.00	0.026
Loss on 22q	47.7%	26.0%	2.56	0.038

Table adapted from Baker *et al.* [96] Supplementary Table 13. P-value adjusted by Benjamini-Hochberg correction.

Next, CNAs more commonly occurring in progressor LGD than nonprogressor LGD were assessed to identify CNA that are associated with progression. The 10 CNAs occurring at the greatest difference rate between progressors and nonprogres-

sors are reported in Table 5.6. The identification of these CNAs is important because their investigation could enable a better understanding of the molecular processes that lead a lesion to progress. Interestingly, among the CNAs most commonly occuring in progressors versus nonprogressor LGD, gains of 8q, loss of 17p and loss of 17q have previously been shown to be important in the progression of adenomas to carcinomas [216]. A better understanding of the association between specific CN gains or losses, progression and their role in carcinogenesis requires more fine mapping of the regions gained and/or lost alongside functional assessment. In the meantime, an investigation of common cancer genes and genes commonly mutated in CA-CRC located within these chromosomal regions was carried out. Several important colorectal cancer genes including *APC*, *RNF43*, *TP53*, *BRAF* and *MYC* were identified (Table 5.7).

Table 5.6: CNAs occurring at the greatest difference rate between progressors and nonprogressors

Chromosomal alteration	Percentage Progressors	Percentage Nonprogressors	Odds ratio	Adjusted* Fisher p-value
Loss within 4q	36%	2.2%	25.1	0.001
Gain within 5p	32%	2.2%	20.5	0.002
Loss within 5q	64%	20%	7.0	0.002
Loss within 6q	41%	11.1%	5.5	0.011
Gain within 7p	32%	4.4%	10.0	0.006
Gain within 7q	23%	4.4%	6.3	0.034
Gain within 8q	27%	2.2%	16.5	0.006
Loss within 17p	59%	6.7%	20.2	<0.001
Loss within 17q	55%	2.2%	51.6	<0.001
Gain within 20p	18%	8.9%	8.1	0.013

Number of progressors in analysis = 22; number of nonprogressors in analysis = 45. P-value adjusted by Benjamini-Hochberg correction.

Located within 8q is *MYC*, pro-oncogene involved in cell proliferation and survival. CN gain of 8q leads to overexpression of *MYC*, which is thought to potentiate tumorigenesis. Interestingly, loss of 8p (common in both progressors and nonprogressors LGD) is frequently seen to co-occur with gains of 8q, with co-occurrence of these two CNAs being associated with an increase risk of cancer development. Furthermore, several tumour suppressor genes, including Dlc1 Rho GTPase activating

protein (*DLC1*) [216], are found within 8p [216]. Of particular interest is the loss of both arms of chromosome 17, a common event in progressor LGD. Located in 17p13 is *TP53*. *TP53* mutations in CA-CRC occur in nondysplastic mucosa and are thought to be a very early event in CA-CRC carcinogenesis [217]. Loss of chromosome 17, could be the second hit in *TP53* that leads to loss of heterozygosity. Other CNA of interest that should be further studied are loss of 5q and loss of 17q, chromosomal location of *APC* and *RNF43*, two negative regulators of the WNT signalling pathway. Assessing the extent of the loss in these chromosomal arms would help further elucidate the WNT signalling landscape in CA-CRC. This would be of particular interest for the colitis-associated lesions with no known driver WNT gene mutation, as the ones found in chapter 4 of this thesis.

Table 5.7: Genes within chromosomal regions affected by CNA

Chromosomal alteration	Genes						
Loss within 4q	*FBXW7*	*FAT4*					
Allele	4q.31.3	4q28.1					
Gain within 5p							
Allele							
Loss within 5q	*APC*	*PIK3R1*	*MCC*	*ANXA6*	*MAP3K1*		
Allele	5q22.2	5q13.1	5q22.2	5q33.1	5q11.2		
Loss within 6q	*LATS1*	*ARID1B*					
Allele	6q25.1	6q25.3					
Gain within 7p	*EGFR*	*PMS2*	*RADIL**	*CARD11*			
Allele	7p11.2	7p22	7p22.1	7p22.2			
Gain within 7q	*BRAF*	*PTPN12*	*TRRAP**	*PAXIP1**	*KMT2C*		
Allele	7q34	7q11.23	7q22.1	7q36.2	7q36.1		
Gain within 8q	*c-MYC*	*RAD54B*	*EIF3E*				
Allele	8q24.21	8q22.1	8q23.1				
Loss within 17p	*TP53**	*FLCN*	*MAP2K4*	*USP6*			
Allele	17p13.1	17p11.2	17p12	17p13			
Loss within 17q	*RNF43**	*NF1*	*CDK12*	*AXIN2*	*SLC39A11**	*ERBB2*	*SOX9*
Allele 17q22	17q11.2	17q12	17q24.1	17q24.3	17q12	17q24	
Gain within 20p	*RRBP1**	*SIRPA*					
Allele 20p12.2	20p13						

Key: * genes commonly mutated in CA-CRC

Collectively, the assessment of CNA in progressors and nonprogressor LGD has

revealed CNA burden to be a strong predictor of cancer that can be used to confidently stratify patients. Furthermore, the identification of specific CN gains and losses common to both progressor and nonprogressor colitis-associated LGD and those most commonly occuring in progressor LGD prompts future work to fine map these alterations, determine which genes they specifically affect and investigate their role in CA-CRC carcinogenesis.

5.3.3 Digital cytohistological assessment of colitis-associated progressor and nonprogressor dysplasia

Image cytometry data show that aneuploidy can be detected through changes in nuclear features, such as nuclear staining intensity [218]. Following the finding that progressor LGD have a significantly greater CNA burden than nonprogressor LGD, a cytohistological analysis of nonprogressor and progressor LGD was undertaken, using the digital pathology platform HALO. The purpose of this analysis was to determine whether differences in the large chromosomal defects detected between progressor and nonprogressor LGD were reflected at the histological level. If detectable histological differences can be identified, it may be possible for digital pathology driven cytohistological analysis to potentially act as a surrogate measure of molecular change and ultimately negate the need for lesion molecular testing to determine cancer risk.

5.3.3.1 Demographics and lesion details

Digital pathology analysis was performed on 179 histological regions (94 nonprogressors and 85 progressors), derived from 97 lesions from 64 patients (Figure 5.17). Clinical and endoscopic details are outlined in Table 5.8.

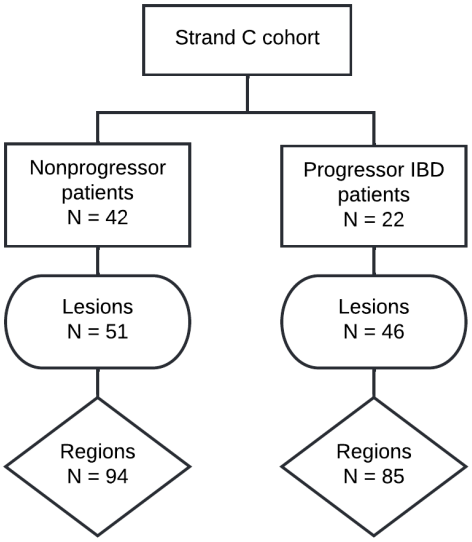


Figure 5.17: Breakdown of number of patients, lesions and regions in strand C cohort.

Table 5.8: Clinical and endoscopic details of cohort in strand C

Features		Nonprogressors	Progressors
Number of patients		42	22
Total number of lesions		51	46
Total number of regions		94	85
Gender	Female	11	9
	Male	31	13
Age	Median	61 yrs	56.95 yrs
	Range	28-78 yrs	34-77 yrs
Disease duration	Median	27 yrs	18 yrs
	Range	0-57 yrs	8-48 yrs
Lesion location	Left	60	59
	Right	8	21
	Transverse	26	5
Lesion size	<10 mm	38	26
	>10 mm	54	55
	Not specified	2	4
Lesion shape	Invisible	5	12
	Polypoid	49	39
	Non-polypoid	38	30
	Not specified	2	4

Location, size and shape reported on per-region basis. Key: yrs - years

5.3.3.2 Digital pathology analysis results

Annotated scans of haematoxylin stained sections of nonprogressor and progressor LGD were assessed on HALO, the gold standard image analysis platform for quantitative tissue analysis in digital pathology. Tissue within each region was segmented into epithelial and stromal compartment and assessed using a pre-established Cytonuclear v1.4 algorithm, which recorded values for staining intensity for haematoxylin, nucleus size and nucleus roundness (details in 2.4.4). The data generated were organised into summary data, with values per tissue area, and single cell data, with values per single cell per tissue area. A breakdown of the results from the analysis is provided in Figure 5.18.

5.3.3.3 Exploratory analysis of digital pathology data

Given the extensiveness and complexity of the data, an initial exploratory analysis was performed using principal component analysis (PCA), an unsupervised multivariate statistical analysis that reduces correlated variation in multiple variables to fewer variables or principal components. This allows for the examination of patterns within a dataset by reducing data dimensionality. First, the summary and single cell data of all tissue compartments were assessed. Next, based on the knowledge that CRC arises from genetic alterations in epithelial cells and that the molecular analysis was performed on epithelium enriched DNA (for the same reason), and assessment of the summary and single cell data of solely the epithelial compartment was carried out. No clear separation of the groups was found following multivariate analysis. Instead, substantial overlap between progressor and nonprogressors cells was seen (Figure 5.19).

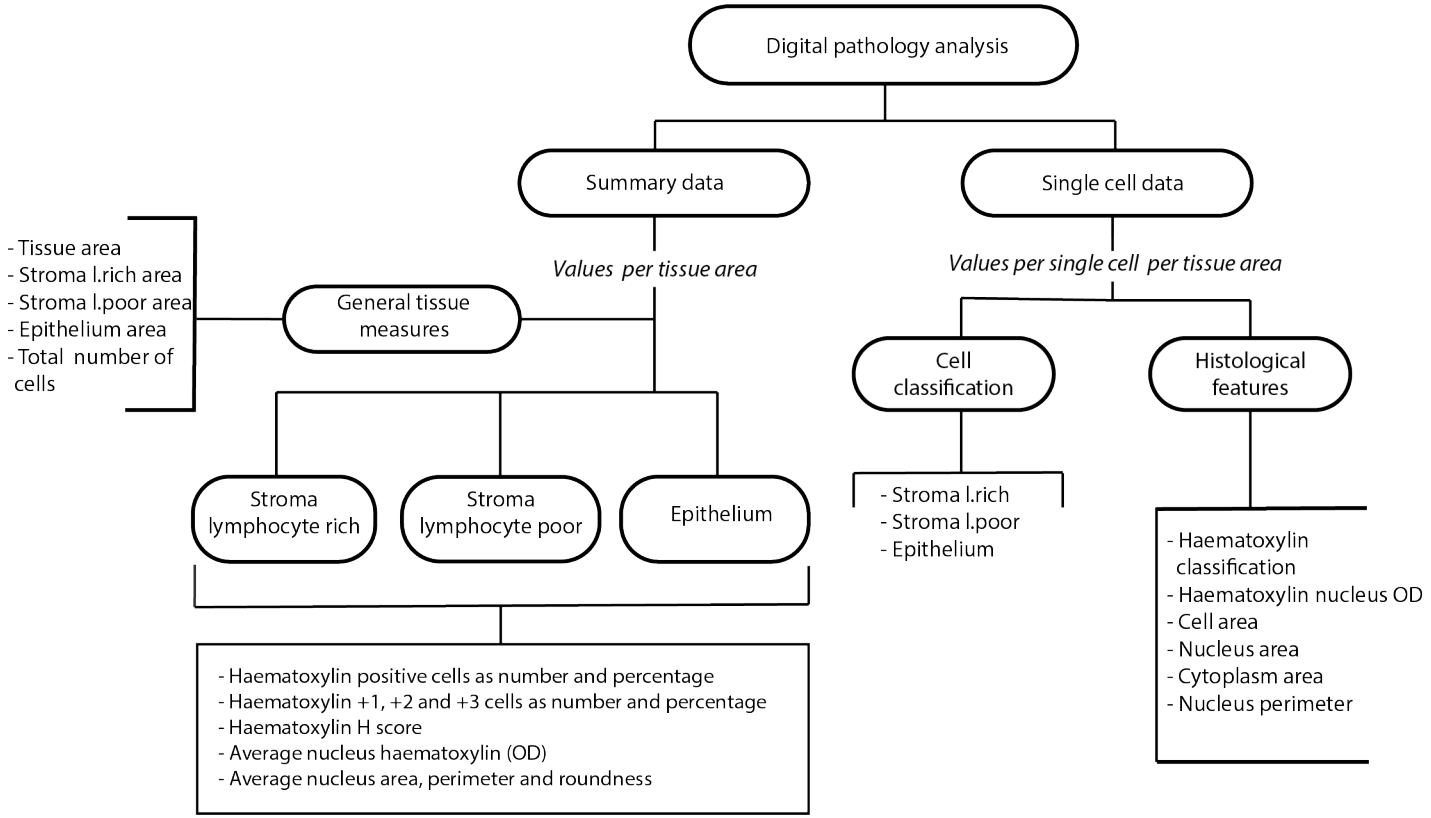


Figure 5.18: Outline of the data gathered from digital histopathology analysis. The analysis generated data divided into summary data and single cell data. The summary data, comprised average values for the tissue area assessed, was divided into stroma lymphocyte-rich, stroma lymphocyte-poor and epithelium. For each of these compartments, average nucleus area, perimeter and roundness values and average measures of staining for haematoxylin were obtained. The single cell data consisted in single cell measures for each cell within the tissue area assessed. For each cell, its classification (stroma lymphocyte-rich, stroma lymphocyte-poor and epithelium) and measures of its total cell area, nucleus area, cytoplasm area and nucleus perimeter, together with categorical and numeric measures of haematoxylin staining intensity were provided.

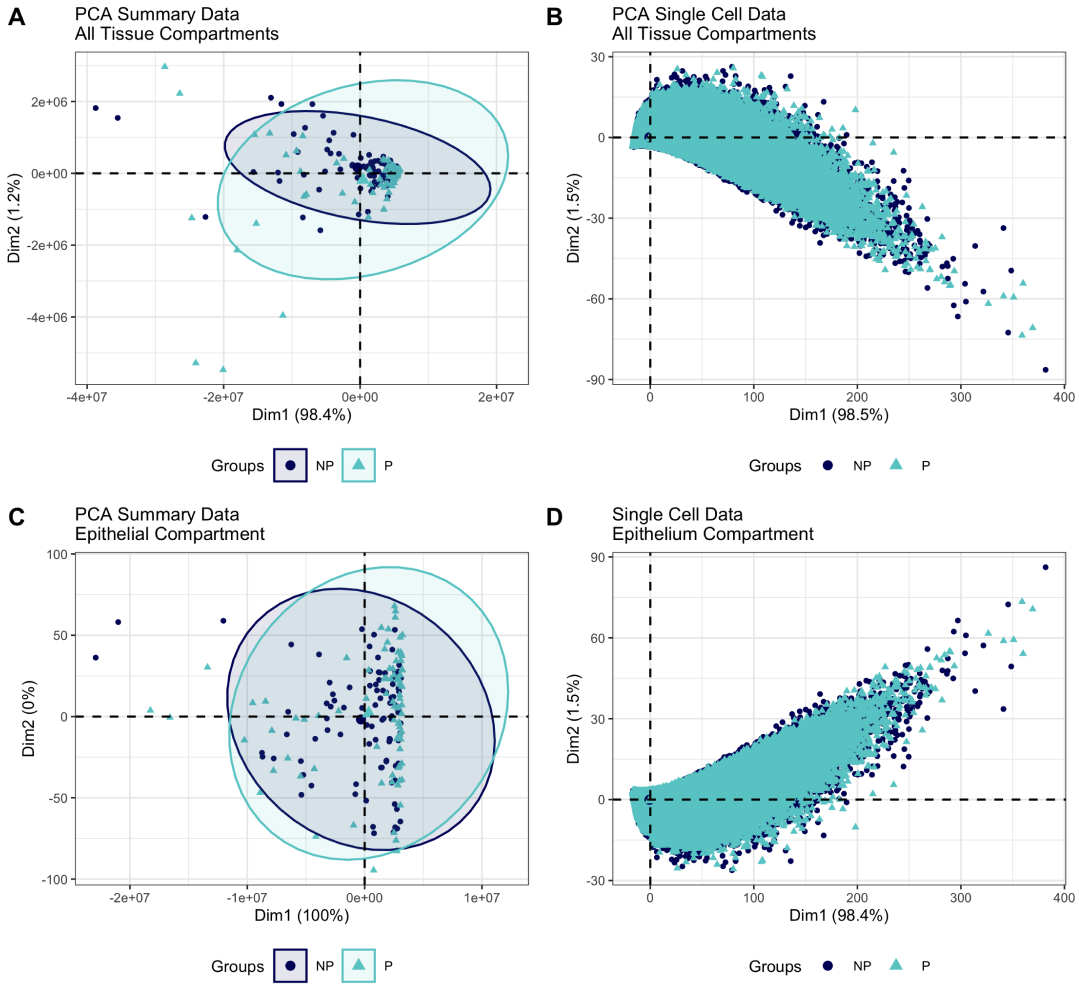


Figure 5.19: Exploratory analysis was undertaken to first explore the data generated from the digital analysis of the LGD. **A.** PCA performed on data for all tissue compartment using the summary data. **B.** PCA performed on data for all tissue compartment using the single cell data. **C.** PCA performed on data for only the epithelial compartment using the summary data. **D.** PCA performed on data for only the epithelial compartment using the single cell data. No clear separation between progressor and nonprogressor regions/cells can be seen.

5.3.3.4 Analysis of the histological features of the epithelium

Next, focusing on the epithelium, three nuclear features indicative of cell integrity: nucleus area, nucleus roundness and nuclear staining intensity were interrogated. An abnormal/cancer cell presents as rounder, larger and with an intensely staining nucleus due to its higher chromatin content. Based on this, the analysis aimed at assessing whether the greater number of aneuploid changes found in progressor LGD could be detected by alteration in these features. First, the summary data of the epithelium

compartment, comprising average tissue values for each region, were investigated. No clear separation between the nonprogressors and progressors was seen based on all three histological features using average tissue values. Statistical comparison of the three cellular features revealed significant results for nucleus roundness. However, a closer look at the distribution of the samples shows that there is a large overlap between the two groups, which suggests this feature would not be a strong discriminator (Figure 5.20 A). Following, the single cell data were assessed. Specifically, the cells within the top quartile (25% worse cells per region) based on nucleus area first, then nuclear staining intensity and finally nucleus roundness were investigated. The choice to focus on specifically these cells within each region was dictated by two factors:

- Normal cells within the dysplastic tissue areas might be providing noise to the signal;
- Clinical pathology diagnoses are usually based on maximal detected severity.

However, even after the attempt to lower the "noise" from normal cells and maximise the signal from cytologically "abnormal" cells, no clear segregation of nonprogressor and progressor cells was seen, despite the t-test for nucleus roundness being statistically significant (Figure 5.20 B).

Lastly, to more directly assess the association between CNAs and changes in nuclear features, the 25% worse cells of each patient's region with the highest CNA burden was examined. For this assessment, single cell values of epithelial cells in the top quartile based nucleus area, nucleus roundness and nuclear staining intensity in the worst nonprogressor and progressor regions based on CNA were compared. In this case no statistically significant difference between nonprogressors and progressors in all three features was detected (Figure 5.21).

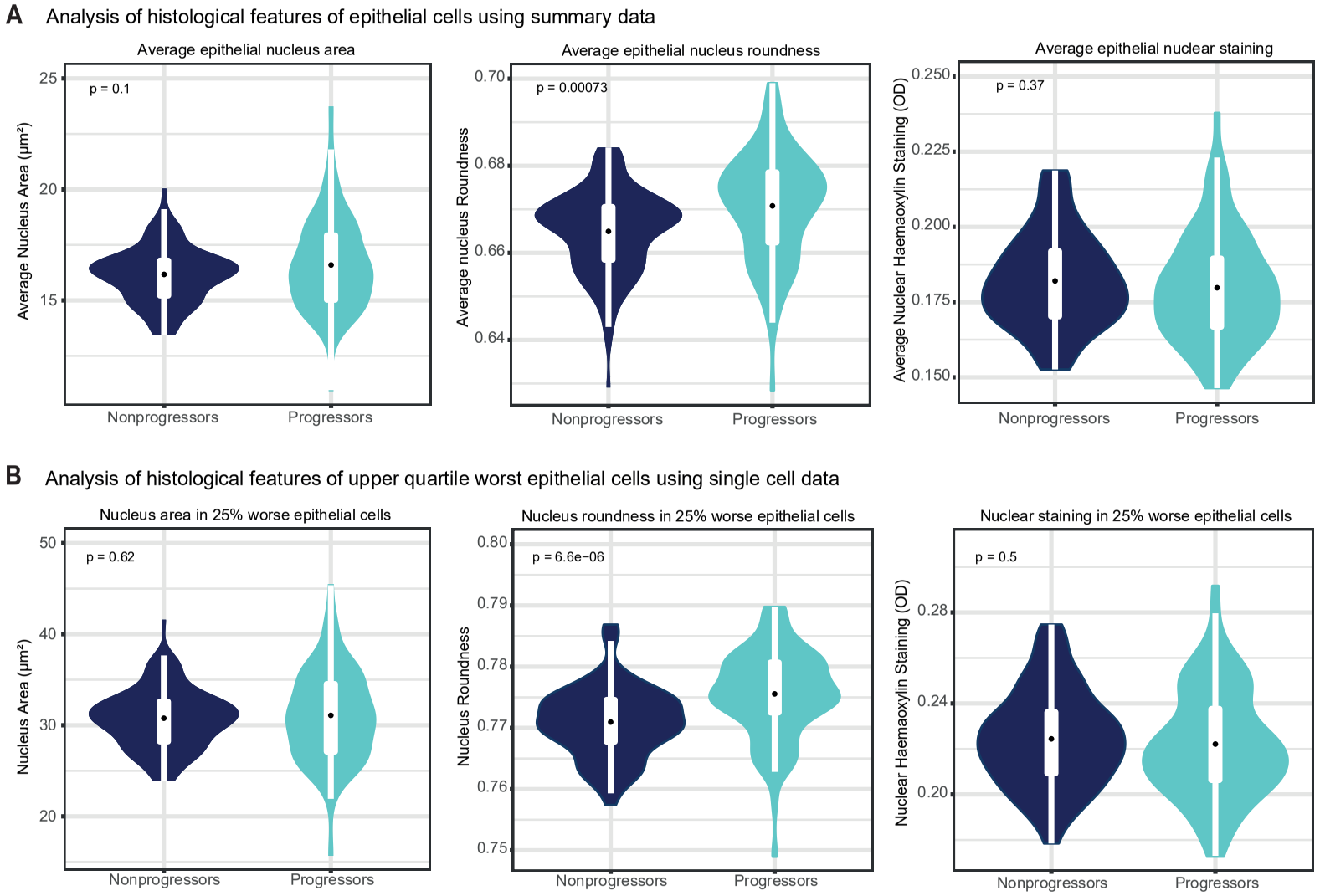


Figure 5.20: Assessment of histological features on the epithelial compartment. Nucleus area, roundness and staining intensity were assessed in the epithelial compartment. Nuclear area is expressed in μm^2 Nuclear staining intensity is reported in OD, a measure of average staining intensity of a cell, calculated as log(max intensity/mean intensity). Values for nucleus roundness vary from 1 to <1, with 1 indicating complete roundness. **Top.** Assessment of the three histological features in the epithelial compartment using the summary data. No clear separation between progressor and nonprogressor LGD can be seen. **Bottom.** Assessment of the epithelial cells in the top quartile based on nucleus area (left), nucleus roundness (centre) and nuclear staining intensity (right).

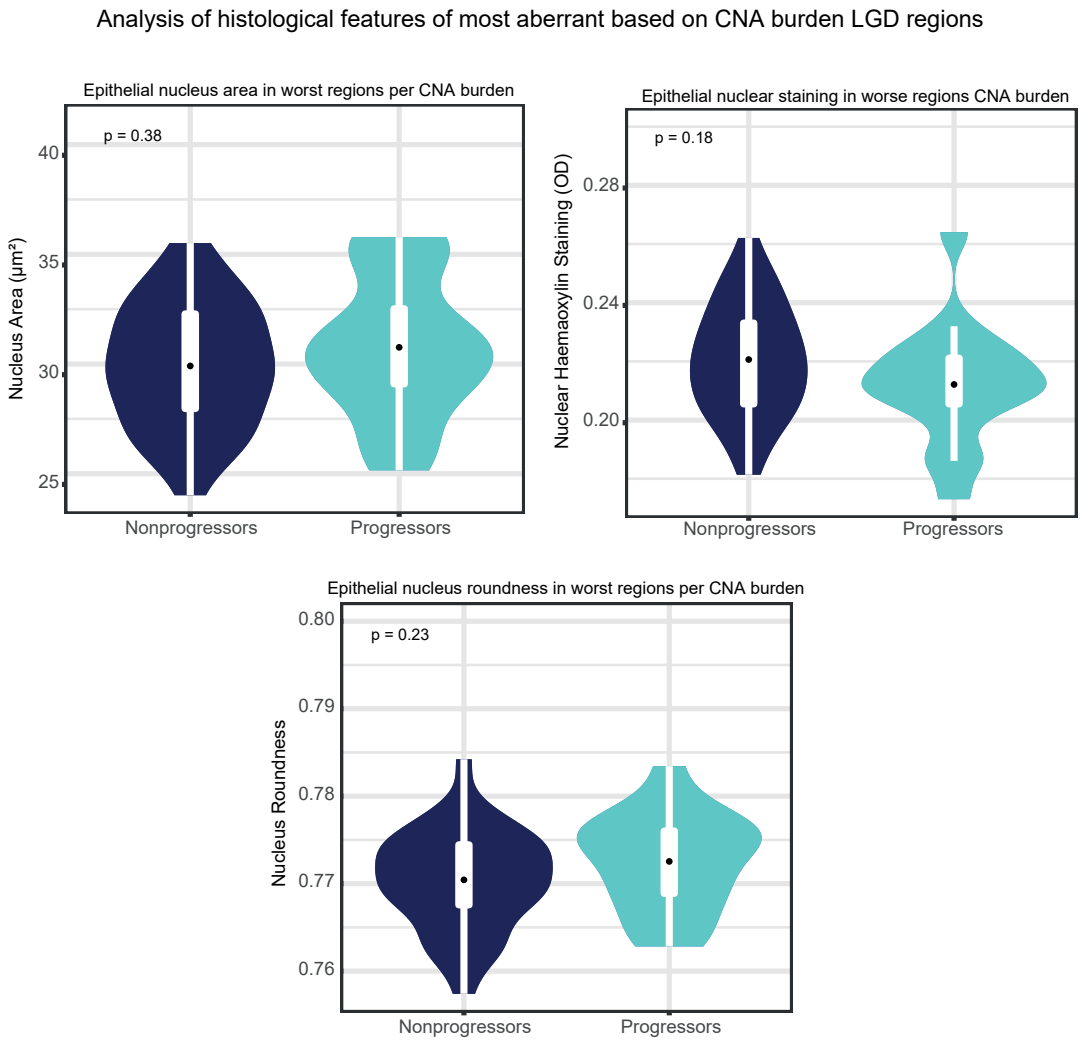


Figure 5.21: Assessment of histological features in the most aberrant regions per patient (i.e. one region per patient) based on CNA burden. Comparison of the epithelial cells in the top quartile based on nucleus area (top left), nucleus roundness (bottom) and nuclear staining intensity (top right) from the histological regions with the highest CNAs burden.

In conclusion, I could not find cytological surrogate measures of the detected differences in CNA burden between progressor and nonprogressor patients. This suggests that there is no evident correlation between genotype and histological phenotype, at least for the nuclear features analysed, or that it is drowned out by "normal noise" from other cells, as in the case of nucleus roundness. Furthermore, the inability to detect a difference between progressor and nonprogressors using digital pathology analysis evidences the difficulty experienced by pathologist in assessing LGD and further supports the use for a molecular biomarker instead.

Collectively, from the histological, molecular and cytohistological assessment of the colitis-associated LGD, CNA burden was the only cancer risk predictor capable of discriminating between LGDs that are likely to progress and LGDs that are unlikely to.

5.4 Key findings

- Poor consensus was obtained in the classification of colitis-associated dysplasia using the Choi *et al.* classification system.
- In this cohort of colitis-associated LGD, no association between nonconventional dysplasia and progression was found.
- CNA burden is statistically higher in progressor lesions than nonprogressor colitis-associated LGD.
- No cytological surrogate measures of CNA burden were found with the digital pathology approach tested.
- No correlation between genotype and nucleus area, roundness and intensity was found in this cohort.
- From the stratification tools assessed, CNA burden was the only cancer risk predictor capable of discriminating between LGDs that are likely to progress and LGDs that are unlikely to.

5.5 Discussion

Determination of cancer risk is pivotal in decision-making, particularly when considering management strategies. Currently, upon LGD detection, IBD patients are offered the possibility to undergo continued and frequent surveillance or preventive colectomy, without supportive individualised cancer risk estimates. On one hand, some patients undergo potentially unnecessary prophylactic colectomy due to fears of developing CRC, which inevitably has an impact on their quality of life. On the other,

some patients opt out of surgery due to concerns about negative impacts of complications, stoma or ileoanal pouch function on their quality of life, and are later diagnosed with CRC, often at a late stage [208]. Therefore, there is a clear clinical need for better evidence-based stratification tools in colitis-associated dysplasia management. Evidence-based and invidualised numerical CRC risk predictions would improve the effectiveness of IBD patients' management. They would facilitate communication and decision-making by aiding in the stratification of patients into low-risk, who could be reassured by undergoing surveillance, and high-risk, who would be more confident of their benefit from preventive colectomy.

Here, colitis-associated LGD were histologically, molecularly and cytohistologically assessed to test and identify a cancer risk predictor capable of discriminating between LGD that are likely to progress and LGD that are unlikely to.

The histological phenotype-based stratification tool proposed by Choi *et al.* stratified dysplasia on the basis of specific histological features and identified their association with CRC. This provided an increased layer of specificity and information regarding individual LGD cancer risk, which the current rather simplistic histological assessment strategy based on grading lacks. Furthermore, the tool had the benefit of being easily applicable in the clinical setting, not requiring novel sample processing techniques to be implemented or additional costs. However, its assessment in the 243 regions with six expert pathologists highlighted the difficulty in assessing and classifying colitis-associated dysplasia with high fidelity. The inter-observer agreement is believed to have been negatively impacted by the complexity of the classification criteria, characterised by several categories, some of which only minimally different and others allowing space for interpretation. In support of this is the fact that inter-rater agreement was better in the grading of all regions (colitis and sporadic). This is to be added to the already known difficulty in assessing colitis-associated LGD, evidenced by the better inter-rater agreement in the scoring of the sporadic regions. Moreover, the association between nonconventional dysplasia and progression reported by the Choi group could not be validated in this cohort. Although, this could have been hindered by the difficulty in using such a complex classification system. The tool could poten-

tially be improved in the future by simplifying some similar categories (both histologically and in terms of cancer risk) into a single category. Given the poor correlation of nonconventional dysplasia with prospective cancer outcome and the complexity of the classification system, its practical use as a clinically applicable tool is likely to be limited.

In terms of molecular-based stratification, CNA burden was shown to be a highly specific cancer-risk predictor. LGD have been found to be diverse in CNA burden, with high-risk/progressor LGD typically bearing a significantly higher CNA burden. In addition to being able to predict cancer with high specificity, the use of CNA as a stratification tool has the benefit of being an objective and quantitative tool, not hindered by inter-rater variability. Furthermore, lpWGS, which was used to determine CNA burden, is a cost-effective technique that can be applied to archival diagnostic (FFPE) tissue, favouring the tool's integration in the clinic.

The use of cytohistological assessment as a stratification tool had the benefit of being based on a strategy already used in clinic i.e. on-slide histological assessment of lesions. Furthermore, the current efforts to introduce digital histopathology in the health care [219] meant that it could be implementable in the clinical care setting. Nonetheless, the striking genomic differences observed between nonprogressor and progressor LGD did not translate into clear differences at the cytohistological level. This suggests that there are no detectable differences in the parameters assessed that can be identified with current technologies that would negate the need for CNA molecular testing.

Both the histological phenotype-based stratification tool and the cytohistological assessment-based stratification tool could potentially be improved in the future through the application of AI algorithms. In terms of the histological phenotype-based stratification tool, AI algorithms could be trained to recognise patterns that could distinguish between the different categories. The use of AI algorithms might also aid in the identification of subtle cellular discriminatory features. However, in both instances large discovery and training sets are required. This would be particularly difficult in CA-CRC due to the fact that only a small proportion of IBD patients develop dysplasia and therefore would require a multicentre, multination effort.

The assessment of the three stratification tools led to the emergence of CNA burden as the only strong cancer risk prediction tool. Interestingly, in Choi *et al.*'s sophisticated histological assessment of dysplasia, a high rate of aneuploidy was found in all histological subclasses that were more likely associated with HGD and CRC, except for SSL-like dysplasia. This is concordant with CNA burden being significantly higher in progressor LGD, but also highlights a potential blind spot of this molecular cancer risk predictor: serrated lesions. Serrated lesions, particularly SSL, are known to be enriched for MSI. MSI lesions have a defective DNA mismatch repair system, thus acquire multiple frameshift SNA in microsatellite regions. Because of the above they rarely acquire CIN as a cancer driver. Hence, the use of CNA as a cancer prediction tool would not be adequate in these lesions. It follows that CNA burden is a strong molecular cancer risk predictor in all IBD patients, except those with serrated lesions. Future work should be undertaken to determine colitis-associated serrated lesions' cancer risk in a larger cohort comprising more serrated lesions. In fact, the increased cancer risk found in SSL-like lesions by the Choi group could not be confirmed in this study, but this is very likely to have been impacted by the number of lesions available (n = 1). If SSL-like dysplasia were to be confirmed to be associated with increased cancer risk, then cancer risk for this specific lesion type would be attributed histologically.

Based on the findings in this chapter, I propose that the current surveillance program be altered to a two-step combined morphological and molecular test process. Firstly, histological assessment is used to determine whether a lesion is serrated or not. All non-serrated lesions are then subjected to CNA molecular testing to predict their cancer risk. Whilst for the serrated lesions, if SSL-like dysplasia is confirmed as a histologically detectable risk, then risk can be attributed histologically. If not, they would need to be monitored as histologically low-risk lesions. However, until their cancer risk is established, serrated lesions should be monitored closely as potentially high-risk lesions.

5.6 Limitations and Future work

Several limitations have been identified in this study. These are outlined below:

- All three analyses were performed on limited size cohorts. Despite the relative small number of patients and lesions, the striking genomic differences between progressor and nonprogressor colitis-associated LGD could be detected. Nevertheless, a larger cohort would have been preferred, particularly for the histological and cytohistological assessments, where in the presence of a large cohort machine learning approaches could have been attempted. However, this was not possible within the time constraints of a DPhil due the rarity of colitis-associated LGD and the difficulty of obtaining archival tissue under the Human Tissue Act.

- In this study, a direct comparison in CNA burden and profiles between colitis-associated LGD with sporadic adenomas was lacking.

- A fine mapping of the chromosomal regions and genes involved in the most common CNA detected was not undertaken. This was out of the scope of this chapter, however would have contributed to the understanding of the differences between progressor and nonprogressor LGD.

Based on the findings in this chapter, I propose two main future investigations to further improve IBD patient cancer risk prediction and the molecular tool validated:

- Currently, although associated with a low-risk of progression, sporadic lesions in IBD patients cannot be discerned from true colitis-associated LGD. Assessment of CNAs could lead to the identification of CNA signatures specific to one or the other that could aid in distinguishing the two and thereby better inform cancer risk. Identification of differences in CNA patterns between colitis-associated LGD and sporadic LGD could also help in elucidating the molecular differences in the early carcinogenesis events of S-CRC and CA-CRC.

- The current tool relies on the detection of LGD. Despite, the advances in endoscopic techniques and improved detection rates of LGD, this can still be difficult,

particularly when dealing with flat lesions. The tool should be tested on random biopsies from progressors and nonprogressor patients as well as surrounding normal mucosa to determine if we can differentiate between these two patient groups on the basis of CNA even in these samples, without the need for idenfication of dysplasia.

Chapter 6

Thesis discussion

6.1 Summary of results and clinical relevance

In this thesis, I have tried to identify and exploit biological and molecular differences between sporadic and colitis-associated carcinogenesis in order to better elucidate the carcinogenesis of CA-CRC and ultimately improve patient management. Three aspects were specifically tackled: the stem cell dynamics in S-CRC and CA-CRC, the role of the WNT signalling pathway in CA-CRC and the early detection of CA-CRC, with all three lines of investigation having a translational potential.

The investigation of the stem cell dynamics in S-CRC and CA-CRC revealed that tumours are supported by an admixture of CBC and RSC, which can be measured with the stem cell index. The degree of CBC-RSC cell admixture could be mapped across CRC subtypes, with colitis-associated precursor and malignant lesions having a predominantly RSC stem cell phenotype. The degree of CBC-RSC, referred to as stem cell phenotype, was associated with both epithelial mutation and micro-environmental signalling disruption. Furthermore, the stem cell phenotype was found to be adaptive, showing a dynamic response to selective pressure, such as targeted Lgr5-positive cells ablation and chemotherapy. The ability of tumours to adapt to selective pressure was associated with poor response to chemotherapy. The understanding of the CRC stem cell phenotype and the factors influencing it opens the opportunity for the development of novel drugs that can specifically target both stem cell phenotypic peaks. Sim-

ilarly, knowledge of the signalling pathways involved in stem cell phenotype shifting upon selective pressure allows for the development of agents that can inhibit activation of these pathways, to thereby improve treatment success, prevent chemoresistance and relapse. In the meantime, the stem cell index is a simple and applicable tool and its ability to detect shifts in stem cell phenotype renders it an attractive tool to assess response to treatment and inform scheduling of current therapeutic strategies.

The investigation of the WNT landscape in CA-CRC suggested a preferential selection of LD alterations in colitis-associated LGD, with LD alterations being found in 56.4% of the small cohort of lesions with established molecular ground truth. To understand the reason behind the possible preferential selection of LD alterations, the effects of Rspo3 overexpression on cells carrying the alteration and their neighbouring cells was investigated in murine models of WNT disruption. Rspo3 overexpression in Lgr5-negative cells was found to favour both cell endogenous and exogenous skewing of secretory cell fate towards Paneth cells. Increasing Paneth cell proportions may encourage regenerative healing through bactericidal effects from secreted proteins and by encouraging stem cell behaviour in residual cells through local WNT ligand expression. Based on this observation, cells carrying *RSPO* fusions were proposed to provide a fitness advantage in a regenerative milieu, through both cell autonomous and non-cell autonomous skewing of residual cell behaviour. The finding that colitis-associated lesions might be selecting for LD alterations opens the possibility for targeted treatment with WNT inhibitors. Furthermore, where future work was to confirm increased Paneth cell differentiation (which can be detected by routine H&E) as a histological marker for *RSPO* alterations, this could be an inexpensive means of triaging lesions for more expensive molecular testing.

In the chapter investigating the early detection of CA-CRC, CNA burden was shown to be an effective and highly specific cancer-risk predictor for all colitis-associated LGD except for serrated colitis-associated LGD. This led to the proposal of a two-step combined morphological and molecular test process. According to this, morphological assessment is to be used to distinguish between serrated and non-serrated colitis-associated LGD. Following, non-serrated colitis-associated are subjected to CNA molec-

ular testing to predict cancer risk, with high CNA representing high risk of progression. Whereas serrated lesions are to be monitored as potentially high-risk, until future work can confirm their specific cancer risk. These findings aim directly at improving an expensive and ineffective national surveillance programme. The proposed two-step combined morphological and molecular test process would allow for low-risk patients to be reassured and subjected to fewer invasive procedures. Whereas high-risk patients would have stronger proof of the benefit of undergoing colectomy or, where they opted against it, they could be more closely monitored to enable detection of early signs of cancer. This would improve the effectiveness of the surveillance programme, potentially leading to at least fewer interval cancers as well as lower costs to the NHS.

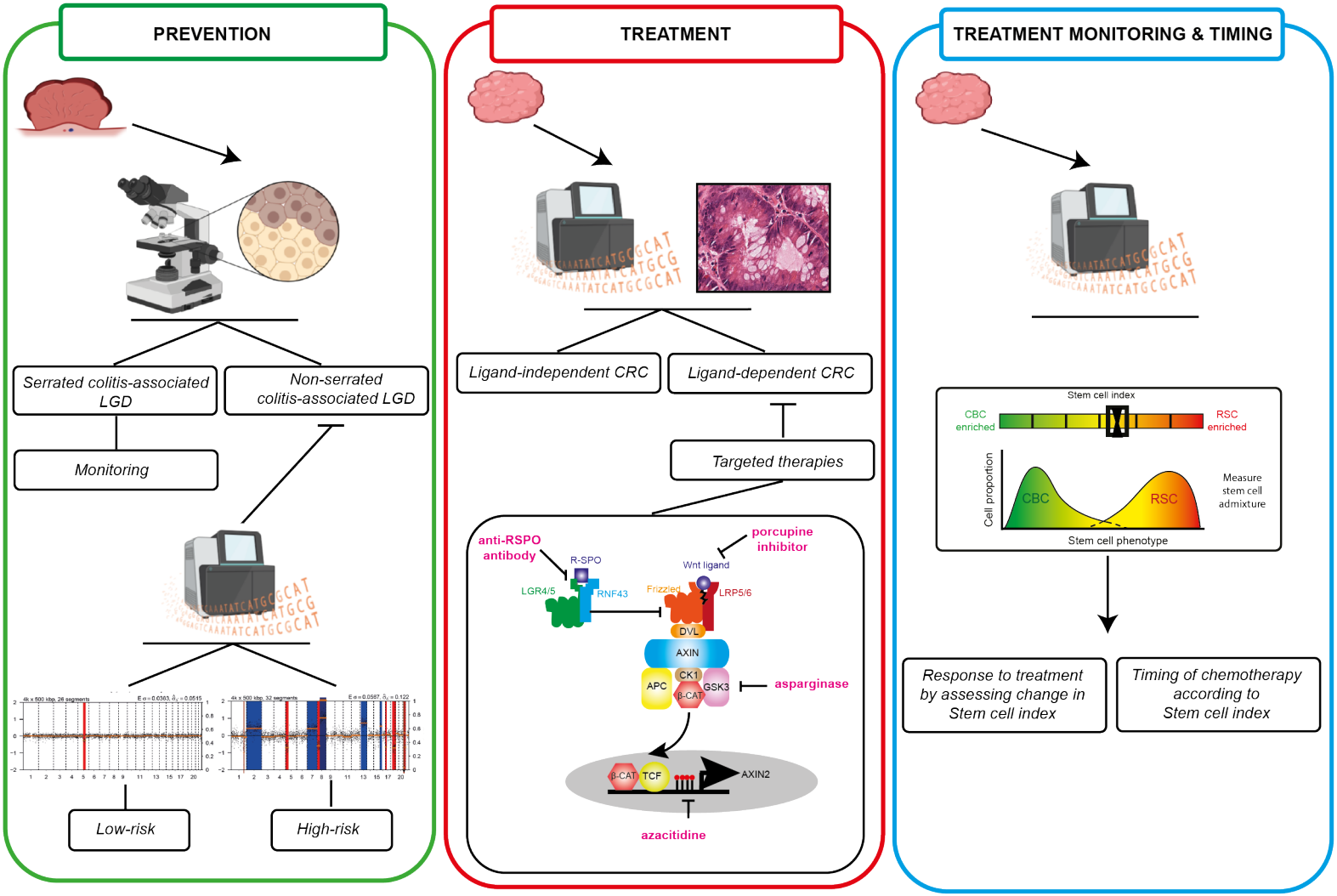


Figure 6.1: Work in this thesis has the potential of improving three aspects of IBD cancer-risk and treatment management: **prevention**, through the proposed two-stem combined morphological and molecular test process, **treatment**, through targetted treatment of LD CA-CRC and **treatment monitoring and timing**, by assessing changes in stem cell phenotype using the stem cell index.

6.2 Main discussion

6.2.1 Colitis-associated colorectal cancer: a molecularly heterogeneous group

The molecular heterogeneity of S-CRC is a well-recognised aspect of the condition, which has been subject of extensive research and led to the development of at least two main S-CRC classification systems of biological and clinical significance. On the other hand, CA-CRC is still largely seen and studied as a single group. Work in this thesis highlights the heterogeneity within colitis-associated precancerous and malignant lesions. In addition to the histological differences between colitis-associated dysplasia described by Choi *et al.* [141], the investigations of the stem cell dynamics, the WNT signalling landscape and analysis of CNA alteration burden in colitis-associated LGD revealed molecular differences within this form of CRC.

Despite colitis-associated lesions having a predominantly RSC stem cell phenotype, a small proportion were predominantly CBC. This is supported by previous identifications of heterogeneous stem cell populations in colitic tumours in murine models [220]. Similarly, the majority of colitis-associated lesions were predicted and/or confirmed to be LD. However, a smaller proportion of colitis-associated LD were predicted/confirmed to be LI. Furthermore, CNA burden analysis highlighted the molecular difference between non-serrated colitis-associated LGD, which progress through genomic instability, and serrated colitis-associated LGD, which tend to be MSI. These observations are suggestive of the fact that, similarly to sporadic lesions, colitis-associated lesions are not a single homogenous type of CRC, but rather a heterogeneous group that needs to be better characterised and molecularly phenotyped. Another postulation to explain the heterogeneity amongst colitis-associated lesions is that the lesions having a predominantly CBC phenotype and/or carrying LI alterations are in fact sporadic lesions occurring in IBD patients. However, the latter would be difficult to prove as there are currently no means to distinguish between sporadic lesions occurring in IBD patients and true colitis-associated lesions [90].

Regardless of the explanation behind the heterogeneity, future work should focus in classifying CA-CRC into molecular subtypes with clear distinguishing features and

biological interpretability, which can be used for clinical stratification and targeted intervention. This would require a multicentre cooperation as CA-CRC are rare. However, a starting point could be molecularly differentiating between lesions that behave as "true colitis lesions" and those that behave "like sporadic lesions" and assessing whether there is a difference in terms of cancer risk.

6.2.2 Colitis-associated colorectal cancer: cancer stem cell and cancer cell of origin

The identification of the cells that initiate and sustain tumorigenesis is key to addressing tumour cell heterogeneity and understanding the tumour biology of cancers [221]. With regard to the cancer stem cell, in this thesis colitis-associated dysplasia and cancers were found to predominantly have a RSC stem cell phenotype. This is in line with recent findings from Iwaya *et al.*, who reported a lack of LGR5 expression in human CA-CRCs [222]. The reason behind the lower LGR5 expression and a predominantly RSC stem cell profile can be postulated to lie in the association between CA-CRC, chronic inflammation and the regenerative nature of the RSC. Under non-cancerous conditions, RSC have been shown to arise following damage through the dedifferentiation of committed cells to enable regeneration of the epithelium. This is a process associated with the upregulation of several key pathways, including the YAP1/TAZ signalling pathway [30]. Cycles of inflammation and regeneration are key features of IBD. They are also believed to be the driving force of CA-CRC carcinogenesis, causing accumulation of ageing-associated mutations and activation of key pro-tumorigenic pathways, which has been previously shown to cause an increase in some RSC populations [96, 104, 105]. In light of this, it can be postulated that the injury-induced cell turnover required for regeneration in IBD relies on RSCs, which then harbour ageing-associated alterations due to the chronic and cyclic nature of the condition. This eventually leads to the initiation and development of colitis-associated neoplasia. Based on this postulation, RSC could be both the predominant cancer stem cell population and the stem cell of origin of CA-CRC. If confirmed it would be congruent with the observations on the sporadic lesions i.e. TVA having a predominantly CBC stem cell

phenotype and CBC as cancer-initiating cell. Determining the identity of the cancer cell of origin could enable a better understanding of the genetic changes, including the identification of truncal driver mutations implicated in tumorigenesis and progression and aid in the identification of early disease biomarkers. This could have a great impact on the development of preventive and therapeutic measures, specifically those aiming at the suppression or reversal of the initial phase of disease.

Another factor to be taken into account with regard to cancer stem cells and the cancer cell of origin is the role of the WNT signalling pathway. In chapter 3, I showed that the CBC phenotype is associated with *APC* mutations, whilst the RSC phenotype is associated with LD alterations. In chapter 4, I outlined that WNT alterations are selectively and preferentially acquired in different precursor and malignant CRC subtypes, with TVAs selecting for LI mutations, whilst serrated precursor lesions, and perhaps colitis-associated precursor lesions, preferentially selecting for LD alterations. To clarify with an example, TVA, which are the classical type of precursor lesions and have a predominantly CBC phenotype, arise from *APC* mutation in CBC cells [18]. This is in line with the hypothesis that strong constitutive WNT signalling is required to induce and maintain neoplastic growth from CBC cells [176]; whilst a more contained WNT signalling may be favoured in RSCs lesions, which are supported by other drivers, including the activation of key microenvironmental pro-tumorigenic pathways.

6.2.3 WNT signalling alterations selfish vs altruistic model

It is now well-recognised that WNT hyperactivation can be achieved through different LI and LD alterations. But, how do these different alterations contribute to the general fitness of mutant cells, to their evolution to malignant cells and ultimately to the characteristics of a given tumour? Assessment of the cell dynamic between WNT mutant cells and their neighbouring cells has aided in understanding that.

Recent studies have shown that *Apc* mutations in Lgr5-positive cells give rise to super-competitors, whereby mutant cells hold an endogenous fitness advantage provided by the intrinsic and constitutive WNT activation, coupled with suppression of neighbouring wild-type cells' stemness and promotion of differentiation through se-

cretion of Notum. This allows *Apc*-mutant cells to "fix" and outcompete wild-type cells. In this thesis, *Rspo3*-mutant Atoh1-positive cells were shown not to secrete Notum, indicating that they might not suppress their neighbouring cells' stemness. Instead, *Rspo3*-mutant cells secreted Rspo3. Rspo3 expression from *Rspo3*-mutant cells was seen to promote endogenous and exogenous skewing of secretory cell fate towards Paneth cells. Given their role as a source of WNT ligands, an increase in Paneth cells, coupled with the presence of elevated Rspo3, is believed to create a WNT-high environment that could potentially promote stemness in all (mutant and wild-type) cells within the milieu. Based on these observations a selfish versus altruistic cell dynamics model is proposed. According to this, *Apc* mutations in Lgr5-positive cells give rise to selfish super-competitors, which act by promoting their own stemness while suppressing that of their neighbouring cells. On the contrary, *Rspo3*-mutant cells' expression of Rspo3 supports their own and their neighbouring cells' stemness by potentiating WNT signalling, through direct effect of secreted Rspo3 and an expansion of Paneth cells. The essence of the difference in behaviour between *Apc*-mutant and *Rspo3*-mutant is encapsuled in the difference in key mediator; while *Apc*-mutant cells secrete a WNT suppressor, which causes a depletion of WNT ligands for surrounding cells, *Rspo3*-mutant cells secrete a WNT potentiator, which enhances and amplifies WNT ligand availability for all cells in the milieu.

Given the expression of WNT negative regulators and mechanism of WNT activation of LI lesions the "selfish super-competitors" model of *Apc*-mutant cells is expected to be valid for *Ctnnb1*-mutant cells too. The "altruistic" model of *Rspo*-mutant cells is a little more complex as it relies on the expression of Rspo3 to potentiate WNT signalling. Consequently, although it has the potential of being "applicable" to *Rnf43*-mutant cells given that both cell types do not express WNT negative regulators, thus lack WNT inhibitors for the neighbouring cells, the mechanism by which *Rnf43*-mutant cells would hypothetically promote neighbouring cells' stemness is expected to be different. To link this back to the suggested preferential selection of LD in CA-CRC, these *Rspo*-mutant altruistic cells might be favoured in a colitic environment, where high proliferative potential is required to regenerate the epithelium. However, this is just a

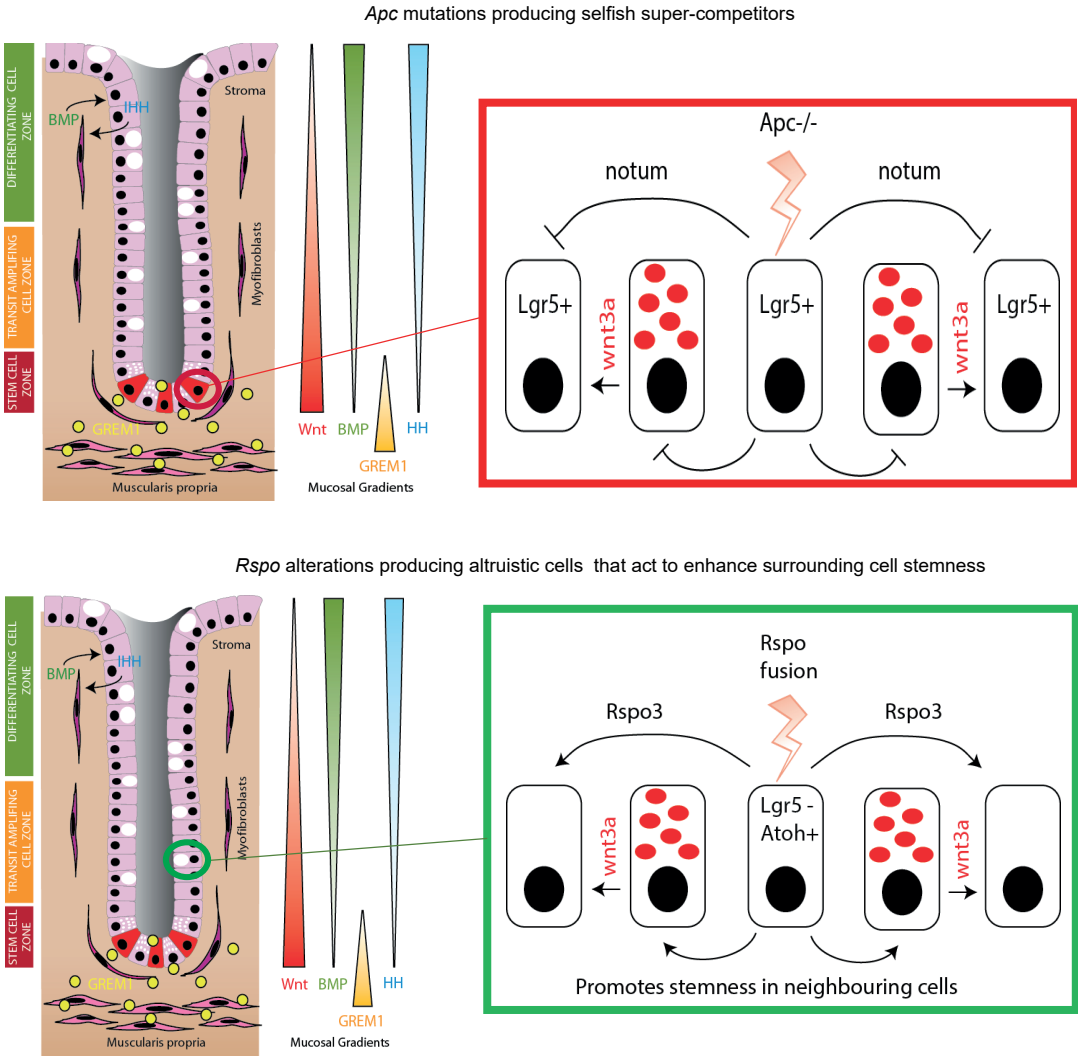


Figure 6.2: Model outlining the cell dynamics between *Apc*-mutant and *Rspo3*-mutant cells and their neighbouring cells. *Apc*-mutant Lgr5-positive cells have constitutive activation of the WNT signalling pathway, which provides them with a fitness advantage, and express/secrete Notum as part of the negative feedback loop to constitutive WNT signalling. The expressed Notum has a paracrine effect on the wild-type neighbouring cells, which results in inhibition of their stemness and promotion of their differentiation. *Rspo3*-mutant Lgr5-negative cells express/secrete Rspo3. The secreted Rspo3 promotes their own and their neighbouring cells' stemness through a direct paracrine effect and by causing an expansion of Paneth cells.

conceptual framework and substantial measurement of LD/LI mutations on stem cell fate is needed for it to be proven.

6.2.4 From *TP53* mutation to aneuploidy

P53, encoded by *TP53* in humans and *Trp53* in mice, is a key tumour suppressor, often referred to as the "guardian of the genome". It plays a pivotal role in preserving genomic integrity and DNA fidelity through several biological processes, including inducing DNA repair, cell arrest, senescence and apoptosis in cells with damaged genomes or exposed to harsh environments, including severe metabolic and oxidative stress. *TP53*, located in 17p13, is commonly mutated in nondysplastic mucosa of IBD patients, representing a very early event in colitis-associated carcinogenesis [217]. *TP53* mutation in IBD is believed to promote cell survival in the inflamed bowel, thereby providing a fitness advantage against the harsh environment of the colitic mucosa. A consistent finding across cancer types is a strong positive correlation between *TP53* mutation and aneuploidy, with mutations in *TP53* being associated with development of aneuploidy. This, together with the observation of other oncogenic phenotypes in association with *TP53* mutation, has led to the concept of mutant *TP53* gain-of-function, whereby mutant *TP53* loses its tumour suppressor function to gain novel oncogenic functions (including induction of aneuploidy) [223].

Aneuploidy is a common early finding in IBD and in this thesis increased CNA burden was found to be associated with progression, with loss of chromosome 17p being one of the CNA occurring at highest rate in progressor LGD compared to non-progressor LGD. If the postulation that aneuploidy is the by-product of *TP53* mutation is considered, it can be hypothesised that *TP53* mutation, acquired to promote cell survival in the inflamed bowel, is the first hit that initiates the early carcinogenesis and consequently genomic instability. Mutation of one allele has previously been shown to drive loss of the wild-type. In the case of CA-CRC, the second hit in *TP53* that leads to loss of heterozygosity could be acquired through loss of chromosome 17. This in turn would cause a bi-allelic inactivation of *TP53*, which together with the acquisition of other CNA increases genomic instability [223–225], leading to a punctuated accruing

of CNA and eventually progression from LGD to HGD. This could potentially explain the increase in CNA in progressor LGD compared to nonprogressor LGD, and the dramatic increase in CNA in the progression from LGD to HGD.

While being a "by-product" or better a consequence of *TP53* mutation, aneuploidy could also allow for the rapid acquisition of adaptive phenotypes to enhance fitness and promote cell survival in the inflamed mucosa. Moreover, a recent study has described aneuploidy as the mechanism driving the gain-of-function phenotypes associated with *TP53* mutation [225]. Specific to CA-CRC, in chapter 5 of this thesis, several chromosomal gains and losses associated with colitis-associated lesions in general and with progression were identified, which could be linked with specific oncogenic phenotypes, including increased proliferation. To exemplify, a CNA most commonly found in progressor LGD was gain of 8q. This was previously discussed to be the chromosomal location of *MYC*, pro-oncogene involved in cell proliferation and survival, and thereby associated with overexpression of *MYC*. Similarly, the chromosomal locations of the WNT negative regulators, *APC* and *RNF43*, had also been found to be most commonly lost in progressor LGD, suggesting another means for WNT hyperactivation. All of these promote increased cell proliferation. The recurrence of these genomic alterations, together with their association with important oncogenic phenotypes, would suggest that aneuploidy is not just random genomic chaos initiated by *TP53* mutation, but rather allows for the acquisition of specific gains and losses, each associated with specific relevant pathways that enable the acquisition of fitness advantages and/or oncogenic phenotypes. On this matter, Baslan *et al.* found that the acquisition of genomic alterations following *Trp53* inactivation in the setting of pancreatic transformation is not random, but rather subjected to deterministic patterns that contribute to the genomic and biological hallmarks of *Trp53*-mutant tumours. In their study, progression from *Trp53* inactivation followed four phases, *Trp53* loss of heterozygosity, accumulation of deletion, genome doubling and the emergence of gains and amplifications, each associated with specific histological stages across the premalignant and malignant spectrum and with the deletion events following *Trp53* inactivation functionally targeting relevant pathways that can shape genomic evolu-

tion. Furthermore, the authors found that loss of tumour suppressor functions and the ensuing genomic instability functionally contributed to the emergence of aggressive cancer in the pancreatic ductal adenocarcinoma model [223], in congruence with CA-CRCs being more aggressive that S-CRCs. Lastly, the co-occurrence of *TP53* mutation, loss of 17p and accruing of CNA so early in the carcinogenesis of CA-CRC, when it occurs in advanced stages in S-CRC, supports the notion that CA-CRC evolves down an accelerated evolutionary trajectory distinct to that of S-CRC.

6.2.5 Colitis-associated colorectal cancer carcinogenesis: how does work from this thesis shape our current understanding

Understanding the molecular mechanisms underlying colitis-associated carcinogenesis is key to improving prevention and treatment of this form of CRC. With the aim to better characterise CA-CRC, work in this thesis progressed from interrogating the stem cell dynamics in S-CRC and CA-CRC, to investigating the role of WNT signalling in CA-CRC, to finish with assessing histological and molecular markers of progression in colitis-associated LGD. Each investigation contributing to our understanding of CA-CRC and of the differences between S-CRC and CA-CRC.

The key to understanding the difference between S-CRC and CA-CRC, as obvious as it seems, is the central role played by inflammation in CA-CRC. Work in this thesis had led to the identification of RSC as the predominant cancer stem cell and to its suggestion as cancer cell of origin. This was extensively discussed above and postulated to be associated with the regenerative nature of RSC and the fact that CA-CRC arises as a result of chronic inflammation. Furthermore, CA-CRC was suggested to preferentially select for LD alterations due to the postulated regenerative advantage that LD alterations, specifically *RSPO* alterations, might provide. Moreover, the acquisition of *TP53* mutation, with subsequent loss of heterozygosity and resulting accruing of CNA, was associated with the need to promote cell survival in the inflamed bowel.

Based on the findings in this thesis, particularly the ones highlighted above, I propose that inflammation is the shaping force behind the features of CA-CRC. In other words, the tumour behaviour, the mutational signatures and other characteristics as-

sociated with CA-CRC are shaped by inflammation. They are either a by-product of chronic inflammation or the result of an adaptive response to chronic inflammation. For instance, the ageing-mutational signatures detected in CA-CRC by Baker *et al.* [96] are a by-product of injury-induced cell turnover. On the other hand, the preferential selection of LD alteration and *TP53* mutation are an adaptive response, with the former supporting regeneration and the latter supporting cell survival under the harsh conditions of the inflamed mucosa. Baker *et al.* proposed that chronic inflammation is the drive that provides a distinctive mutational selective pressure [96]. Similarly, in my proposed working theory, inflammation is not only the causative force, but also the selective force that shapes the features of CA-CRC.

6.3 Final remarks

In conclusion, in this thesis, which aimed at better characterising CA-CRC, several aspects of this form of CRC were elucidated. Key findings include the identification of RSC as the predominant stem cell population in CA-CRC, the suggestion that CA-CRC might acquire WNT hyperactivation in predominantly a LD manner and the association between high CNA and progression in colitis-associated LGD. Furthermore, findings in this thesis have led to the proposition of three ways in which prevention, treatment and timing and monitoring of treatment could be improved.

Appendix A

A.1 Chapter 3 specific materials and methods: Mouse cohort

Table A.1: Details of Mouse models and derived organoids used in the study

Short name	*Model*	*Reference*	*Cohort*
Alk4T	*Tgfbr2*$^{fl/fl}$	Levéen *et al.* 2002 [226]	ACRCelerate
Apc^{Min}	*C57BL/6J-Apc*^{Min}*/J*	Moser *et al.* 1990 [118]	Oxford
Apc^{Min};Lgr5^{DTR}	*Lgr5tm2*$^{(DTR/EGFP)}$*Fjs;C57BL/6J-Apc*^{Min}*/J*	Moser *et al.* 1990 [118];Tian *et al.* 2011 [122];	Oxford
A	*Vil1-Cre*^{ERT2}*; Apc*$^{fl/+}$	Shibata *et al.* 1997 [117]	ACRCelerate
AK	*Vil1-Cre*^{ERT2}*;Apc*$^{fl/+}$*;Kras*^{G12D}	Jackstadt *et al.* 2019 [117];	ACRCelerate
AKPT	*Vil1-Cre*^{ERT2}*;Apc*$^{fl/+}$*;Kras*^{G12D}*;Trp53*$^{fl/fl}$*;Tgfbr1*$^{fl/fl}$	Jackstadt *et al.* 2019 [121]	ACRCelerate
BT	*Vil1-Cre*^{ERT2}*;Braf*^{V600E}*;Tgfbr1*$^{fl/fl}$	Mercer *et al.* 2005 [227]	ACRCelerate
BPN	*Vil1-Cre*^{ERT2}*;Braf*^{V600E}*;p53*$^{fl/fl}$*;Rosa*$^{26N1icd/+}$	Mercer *et al.* 2005 [227], Marino *et al.* 2000 [125], Murtaugh *et al.* 2003 [228]	ACRCelerate
KP	*Vil1-Cre*^{ERT2}*; Kras*$^{G12D-}$*;p53*$^{fl/fl}$	Jackstadt *et al.* 2019 [121]	ACRCelerate
KPN	*Vil1-Cre*^{ERT2}*;Kras*$^{G12D-}$*;p53*$^{fl/fl}$*;Rosa*$^{26N1icd/+}$	Jackstadt *et al.* 2019 [121]	ACRCelerate and Oxford

Appendix B

B.1 The stem cell population dynamics in sporadic colorectal cancer and colitis-associated colorectal cancer

Table B.1: RSC gene list

Mmusculus ensembl gene	Mmusculus gene name	Hsapiens homolog ensembl gene	Hsapiens homolog gene name
ENSMUSG00000026831	1700007K13Rik	ENSG00000160345	C9orf116
ENSMUSG00000073598	1700066B19Rik	ENSG00000283288	SMIM33
ENSMUSG00000054951	9130008F23Rik	ENSG00000197261	C6orf141
ENSMUSG00000028970	Abcb1b	ENSG00000085563	ABCB1
ENSMUSG00000018381	Abi3	ENSG00000108798	ABI3
ENSMUSG00000030935	Acsm3	ENSG00000005187	ACSM3
ENSMUSG00000061758	Akr1b10	ENSG00000198074	AKR1B10
ENSMUSG00000001642	Akr1b3	ENSG00000085662	AKR1B1
ENSMUSG00000029762	Akr1b8	ENSG00000198074	AKR1B10
ENSMUSG00000022636	Alcam	ENSG00000170017	ALCAM
ENSMUSG00000024659	Anxa1	ENSG00000135046	ANXA1
ENSMUSG00000029484	Anxa3	ENSG00000138772	ANXA3
ENSMUSG00000027712	Anxa5	ENSG00000164111	ANXA5
ENSMUSG00000054702	Ap1s3	ENSG00000152056	AP1S3
ENSMUSG00000033389	Arhgap44	ENSG00000006740	ARHGAP44
ENSMUSG00000037509	Arhgef4	ENSG00000136002	ARHGEF4
ENSMUSG00000098207	Arl14	ENSG00000179674	ARL14
ENSMUSG00000049866	Arl4c	ENSG00000188042	ARL4C
ENSMUSG00000018821	Avpi1	ENSG00000119986	AVPI1
ENSMUSG00000018126	Baiap2l2	ENSG00000128298	BAIAP2L2
ENSMUSG00000045763	Basp1	ENSG00000176788	BASP1
ENSMUSG00000026278	Bok	ENSG00000176720	BOK
ENSMUSG00000020785	Camkk1	ENSG00000004660	CAMKK1
ENSMUSG00000041570	Camsap2	ENSG00000118200	CAMSAP2

ENSMUSG00000021373	Cap2	ENSG00000112186	CAP2
ENSMUSG00000035547	Capn5	ENSG00000149260	CAPN5
ENSMUSG00000032373	Car12	ENSG00000074410	CA12
ENSMUSG00000027562	Car2	ENSG00000104267	CA2
ENSMUSG00000037060	Cavin3	ENSG00000170955	CAVIN3
ENSMUSG00000037979	Ccdc92	ENSG00000119242	CCDC92
ENSMUSG00000035385	Ccl2	ENSG00000108691	CCL2
ENSMUSG00000019997	Ccn2	ENSG00000118523	CCN2
ENSMUSG00000051439	Cd14	ENSG00000170458	CD14
ENSMUSG00000005087	Cd44	ENSG00000026508	CD44
ENSMUSG00000018774	Cd68	ENSG00000129226	CD68
ENSMUSG00000002107	Celf2	ENSG00000048740	CELF2
ENSMUSG00000032232	Cgnl1	ENSG00000128849	CGNL1
ENSMUSG00000037725	Ckap2	ENSG00000136108	CKAP2
ENSMUSG00000001270	Ckb	ENSG00000166165	CKB
ENSMUSG00000074195	Clca4b	ENSG00000016602	CLCA4
ENSMUSG00000022512	Cldn1	ENSG00000163347	CLDN1
ENSMUSG00000047501	Cldn4	ENSG00000189143	CLDN4
ENSMUSG00000015093	Clic3	ENSG00000169583	CLIC3
ENSMUSG00000022037	Clu	ENSG00000120885	CLU
ENSMUSG00000004665	Cnn2	ENSG00000064666	CNN2
ENSMUSG00000019866	Crybg1	ENSG00000112297	CRYBG1
ENSMUSG00000032359	Ctsh	ENSG00000103811	CTSH
ENSMUSG00000029154	Cwh43	ENSG00000109182	CWH43
ENSMUSG00000029380	Cxcl1	ENSG00000163739	CXCL1
ENSMUSG00000034855	Cxcl10	ENSG00000169245	CXCL10
ENSMUSG00000018920	Cxcl16	ENSG00000161921	CXCL16
ENSMUSG00000058427	Cxcl2	ENSG00000081041	CXCL2
ENSMUSG00000029371	Cxcl5	ENSG00000163735	CXCL5
ENSMUSG00000029417	Cxcl9	ENSG00000138755	CXCL9
ENSMUSG00000021559	Dapk1	ENSG00000196730	DAPK1
ENSMUSG00000028194	Ddah1	ENSG00000153904	DDAH1
ENSMUSG00000016526	Dyrk3	ENSG00000143479	DYRK3
ENSMUSG00000030208	Emp1	ENSG00000134531	EMP1
ENSMUSG00000022505	Emp2	ENSG00000213853	EMP2
ENSMUSG00000022995	Enah	ENSG00000154380	ENAH
ENSMUSG00000037419	Endod1	ENSG00000149218	ENDOD1
ENSMUSG00000006445	Epha2	ENSG00000142627	EPHA2
ENSMUSG00000026235	Epha4	ENSG00000116106	EPHA4
ENSMUSG00000010080	Epn3	ENSG00000049283	EPN3
ENSMUSG00000043439	Epop	ENSG00000273604	EPOP
ENSMUSG00000006154	Eps8l1	ENSG00000131037	EPS8L1
ENSMUSG00000026048	Ercc5	ENSG00000134899	ERCC5
ENSMUSG00000029377	Ereg	ENSG00000124882	EREG
ENSMUSG00000013089	Etv5	ENSG00000244405	ETV5

ENSMUSG00000017897	Eya2	ENSG00000064655	EYA2
ENSMUSG00000027533	Fabp5	ENSG00000164687	FABP5
ENSMUSG00000054200	Ffar4	ENSG00000186188	FFAR4
ENSMUSG00000008136	Fhl2	ENSG00000115641	FHL2
ENSMUSG00000029781	Fkbp9	ENSG00000122642	FKBP9
ENSMUSG00000031328	Flna	ENSG00000196924	FLNA
ENSMUSG00000036053	Fmnl2	ENSG00000157827	FMNL2
ENSMUSG00000024912	Fosl1	ENSG00000175592	FOSL1
ENSMUSG00000055978	Fut2	ENSG00000176920	FUT2
ENSMUSG00000057092	Fxyd3	ENSG00000089356	FXYD3
ENSMUSG00000015312	Gadd45b	ENSG00000099860	GADD45B
ENSMUSG00000027955	Gask1b	ENSG00000164125	GASK1B
ENSMUSG00000027199	Gatm	ENSG00000171766	GATM
ENSMUSG00000038843	Gcnt1	ENSG00000187210	GCNT1
ENSMUSG00000032226	Gcnt3	ENSG00000140297	GCNT3
ENSMUSG00000042367	Gjb3	ENSG00000188910	GJB3
ENSMUSG00000056888	Glipr1	ENSG00000139278	GLIPR1
ENSMUSG00000029816	Gpnmb	ENSG00000136235	GPNMB
ENSMUSG00000021306	Gpr137b	ENSG00000077585	GPR137B
ENSMUSG00000042808	Gpx2	ENSG00000176153	GPX2
ENSMUSG00000022385	Gtse1	ENSG00000075218	GTSE1
ENSMUSG00000045502	Hcar2	ENSG00000182782	HCAR2
ENSMUSG00000005413	Hmox1	ENSG00000100292	HMOX1
ENSMUSG00000090877	Hspa1b	ENSG00000204388	HSPA1B
ENSMUSG00000059970	Hspa2	ENSG00000126803	HSPA2
ENSMUSG00000004951	Hspb1	ENSG00000106211	HSPB1
ENSMUSG00000029657	Hsph1	ENSG00000120694	HSPH1
ENSMUSG00000010051	Hyal1	ENSG00000114378	HYAL1
ENSMUSG00000062995	Ica1	ENSG00000003147	ICA1
ENSMUSG00000037405	Icam1	ENSG00000090339	ICAM1
ENSMUSG00000020644	Id2	ENSG00000115738	ID2
ENSMUSG00000025492	Ifitm3	ENSG00000142089	IFITM3
ENSMUSG00000043088	Il17re	ENSG00000163701	IL17RE
ENSMUSG00000026981	Il1rn	ENSG00000136689	IL1RN
ENSMUSG00000024810	Il33	ENSG00000137033	IL33
ENSMUSG00000022900	Ildr1	ENSG00000145103	ILDR1
ENSMUSG00000048458	Inka2	ENSG00000197852	INKA2
ENSMUSG00000018899	Irf1	ENSG00000125347	IRF1
ENSMUSG00000015533	Itga2	ENSG00000164171	ITGA2
ENSMUSG00000027087	Itgav	ENSG00000138448	ITGAV
ENSMUSG00000020758	Itgb4	ENSG00000132470	ITGB4
ENSMUSG00000095115	Itpripl2	ENSG00000205730	ITPRIPL2
ENSMUSG00000054342	Kcnn4	ENSG00000104783	KCNN4
ENSMUSG00000031788	Kifc3	ENSG00000140859	KIFC3
ENSMUSG00000025959	Klf7	ENSG00000118263	KLF7

ENSMUSG00000054146	Krt15	ENSG00000171346	KRT15
ENSMUSG00000023043	Krt18	ENSG00000111057	KRT18
ENSMUSG00000006777	Krt23	ENSG00000108244	KRT23
ENSMUSG00000024421	Lama3	ENSG00000053747	LAMA3
ENSMUSG00000028581	Laptm5	ENSG00000162511	LAPTM5
ENSMUSG00000026822	Lcn2	ENSG00000148346	LCN2
ENSMUSG00000032193	Ldlr	ENSG00000130164	LDLR
ENSMUSG00000068220	Lgals1	ENSG00000100097	LGALS1
ENSMUSG00000027134	Lpcat4	ENSG00000176454	LPCAT4
ENSMUSG00000037095	Lrg1	ENSG00000171236	LRG1
ENSMUSG00000020184	Mdm2	ENSG00000135679	MDM2
ENSMUSG00000059149	Mfsd4a	ENSG00000174514	MFSD4A
ENSMUSG00000074217	Misp3	ENSG00000141854	MISP3
ENSMUSG00000032470	Mras	ENSG00000158186	MRAS
ENSMUSG00000039693	Msantd3	ENSG00000066697	MSANTD3
ENSMUSG00000063011	Msln	ENSG00000102854	MSLN
ENSMUSG00000031207	Msn	ENSG00000147065	MSN
ENSMUSG00000032584	Mst1r	ENSG00000164078	MST1R
ENSMUSG00000052105	Mtcl1	ENSG00000168502	MTCL1
ENSMUSG00000094840	Muc3a	ENSG00000169894	MUC3A
ENSMUSG00000079620	Muc4	ENSG00000145113	MUC4
ENSMUSG00000022443	Myh9	ENSG00000100345	MYH9
ENSMUSG00000029832	Nfe2l3	ENSG00000050344	NFE2L3
ENSMUSG00000021025	Nfkbia	ENSG00000100906	NFKBIA
ENSMUSG00000035356	Nfkbiz	ENSG00000144802	NFKBIZ
ENSMUSG00000020826	Nos2	ENSG00000007171	NOS2
ENSMUSG00000019320	Noxo1	ENSG00000196408	NOXO1
ENSMUSG00000053765	Oas1f	ENSG00000089127	OAS1
ENSMUSG00000037606	Osbpl5	ENSG00000021762	OSBPL5
ENSMUSG00000032860	P2ry2	ENSG00000175591	P2RY2
ENSMUSG00000048779	P2ry6	ENSG00000171631	P2RY6
ENSMUSG00000025931	Paqr8	ENSG00000170915	PAQR8
ENSMUSG00000028716	Pdzk1ip1	ENSG00000162366	PDZK1IP1
ENSMUSG00000020205	Phlda1	ENSG00000139289	PHLDA1
ENSMUSG00000010760	Phlda2	ENSG00000181649	PHLDA2
ENSMUSG00000020808	Pimreg	ENSG00000129195	PIMREG
ENSMUSG00000021822	Plau	ENSG00000122861	PLAU
ENSMUSG00000046223	Plaur	ENSG00000011422	PLAUR
ENSMUSG00000020937	Plcd3	ENSG00000161714	PLCD3
ENSMUSG00000035818	Plekhs1	ENSG00000148735	PLEKHS1
ENSMUSG00000021701	Plk2	ENSG00000145632	PLK2
ENSMUSG00000038400	Pmepa1	ENSG00000124225	PMEPA1
ENSMUSG00000024538	Ppic	ENSG00000168938	PPIC
ENSMUSG00000039457	Ppl	ENSG00000118898	PPL
ENSMUSG00000068744	Psrc1	ENSG00000134222	PSRC1

ENSMUSG00000032487	Ptgs2	ENSG00000073756	PTGS2
ENSMUSG00000051343	Rab11fip5	ENSG00000135631	RAB11FIP5
ENSMUSG00000021062	Rab15	ENSG00000139998	RAB15
ENSMUSG00000020372	Rack1	ENSG00000204628	RACK1
ENSMUSG00000037110	Ralgapa2	ENSG00000188559	RALGAPA2
ENSMUSG00000026970	Rbms1	ENSG00000153250	RBMS1
ENSMUSG00000002458	Rgs19	ENSG00000171700	RGS19
ENSMUSG00000041845	Rhod	ENSG00000173156	RHOD
ENSMUSG00000041135	Ripk2	ENSG00000104312	RIPK2
ENSMUSG00000092837	Rpph1	ENSG00000277209	RPPH1
ENSMUSG00000022952	Runx1	ENSG00000159216	RUNX1
ENSMUSG00000042306	S100a14	ENSG00000189334	S100A14
ENSMUSG00000001020	S100a4	ENSG00000196154	S100A4
ENSMUSG00000001025	S100a6	ENSG00000197956	S100A6
ENSMUSG00000060487	Samd5	ENSG00000203727	SAMD5
ENSMUSG00000088689	Scarna17	ENSG00000251992	SCARNA17
ENSMUSG00000022123	Scel	ENSG00000136155	SCEL
ENSMUSG00000038264	Sema7a	ENSG00000138623	SEMA7A
ENSMUSG00000021403	Serpinb9b	ENSG00000170542	SERPINB9
ENSMUSG00000032261	Sh3bgrl2	ENSG00000198478	SH3BGRL2
ENSMUSG00000040918	Slc19a2	ENSG00000117479	SLC19A2
ENSMUSG00000025986	Slc39a10	ENSG00000196950	SLC39A10
ENSMUSG00000057193	Slc44a2	ENSG00000129353	SLC44A2
ENSMUSG00000040010	Slc7a5	ENSG00000103257	SLC7A5
ENSMUSG00000038963	Slco4a1	ENSG00000101187	SLCO4A1
ENSMUSG00000017002	Slpi	ENSG00000124107	SLPI
ENSMUSG00000064637	Snora20	ENSG00000207392	SNORA20
ENSMUSG00000064387	Snora73a	ENSG00000274266	SNORA73A
ENSMUSG00000065353	Snora73b	ENSG00000200087	SNORA73B
ENSMUSG00000065126	Snord104	ENSG00000199753	SNORD104
ENSMUSG00000064899	Snord118	ENSG00000200463	SNORD118
ENSMUSG00000105167	Snord43	ENSG00000263764	SNORD43
ENSMUSG00000093289	Snord92	ENSG00000264994	SNORD92
ENSMUSG00000053113	Socs3	ENSG00000184557	SOCS3
ENSMUSG00000049313	Sorl1	ENSG00000137642	SORL1
ENSMUSG00000034401	Spata6	ENSG00000132122	SPATA6
ENSMUSG00000029304	Spp1	ENSG00000118785	SPP1
ENSMUSG00000039232	Stx11	ENSG00000135604	STX11
ENSMUSG00000006800	Sulf2	ENSG00000196562	SULF2
ENSMUSG00000031098	Syt8	ENSG00000149043	SYT8
ENSMUSG00000056313	Tcim	ENSG00000176907	TCIM
ENSMUSG00000001750	Tcirg1	ENSG00000110719	TCIRG1
ENSMUSG00000035493	Tgfbi	ENSG00000120708	TGFBI
ENSMUSG00000037820	Tgm2	ENSG00000198959	TGM2
ENSMUSG00000017466	Timp2	ENSG00000035862	TIMP2

ENSMUSG00000028776	Tinagl1	ENSG00000142910	TINAGL1
ENSMUSG00000039005	Tlr4	ENSG00000136869	TLR4
ENSMUSG00000031617	Tmem184c	ENSG00000164168	TMEM184C
ENSMUSG00000030095	Tmem43	ENSG00000170876	TMEM43
ENSMUSG00000024401	Tnf	ENSG00000232810	TNF
ENSMUSG00000019850	Tnfaip3	ENSG00000118503	TNFAIP3
ENSMUSG00000062210	Tnfaip8	ENSG00000145779	TNFAIP8
ENSMUSG00000028599	Tnfrsf1b	ENSG00000028137	TNFRSF1B
ENSMUSG00000035678	Tnfsf9	ENSG00000125657	TNFSF9
ENSMUSG00000044162	Tnip3	ENSG00000050730	TNIP3
ENSMUSG00000031097	Tnni2	ENSG00000130598	TNNI2
ENSMUSG00000026414	Tnnt2	ENSG00000118194	TNNT2
ENSMUSG00000017607	Tns4	ENSG00000131746	TNS4
ENSMUSG00000041779	Tram2	ENSG00000065308	TRAM2
ENSMUSG00000050747	Trim15	ENSG00000204610	TRIM15
ENSMUSG00000073399	Trim40	ENSG00000204614	TRIM40
ENSMUSG00000025511	Tspan4	ENSG00000214063	TSPAN4
ENSMUSG00000067377	Tspan6	ENSG00000000003	TSPAN6
ENSMUSG00000072235	Tuba1a	ENSG00000167552	TUBA1A
ENSMUSG00000073002	Vamp5	ENSG00000168899	VAMP5
ENSMUSG00000026728	Vim	ENSG00000026025	VIM
ENSMUSG00000037440	Vnn1	ENSG00000112299	VNN1
ENSMUSG00000017723	Wfdc2	ENSG00000101443	WFDC2
ENSMUSG00000027803	Wwtr1	ENSG00000018408	WWTR1
ENSMUSG00000011267	Zfp296	ENSG00000170684	ZNF296
ENSMUSG00000085795	Zfp703	ENSG00000183779	ZNF703
ENSMUSG00000027663	Zmat3	ENSG00000172667	ZMAT3
ENSMUSG00000073421	H2-Ab1	ENSG00000179344	HLA-DQB1
ENSMUSG00000055413	H2-Q5	ENSG00000204632	HLA-G
ENSMUSG00000105743	Mir1949		
ENSMUSG00000085972	1110028F11Rik		
ENSMUSG00000097651	4930461G14Rik		
ENSMUSG00000081793	5730416F02Rik		
ENSMUSG00000092517	Art2a		
ENSMUSG00000075010	AW112010		
ENSMUSG00000095562	Erdr1		
ENSMUSG00000035778	Ggta1		
ENSMUSG00000069804	Gm10277		
ENSMUSG00000087342	Gm12238		
ENSMUSG00000073555	Gm4951		
ENSMUSG00000094790	Gm6625		
ENSMUSG00000036594	H2-Aa		
ENSMUSG00000026535	Ifi202b		
ENSMUSG00000054072	Iigp1		
ENSMUSG00000043165	Lor		

ENSMUSG00000075602	Ly6a		
ENSMUSG00000079018	Ly6c1		
ENSMUSG00000022584	Ly6c2		
ENSMUSG00000022583	Ly6f		
ENSMUSG00000022226	Mcpt2		
ENSMUSG00000093336	Mir1954		
ENSMUSG00000076948	Mir467e		
ENSMUSG00000071203	Naip5		
ENSMUSG00000079092	Prl2c2		
ENSMUSG00000056457	Prl2c3		
ENSMUSG00000037071	Scd1		
ENSMUSG00000064968	Snord47		
ENSMUSG00000093880	Tmem181c-ps		
ENSMUSG00000079451	Tmprss11g		
ENSMUSG00000000983	Wfdc18		
ENSMUSG00000051341	Zfp52		

Table B.2: Statistical analyses results of stem cell index across human premalignant and malignant subtypes

Stem cell index across human premalignant and malignant subtypes					
Stem cell index in premalignant (Figure 3.4 A)					
group1	group2	p	p.adj	p.signif	method
IBD-LGD	SSL	0.175693	0.7	ns	Wilcoxon
IBD-LGD	TSA	0.698202	0.75	ns	Wilcoxon
IBD-LGD	TVA	0.007794	0.039	**	Wilcoxon
SSL	TSA	0.31484	0.75	ns	Wilcoxon
SSL	TVA	0.000532	0.0032	***	Wilcoxon
TSA	TVA	0.250371	0.75	ns	Wilcoxon
Stem cell index in CMS subtypes (Figure 3.5 A)					
group1	group2	p	p.adj	p.signif	method
CA-CRC	CMS4	4.53E-01	1	ns	Wilcoxon
CA-CRC	CMS3	5.92E-01	1	ns	Wilcoxon
CA-CRC	CMS2	3.81E-02	0.27	*	Wilcoxon
CA-CRC	CMS1	2.45E-01	1	ns	Wilcoxon
CMS4	CMS3	6.01E-01	1	ns	Wilcoxon
CMS4	CMS2	1.44E-05	0.00013	****	Wilcoxon
CMS4	CMS1	8.39E-01	1	ns	Wilcoxon
CMS3	CMS2	5.62E-03	0.045	**	Wilcoxon
CMS3	CMS1	5.21E-01	1	ns	Wilcoxon
CMS2	CMS1	1.97E-06	0.00002	****	Wilcoxon
Stem cell index in CRIS subtypes (Figure 3.5 B)					
group1	group2	p	p.adj	p.signif	method
CA-CRC	CRIS-E	1.15E-01	6.90E-01	ns	Wilcoxon
CA-CRC	CRIS-D	3.35E-01	1.00E+00	ns	Wilcoxon
CA-CRC	CRIS-C	2.66E-02	2.40E-01	*	Wilcoxon
CA-CRC	CRIS-B	5.90E-02	4.70E-01	ns	Wilcoxon
CA-CRC	CRIS-A	2.49E-01	9.90E-01	ns	Wilcoxon
CRIS-E	CRIS-D	5.64E-01	1.00E+00	ns	Wilcoxon
CRIS-E	CRIS-C	6.19E-01	1.00E+00	ns	Wilcoxon
CRIS-E	CRIS-B	6.57E-06	8.50E-05	****	Wilcoxon
CRIS-E	CRIS-A	2.57E-04	2.80E-03	***	Wilcoxon
CRIS-D	CRIS-C	1.33E-01	6.90E-01	ns	Wilcoxon
CRIS-D	CRIS-B	7.22E-05	8.70E-04	****	Wilcoxon
CRIS-D	CRIS-A	1.61E-03	1.60E-02	**	Wilcoxon
CRIS-C	CRIS-B	2.92E-07	4.40E-06	****	Wilcoxon
CRIS-C	CRIS-A	4.61E-07	6.50E-06	****	Wilcoxon
CRIS-B	CRIS-A	8.21E-02	5.70E-01	ns	Wilcoxon

Table B.3: Statistical analyses results of stem cell index across murine intestinal neoplasia models

Stem cell index across mouse autochthonous tumours (Figure 3.6 A)					
group1	group2	p	p.adj	p.signif	method
APCmin	A	0.523699	1	ns	Wilcoxon
APCmin	AK	0.012121	0.28	*	Wilcoxon
APCmin	AKPT	0.012121	0.28	*	Wilcoxon
APCmin	BT	0.00404	0.097	**	Wilcoxon
APCmin	BPN	0.00373	0.093	**	Wilcoxon
APCmin	KP	0.012121	0.28	*	Wilcoxon
APCmin	KPN	1.59E-05	0.00044	****	Wilcoxon
A	AK	0.035714	0.68	*	Wilcoxon
A	AKPT	0.035714	0.68	*	Wilcoxon
A	BT	0.015873	0.32	*	Wilcoxon
A	BPN	0.002525	0.066	**	Wilcoxon
A	KP	0.035714	0.68	*	Wilcoxon
A	KPN	0.000323	0.0087	***	Wilcoxon
AK	AKPT	0.7	1	ns	Wilcoxon
AK	BT	0.628571	1	ns	Wilcoxon
AK	BPN	0.266667	1	ns	Wilcoxon
AK	KP	0.1	1	ns	Wilcoxon
AK	KPN	0.448352	1	ns	Wilcoxon
AKPT	BT	0.4	1	ns	Wilcoxon
AKPT	BPN	1	1	ns	Wilcoxon
AKPT	KP	0.1	1	ns	Wilcoxon
AKPT	KPN	0.18022	1	ns	Wilcoxon
BT	BPN	0.109091	1	ns	Wilcoxon
BT	KP	0.057143	0.86	ns	Wilcoxon
BT	KPN	0.683516	1	ns	Wilcoxon
BPN	KP	0.066667	0.93	ns	Wilcoxon
BPN	KPN	0.035842	0.68	*	Wilcoxon
KP	KPN	0.136264	1	ns	Wilcoxon
Stem cell index across mouse derived organoids (Figure 3.6 B)					
group1	group2	p	p.adj	p.signif	method
A	AK	5.66E-03	0.062	**	Wilcoxon
A	AKPT	9.49E-01	1	ns	Wilcoxon
A	BPN	1.47E-03	0.021	**	Wilcoxon
A	KP	3.85E-02	0.31	*	Wilcoxon
A	KPN	2.65E-05	0.0004	****	Wilcoxon
AK	AKPT	1.96E-02	0.2	*	Wilcoxon
AK	BPN	8.25E-01	1	ns	Wilcoxon
AK	KP	7.27E-01	1	ns	Wilcoxon
AK	KPN	2.36E-01	1	ns	Wilcoxon
AKPT	BPN	2.86E-02	0.26	*	Wilcoxon
AKPT	KP	1.14E-01	0.8	ns	Wilcoxon
AKPT	KPN	4.04E-03	0.053	**	Wilcoxon
BPN	KP	6.29E-01	1	ns	Wilcoxon
BPN	KPN	4.04E-03	0.053	**	Wilcoxon
KP	KPN	7.76E-01	1	ns	Wilcoxon

Table B.4: Statistical analyses results of Apc^{Min};$Lgr5^{DTR}$ mice assessment

DTR stem cell index (Figure 3.12 B)					
group1	group2	p	p.adj	p.signif	method
Apc^{Min};$Lgr5^{DTR}$; Day5 post DT	Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	0.001851	0.011	**	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day5 post DT	Apc^{Min};$Lgr5^{DTR}$; No DT	0.034987	0.1	*	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day5 post DT	Apc^{Min}	0.931427	0.93	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	Apc^{Min};$Lgr5^{DTR}$; No DT	0.1333	0.27	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	Apc^{Min}	0.002756	0.014	**	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; No DT	Apc^{Min}	0.005672	0.023	**	Wilcoxon
DTR lesion size (Figure 3.13 B)					
group1	group2	p	p.adj	p.signif	method
Apc^{Min}	Apc^{Min};$Lgr5^{DTR}$; No DT	0.638562	1	ns	Wilcoxon
Apc^{Min}	Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	0.957876	1	ns	Wilcoxon
Apc^{Min}	Apc^{Min};$Lgr5^{DTR}$; Day3 post DT	0.473499	1	ns	Wilcoxon
Apc^{Min}	Apc^{Min};$Lgr5^{DTR}$; Day5 post DT	0.29026	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; No DT	Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	1	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; No DT	Apc^{Min};$Lgr5^{DTR}$; Day3 post DT	0.402556	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; No DT	Apc^{Min};$Lgr5^{DTR}$; Day5 post DT	0.53733	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	Apc^{Min};$Lgr5^{DTR}$; Day3 DT	0.842105	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	Apc^{Min};$Lgr5^{DTR}$; Day5 DT	0.639899	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day3 post DT	Apc^{Min};$Lgr5^{DTR}$; Day5 DT	0.177495	1	ns	Wilcoxon
DTR staining (Figure 3.13 C)					
group1	group2	p	p.adj	p.signif	method
Apc^{Min}	Apc^{Min};$Lgr5^{DTR}$; No DT	0.571358	1	ns	Wilcoxon
Apc^{Min}	Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	0.382284	1	ns	Wilcoxon
Apc^{Min}	Apc^{Min};$Lgr5^{DTR}$; Day3 post DT	0.949717	1	ns	Wilcoxon
Apc^{Min}	Apc^{Min};$Lgr5^{DTR}$; Day5 post DT	0.382284	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; No DT	Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	0.571358	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; No DT	Apc^{Min};$Lgr5^{DTR}$; Day3 post DT	0.616462	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; No DT	Apc^{Min};$Lgr5^{DTR}$; Day5 post DT	0.208288	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	Apc^{Min};$Lgr5^{DTR}$; Day3 post DT	0.490843	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day1 post DT	Apc^{Min};$Lgr5^{DTR}$; Day5 post DT	0.234499	1	ns	Wilcoxon
Apc^{Min};$Lgr5^{DTR}$; Day3 post DT	Apc^{Min};$Lgr5^{DTR}$; Day5 post DT	0.413586	1	ns	Wilcoxon

Table B.5: Statistical analyses results of FOxTROT analysis

Proliferation score FoxTROT (Figure 3.14 C)					
group1	group2	p	p.adj	p.signif	method
CBC to RSC	Static CBC	0.260125	1	ns	Wilcoxon
CBC to RSC	Static RSC	0.487592	1	ns	Wilcoxon
CBC to RSC	RSC to CBC	0.286893	1	ns	Wilcoxon
Static CBC	Static RSC	0.659828	1	ns	Wilcoxon
Static CBC	RSC to CBC	0.894141	1	ns	Wilcoxon
fREG	RSC to CBC	0.77548	1	ns	Wilcoxon
Stem cell index FoxTROT (Figure 3.14 C)					
group1	group2	p	p.adj	p.signif	method
CBC to RSC	Static CBC	7.10E-07	3.50E-06	****	Wilcoxon
CBC to RSC	Static RSC	8.84E-05	3.50E-04	****	Wilcoxon
CBC to RSC	RSC to CBC	2.44E-09	1.50E-08	****	Wilcoxon
Static CBC	Static RSC	6.60E-01	6.60E-01	ns	Wilcoxon
Static CBC	RSC to CBC	1.47E-04	3.50E-04	***	Wilcoxon
Static RSC	RSC to CBC	1.12E-04	3.50E-04	***	Wilcoxon

Appendix C

C.1 The WNT signalling pathway in colitis-associated colorectal cancer: Phenotyping of colitis-associated precursor lesion

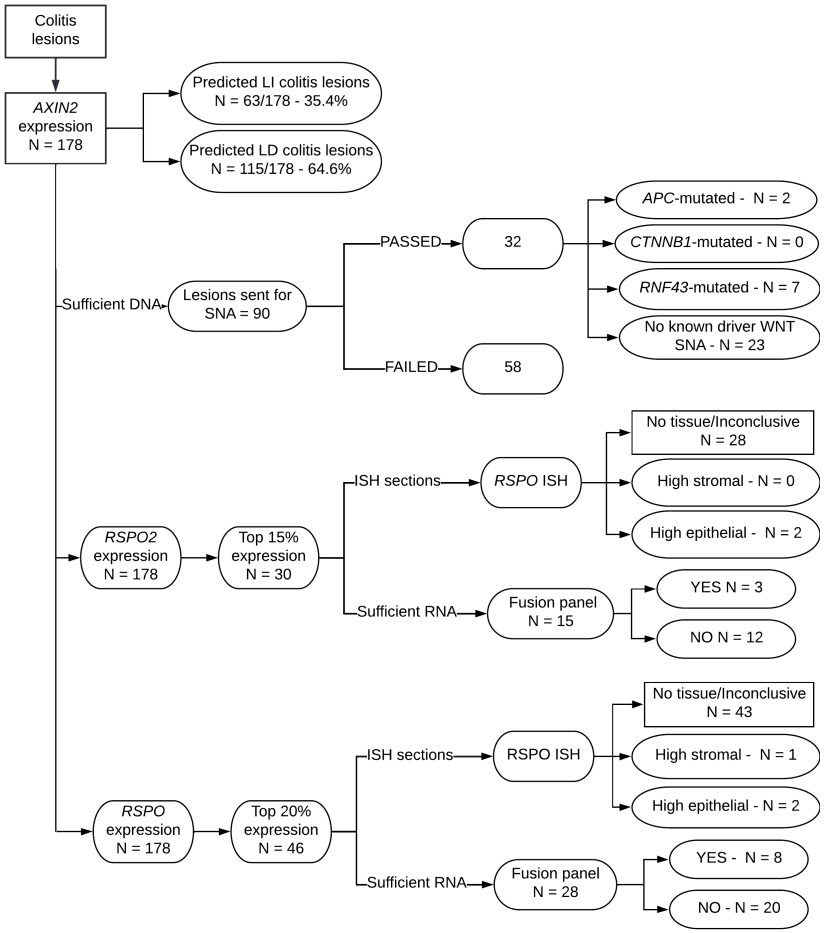


Figure C.1: Breakdown of the multifaceted phenotyping of the colitis-associated lesions.

Table C.1: Statistical analyses results of investigations carried in murine models of WNT disruption

Notum expression in murine model of WNT disruption (Figure 4.13)					
group1	group2	p	p.adj	p.signif	method
ApcVil	Rspo3Vil	0.0636	0.32	ns	Wilcoxon
ApcVil	Rspo3Atoh	0.0636	0.32	ns	Wilcoxon
ApcVil	Rspo3Gli	0.0765	0.32	ns	Wilcoxon
Paneth cell proportions across in mouse models of WNT (Figure 4.15 B)					
ApcVil	Rspo3Vil	0.000288	0.0017	**	Wilcoxon
ApcVil	Rspo3Atoh	0.007051	0.028	*	Wilcoxon
ApcVil	Rspo3Gli	0.000536	0.0027	**	Wilcoxon
Paneth cell proportions in case-control study (Figure 4.16 B)					
tdTomAtoh 30hr	Rspo3Atoh 30hr	4.12E-06	3.70E-05	****	Wilcoxon
tdTomAtoh Day4	Rspo3Atoh Day4	2.83E-09	3.40E-08	****	Wilcoxon
tdTomAtoh Day30	Rspo3Atoh Day30	1.41E-09	1.80E-08	****	Wilcoxon
Goblet cells proportions in case-control study (Figure 4.17 B)					
tdTomAtoh 30hr	Rspo3Atoh 30hr	8.01E-09	9.60E-08	****	Wilcoxon
tdTomAtoh Day4	Rspo3Atoh Day4	8.01E-09	9.60E-08	****	Wilcoxon
tdTomAtoh Day30	Rspo3Atoh Day30	8.01E-09	9.60E-08	****	Wilcoxon
Proportion of lineage tracing Paneth cells (Figure 4.18 A)					
tdTomAtoh 30hr	Rspo3Atoh 30hr	3.29E-05	3.00E-04	***	Wilcoxon
tdTomAtoh Day4	Rspo3Atoh Day4	3.93E-07	4.80E-06	****	Wilcoxon
tdTomAtoh Day30	Rspo3Atoh Day30	1.06E-07	1.60E-06	****	Wilcoxon
Proportion of non-lineage tracing Paneth cells (Figure 4.18 B)					
tdTomAtoh 30hr	Rspo3Atoh 30hr	9.57E-01	1.00E+00	ns	Wilcoxon
tdTomAtoh Day4	Rspo3Atoh Day4	8.23E-01	1.00E+00	ns	Wilcoxon
tdTomAtoh Day30	Rspo3Atoh Day30	6.53E-08	9.80E-07	****	Wilcoxon

Bibliography

1. Sensoy, I. A review on the food digestion in the digestive tract and the used in vitro models. *Current Research in Food Science* **4,** 308–319. ISSN: 2665-9271. https://www.sciencedirect.com/science/article/pii/S2665927121000307 (2021).
2. Azzouz, L. & Sharma, S. *Physiology, Large Intestine* (StatPearls Publishing, Treasure Island (FL), USA, 2021).
3. Mey, J. R. D. & Freund, J.-N. Understanding epithelial homeostasis in the intestine. *Tissue Barriers* **1.** PMID: 24665395, e24965. https://doi.org/10.4161/tisb.24965 (2013).
4. Beumer, J. & Clevers, H. Regulation and plasticity of intestinal stem cells during homeostasis and regeneration. *Development* **143,** 3639–3649. https://pubmed.ncbi.nlm.nih.gov/27802133/ (2016).
5. Krausova, M. & Korinek, V. Wnt signaling in adult intestinal stem cells and cancer. *Cellular Signalling* **26,** 570–579. https://www.nature.com/articles/s41575-019-0161-7#citeas (2019).
6. Cleynen, I. & Laukens, D. Cellular diversity in the colon: Another brick in the wall. *Nature Reviews Gastroenterology and Hepatology* **16,** 391–392. https://doi.org/10.1038/s41575-019-0161-7 (2019).
7. Sasaki, N. *et al.* Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. *Proceedings of the National Academy of Sciences* **113,** E5399–E5407. https://pubmed.ncbi.nlm.nih.gov/27573849/ (2016).
8. Barker, N. Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nature reviews Molecular cell biology* **15,** 19–33. https://www.nature.com/articles/nrm3721#citeas (2014).
9. Biswas, S. *et al.* Microenvironmental control of stem cell fate in intestinal homeostasis and disease. *The Journal of Pathology* **237,** 135–145. https://onlinelibrary.wiley.com/doi/abs/10.1002/path.4563 (2015).
10. Pennings, S., Liu, K. J. & Qian, H. The stem cell niche: interactions between stem cells and their environment. *Stem cells international* **2018.** https://www.hindawi.com/journals/sci/2018/4879379/ (2018).
11. Santos, A. J., Lo, Y.-H., Mah, A. T. & Kuo, C. J. The intestinal stem cell niche: homeostasis and adaptations. *Trends in cell biology* **28,** 1062–1078. https://www.cell.com/trends/cell-biology/fulltext/S0962-8924(18)30139-9?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0962892418301399%3Fshowall%3Dtrue (2018).

12. de Sousa e Melo, F. & de Sauvage, F. J. Cellular Plasticity in Intestinal Homeostasis and Disease. *Cell Stem Cell* **24,** 54–64. ISSN: 1934-5909. https://www.sciencedirect.com/science/article/pii/S1934590918305551 (2019).

13. Greicius, G. *et al.* PDGFRα+ pericryptal stromal cells are the critical source of Wnts and RSPO3 for murine intestinal stem cells in vivo. *Proceedings of the National Academy of Sciences* **115,** E3173–E3181. https://www.pnas.org/doi/abs/10.1073/pnas.1713510115 (2018).

14. Eggington, H. R., Mulholland, E. J. & Leedham, S. J. Morphogen regulation of stem cell plasticity in intestinal regeneration and carcinogenesis. *Developmental Dynamics* **251,** 61–74. https://anatomypubs.onlinelibrary.wiley.com/doi/abs/10.1002/dvdy.434 (2022).

15. Aoki, R. *et al.* Foxl1-Expressing Mesenchymal Cells Constitute the Intestinal Stem Cell Niche. *Cellular and Molecular Gastroenterology and Hepatology* **2,** 175–188. ISSN: 2352-345X. https://www.sciencedirect.com/science/article/pii/S2352345X15001988 (2016).

16. Koch, S. Extrinsic control of Wnt signaling in the intestine. *Differentiation* **97,** 1–8. ISSN: 0301-4681. https://www.sciencedirect.com/science/article/pii/S0301468117300774 (2017).

17. Krausova, M. & Korinek, V. Wnt signaling in adult intestinal stem cells and cancer. *Cellular Signalling* **26,** 570–579. ISSN: 0898-6568. https://www.sciencedirect.com/science/article/pii/S089865681300363X (2014).

18. Kleeman, S. O. *et al.* Exploiting differential Wnt target gene expression to generate a molecular biomarker for colorectal cancer stratification. *Gut* **69,** 1092–1103. ISSN: 0017-5749. https://gut.bmj.com/content/69/6/1092 (2020).

19. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes & development* **17,** 1709–1713. http://genesdev.cshlp.org/content/17/14/1709.short (2003).

20. Sato, T. *et al.* Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium. *Gastroenterology* **141,** 1762–1772. ISSN: 0016-5085. https://www.sciencedirect.com/science/article/pii/S0016508511011085 (2011).

21. Salazar, V. S., Gamer, L. W. & Rosen, V. BMP signalling in skeletal development, disease and repair. *Nature Reviews Endocrinology* **12,** 203–221. https://www.nature.com/articles/nrendo.2016.12 (2016).

22. Bray, S. J. Notch signalling: a simple pathway becomes complex. *Nature reviews Molecular cell biology* **7,** 678–689. https://www.nature.com/articles/nrm2009 (2006).

23. Koveitypour, Z. *et al.* Signaling pathways involved in colorectal cancer progression. *Cell Biosci* **9.** https://cellandbioscience.biomedcentral.com/articles/10.1186/s13578-019-0361-4#citeas (2019).

24. Kriz, V. & Korinek, V. Wnt, RSPO and Hippo signalling in the intestine and intestinal stem cells. *Genes* **9,** 20. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5793173/# (2018).

25. Fallah, S. & Beaulieu, J.-F. The Hippo pathway effector YAP1 regulates intestinal epithelial cell differentiation. *Cells* **9,** 1895. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7463744/ (2020).

26. Gjorevski, N. *et al.* Designer matrices for intestinal stem cell and organoid culture. *Nature* **539,** 560–564. https://pubmed.ncbi.nlm.nih.gov/27851739/ (2016).

27. Zhao, J. *et al.* R-Spondin1 protects mice from chemotherapy or radiation-induced oral mucositis through the canonical Wnt/β-catenin pathway. *Proceedings of the National Academy of Sciences* **106,** 2331–2336. https://www.pnas.org/doi/abs/10.1073/pnas.0805159106 (2009).

28. Miyoshi, H., Ajima, R., Luo, C. T., Yamaguchi, T. P. & Stappenbeck, T. S. Wnt5a potentiates TGF-β signaling to promote colonic crypt regeneration after tissue injury. *Science* **338,** 108–113. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3706630/ (2012).

29. Harnack, C. *et al.* R-spondin 3 promotes stem cell recovery and epithelial regeneration in the colon. *Nature communications* **10,** 1–15. https://www.nature.com/articles/s41467-019-12349-5 (2019).

30. Yui, S. *et al.* YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell* **22,** 35–49.e7. ISSN: 1934-5909. https://www.sciencedirect.com/science/article/pii/S1934590917304538 (2018).

31. Koppens, M. A. *et al.* Bone Morphogenetic Protein Pathway Antagonism by Grem1 Regulates Epithelial Cell Fate in Intestinal Regeneration. *Gastroenterology* **161,** 239–254.e9. ISSN: 0016-5085. https://www.sciencedirect.com/science/article/pii/S0016508521005795 (2021).

32. Ye, Y., Pang, Z., Chen, W., Ju, S. & Zhou, C. The epidemiology and risk factors of inflammatory bowel disease. *Internation Journal of Clinical and Experimental Medicine* **8,** 22529–22542. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4730025/ (2015).

33. Kobayashi, T. *et al.* Ulcerative colitis. *Nature Reviews Disease Primers* **6.** https://www.nature.com/articles/s41572-020-0205-x#citeas (2020).

34. Alatab, S. *et al.* The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet gastroenterology & hepatology* **5,** 17–30. https://www.thelancet.com/journals/langas/article/PIIS2468-1253(19)30333-4/fulltext (2020).

35. Ananthakrishnan, A. N. Epidemiology and risk factors for IBD. *Nature Reviews Gastroenterology and Hepatology* **12,** 205–217. https://www.nature.com/articles/nrgastro.2015.34#citeas (2015).

36. Pasvol, T. J. *et al.* Incidence and prevalence of inflammatory bowel disease in UK primary care: a population-based cohort study. *BMJ Open* **10.** ISSN: 2044-6055. https://bmjopen.bmj.com/content/10/7/e036584 (2020).

37. Kaplan, G. G. The global burden of IBD: From 2015 to 2025. *Nature Reviews Gastroenterology and Hepatology* **12,** 720–727. https://www.nature.com/articles/nrgastro.2015.150 (2015).

38. Ng, S. C. *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *The Lancet* **390,** 2769–2778. ISSN: 0140-6736. https://www.sciencedirect.com/science/article/pii/S0140673617324480 (2017).

39. Alatab, S. *et al.* The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology and Hepatology* **5,** 17–30. ISSN: 2468-1253. https://www.sciencedirect.com/science/article/pii/S2468125319303334 (2020).

40. Johnston, R. D. & Logan, R. F. What is the peak age for onset of IBD? *Inflammatory Bowel Diseases* **14,** S4–S5. https://onlinelibrary.wiley.com/doi/abs/10.1002/ibd.20545 (2008).

41. Prelipcean, C. C., Mihai, C., Gogalniceanu, P. & Mihai, B. What is the impact of age on adult patients with inflammatory bowel disease. *Clujul Med* **86,** 3–9. https://pubmed.ncbi.nlm.nih.gov/26527906/ (2013).

42. Zhang, Y.-Z. & Li, Y.-Y. Inflammatory bowel disease: Pathogenesis. *World J Gastroenterology* **20,** 91–99. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3886036/#!po=3.57143 (2014).

43. Rutgeerts, P., Vermeire, S. & Van Assche, G. Mucosal healing in inflammatory bowel disease: impossible ideal or therapeutic target? *Gut* **56,** 453–455. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1856849/pdf/453.pdf (2007).

44. Al Bakir, I., Curtius, K. & Graham, T. A. From colitis to cancer: an evolutionary trajectory that merges maths and biology. *Frontiers in immunology* **9,** 2368. https://www.frontiersin.org/articles/10.3389/fimmu.2018.02368/full (2018).

45. Dvorak, H. F. Tumors: wounds that do not heal—redux. *Cancer immunology research* **3,** 1–11. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4288010/ (2015).

46. Dvorak, H. F. Tumors: wounds that do not heal. *New England Journal of Medicine* **315,** 1650–1659. https://www.nejm.org/doi/pdf/10.1056/NEJM198612253152606 (1986).

47. WHO. *Cancer* online report. 2022. https://www.who.int/news-room/fact-sheets/detail/cancer.

48. Rawla, P., Sunkara, T. & Barsouk, A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Gastroenterology Review/Przegląd Gastroenterologiczny* **14,** 89–103. ISSN: 1895-5770. http://dx.doi.org/10.5114/pg.2018.81072 (2019).

49. CRUK. *Bowel cancer statistics* online report. 2022. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer#heading-One..

50. Rawla, P., Sunkara, T. & Barsouk, A. Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Prz Gastroenterology* **14,** 89–103. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6791134/ (2019).

51. Aran, V., Victorino, A. P., Thuler, L. C. & Ferreira, C. G. Colorectal Cancer: Epidemiology, Disease Mechanisms and Interventions to Reduce Onset and Mortality. *Clinical Colorectal Cancer* **15,** 195–203. ISSN: 1533-0028. https://www.sciencedirect.com/science/article/pii/S1533002816300202 (2016).

52. Wan, M.-l. *et al.* Colorectal cancer (CRC) as a multifactorial disease and its causal correlations with multiple signaling pathways. *Bioscience Reports* **40.** BSR20200265. ISSN: 0144-8463. eprint: https://portlandpress.com/bioscirep/article-pdf/40/3/BSR20200265/869866/bsr-2020-0265.pdf. https://doi.org/10.1042/BSR20200265 (Mar. 2020).

53. Tiwari, A., Saraf, S., Verma, A., Panda, P. K. & Jain, S. K. Novel targeting approaches and signaling pathways of colorectal cancer: An insight. *World journal of gastroenterology* **24,** 4428. https://www.wjgnet.com/1007-9327/full/v24/i39/4428.htm (2018).

54. Koveitypour, Z. *et al.* Signaling pathways involved in colorectal cancer progression. *Cell & bioscience* **9,** 1–14. https://cellandbioscience.biomedcentral.com/articles/10.1186/s13578-019-0361-4 (2019).

55. Camargo, F. D. *et al.* YAP1 Increases Organ Size and Expands Undifferentiated Progenitor Cells. *Current Biology* **17,** 2054–2060. ISSN: 0960-9822. https://www.sciencedirect.com/science/article/pii/S0960982207021367 (2007).

56. Zhou, D. *et al.* Mst1 and Mst2 protein kinases restrain intestinal stem cell proliferation and colonic tumorigenesis by inhibition of Yes-associated protein (Yap) overabundance. *Proceedings of the National Academy of Sciences* **108,** E1312–E1320. https://www.pnas.org/doi/abs/10.1073/pnas.1110428108 (2011).

57. Dong, T. *et al.* High cytoplasmic YAP1 expression predicts a poor prognosis in patients with colorectal cancer. *PeerJ* **8,** e10397. https://peerj.com/articles/10397/ (2020).

58. Yuen, H.-F. *et al.* TAZ Expression as a Prognostic Indicator in Colorectal Cancer. *PLOS ONE* **8,** 1–17. https://doi.org/10.1371/journal.pone.0054211 (2013).

59. Hong, A. W., Meng, Z. & Guan, K.-L. The Hippo pathway in intestinal regeneration and disease. *Nature reviews Gastroenterology & hepatology* **13,** 324–337. https://www.nature.com/articles/nrgastro.2016.59 (2016).

60. Imperiale, T. F., Prabhu, L., Loehrer, P. J. & Lu, T. Genetic pathways, prevention, and treatment of sporadic colorectal cancer. *Oncoscience* **1,** 400–406. https://www.oncoscience.us/article/59/text/ (2014).

61. Malki, A. *et al.* Molecular Mechanisms of Colon Cancer Progression and Metastasis: Recent Insights and Advancements. *International Journal of Molecular Sciences* **22.** ISSN: 1422-0067. https://www.mdpi.com/1422-0067/22/1/130 (2021).

62. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61,** 759–767. ISSN: 0092-8674. https://www.sciencedirect.com/science/article/pii/009286749090186I (1990).

63. Colussi, D., Brandi, G., Bazzoli, F. & Ricciardiello, L. Molecular Pathways Involved in Colorectal Cancer: Implications for Disease Behavior and Prevention. *International Journal of Molecular Sciences* **14,** 16365–16385. ISSN: 1422-0067. https://www.mdpi.com/1422-0067/14/8/16365 (2013).

64. Chen, B. *et al.* Differential pre-malignant programs and microenvironment chart distinct paths to malignancy in human colorectal polyps. *Cell* **184,** 6262–6280.e26. ISSN: 0092-8674. https://www.sciencedirect.com/science/article/pii/S0092867421013817 (2021).

65. Joanito, I. *et al.* Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. *Nature genetics* **54,** 963–975. https://www.nature.com/articles/s41588-022-01100-4 (2022).

66. Carethers, J. M. & Jung, B. H. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology* **149.** Genetics, Genetic Testing, and Biomarkers of Digestive Diseases, 1177–1190.e3. ISSN: 0016-5085. https://www.sciencedirect.com/science/article/pii/S0016508515010045 (2015).

67. Sameer, A. Colorectal Cancer: Molecular Mutations and Polymorphisms. *Frontiers in Oncology* **3.** ISSN: 2234-943X. https://www.frontiersin.org/article/10.3389/fonc.2013.00114 (2013).

68. Sawicki, T. *et al.* A Review of Colorectal Cancer in Terms of Epidemiology, Risk Factors, Development, Symptoms and Diagnosis. *Cancers* **13.** ISSN: 2072-6694. https://www.mdpi.com/2072-6694/13/9/2025 (2021).

69. Hnatyszyn, A. *et al.* Colorectal carcinoma in the course of inflammatory bowel diseases. *Hereditary Cancer in Clinical Practice* **17.** https://hccpjournal.biomedcentral.com/articles/10.1186/s13053-019-0118-4#citeas (2019).

70. Castro, Á. G. & Iglesias, H. L. *Chapter 14 - Histopathological diagnosis of CRC* 149–156. ISBN: 978-0-323-90055-3. https://www.sciencedirect.com/science/article/pii/B9780323900553000508 (Academic Press, 2022).

71. Nagtegaal, I. D. *et al.* The 2019 WHO classification of tumours of the digestive system. *Histopathology* **76,** 182–188. https://onlinelibrary.wiley.com/doi/abs/10.1111/his.13975 (2019).

72. Galuppini, F. *et al.* The histomorphological and molecular landscape of colorectal adenomas and serrated lesions. *Pathologica* **113,** 218. https://www.pathologica.it/article/view/270 (2021).

73. Okita, A. *et al.* Consensus molecular subtypes classification of colorectal cancer as a predictive factor for chemotherapeutic efficacy against metastatic colorectal cancer. *Oncotarget* **9,** 18698–18711. ISSN: 1949-2553. https://www.oncotarget.com/article/24617/ (2018).

74. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nature Medicine* **21,** 1350–1356. https://www.nature.com/articles/nm.3967#citeas (2015).

75. Buikhuisen, J. Y., Torang, A. & Medema, J. P. Exploring and modelling colon cancer inter-tumour heterogeneity: opportunities and challenges. *Oncogenesis* **9,** 1–15. https://www.nature.com/articles/s41389-020-00250-6 (2020).

76. Lenz, H.-J. *et al.* Impact of Consensus Molecular Subtype on Survival in Patients With Metastatic Colorectal Cancer: Results From CALGB/SWOG 80405 (Alliance). *Journal of Clinical Oncology* **37.** PMID: 31042420, 1876–1885. https://doi.org/10.1200/JCO.18.02258 (2019).

77. Isella, C. *et al.* Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nature Communications* **8.** https://www.nature.com/articles/ncomms15107#citeas (2017).

78. Oliveira, R. C., Abrantes, A. M., Tralhão, J. G. & Botelho, M. F. The role of mouse models in colorectal cancer research—The need and the importance of the orthotopic models. *Animal Models and Experimental Medicine* **3,** 1–8. https://onlinelibrary.wiley.com/doi/abs/10.1002/ame2.12102 (2020).

79. Heijstek, M., Kranenburg, O. & Rinkes, I. B. Mouse models of colorectal cancer and liver metastases. *Digestive surgery* **22,** 16–25. https://www.karger.com/Article/FullText/85342# (2005).

80. Sugimoto, S. & Sato, T. Organoid vs In Vivo Mouse Model: Which is Better Research Tool to Understand the Biologic Mechanisms of Intestinal Epithelium? *Cellular and Molecular Gastroenterology and Hepatology* **13,** 195–197. ISSN: 2352-345X. https://www.sciencedirect.com/science/article/pii/S2352345X21002009 (2022).

81. Bürtin, F., Mullins, C. S. & Linnebacher, M. Mouse models of colorectal cancer: Past, present and future perspectives. *World Journal of Gastroenterology* **26,** 1394. https://www.wjgnet.com/1007-9327/full/v26/i13/1394.htm (2020).

82. Tong, Y., Yang, W. & Koeffler, H. P. Mouse models of colorectal cancer. *Chinese journal of cancer* **30,** 450. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4013420/ (2011).

83. Kar, S. K. *et al.* Organoids: a promising new in vitro platform in livestock and veterinary research. *Veterinary Research* **52,** 1–17 (2021).

84. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459,** 262–265. https://www.nature.com/articles/nature07935#citeas (2009).

85. Luo, L., Ma, Y., Zheng, Y., Su, J. & Huang, G. Application Progress of Organoids in Colorectal Cancer. *Frontiers in Cell and Developmental Biology* **10.** ISSN: 2296-634X. https://www.frontiersin.org/article/10.3389/fcell.2022.815067 (2022).

86. Ji, D.-B. & Wu, A.-W. Organoid in colorectal cancer: progress and challenges. *Chinese medical journal* **133,** 1971–1977. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7462208/ (2020).

87. Zhen, Y., Luo, C. & Zhang, H. Early detection of ulcerative colitis-associated colorectal cancer. *Gastroenterology Report* **6,** 83–92. ISSN: 2052-0034. https://doi.org/10.1093/gastro/goy010 (2018).

88. Eaden, J. A., Abrams, K. R. & Mayberry, J. F. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* **48,** 526–535. ISSN: 0017-5749. https://gut.bmj.com/content/48/4/526 (2001).

89. Canavan, C., Abrams, K. R. & Mayberry, J. Meta-analysis: colorectal and small bowel cancer risk in patients with Crohn's disease. *Alimentary Pharmacology & Therapeutics* **23,** 1097–1104. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2036.2006.02854.x (2006).

90. Yalchin, M., Baker, A.-M., Graham, T. A. & Hart, A. Predicting Colorectal Cancer Occurrence in IBD. *Cancers* **13.** ISSN: 2072-6694. https://www.mdpi.com/2072-6694/13/12/2908 (2021).

91. Castaño-Milla, C., Chaparro, M. & Gisbert, J. Systematic review with meta-analysis: the declining risk of colorectal cancer in ulcerative colitis. *Alimentary pharmacology & therapeutics* **39,** 645–659. https://onlinelibrary.wiley.com/doi/full/10.1111/apt.12651 (2014).

92. Lutgens, M. W. M. D. *et al.* Declining Risk of Colorectal Cancer in Inflammatory Bowel Disease: An Updated Meta-analysis of Population-based Cohort Studies. *Inflammatory Bowel Diseases* **19,** 789–799. ISSN: 1078-0998. https://doi.org/10.1097/MIB.0b013e31828029c0 (2013).

93. Choi, C.-H. R. *et al.* Forty-year analysis of colonoscopic surveillance program for neoplasia in ulcerative colitis: An updated overview. *American Journal of Gastroenterology* **110,** 1022–1034. https://journals.lww.com/ajg/Fulltext/2015/07000/Forty_Year_Analysis_of_Colonoscopic_Surveillance.19.aspx (2015).

94. Desai, D. *et al.* Colorectal cancers in ulcerative colitis from a low-prevalence area for colon cancer. *World J. Gastroenterology: WJG* **21,** 3644–3649. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4375589/ (2015).

95. Watanabe, T. *et al.* Ulcerative colitis-associated colorectal cancer shows a poorer survival than sporadic colorectal cancer: A nationwide Japanese study. *Inflammatory Bowel Diseases* **17,** 802–808. ISSN: 1078-0998. https://doi.org/10.1002/ibd.21365 (2010).

96. Baker, A.-M. *et al.* Evolutionary history of human colitis-associated colorectal cancer. *Gut* **68,** 985–995. ISSN: 0017-5749. https://gut.bmj.com/content/68/6/985 (2019).

97. Fantini, M. C. & Guadagni, I. From inflammation to colitis-associated colorectal cancer in inflammatory bowel disease: Pathogenesis and impact of current therapies. *Digestive and Liver Disease* **53,** 558–565. ISSN: 1590-8658. https://www.sciencedirect.com/science/article/pii/S1590865821000189 (2021).

98. Ou, B., Zhao, J., Guan, S. & Lu, A. Survival of colorectal cancer in patients with or without inflammatory bowel disease: a meta-analysis. *Digestive diseases and sciences* **61,** 881–889. https://link.springer.com/article/10.1007/s10620-015-3940-1 (2016).

99. Rothwell, P. M. *et al.* Effect of daily aspirin on long-term risk of death due to cancer: analysis of individual patient data from randomised trials. *The Lancet* **377,** 31–41. ISSN: 0140-6736. https://www.sciencedirect.com/science/article/pii/S0140673610621101 (2011).

100. Satoh-Takayama, N. *et al.* Microbial Flora Drives Interleukin 22 Production in Intestinal NKp46+ Cells that Provide Innate Mucosal Immune Defense. *Immunity* **29,** 958–970. ISSN: 1074-7613. https://www.sciencedirect.com/science/article/pii/S1074761308005049 (2008).

101. Sugimoto, K. *et al.* IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. *The Journal of Clinical Investigation* **118,** 534–544. https://doi.org/10.1172/JCI33194 (2008).

102. Kamanaka, M. *et al.* Memory/effector (CD45RBlo) CD4 T cells are controlled directly by IL-10 and cause IL-22–dependent intestinal pathology. *Journal of Experimental Medicine* **208,** 1027–1040. ISSN: 0022-1007. https://doi.org/10.1084/jem.20102149 (2011).

103. Pickert, G. *et al.* STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. *Journal of Experimental Medicine* **206,** 1465–1472. ISSN: 0022-1007. https://doi.org/10.1084/jem.20082683 (2009).

104. Kirchberger, S. *et al.* Innate lymphoid cells sustain colon cancer through production of interleukin-22 in a mouse model. *Journal of Experimental Medicine* **210,** 917–931. ISSN: 0022-1007. https://doi.org/10.1084/jem.20122308 (2013).

105. Schwitalla, S. *et al.* Intestinal Tumorigenesis Initiated by Dedifferentiation and Acquisition of Stem-Cell-like Properties. *Cell* **152,** 25–38. ISSN: 0092-8674. https://www.sciencedirect.com/science/article/pii/S0092867412014997 (2013).

106. Itzkowitz, S. H. & Yio, X. Inflammation and Cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **287.** PMID: 15194558, G7–G17. https://doi.org/10.1152/ajpgi.00079.2004 (2004).

107. Vogelstein, B. *et al.* Genetic Alterations during Colorectal-Tumor Development. *New England Journal of Medicine* **319.** PMID: 2841597, 525–532. https://doi.org/10.1056/NEJM198809013190901 (1988).

108. Curtius, K., Wright, N. A. & Graham, T. A. An evolutionary perspective on field cancerization. *Nature Reviews Cancer* **18,** 19–32. https://www.nature.com/articles/nrc.2017.102 (2018).

109. Leedham, S. J. *et al.* Clonality, founder mutations, and field cancerization in human ulcerative colitis–associated neoplasia. *Gastroenterology* **136,** 542–550. https://www.sciencedirect.com/science/article/pii/S0016508508020398?casa_token=fHCoGacFzcIAAAAA:Cc8hTxSNpESm-dK9O6qCJl6kSvWdYpN07Oz6cXHVAdaXDmfiCFQfk-5eZcxmMrDRmx1lmembxg (2009).

110. Fujita, M. *et al.* Genomic landscape of colitis-associated cancer indicates the impact of chronic inflammation and its stratification by mutations in the Wnt signaling. *Oncotarget* **9,** 969–981. ISSN: 1949-2553. https://www.oncotarget.com/article/22867/ (2018).

111. Yaeger, R. *et al.* Genomic Alterations Observed in Colitis-Associated Cancers Are Distinct From Those Found in Sporadic Colorectal Cancers and Vary by Type of Inflammatory Bowel Disease. *Gastroenterology* **151,** 278–287.e6. ISSN: 0016-5085. https://www.sciencedirect.com/science/article/pii/S0016508516300531 (2016).

112. Chatila, W. K. *et al.* Genomic alterations in colitis-associated cancers in comparison to those found in sporadic colorectal cancer and present in precancerous dysplasia. *Journal of Clinical Oncology* **38,** 191–191. https://doi.org/10.1200/JCO.2020.38.4_suppl.191 (2020).

113. Beaugerie, L. & Itzkowitz, S. H. Cancers Complicating Inflammatory Bowel Disease. *New England Journal of Medicine* **372.** PMID: 25853748, 1441–1452. https://doi.org/10.1056/NEJMra1403718 (2015).

114. Rajamäki, K. *et al.* Genetic and Epigenetic Characteristics of Inflammatory Bowel Disease-Associated Colorectal Cancer. *Gastroenterology* **161,** 592–607. ISSN: 0016-5085. https://www.sciencedirect.com/science/article/pii/S0016508521006806 (2021).

115. Soh, J. S. *et al.* Immunoprofiling of colitis-associated and sporadic colorectal cancer and its clinical significance. *Scientific reports* **9,** 1–10. https://www.nature.com/articles/s41598-019-42986-1.#citeas (2019).

116. Porter, R. J., Arends, M. J., Churchhouse, A. M. D. & Din, S. Inflammatory Bowel Disease-Associated Colorectal Cancer: Translational Risks from Mechanisms to Medicines. *Journal of Crohn's and Colitis* **15,** 2131–2141. ISSN: 1873-9946. https://doi.org/10.1093/ecco-jcc/jjab102 (2021).

117. Shibata, H. *et al.* Rapid Colorectal Adenoma Formation Initiated by Conditional Targeting of the Apc Gene. *Science* **278,** 120–123. https://www.science.org/doi/abs/10.1126/science.278.5335.120 (1997).

118. Moser, A. R., Pitot, H. C. & Dove, W. F. A Dominant Mutation That Predisposes to Multiple Intestinal Neoplasia in the Mouse. *Science* **247,** 322–324. https://www.science.org/doi/abs/10.1126/science.2296722 (1990).

119. Tomic, G. *et al.* Phospho-regulation of ATOH1 Is Required for Plasticity of Secretory Progenitors and Tissue Regeneration. *Cell Stem Cell* **23,** 436–443.e7. ISSN: 1934-5909. https://www.sciencedirect.com/science/article/pii/S1934590918303382 (2018).

120. Ahn, S. & Joyner, A. L. Dynamic Changes in the Response of Cells to Positive Hedgehog Signaling during Mouse Limb Patterning. *Cell* **118,** 505–516. ISSN: 0092-8674. https://www.sciencedirect.com/science/article/pii/S0092867404007111 (2004).

121. Jackstadt, R. *et al.* Epithelial NOTCH Signaling Rewires the Tumor Microenvironment of Colorectal Cancer to Drive Poor-Prognosis Subtypes and Metastasis. *Cancer Cell* **36,** 319–336.e7. ISSN: 1535-6108. https://www.sciencedirect.com/science/article/pii/S153561081930371X (2019).

122. Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478,** 255–259. https://pubmed.ncbi.nlm.nih.gov/21927002/ (2011).

123. Hilkens, J. *et al.* RSPO3 expands intestinal stem cell and niche compartments and drives tumorigenesis. *Gut* **66,** 1095–1105. https://gut.bmj.com/content/66/6/1095 (2017).

124. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature neuroscience* **13,** 133–140. https://pubmed.ncbi.nlm.nih.gov/20023653/ (2010).

125. Marino, S., Vooijs, M., van der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes & Development* **14,** 994–1004. http://genesdev.cshlp.org/content/14/8/994.abstract (2000).

126. El Marjou, F. *et al.* Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *genesis* **39,** 186–193. https://onlinelibrary.wiley.com/doi/abs/10.1002/gene.20042 (2004).

127. Rudling, R., Hassan, A. B., Kitau, J., Mandir, N. & Goodlad, R. A. A simple device to rapidly prepare whole mounts of murine intestine. *Cell Proliferation* **39,** 415–420. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2184.2006.00391.x (2006).

128. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9,** 671–675. https://www.nature.com/articles/nmeth.2089 (2012).

129. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Scientific reports* **7,** 1–7. https://www.nature.com/articles/s41598-017-17204-5 (2017).

130. Ubink, I. *et al.* A Novel Diagnostic Tool for Selecting Patients With Mesenchymal-Type Colon Cancer Reveals Intratumor Subtype Heterogeneity. *JNCI: Journal of the National Cancer Institute* **109.** ISSN: 0027-8874. https://doi.org/10.1093/jnci/djw303 (2017).

131. Lee, H.-O. *et al.* Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nature Genetics* **52,** 594–603. https://www.nature.com/articles/s41588-020-0636-z (2020).

132. Muñoz, J. *et al.* The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *The EMBO Journal* **31,** 3079–3091. https://www.embopress.org/doi/abs/10.1038/emboj.2012.166 (2012).

133. Mustata, R. C. *et al.* Identification of Lgr5-Independent Spheroid-Generating Progenitors of the Mouse Fetal Intestinal Epithelium. *Cell Reports* **5,** 421–432. ISSN: 2211-1247. https://www.sciencedirect.com/science/article/pii/S2211124713005135 (2013).

134. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* **14,** 1–15. https://link.springer.com/article/10.1186/1471-2105-14-7 (2013).

135. Gregorieff, A., Liu, Y., Inanlou, M. R., Khomchuk, Y. & Wrana, J. L. Yap-dependent reprogramming of Lgr5+ stem cells drives intestinal regeneration and cancer. *Nature* **526,** 715–718. https://www.nature.com/articles/nature15382 (2015).

136. Sansom, O. Reed KR, Hayes AJ, Ireland H, Brinkmann H, Newton IP, Batlle E, Simon-Assmann P, Clevers H, Nathke IS, Clarke AR, Winton DJ. *Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. Genes Dev* **18,** 1385–1390. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC423189/ (2004).

137. Van der Flier, L. G. *et al.* The Intestinal Wnt/TCF Signature. *Gastroenterology* **132,** 628–632. ISSN: 0016-5085. https://www.sciencedirect.com/science/article/pii/S0016508506017999 (2007).

138. Loboda, A. *et al.* A gene expression signature of RAS pathway dependence predicts response to PI3K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors. *BMC medical genomics* **3,** 1–11. https://bmcmedgenomics.biomedcentral.com/articles/10.1186/1755-8794-3-26#citeas (2010).

139. Ayers, M. *et al.* IFN-γ–related mRNA profile predicts clinical response to PD-1 blockade. *The Journal of Clinical Investigation* **127,** 2930–2940. https://www.jci.org/articles/view/91190 (2017).

140. Calon, A. *et al.* Dependency of Colorectal Cancer on a TGF-β-Driven Program in Stromal Cells for Metastasis Initiation. *Cancer Cell* **22,** 571–584. ISSN: 1535-6108. https://www.sciencedirect.com/science/article/pii/S1535610812003558 (2012).

141. Choi, W.-T. *et al.* Nonconventional dysplasia in patients with inflammatory bowel disease and colorectal carcinoma: a multicenter clinicopathologic study. *Modern Pathology* **33,** 933–943. https://pubmed.ncbi.nlm.nih.gov/31822800/ (2020).

142. Lee, H., Rabinovitch, P. S., Mattis, A. N., Lauwers, G. Y. & Choi, W.-T. Non-conventional dysplasia in inflammatory bowel disease is more frequently associated with advanced neoplasia and aneuploidy than conventional dysplasia. *Histopathology* **78,** 814–830. https://onlinelibrary.wiley.com/doi/abs/10.1111/his.14298.

143. Nusse, Y. M. *et al.* Parasitic helminths induce fetal-like reversion in the intestinal stem cell niche. *Nature* **559,** 109–113. https://www.nature.com/articles/s41586-018-0257-1 (2018).

144. Van Es, J. H. *et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. *Nature cell biology* **14,** 1099–1104. https://www.nature.com/articles/ncb2581 (2012).

145. Buczacki, S. J. *et al.* Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* **495,** 65–69. https://www.nature.com/articles/nature11965 (2013).

146. Westphalen, C. B. *et al.* Long-lived intestinal tuft cells serve as colon cancer–initiating cells. *The Journal of clinical investigation* **124,** 1283–1295. https://pubmed.ncbi.nlm.nih.gov/24487592/ (2014).

147. Tetteh, P. W. *et al.* Replacement of lost Lgr5-positive stem cells through plasticity of their enterocyte-lineage daughters. *Cell stem cell* **18,** 203–213. https://pubmed.ncbi.nlm.nih.gov/26831517/ (2016).

148. Yu, S. *et al.* Paneth cell multipotency induced by notch activation following injury. *Cell stem cell* **23,** 46–59. https://www.sciencedirect.com/science/article/pii/S1934590918302200?via%3Dihub (2018).

149. Schmitt, M. *et al.* Paneth cells respond to inflammation and contribute to tissue regeneration by acquiring stem-like features through SCF/c-Kit signaling. *Cell reports* **24,** 2312–2328. https://www.cell.com/cell-reports/fulltext/S2211-1247(18)31203-8?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2211124718312038%3Fshowall%3Dtrue (2018).

150. Varga, J. & Greten, F. Cell plasticity in epithelial homeostasis and tumorigenesis. *Nature Cell Biology* **19,** 1133–1141. https://www.nature.com/articles/ncb3611 (2017).

151. Nimmakayala, R. K., Batra, S. K. & Ponnusamy, M. P. Unraveling the journey of cancer stem cells from origin to metastasis. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1871,** 50–63. ISSN: 0304-419X. https://www.sciencedirect.com/science/article/pii/S0304419X18301161 (2019).

152. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature medicine* **3,** 730–737. https://www.nature.com/articles/nm0797-730 (1997).

153. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences* **100,** 3983–3988. https://www.pnas.org/doi/abs/10.1073/pnas.0530291100 (2003).

154. Lee, J. *et al.* Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* **9,** 391–403. ISSN: 1535-6108. https://www.sciencedirect.com/science/article/pii/S1535610806001176 (2006).

155. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445,** 106–110. https://www.nature.com/articles/nature05372 (2007).

156. Hervieu, C., Christou, N., Battu, S. & Mathonnet, M. The Role of Cancer Stem Cells in Colorectal Cancer: From the Basics to Novel Clinical Trials. *Cancers* **13.** ISSN: 2072-6694. https://www.mdpi.com/2072-6694/13/5/1092 (2021).

157. Zhou, Y. *et al.* Cancer stem cells in progression of colorectal cancer. *Oncotarget* **9,** 33403–33415. ISSN: 1949-2553. https://www.oncotarget.com/article/23607/ (2018).

158. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457,** 608–611. https://www.nature.com/articles/nature07602 (2009).

159. Davis, H. *et al.* Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche. *Nature medicine* **21,** 62–70. https://www.nature.com/articles/nm.3750 (2015).

160. Sousa e Melo, F. d. *et al.* A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer. vol. 543. *Nature,* 676–680. https://www.nature.com/articles/nature21713 (2017).

161. Fumagalli, A. *et al.* Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. *Cell Stem Cell* **26,** 569–578.e7. ISSN: 1934-5909. https://www.sciencedirect.com/science/article/pii/S1934590920300618 (2020).

162. Kalimuthu, S. N., Chelliah, A. & Chetty, R. From traditional serrated adenoma to tubulovillous adenoma and beyond. *World journal of gastrointestinal oncology* **8,** 805. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5156846/ (2016).

163. Jackstadt, R. *et al.* Epithelial NOTCH signaling rewires the tumor microenvironment of colorectal cancer to drive poor-prognosis subtypes and metastasis. *Cancer cell* **36,** 319–336. https://pubmed.ncbi.nlm.nih.gov/31526760/ (2019).

164. Lannagan, T. R., Jackstadt, R., Leedham, S. J. & Sansom, O. J. Advances in colon cancer research: in vitro and animal models. *Current opinion in genetics & development* **66,** 50–56. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7985292/ (2021).

165. Merlos-Suárez, A. *et al.* The Intestinal Stem Cell Signature Identifies Colorectal Cancer Stem Cells and Predicts Disease Relapse. *Cell Stem Cell* **8,** 511–524. ISSN: 1934-5909. https://www.sciencedirect.com/science/article/pii/S193459091100110X (2011).

166. Tan, S. H. *et al.* A constant pool of Lgr5+ intestinal stem cells is required for intestinal homeostasis. *Cell Reports* **34,** 108633. https://linkinghub.elsevier.com/retrieve/pii/S2211-1247(20)31622-3 (2021).

167. Flanagan, D. J. *et al.* NOTUM from Apc-mutant cells biases clonal competition to initiate cancer. *Nature* **594,** 430–435. https://www.nature.com/articles/s41586-021-03525-z#citeas (2021).

168. Van Neerven, S. M. *et al.* Apc-mutant cells act as supercompetitors in intestinal tumour initiation. *Nature* **594,** 436–441. https://www.nature.com/articles/s41586-021-03558-4 (2021).

169. Han, T. *et al.* Lineage Reversion Drives WNT Independence in Intestinal Cancer. *Cancer Discovery* **10,** 1590–1609. ISSN: 2159-8274. https://doi.org/10.1158/2159-8290.CD-19-1536 (2020).

170. Solé, L. *et al.* p53 wild-type colorectal cancer cells that express a fetal gene signature are associated with metastasis and poor prognosis. *Nature Communications* **13,** 1–15. https://www.nature.com/articles/s41467-022-30382-9#citeas (2022).

171. Álvarez-Varela, A. *et al.* Mex3a marks drug-tolerant persister colorectal cancer cells that mediate relapse after chemotherapy. *Nature Cancer,* 1–19. https://www.nature.com/articles/s43018-022-00402-0 (2022).

172. Touil, Y. *et al.* Colon cancer cells escape 5FU chemotherapy-induced cell death by entering stemness and quiescence associated with the c-Yes/YAP axis. *Clinical cancer research* **20,** 837–846. https://pubmed.ncbi.nlm.nih.gov/24323901/ (2014).

173. Gramlich, M. *et al.* HDX-MS for Epitope Characterization of a Therapeutic ANTIBODY Candidate on the Calcium-Binding Protein Annexin-A1. *Antibodies* **10.** ISSN: 2073-4468. https://www.mdpi.com/2073-4468/10/1/11 (2021).

174. Bai, F. *et al.* Targeting ANXA1 abrogates Treg-mediated immune suppression in triple-negative breast cancer. *Journal for ImmunoTherapy of Cancer* **8.** https://jitc.bmj.com/content/8/1/e000169 (2020).

175. Dempsey, F. C. *et al.* Efficacy of MDX-124, a novel anti-annexin-A1 antibody, in preclinical models of pancreatic cancer. *Journal of Clinical Oncology* **40,** 590–590. https://doi.org/10.1200/JCO.2022.40.4_suppl.590 (2022).

176. Kleeman, S. O. & Leedham, S. J. Not All Wnt Activation Is Equal: Ligand-Dependent versus Ligand-Independent Wnt Activation in Colorectal Cancer. *Cancers* **12.** ISSN: 2072-6694. https://www.mdpi.com/2072-6694/12/11/3355 (2020).

177. Parsons, M. J., Tammela, T. & Dow, L. E. WNT as a Driver and Dependency in Cancer. *Cancer Discovery* **11,** 2413–2429. ISSN: 2159-8274. https://doi.org/10.1158/2159-8290.CD-21-0190 (2021).

178. Gong, X. *et al.* Aberrant RSPO3-LGR4 signaling in Keap1-deficient lung adenocarcinomas promotes tumor aggressiveness. *Oncogene* **34,** 4692–4701. https://www.nature.com/articles/onc2014417 (2015).

179. Leedham, S. J. & Chetty, R. Wnt disruption in colorectal polyps – the traditional serrated adenoma enters the fray. *The Journal of Pathology* **239,** 387–390. https://onlinelibrary.wiley.com/doi/abs/10.1002/path.4741 (2016).

180. Yum, M. K. *et al.* Tracing oncogene-driven remodelling of the intestinal stem cell niche. *Nature* **594,** 442–447. https://www.nature.com/articles/s41586-021-03605-0 (2021).

181. Phesse, T. J. *et al.* Partial inhibition of gp130-Jak-Stat3 signaling prevents Wnt-B-catenin-mediated intestinal tumor growth and regeneration. *Science Signaling* **7,** ra92–ra92. https://www.science.org/doi/abs/10.1126/scisignal.2005411 (2014).

182. Kim, K.-A. *et al.* R-Spondin Family Members Regulate the Wnt Pathway by a Common Mechanism. *Molecular Biology of the Cell* **19,** 2588–2596. https://doi.org/10.1091/mbc.e08-02-0187 (2008).

183. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488,** 660–664. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3690621/?report=reader (2012).

184. Shinmura, K. *et al.* RSPO fusion transcripts in colorectal cancer in Japanese population. *Molecular biology reports* **41,** 5375–5384. https://pubmed.ncbi.nlm.nih.gov/24847761/ (2014).

185. Hashimoto, T. *et al.* EIF3E–RSPO2 and PIEZO1–RSPO2 fusions in colorectal traditional serrated adenoma. *Histopathology* **75,** 266–273. https://onlinelibrary.wiley.com/doi/abs/10.1111/his.13867 (2019).

186. Gay, D. M. *et al.* Loss of BCL9/9l suppresses Wnt driven tumourigenesis in models that recapitulate human cancer. *Nature communications* **10,** 1–16. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6374445/pdf/41467_2019_Article_8586.pdf (2019).

187. Metidji, A. *et al.* The Environmental Sensor AHR Protects from Inflammatory Damage by Maintaining Intestinal Stem Cell Homeostasis and Barrier Integrity. *Immunity* **49,** 353–362.e5. ISSN: 1074-7613. https://www.sciencedirect.com/science/article/pii/S1074761318303303 (2018).

188. Chan, J. M. *et al.* Rare germline variants in the AXIN2 gene in families with colonic polyposis and colorectal cancer. *Familial Cancer,* 1–15. https://pubmed.ncbi.nlm.nih.gov/34817745/ (2021).

189. Simmonds, N., Furman, M., Karanika, E., Phillips, A. & Bates, A. W. Paneth cell metaplasia in newly diagnosed inflammatory bowel disease in children. *BMC gastroenterology* **14,** 1–6. https://bmcgastroenterol.biomedcentral.com/articles/10.1186/1471-230X-14-93#citeas (2014).

190. Aust, D. E. *et al.* The APC/β-catenin pathway in ulcerative colitis–related colorectal carcinomas. *Cancer* **94,** 1421–1427. https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.1002/cncr.10334 (2002).

191. Tarmin, L. *et al.* Adenomatous polyposis coli gene mutations in ulcerative colitis-associated dysplasias and cancers versus sporadic colon neoplasms. *Cancer research* **55,** 2035–2038. https://pubmed.ncbi.nlm.nih.gov/7743497/ (1995).

192. Madan, B. *et al.* Wnt addiction of genetically defined cancers reversed by PORCN inhibition. *Oncogene* **35,** 2197–2207. https://www.nature.com/articles/onc2015280 (2016).

193. Koo, B.-K., van Es, J. H., van den Born, M. & Clevers, H. Porcupine inhibitor suppresses paracrine Wnt-driven growth of Rnf43;Znrf3-mutant neoplasia. *Proceedings of the National Academy of Sciences* **112,** 7548–7550. https://www.pnas.org/doi/abs/10.1073/pnas.1508113112 (2015).

194. Chartier, C. *et al.* Therapeutic targeting of tumor-derived R-spondin attenuates β-catenin signaling and tumorigenesis in multiple cancer types. *Cancer research* **76,** 713–723. https://aacrjournals.org/cancerres/article/76/3/713/616124/Therapeutic-Targeting-of-Tumor-Derived-R-Spondin (2016).

195. Fischer, M. M. *et al.* RSPO3 antagonism inhibits growth and tumorigenicity in colorectal tumors harboring common Wnt pathway mutations. *Scientific reports* **7,** 1–9. https://pubmed.ncbi.nlm.nih.gov/29127379/ (2017).

196. Janku, F. *et al.* Abstract C45: Phase I study of WNT974, a first-in-class Porcupine inhibitor, in advanced solid tumors. *Molecular Cancer Therapeutics* **14,** C45–C45. https://doi.org/10.1158/1535-7163.TARG-15-C45 (2015).

197. Bendell, J. *et al.* 68 - Initial results from a phase 1a/b study of OMP-131R10, a first-in-class anti-RSPO3 antibody, in advanced solid tumors and previously treated metastatic colorectal cancer (CRC). *European Journal of Cancer* **69,** S29–S30. https://www.sciencedirect.com/science/article/pii/S0959804916326685 (2016).

198. Funck-Brentano, T. *et al.* Porcupine inhibitors impair trabecular and cortical bone mass and strength in mice. *Journal of Endocrinology* **238,** 13–23. https://joe.bioscientifica.com/view/journals/joe/238/1/JOE-18-0153.xml (2018).

199. Garcia, J. S. *et al.* A systematic review of higher-risk myelodysplastic syndromes clinical trials to determine the benchmark of azacitidine and explore alternative endpoints for overall survival. *Leukemia Research* **104,** 106555. ISSN: 0145-2126. https://www.sciencedirect.com/science/article/pii/S0145212621000564 (2021).

200. Perminow, G. *et al.* Defective Paneth Cell—Mediated Host Defense in Pediatric Ileal Crohn's Disease. *American Journal of Gastroenterology* **105,** 452–459. https://journals.lww.com/ajg/Fulltext/2010/02000/Defective_Paneth_Cell_Mediated_Host_Defense_in.32.aspx (2010).

201. Lueschow, S. R. & McElroy, S. J. The Paneth cell: the curator and defender of the immature small intestine. *Frontiers in immunology* **11,** 587. https://www.frontiersin.org/articles/10.3389/fimmu.2020.00587/full (2020).

202. Heuberger, J. *et al.* Shp2/MAPK signaling controls goblet/paneth cell fate decisions in the intestine. *Proceedings of the National Academy of Sciences* **111,** 3472–3477. https://www.pnas.org/doi/abs/10.1073/pnas.1309342111 (2014).

203. Tsai, J.-H. *et al.* RNF43 mutation frequently occurs with GNAS mutation and mucin hypersecretion in intraductal papillary neoplasms of the bile duct. *Histopathology* **70,** 756–765. https://onlinelibrary.wiley.com/doi/abs/10.1111/his.13125 (2017).

204. Ryland, G. L. *et al.* RNF43 is a tumour suppressor gene mutated in mucinous tumours of the ovary. *The Journal of Pathology* **229,** 469–476. https://onlinelibrary.wiley.com/doi/abs/10.1002/path.4134 (2013).

205. Chang, X. Y., Wu, Y., Jiang, Y., Wang, P. Y. & Jie, C. RNF43 mutations in IPMN cases: a potential prognostic factor. *Gastroenterology research and practice* **2020.** https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7479465/ (2020).

206. Li, Y. *et al.* Frequent RNF43 mutation contributes to moderate activation of Wnt signaling in colorectal signet-ring cell carcinoma. *Protein & cell* **11,** 292–298. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7093338/ (2020).

207. Lamb, C. A. *et al.* British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut* **68,** s1–s106. ISSN: 0017-5749. https://gut.bmj.com/content/68/Suppl_3/s1 (2019).

208. Curtius, K. *et al.* Multicentre derivation and validation of a colitis-associated colorectal cancer risk prediction web tool. *Gut* **71,** 705–715. ISSN: 0017-5749. https://gut.bmj.com/content/71/4/705 (2022).

209. Odze, R. D., Farraye, F. A., Hecht, J. L. & Hornick, J. L. Long-term follow-up after polypectomy treatment for adenoma-like dysplastic lesions in ulcerative colitis. *Clinical Gastroenterology and Hepatology* **2,** 534–541. ISSN: 1542-3565. https://www.sciencedirect.com/science/article/pii/S154235650400237X (2004).

210. Kisiel, J. B., Loftus Jr, E. V., Harmsen, S. W., Zinsmeister, A. R. & Sandborn, W. J. Outcome of sporadic adenomas and adenoma-like dysplasia in patients with ulcerative colitis undergoing polypectomy. *Inflammatory bowel diseases* **18,** 226–235. https://academic.oup.com/ibdjournal/article/18/2/226/4607842?login=true (2012).

211. Choi, C.-H. R. *et al.* Low-grade dysplasia in ulcerative colitis: risk factors for developing high-grade dysplasia or colorectal cancer. *The American journal of gastroenterology* **110,** 1461. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4697133/ (2015).

212. Van Schaik, F. D. *et al.* Endoscopic and pathological aspects of colitis-associated dysplasia. *Nature Reviews Gastroenterology & Hepatology* **6,** 671–678. https://www.nature.com/articles/nrgastro.2009.162#citeas (2009).

213. Rubin, C. E. *et al.* DNA aneuploidy in colonic biopsies predicts future development of dysplasia in ulcerative colitis. *Gastroenterology* **103,** 1611–1620. https://pubmed.ncbi.nlm.nih.gov/1426881/ (1992).

214. Hartmann, D.-P., Montgomery, E. A., Carr, N. J., Gupta, P. K. & Azumi, N. Flow Cytometric DNA Analysis of Ulcerative Colitis Using Paraffin–embedded Biopsy Specimens: Comparison with Morphology and DNA Analysis of Fresh Samples. *American Journal of Gastroenterology (Springer Nature)* **90.** https://pubmed.ncbi.nlm.nih.gov/7717317/ (1995).

215. McHugh, M. L. Interrater reliability: the kappa statistic. *Biochemia medica* **22,** 276–282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/ (2012).

216. Tan, E. S. *et al.* Copy Number Alterations as Novel Biomarkers and Therapeutic Targets in Colorectal Cancer. *Cancers* **14.** ISSN: 2072-6694. https://www.mdpi.com/2072-6694/14/9/2223 (2022).

217. Aust, D. E. *et al.* Chromosomal alterations in ulcerative colitis-related and sporadic colorectal cancers by comparative genomic hybridization. *Human Pathology* **31,** 109–114. ISSN: 0046-8177. https://www.sciencedirect.com/science/article/pii/S0046817700802063 (2000).

218. Dunn, J. *et al.* Image cytometry accurately detects DNA ploidy abnormalities and predicts late relapse to high-grade dysplasia and adenocarcinoma in Barrett's oesophagus following photodynamic therapy. *British journal of cancer* **102,** 1608–1617. https://www.nature.com/articles/6605688 (2010).

219. Williams, B. J., Lee, J., Oien, K. A. & Treanor, D. Digital pathology access and usage in the UK: results from a national survey on behalf of the National Cancer Research Institute's CM-Path initiative. *Journal of clinical pathology* **71,** 463–466. https://jcp.bmj.com/content/71/5/463?etoc=&int_source=trendmd&int_medium=cpc&int_campaign=usage-042019 (2018).

220. Ishibashi, F. *et al.* Contribution of ATOH1+ Cells to the Homeostasis, Repair, and Tumorigenesis of the Colonic Epithelium. *Stem Cell Reports* **10,** 27–42. ISSN: 2213-6711. https://www.sciencedirect.com/science/article/pii/S2213671117304927 (2018).

221. Rycaj, K. & Tang, D. G. Cell-of-Origin of Cancer versus Cancer Stem Cells: Assays and Interpretations. *Cancer Research* **75,** 4003–4011. ISSN: 0008-5472. https://doi.org/10.1158/0008-5472.CAN-15-0798 (2015).

222. Iwaya, M. *et al.* Most colitis associated carcinomas lack expression of LGR5: a preliminary study with implications for unique pathways of carcinogenesis compared to sporadic colorectal carcinoma. *BMC cancer* **21,** 1–9. https://bmccancer.biomedcentral.com/articles/10.1186/s12885-021-07835-3#citeas (2021).

223. Baslan, T. *et al.* Ordered and deterministic cancer genome evolution after p53 loss. *Nature,* 1–8. https://www.nature.com/articles/s41586-022-05082-5#citeas (2022).

224. Hanel, W. & Moll, U. M. Links between mutant p53 and genomic instability. *Journal of Cellular Biochemistry* **113,** 433–439. https://onlinelibrary.wiley.com/doi/abs/10.1002/jcb.23400 (2012).

225. Redman-Rivera, L. N. *et al.* Acquisition of aneuploidy drives mutant p53-associated gain-of-function phenotypes. *Nature communications* **12,** 1–15. https://www.nature.com/articles/s41467-021-25359-z#citeas (2021).

226. Levéen, P. *et al.* Induced disruption of the transforming growth factor beta type II receptor gene in mice causes a lethal inflammatory disorder that is transplantable. *Blood* **100,** 560–568. https://doi.org/10.1182/blood.V100.2.560 (July 2002).

227. Mercer, K. *et al.* Expression of Endogenous Oncogenic V600EB-raf Induces Proliferation and Developmental Defects in Mice and Transformation of Primary Fibroblasts. *Cancer Research* **65,** 11493–11500. ISSN: 0008-5472. https://doi.org/10.1158/0008-5472.CAN-05-2211 (2005).

228. Murtaugh, L. C., Stanger, B. Z., Kwan, K. M. & Melton, D. A. Notch signaling controls multiple steps of pancreatic differentiation. *Proceedings of the National Academy of Sciences* **100,** 14920–14925. https://www.pnas.org/doi/abs/10.1073/pnas.2436557100 (2003).