# Host-microbe interactions in inflammation-driven cancer

Alina Joanna Janney




**University of Oxford**

St. Cross College

Kennedy Institute of Rheumatology, NDORMS

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

Michaelmas 2021

Abstract Tanya Sam

Alina Janney | DPhil Thesis Abstract | St Cross College | Michaelmas 2021

**Host-microbe interactions in inflammation-driven cancer**

Colorectal cancer (CRC) is a heterogenous disease of the intestinal epithelium that is characterised by the accumulation of mutations and a dysregulated immune response. A growing body of literature now also illustrates that this is accompanied by an "oncogenic" CRC-associated microbiota, but whether this dysbiosis is a functional hallmark of CRC or merely a bystander effect of disease is unclear. Therefore, the aim of this DPhil project was to identify changes within the microbiome during cancer progression and the functional consequences. To achieve this, we induced dysbiosis in genetically susceptible mice by colonisation with the pathobiont *Helicobacter hepaticus (Hh)* followed by the treatment with the carcinogen azoxymethane (*Hh*AOM).

Using a range of unbiased longitudinal approaches, we found a distal colon-specific neutrophil signature during tumorigenesis and identified 118 genes that were differentially expressed in *Hh*AOM versus control mice, indicative of putative tumour-promoting properties. Amongst these, the enzyme peptidyl arginine deaminase 4 (PAD4) was upregulated in epithelial cells of early and late tumorigenesis. PAD4 was verified to play a functional role in cancer progression because therapeutic *in vivo* inhibition reduced tumour area. Alongside this observation, we discovered that *Hh* resides within established tumours but lacks the capability to drive CAC in gnotobiotic settings.

Collectively, our findings suggest that *Hh*-driven colitis triggers an early neutrophil influx into the distal colon which fosters a microenvironment that promotes tumour progression characterised by the expression of PAD4. These observations open new potential avenues to explore tumour-promoting function of neutrophil subsets in more detail and the role of the citrullination enzymes like PAD4 in CRC progression.

# Acknowledgements

I would like to express my deepest gratitude to my primary supervisor Professor Fiona Powrie who has been continuously supportive, provided excellent scientific guidance and granted me the freedom to explore novel ideas throughout this DPhil.

This gratitude naturally extends to my co-supervisor Dr. Elizabeth Mann. She has not only taught me great scientific and much-needed management skills, but has also continuously supported and encouraged me in the toughest of times. I will forever appreciate the time and energy Beth has invested in me.

Also, I would like to thank Professor Cindy Sears and Professor Simon Leedham for their profound participation in my thesis committee, and Dr. Alejandro Suarez-Bonnet, Dr. Sena Bae and Nadia Nasreddin for their helpful collaborations.

Further, I would like to express my gratitude to each Powrie lab member for their continuous support throughout this terrific experience. I am especially thankful for Fanny who was the first member to take me under her generous wing, alongside Mathilde and Matthias who shared their immense knowledge with me as inspiring colleagues as well as great friends. A special thank you also goes to Claire for her all-round, daily support (+ numerous takedowns!) and to Nick - a bioinformatic mentor without whom parts of thesis would have been unimaginable. Lastly, a big thanks to the lab members Sam, Andrea, Irene, Danny, Lilian, Eli, Matt, Becky, Cornelia and Ali who have been extremely helpful and wonderfully encouraging throughout this thesis.

# Acknowledgements

Moving to Oxford changed my life and I was lucky enough to not only find supportive colleagues and friends, but a true OXFAMily. Anne, Sara, Lion, Felix, Sarah, David D, David T and Luca – thank you.

Another very special member of this family is my partner, Pablo. I am forever thankful for his never-ending support and the love he provides in all areas of my life. His brilliant mind has always inspired me, but it was his encouragement, his heart and his lasagne that kept me going these past years - and for many more to come.

Lastly, I would like to thank the people who encouraged me to pursue this once in a life time opportunity. This includes my great support network in Germany, my wonderful Aunt Sam and Uncle Graham, my loving Dad and stepmother Suzanne and my extraordinary sister Marie. The amount of love, support and inspiration provided by them goes beyond any description.

# Contributions

- Dr. Alejandro Suarez-Bonnet performed the histopathological analysis on murine colon tissue.

- Nadia Nasreddin stained *Lgr5* and *Ly6a* on murine colon tissue.

- Dr. Elizabeth Mann performed the *in silico* analysis of PAD4 expression (RNA and protein) in CRC patients.

- Dr. Sena Bae analysed the created 16S-rRNA Sequencing dataset.

# Table of Contents

ABSTRACT ..... II

ACKNOWLEDGEMENTS ..... IV

CONTRIBUTIONS ..... VI

TABLE OF CONTENTS ..... VII

LIST OF ABBREVIATIONS ..... XIII

1 INTRODUCTION ..... 1

**1.1 The host-microbiota interface ..... 1**

1.1.1 Colon architecture ..... 1

1.1.2 The gut microbiota ..... 3

1.1.3 The epithelial barrier ..... 3

1.1.4 Immune responses in the colon ..... 5

1.1.5 Adaptive Immunity ..... 11

**1.2 The aetiology of CRC and CAC ..... 12**

**1.3 The tumour microenvironment ..... 14**

1.3.1 Immune cells in the TME – a double-edged sword. ..... 15

1.3.2 The ambiguous role of tumour-associated neutrophils ..... 16

1.3.3 IL-22 in CRC – Friend or foe? ..... 19

**1.4 A novel CRC hallmark on the horizon: Dysbiosis ..... 20**

1.4.1 Inflammatory mechanisms underlying microbiota interactions in CRC ..... 23

**1.5 Mouse models recapitulating CRC ..... 25**

**1.6 Research hypotheses and experimental aims ..... 27**

2 MATERIALS AND METHODS ..... 29

**2.1 Mice ..... 29**

**2.2 Mouse models ..... 29**

2.2.1 *Hh*AOM 129SvEv.RAG$^{-/-}$ ..... 29

2.2.2 GSK484 (PAD4 inhibitor) ..... 30

2.2.3 Culturing Hh ..... 30

**2.3 Histology ..... 31**

2.3.1 Definitions of proximal versus distal colon ..... 31

2.3.2 Formalin-fixed and paraffin-embedded (FFPE) swiss rolls ..... 31

2.3.3 Methacarn fixation ..... 32

2.3.4 Histopathological analysis ..... 32

**2.4 Cell isolation ..... 33**

2.4.1 Cell isolation from the spleen ..... 33

2.4.2 Cell isolation from the colon ..... 34

2.4.3 LPL supernatant generation ..... 34

2.4.4 Flow cytometry ..... 35

2.4.5 Surface Staining ..... 35

2.4.6 Acquisition and analysis ..... 36

**2.5 Gene expression analysis ..... 36**

2.5.1 Tissue RNA and DNA extraction ..... 36

2.5.2 qRT-PCR ..... 37

2.5.3 RNA sequencing ..... 38

2.5.4 RNA *in silico* analysis ..... 38

**2.6 Stool sampling ..... 38**

2.6.1 Faecal DNA extraction ..... 38

2.6.2 Faecal supernatant extraction ..... 39

**2.7 Bacterial DNA analysis ..... 39**

2.7.1 Quantification of bacterial colonisation ..... 39

2.7.2 Quantification of *Hh* colonisation ..... 40

2.7.3 16S rRNA Sequencing ..... 40

2.7.4 16S rRNA *in silico* analysis ..... 40

**2.8 Analysis of protein expression ..... 41**

2.8.1 Luminex assay ..... 41

2.8.2 Enzyme-linked immunosorbent assays ..... 41

2.8.3 Proteomics on patient CRC samples ..... 41

**2.9 *In situ* analysis ..... 42**

2.9.1 Immunohistochemistry and HALO quantification ..... 42

2.9.2 Immunofluorescence ..... 43

2.9.3 BugFISH ..... 44

**2.10 Statistical analysis ..... 45**

3 INVESTIGATION OF HOST-MICROBE INTERACTIONS IN A GENETICALLY-SUSCEPTIBLE COLITIS-ASSOCIATED CANCER MODEL ..... 46

**3.1 Introduction ..... 46**

**3.2 Aim ..... 47**

**3.3 Results ..... 47**

3.3.1 Colitis-associated cancer (CAC) occurs in the distal colon in the *Hh*AOM model ..... 47

3.3.2 Tumours are enriched in neutrophil-associated markers. ..... 51

3.3.3 Neutrophil accumulation precedes CAC. ..... 53

3.3.4 Immature and mature neutrophils are increased during colitis and cancer ..... 63

3.3.5 S100A8, IL-1β, IL-17A and IL-22 are increased in the distal colon of HhAOM mice ..... 71

**3.4 Discussion ..... 76**

3.4.1 The distal colon niche supports CRC. ..... 76

3.4.2 The ambiguous role of neutrophils in cancer ..... 77

3.4.3 Neutrophil subsets increase over the course of disease. ..... 78

3.4.4 The HICCS locus drives neutrophil influx ..... 79

3.4.5 A distinct cytokine network in the distal colon ..... 80

**3.5 Conclusion ..... 84**

**3.6 Future outlook and limitations ..... 84**

4 IDENTIFICATION OF THE ENZYME PAD4 AS A POTENTIAL CANCER PROMOTING ENZYME IN CRC. ..... 86

**4.1 Introduction ..... 86**

**4.2 Aim ..... 87**

**4.3 Results ..... 88**

4.3.1 118 genes are differentially expressed in tumorigenesis as compared to colitis ..... 88

4.3.2 *Padi4* mRNA and protein is increased in distal colons of HhAOM mice. ..... 93

4.3.3 Epithelial PAD4 expression coincides with the loss of surface marker EpCAM ..... 98

4.3.4 Comparable expression pattern of Lgr5 and PAD4. ..... 105

4.3.5 *In vivo* PAD4 inhibition results in a decreased tumorigenesis. ..... 106

4.3.6 Lack of PAD4 expression in *Hh*AOM colons treated with anti-IL22. ..... 107

4.3.7 Variable *PADI4* expression in human CRC ..... 109

**4.4 Discussion ..... 110**

4.4.1 PAD4 expression in DNA-damaged epithelium of *Hh*AOM mice. ..... 111

4.4.2 The enzyme PAD4 augments tumorigenesis. ..... 112

4.4.3 Preliminary evidence of PAD4-driven stemness. ..... 114

**4.5 Conclusion ..... 117**

**4.6 Future outlook and limitations ..... 117**

5 THE ROLE OF *HH* IN SHAPING DYSBIOSIS-DRIVEN CANCER ..... 120

**5.1 Introduction ..... 120**

**5.2 Aim ..... 121**

**5.3 Results ..... 121**

**5.4 Discussion ..... 133**

5.4.1 *Hh*-driven dysbiosis is altered during tumorigenesis. ..... 134

5.4.2 The tumour microenvironment is enriched with *Hh* ..... 134

5.4.3 *Hh* does not trigger disease in a gnotobiotic setting. ..... 135

**5.5 Conclusion ..... 136**

**5.6 Future outlook and limitations ..... 136**

6 GENERAL DISCUSSION ..... 138

**6.1 Summary ..... 138**

**6.2 New insights into how inflammation and epigenetics impact CRC progression ..... 139**

**6.3 The importance of TME characterisation and patient stratification ..... 141**

**6.4 Deciphering the role of the microbiome in CRC ..... 143**

**6.5 Conclusion ..... 144**

7 REFERENCES ..... 145

8 APPENDIX ..... 160

**8.1 Link to R Code used for analysis ..... 160**

**8.2 Table of top 50 genes* DESeq (healthy distal colons vs tumours) ..... 160**

**8.3 Table of neutrophil genes* DESeq (untreated distal colons vs tumours) ..... 163**

**8.4 Table of 118 candidate genes* (Tumour versus Hh 15 weeks) ..... 165**

**8.5 Table of 118 candidate genes* (HhAOM distal colon 10 weeks versus Hh 10 weeks) ..... 170**

# List of abbreviations

| | |
|---|---|
| **Alpk1** | Alpha Kinase 1 |
| **AMP** | Antimicrobial Peptide |
| **AOM** | Azoxymethane |
| **APC** | Antigen Presenting Cell |
| **APC** | Adenomatous Polyposis Coli Gene |
| **Areg** | Amphiregulin |
| **BCR** | B Cell Receptor |
| **BMP** | Bone Morphogenetic Protein |
| **BRAF** | B-raf Proto-Oncogene |
| **CAC** | Colitis-Associated Cancer |
| **CAF** | Cancer associated Fibroblast |
| **CD** | Cluster of Differentiation |
| **CDT** | Cytolethal Distending Toxin |
| **CFU** | Colony Forming Units |
| **CMS** | Consensus Molecular Subtypes |
| **cNK** | Conventional Natural Killer |
| **cpm** | Counts per Million |
| **CRC** | Colorectal Cancer |
| **CRIS** | Colorectal Cancer Intrinsic Subtypes |
| **CSC** | Cancer Stem Cell |
| **Cxcr** | Chemokine |

| | |
|---|---|
| **DAMP** | Damage-Associated Molecular Pattern |
| **DC** | Dendritic Cell |
| **DSS** | Dextran Sulfate Sodium |
| ***E. coli*** | *Escherichia coli* |
| **EEC** | Enteroendocrine cells |
| **EGF** | Epidermal Growth Factor |
| **ELISA** | Enzyme-linked Immunosorbent Assay |
| **EMT** | Epithelial-Mesenchymal Transition |
| **EpCAM** | Epithelial Cell-Adhesion Molecule |
| **ETBF** | Enterotoxigenic *Bacteroides fragilis* |
| **FACS** | Fluorescence Activated Cell Sorting |
| **FAP** | Familial Adenomatous Polyposis |
| **FBS** | Fetal Bovine Serum |
| **FISH** | Fluorescence *in situ H*ybridization |
| **G-CSF** | Granulocyte Colony-Stimulating Factor |
| **GEMM** | Genetically-Engineered Mouse Models |
| **GF** | Germ-free |
| **GMP** | Granulocyte-Monocyte Precursors |
| **H&E** | Haematoxylin & Eosin |
| **HGF** | Hepatocyte Growth Factors |
| ***Hh*** | *Helicobacter hepaticus* |
| **HICCS** | *Hh*-induced colitis and associated cancer susceptibility |
| **i.p.** | Intraperitoneal |

| | |
|---|---|
| **IBD** | Inflammatory Bowel Disease |
| **IEC** | Intestinal Epithelial Cell |
| **IEL** | Intraepithelial lymphocyte |
| **IFN** | Interferon |
| **IHC** | Immunohistochemistry |
| **IL** | Interleukin |
| **IL-10R2** | IL-10 β Receptor Rhain |
| **IL-22RA1** | IL-22 Receptor α Chain |
| **ILC** | Innate Lymphoid Cell |
| **KRAS** | Kirsten Rat Sarcoma Virus |
| **Lcn2** | Lipocalin 2 |
| **Lgr5** | Leucine-rich repeat-containing G-protein coupled receptor |
| **MHC** | Major Histocompatibility Complex |
| **MMP** | Matrix Metallopeptidase |
| **MMR** | Mismatch Repair |
| **MPO** | Myeloperoxidase |
| **MSI** | Microsatellite Instable |
| **MSS** | Microsatellite Stable |
| **Muc2** | Mucin-2 |
| **NET** | Neutrophil Extracellular Trap |
| **NLR** | Nuclear Oligomerisation Domain- (NOD-) Like Receptors |
| **NO** | Nitric Oxide |
| **OMM** | Oligo-Mouse-Microbiota |

| | |
|---|---|
| **OSM** | Oncostastin-M |
| **PAD** | Peptidylargine Deiminase |
| **PAD4** | Peptidylargine Deiminase 4 |
| **PAMP** | Pathogen-Associated Molecular Pattern |
| **PBS** | Phosphate Buffered Saline |
| **PCA** | Principal Component Analysis |
| **pks** | Polyketide Synthase |
| **PMN** | Polymorphonuclear Granulocytes |
| **PMN-MDSC** | Polymorphonuclear Myeloid-Derived Suppressor Cell |
| **PRR** | Pattern Recognition Receptor |
| **RAG** | Recombination Activating Gene |
| **RegIIIγ** | Regenerating Islet-Derived Protein IIIγ |
| **RLR** | RIG-I-like Receptor |
| **Rorγt** | RAR-related Orphan Receptor-γt |
| **ROS** | Reactive Oxygen Species |
| **RT-qPCR** | Real Time Quantitative Polymerase Chain Reaction |
| **RVC** | Royal Veterinary College |
| **SEM** | Standard Error of the Mean |
| **SIgA** | Secretory immunoglobulin A |
| **SPF** | Specific Pathogen-Free |
| **SSA** | Sessile Rerrated Adenomas |
| **T-bet** | T-box transcription Factor |
| **TAN** | Tumour-Associated Neutrophils |

| | |
|---|---|
| **TCR** | T Cell Receptor |
| **TGF** | Transforming Growth Factor |
| **TH** | Helper T cell |
| **TLR** | Toll-Like Receptor |
| **TME** | Tumour Microenvironment |
| **TNF** | Tumour Necrosis Factor |
| **Treg** | Regulatory T cell |
| **TSLP** | Thymic Stromal Lymphopoietin |
| **UC** | Ulcerative Colitis |
| **UV** | Ultraviolet |
| **VEGFA** | Vascular Endothelial Growth Factor A |
| **WNT** | Portmanteau of genes *Wingless* and *Int-1* |

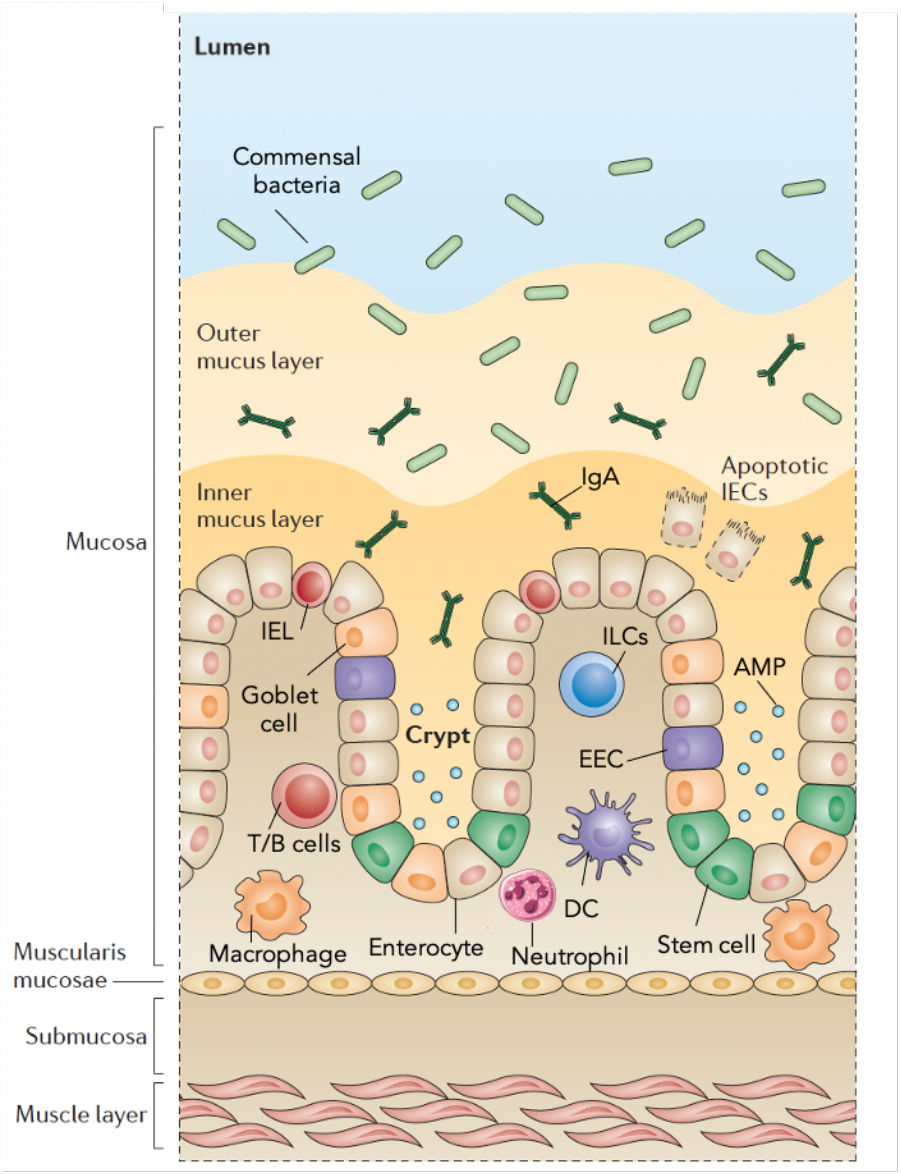
# 1 Introduction

The intestine is home to a large repertoire of immune cells alongside a substantial number of commensal microbes, dietary antigens and potential pathogens[1,2]. It is therefore essential that a balance between immune tolerance and immune response is maintained. Indeed, dysregulation of this host-microbiota interface can contribute to severe inflammatory disorders, one of which is colorectal cancer (CRC)[3].

## 1.1 The host-microbiota interface

### 1.1.1 Colon architecture

The colon is a continuous tube beginning with the ascending (proximal) colon and followed by the transverse colon, the descending (distal) colon, the rectum and the anus. The colon lumen is the main reservoir for trillions of commensal microbes, separated solely from the host lamina propria by a single layer of mucus producing epithelial cells and a thin basement membrane. The lamina propria is composed of connective tissue containing the blood and nervous supply, lymph drainage and most cells of the immune system (**Figure 1.1**). A thin muscle layer underlies the lamina propria ('muscularis mucosa') and is followed by an area of connective tissue known as the submucosa in which a network of parasympathetic nerves lies. Finally, a thicker layer of muscle cells and a serosa provide a compact covering separating the intestine from the peritoneal cavity [4].




**<u>Figure 1.1:</u> Anatomy of the colon**

Schematic of the colon illustrating the major components of tissue architecture. The microbiota resides in the lumen of the colon segregated from the underlying host tissue by two mucus layers. A layer of intestinal epithelial cells (IECs) functions as a barrier to the lamina propria, consisting of stem cells as well as differentiated cells such as enterocytes, goblet cells and enteroendocrine cells (EEC). IECs secrete factors such as mucus and antimicrobial peptides (AMPs) into the lumen and chemokines and cytokines into the lamina propria - in which a variety of immune cells are found. The underlying submucosa mainly contains non-haematopoietic cells.

DC, dendritic cell; IEL, intraepithelial lymphocytes; SIgA, secretory immunoglobulin A.

(Figure adapted from Mowat and Agace, 2014[4])

### 1.1.2 The gut microbiota

Many aspects of intestinal physiology are influenced by complex communities of bacteria, fungi, viruses and archaea residing in the intestinal lumen and collectively termed the 'gut microbiota'. Most studies have focused on investigating bacterial communities due to early advancements in 16S-rRNA gene sequencing and have characterised over 1000 different species in a 'healthy' human gut microbiome[5]. Even though individuals are dominantly colonised by the two phyla of Bacteroidetes and Firmicutes, any one individual only carries an average of approximately 160 out of the 1000 species[5]. Thus, a vast intra-individual heterogeneity makes the species-level definition of one universal healthy microbiome practically impossible. However, studies are now deciphering the metabolic output of intestinal communities, hypothesising that the "functional core" will be similar in different healthy colons but will not necessarily be provided by the same organisms[5]. These functions can range from nutrient breakdown and storage to immune modulation and pathogen restriction due to niche competition[2].

### 1.1.3 The epithelial barrier

Despite consisting of a single-cell layer, the intestinal epithelium can withstand repeated mechanical and biological insults. The large surface area of the crypt enables cells to act not only as a physical barrier, but also as an active contributor to nutrient uptake and defence[6]. The epithelial barrier is continuously renewed by multipotent stem cells (or crypt base columnar cells; CBC) residing in the base of intestinal crypts, an area that is protected from bacterial metabolites by the invagination of the intestinal wall[7]. Stem cells give rise to rapidly proliferating progenitor, or transit-amplifying, cells that mature into differentiated intestinal

epithelial cells (IECs) dictated by the surrounding intestinal stem cell niche. Other epithelial and mesenchymal cells within this niche provide key signals that regulate cellular maturation, including WNT (portmanteau of genes *Wingless* and *Int-1*), Notch, epidermal growth factor (EGF) and bone morphogenetic protein (BMP) signalling[6].

To date, six different mature IEC types have been defined, classified as absorptive (enterocytes and M-cells) or secretory (Paneth, goblet, enteroendocrine and tuft cell) lineages[6]. Of note, Paneth cells are absent in the colon and instead a greater number of goblet cells are found in comparison to the small intestine[8]. More recent literature also refers to enterocytes of the colon as 'colonocytes'[9].

The high microbial density in the colon makes goblet cells indispensable due to their mucin-secreting function. Two layers of protective mucus line the colon, one dense and adherent to the epithelial barrier and another loose layer populated by commensal microbes. It is widely hypothesised that the dense layer serves as a first line of defence whereas the second provides an energy source for bacteria[10]. Mice lacking goblet cell-derived mucin-2 (Muc2) spontaneously develop severe inflammatory disorders such as colitis, illustrating the importance of an intact mucus layers[11]. Similarly, the secretion of antimicrobial peptides (AMPs) also reinforces an intact intestinal barrier. Colonocytes are capable of producing AMPs, such as the C-type lectin regenerating islet-derived protein IIIγ (RegIIIγ), however the Paneth cells of the small intestine are the IECs uniquely adapted to secrete a variety of AMPs including defensins, cathelicidins and lysozyme[12,13].

Importantly, IECs are also able to sense the microbiota directly by expressing pattern recognition receptors (PRRs) including toll-like receptors (TLRs), nuclear oligomerisation

domain- (NOD-) like receptors (NLRs) and RIG-I-like receptor (RLRs). Stimulation of these receptors by pathogen-associated molecular patterns (PAMPs) or host-derived damage-associated molecular patterns (DAMPS) promotes epithelial barrier repair, upregulation of tight junction proteins and triggers production of cytokines that signal to and attract immune cells of the underlying lamina propria[13].

### 1.1.4 Immune responses in the colon

Following cellular damage caused by infection or cell death ('sterile' injury), the hosts' biological reaction is the complex and well-coordinated response of inflammation[14]. If, for example, harmful microbes breach the epithelial host barrier, a fine-tuned interplay of the innate and the adaptive immune system results in pathogen recognition and elimination with subsequent initiation of tissue regeneration and healing[15,16].

#### *1.1.4.1 Innate Immunity*

Evolutionarily conserved innate immune cells comprise an extensive range of myeloid and lymphoid cell types that derive from a haematopoietic stem cell and which function as the first line of defence following barrier breach. Comparable to IECs, innate cells express PRRs to detect PAMPs and DAMPs as well as respond to cytokines secreted by other cell types. Subsequent responses to restrain pathogen expansion are mediated through cell-dependent mechanisms (i.e., phagocytosis and/or killing) or by secreted factors including AMPs, cytokines, alarmins and a range of other molecules[17].

#### 1.1.4.1.1 Neutrophils

Neutrophils are a subset of myeloid cells that develop from granulocyte-monocyte precursors (GMPs) in the bone marrow and rapidly migrate to distant tissue upon microbial challenge. They exit the bone marrow and enter the peripheral blood when decreased chemokine receptor 4 (CXCR4) expression coincides with the increase of chemokine receptor 2 (CXCR2) signalling. Upregulation of CXCR4 on 'aged' neutrophils in the periphery then results in neutrophil homing back to the bone marrow to be eliminated by phagocytosis[18]. Neutrophils are professional phagocytes that produce antimicrobial peptides alongside reactive oxygen and nitrogen species to destroy invading pathogens[19,20].

Historically, neutrophils were defined to undergo a multitude of sequential maturation steps ranging from committed mitotic pools of precursors, to a non-proliferative pool of banded cells up to segmented neutrophils[19]. Recently, single cell sequencing and cytometry studies have redefined neutrophil pre-cursors[21-23]. On this basis, a working model has been proposed clustering mouse and human neutrophils into three subsets: a proliferative and committed pre-neutrophil subset alongside non-proliferative immature and mature neutrophil subsets[19]. The earliest neutrophil progenitors originate from a granulocyte- monocyte progenitor pool (GMP) and are named 'early neutrophil progenitors' in humans and 'proneutrophils' in mice. These proliferative cells give rise to myelocytes and metamyelocytes which in turn bring about immature neutrophils (banded) and mature neutrophils[24]. Neutrophil maturation is largely driven by granulocyte colony-stimulating factor (G-CSF) signalling due to deficiency in this receptor resulting in neutropenia[25,26]. However, other mediators include granulocyte-macrophage stimulating factor (GM-CSF) and IL-6[20,27,28]. The retention and release of neutrophils is largely regulated by the chemokine CXCL12 and its corresponding receptor

CXCR4 expressed by neutrophils. Thus, downregulation of CXCR4 leads to the release of neutrophils from the bone marrow into the periphery[29]. This release is not constant, but fluctuates according to the circadian rhythm, with most neutrophils mobilised during active states, i.e. the day time in humans and night time in nocturnal mice[18,30]. Notably, the production of neutrophils in the bone marrow is linked to the apoptosis of tissue neutrophils in a feedback loop. Phagocytosis of dead neutrophils leads to an increase of the cytokine IL-23 at distant sites[31]. This secretion in turn triggers the production of IL-17A from a range of T-cells and innate lymphoid 3 cells (ILC3s)[32] which itself stimulates G-CSF production[33]. Neutrophil maturation predominantly takes place in the bone marrow, however extramedullary haematopoiesis can occur in the spleen and liver, often during active immune responses to pathogens[34,35]. At the end of their life span, neutrophils upregulate CXCR4 and home back to the bone marrow for clearance by stromal macrophages[36,37]. However, it has also been shown that aged neutrophils can be cleared by spleen and liver[38,39].

Circulating neutrophils express high levels of CXCR1 and CXCR2[40]. In humans, seven different corresponding chemokines are known to attract neutrophils to peripheral sites of infection, whereas only three analogues CXCL1, CXCL2 and CXCL5 exist in mice[40]. During an immune response, neutrophils must extravasate along a chemokine gradient to encounter the invading pathogen. In brief, the mechanism consists of capture by endothelial cells with subsequent cell rolling (mediated by P-, E- and L-selectins as well as P-selectin glycoprotein ligand 1), activation (mediated by chemokines such as CXCL4 and CXCL5), arrest (mediated by among others, β1-integrin VLA4 and the β2-integrin LFA1) and subsequent trans-endothelial migration (mediated by among others, intercellular adhesion molecule 1, junction adhesion molecules and platelet/endothelial-cell adhesion molecule 1)[41].

Peripheral neutrophils are heterogeneous and have recently been broadly classified into immature, mature, aged and interferon-signature bearing subsets[42] (**Figure 1.2).** Studies are now expanding upon this to investigate how defined neutrophil subsets differ at a functional level. Evrard and colleagues revealed immature and mature neutrophils to have a similar migratory capacity, but mature neutrophils to be more prone to phagocytosis and ROS production[21]. Furthermore, neutrophil aging triggered by the microbiota correlated with enhanced neutrophil extracellular trap (NET) formation (NETosis), a crucial function for pathogen entrapment and killing[43]. NETs are web-like structures of chromatin and cytosolic and/or granule proteins such as AMPs which are released by neutrophils upon cell death or non-lytic expulsion[44]. Further, a very recent study has identified circulating neutrophils in mice and humans with a set of interferon-stimulated genes and postulated these to be neutrophils primed to fight infection[45]. Notable evidence also supports neutrophils to be transcriptionally active with tissue-derived signals imprinting phenotypic and functional properties[46]. Most recently, a study identified key genomic regulators of neutrophils, pinpointing transcription factors such as RUNX1 in controlling neutrophil maturation, RELB affecting neutrophil survival and IRF5 governing neutrophil effector functions[47].

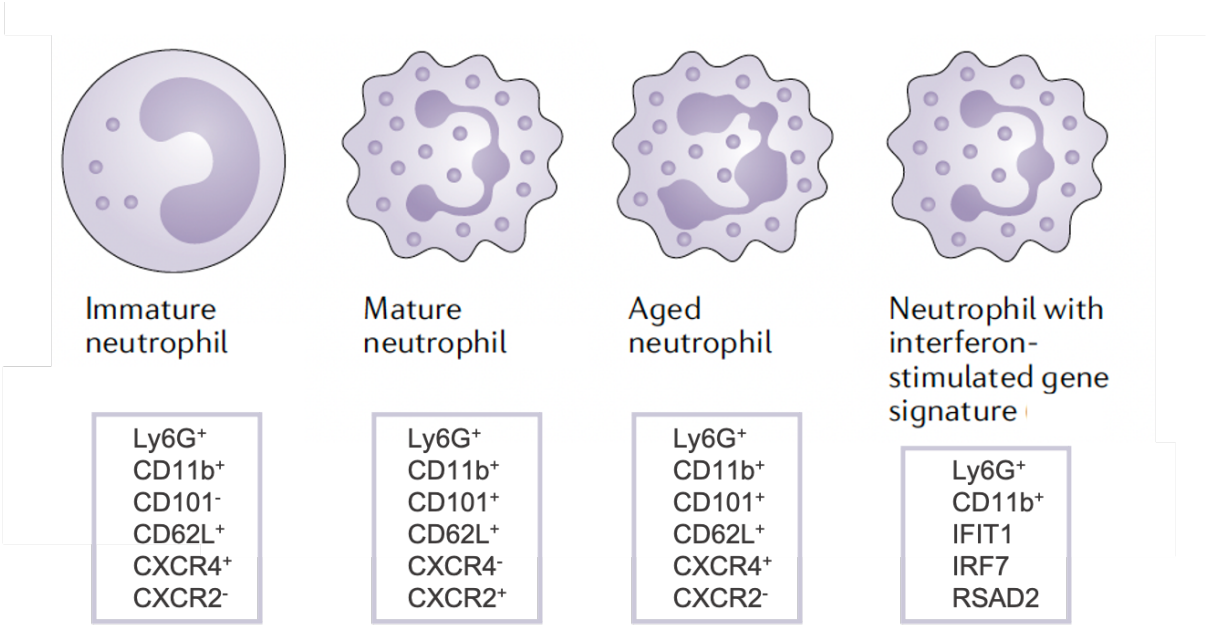



**Figure 1.2 Neutrophil subsets**
Due to recent transcriptomic and proteomic advances, peripheral neutrophils are now broadly classified into immature, mature, aged and interferon-signature bearing subsets with distinct sets of phenotypic markers.

(Figure adapted from Jaillon et al., 2020[42])

#### 1.1.4.1.2 ILC3s

Innate lymphoid cells (ILCs) are a more recently discovered group of immune cells which are enriched at parenchymal tissues, especially at mucosal surfaces such as the intestine[48]. They have four main features which are the absence of antigen receptors dependent on the recombination activating gene (RAG), a lack of myeloid cell markers, a typical lymphocyte morphology and development from the common lymphoid progenitor. While lacking antigen-specific receptors, ILCs react to pathogens indirectly by sensing soluble mediators and in turn secrete cytokines[48]. They play important roles in tissue homeostasis, epithelial maintenance and tertiary lymphoid structure formation[49]. The first ILC to be discovered was the conventional natural killer cells (cNK)[50], but since then three ILC subsets (ILC1, ILC2 & ILC3) resembling $CD4^+$ T 'helper' ($T_H$) cell subsets were added **(Figure 1.3)**. ILC1s are analogous to

$T_H1$ cells as they depend on the transcription factor T-bet, whereas GATA3$^+$ ILC2s represent $T_H2$ cells and RORγt$^+$ ILC3s are analogous to $T_H17$ cells[48,51].

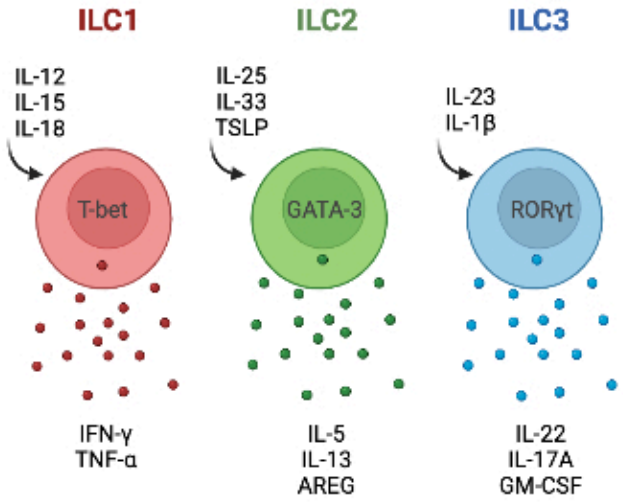



**Figure 1.3 ILC subsets**
Innate lymphoid cells are categorised in three subsets analogous in transcription factors and expressed cytokines to CD4$^+$ T 'helper' ($T_H$) cell subsets. ILC1s are dependent on the transcription factor T-bet, whereas ILC2s on GATA3$^+$ and ILC3s on RORγt$^+$.

ILC, innate lymphoid cell; IL, interleukin; IFN, interferon; GM-CSF, Granulocyte-macrophage colony-stimulating factor, TNF, Tumour necrosis growth, AREG, amphiregulin; TSLP, lymphopoietin

(Created with BioRender.com)

ILC3s play crucial roles in defence against microbial and fungal stimuli at mucosal barrier sites. They are predominantly activated by IL-23, IL-1β and in turn secrete a range of cytokines including IL-22, IL-17A, and GM-CSF[49]. The function of ILC3s in disease has been demonstrated in different infection models with contrasting results. IL-22 secretion by ILC3s promoted resistance against *Citrobacter rodentium* and Rotavirus infections[52,53], however colonisation with the pathobiont *Helicobacter hepaticus* (*Hh*) resulted in aggravated intestinal inflammation driven by IL-17A and GM-CSF[54,55]. A similar complexity is observed during tumorigenesis and discussed in the upcoming chapter sub-sections.

#### 1.1.4.1.3 Other innate cells

Other innate cells are also crucial in orchestrating an appropriate immune response. Mononuclear phagocytes (which consist of macrophages and dendritic cells) are the most abundant leukocytes in the healthy intestine. They have important functions ranging from cytokine production and phagocytosis to the uptake and presentation of antigen (APCs) and subsequent migration to lymph nodes to instruct the adaptive immune compartment[4,56]. It is widely accepted; intestinal macrophages are continuously replenished by infiltrating blood monocytes that differentiate under the influence of local, microbial stimuli[57]. However, this paradigm has been challenged with more recent research showing a subset of intestinal macrophages to be maintained locally and independent from the monocyte pool[58]. Further, in addition to neutrophils, polymorphonuclear granulocytes (PMN) consist of eosinophils, basophils and mast cells which play important roles in allergic reactions or parasite and other infections[4].

### 1.1.5 Adaptive Immunity

The adaptive immune system consists of antigen-specific B and T lymphocytes. All lymphocytes are equipped with highly diverse antigen receptors generated by random recombination of gene segments encoding the different sections of T-cell antigen receptors (TCR) and B-cell antigen receptors (BCR)[59]. The recombinases RAG1 and RAG2 (encoded by the *Rag1* and *Rag2* genes) are critical in the genetic recombination of BCR and TCR gene segments and the differentiation of immature lymphocytes into competent mature cells capable of antigen recognition. Thus, mice lacking either *Rag* gene fail to develop a repertoire

of antigen-specific B and T cells capable of cycling between blood, the lymphatic system and peripheral tissues in search for cognate signals[56,60]. Suitable antigen triggers the TCR and BCR on naïve lymphocytes leading to cell activation, clonal expansion and memory cell generation[56]. The latter results in highly specific and long-lasting immunity against encountered pathogens[61].

The predominant pool of peripheral T cells is represented by two lineages, expressing the co-receptors CD8 or CD4, which recognise antigen presented on major histocompatibility complex (MHC) class I or II, respectively[62]. MHCI is thereby expressed by all nucleated cells whereas MHCII is solely expressed by professional APCs[63]. Upon activation, cytotoxic $CD8^+$ T cells kill other cells infected with pathogens, whereas $CD4^+$ T cells provide helper signals in the form of cytokines and regulatory T cells prevent autoreactive immune responses[56,64]. Upon antigen encounter, B cells differentiate into antibody-secreting plasma cells which produce high amounts of antibodies of a certain specificity[56]. In particular the secretion of bacteria coating immunogloblin A (IgA) is of great importance in the homeostasis of the intestinal mucosa[65].

## 1.2 The aetiology of CRC and CAC

Colorectal cancer (CRC) is a disease of intestinal stem cells and is characterised by the multifaceted dysregulation of intestinal components including immune cells[3]. Inherited genetic conditions, such as familial adenomatous polyposis (FAP) or Lynch syndrome, account for only 2-5% of CRC patients[66]. In FAP, patients inherit an autosomal dominant germline

mutation in the adenomatous polyposis coli (APC) gene, whereas germline mutations in mismatch repair (MMR) genes drive Lynch syndrome[67].

The majority of CRCs are hypothesised to be triggered by sporadic factors, possibly on the background of genetic susceptibility. Indeed, up to 90% of disease risk is attributed to environmental factors, including low fibre diets or high red meat intake[68,69]. Of relevance to this thesis, a small subset of about 1-2% of CRC also arises in the context of chronic inflammation. For example, patients with ulcerative colitis have a 2.4-fold increase in developing colitis-associated cancer (CAC)[70]. Indeed, the pathology of ulcerative colitis (UC) is characterised by long-lasting inflammation and a disrupted barrier[71], both of which are also considered to shape colon tumorigenesis[3]. However, despite this knowledge the exact aetiology of UC is still unknown with genetic susceptibility and environmental factors considered risk factors[71].

Regardless of what underlies initiation, colon tumorigenesis depends upon the succession of genomic alterations in intestinal stem cells, referred to as the 'adenoma-to-carcinoma' sequence[72]. Pre-malignant lesions termed polyps or adenomas evolve into invasive adenocarcinomas by the accumulation of driver mutations which grant cells a selective advantage[72,73]. A range of such driver mutations have been characterised, amongst which tumour suppressor genes *APC* and *TP53* alongside oncogenes *KRAS*, *PIK3CA* and *BRAF* are the most common[73,74].

In 2015, a large consortium project advanced our understanding of the genetic and molecular heterogeneity in colorectal cancer by defining four consensus molecular subtypes (CMS)

based on transcriptomic signatures of bulk RNA-sequencing of tumour tissue[75]. CMS1 is characterised by microsatellite instability (MSI) and immune activation (14%), CMS2 by epithelial WNT and MYC-signalling (37%), CMS3 by metabolic de-regulation (13%) and CMS4 by angiogenesis and stromal invasion. The remaining 13% represent an unclassified group with mixed features. Each CMS subgroup was shown to have a different prognostic value with CMS4 tumours presenting the worst prognosis[75]. In a similar vein, a Colorectal Cancer Intrinsic Subtypes (CRIS) classification system was published in 2017 which clusters patients based on their intrinsic epithelial expression profile rather than the bulk tumour profiling which CMS is based upon [76].

## 1.3 The tumour microenvironment

As illustrated by the aetiology and CMS classification above, CRC is not solely driven by intrinsic alterations of stems cells but also by extrinsic signals provided by various neighbouring cell types. The importance of a tumour microenvironment (TME) was already postulated over a century ago by the fundamental seed (cancer cells) and soil (environment) hypothesis[77]. Today, it is well appreciated that many of the core hallmarks of cancer, including proliferation, angiogenesis and invasion by epithelial-to-mesenchymal transition (EMT)[78,79], are promoted by the interaction of malignant cancer cells with surrounding non-transformed cells and the extracellular matrix (ECM)[80]. Three over-arching groups of cells have been characterised, namely angiogenic vascular cells (i.e. endothelial cells and pericytes), infiltrating immune cells and cancer-associated fibroblasts (CAFs)[78]. Despite all having crucial roles in CRC progression, this thesis has focused on neighbouring immune cells and their interactions with the epithelium during inflammation and CRC.

### 1.3.1 Immune cells in the TME – a double-edged sword.

The presence of immune cells has been long known in cancer with the historic belief being that they are protective through tumour surveillance and antitumour immunity[81]. To this end, somatic mutations in cancer are known to result in the expression of tumour-specific antigens (neoantigens) that can activate immunity and mediate rejection – especially in early stages of disease. This is predominantly executed by IFNγ$^+$ T cells and natural killer (NK) cells[82]. Similar to foreign microbial products, processed tumour antigen is cross-presented by specialised antigen-presenting myeloid cells and engagement with cognate CD8$^+$ T cells induces cytotoxic attacks[82]. In CRC, the numbers of *in situ* T cells indeed correlates with a beneficial patient outcome[83]. Importantly, the functionality of conventional T cells depends on priming by antigen presenting cells, whereas NK cells require no prior activation to kill. NK cells thereby recognize the downregulation of MHCI which occurs in cancer or infection settings and respond by cell killing[84]. Similarly, gamma delta (γδ) T cells, which are distinct from conventional T cells harbouring a TCR consistent of an alpha (α) and beta (β) chain, recognize ligands independent from antigen-processing or MHC presentation. Once activated, these cells also have the capacity to be potent tumour cell killers[85].

Today, many novel therapeutic strategies harness an antigen-specific elimination strategy, but a majority of patients remain resistant[86], especially in CRC [87]. For example, tumour cells can become undetectable to the immune system and escape elimination by immunoediting their tumour antigens or upregulating checkpoint molecules[88]. Similarly, although phagocytic neutrophils and macrophages are crucial accomplices in cancer cell cytotoxicity and tumour

clearance, they can be hijacked by the tumours to suppress anti-tumour responses[42,89,90]. T cells and/or ILCs may also not always be protective, with evidence that $T_H17$-associated cytokines can drive many hallmarks of cancer[91]. Hence, immune cells within a TME have disparate roles which are often modified by the evolving tumour. Along these lines, previously introduced mechanisms that are crucial in colonic tissue healing and host protection (**1.1.3**), such as proliferation or cytokine secretion, can promote tumour progression and invasion in late-stage disease when out of control[74,92]. This is often mediated by complex cellular networks[74] of which the most thesis relevant are explained in more detail later.

### 1.3.2 The ambiguous role of tumour-associated neutrophils

Neutrophils are among the immune cells with the highest correlation with adverse outcomes in lung, colon and pancreatic cancer[93]. Indeed, tumour-associated neutrophils (TANs) have been shown to affect genetic stability, proliferation, angiogenesis and the suppression of adaptive immune cells[42] (**Figure 1.3**).

The secretion of high quantities of ROS is a fundamental property of neutrophils[94] which is critical in antimicrobial host defence. However, ROS can also lead to DNA damage and genetic instability in cancer[95]. ROS and neutrophil-derived inducible nitric oxide synthase (iNOS)-dependent nitric oxide (NO) are also able to suppress T-cell activation in models of breast cancer[96,97]. TANs can additionally secrete molecules associated with EMT and therapy-resistance, such as hepatocyte growth factors (HGFs)[98,99] and angiogenesis-promoting vascular endothelial growth factor A (VEGFA)[100,101]. Moreover, NETs have been observed in a range of tumour types including cancers of the small intestine[102-105]. A recent study utilising

intravital imaging revealed that NETs physically barricade tumour cells from cytotoxic T and NK cells in a melanoma model[105].

Neutrophils and NETs play essential roles in shaping the metastatic niche. Notably, proteases contained in NETs can trigger extracellular remodelling and awaken dormant cancer cells during lung inflammation by cleaving ECM proteins[106]. Additionally, genetically-engineered mouse models (GEMMs) of metastatic mammary ('KEP' model) or metastatic intestinal cancer ('KPN' model) have revealed that neutrophils can drive metastasis by suppressing T cell responses[96,107]. In the KPN model, neutrophil-driven metastasis was shown to be largely dependent on TGFβ signalling[107]. This is in line with another pivotal publication identifying TGFβ as a major inducer of neutrophil-mediated T cell suppression in the TME of murine flank and orthotopic tumours[108].

Despite the numerous destructive properties of neutrophils, they remain crucial in anti-tumour immunity (**Figure 1.3**). In particular the secretion of ROS, NO, tumour necrosis factor (TNF) and TNF-related apoptosis inducing ligand (TRAIL) are well known mediators of tumour cell killing[42]**.** Furthermore, neutrophils can express T-cell recruiting CXCL10 mRNA[109] and are able to augment $CD4^+$ and $CD8^+$ T-cell proliferation in early non-small cell lung cancer[110,111]**.** Notably, anti-tumorigenic properties of T-cells can be amplified by IFNβ[112] or a combination of IFNγ and GM-CSF[110]. This correlated with human leukocyte antigen presenting (HLA)-DR and CD86 expression on neutrophils, which is a sign of enhanced antigen presentation with the required co-stimulation[110]. Additionally, neutrophils isolated from CRC biopsies increased proliferation and IFNγ secretion of TCR-stimulated $CD8^+$ T cells in co-culture experiments[113].

To date, immunosuppression is the key functional property utilised to divide TANs into protective 'N1' and destructive 'N2' neutrophils[42,108]. Of note, these debated terms as, well as other such as polymorphonuclear myeloid-derived suppressor cell (PMN-MDSC), describe immunosuppressive neutrophils with consensus on nomenclature outstanding[89]. Previously described immature and mature cells have both shown to bear immunosuppressive functions in certain contexts[42], however it was particularly $CD101^-$ immature neutrophils that positively correlated with tumour burden in an orthotopic pancreas tumour model[21]. In human and mouse lung cancer, a subset of neutrophils with a set of interferon stimulated genes was detected[114].

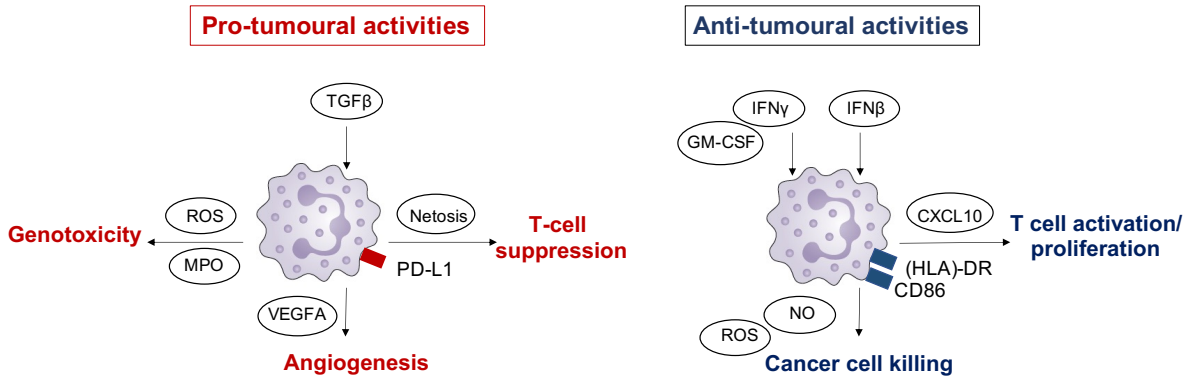



**Figure 1.3: Neutrophil effects on tumorigenesis**

Neutrophils have pro- and anti-tumoural functions within the TME. Neutrophils can initiate tumourigenesis by expressing DNA damaging ROS or MPO and further drive disease by promoting angionesis by VEGFA expression. Further, the cytokine TGFβ is a major inducer of pro-tumour neutrophil-mediated T cell suppression. In line with this, neutrophils can also express T cell supressing PD-L1. Anti-tumour functions can be are triggered by IFNβ or a combination of IFNγ and GM-CSF which include the attraction (CXCL10), proliferation and activation (HLA-DR, CD86) of T cells. The DNA damaging capabilities of neutrophil-derived ROS and NO are also beneficial as they aid in cancer cell killing.

IFN, interferon; GM-CSF, Granulocyte-macrophage colony-stimulating factor, TGF, Transforming growth factor, ROS, reactive oxygen species; MPO, myeloperoxidase; VEGFA, Vascular Endothelial Growth Factor; PD-L, Programmed death-ligand; CXCL, chemokine (C-X-C motif) ligand; HLA-DR, Human Leukocyte Antigen – DR isotype; CD, cluster of differentiation

(Figure adapted from Jaillon et al., 2020[42] and Hedrick et al., 2022[24])

Overall, it appears that the disease stage, tumour type and tissue context are key determinants of the functional properties of neutrophils in cancer. Dissecting the heterogeneity of neutrophils in the evolving TME is still on-going, with limited studies in primary CRC and even fewer probing longitudinally over time.

### 1.3.3 IL-22 in CRC – Friend or foe?

IL-22 is a key mediator of intestinal barrier integrity by promoting epithelial cell proliferation and survival alongside the secretion of crucial antimicrobial peptides and mucins[115,116]. However, this pleiotropic cytokine plays disparate roles in inflammation and CRC, possibly related to the stage of disease[116].

The cellular sources of IL-22 are of hematopoietic origin and include $CD4^+$ $T_H17$ cells, γδ T cells, ILC3s[115] **(1.1.4.1.2)** and most recently, neutrophils[117]. IL-22 signals through the heterodimeric IL-22 receptor (IL-22R) composed of the IL-22 receptor α chain (IL-22RA1) and the IL-10 β receptor chain (IL-10R2) to mediate leukocyte-epithelial crosstalk. While IL-10R2 is ubiquitously expressed, the IL-22RA1 chain is restricted to non-hematopoietic cells including the intestinal epithelium[116,118]. Upon binding, IL-22 predominantly signals via the JAK1/TYK2/STAT3[116] pathway which leads to the secretion of, for example, antimicrobial lectins RegIIIβ and RegIIIγ[119].

Although necessary for barrier integrity and repair, IL-22 was also shown to promote colitis by triggering the secretion of pro-inflammatory cytokines and augmenting hyperplasia[120]. Further, elevated IL-22 levels are also detected in the serum[121] and blood of IBD

patients[122,123], but whether this is cause or effect is not clear. IL-22 also plays contrasting roles in CRC. Murine models of disease have illustrated detrimental roles for IL-22 which are often linked to IL-22-driven proliferation of transformed cells during cancer progression[124,125]. In our lab, therapeutic blockade of IL-22 *in vivo* significantly reduced dysplasia in a bacterially-driven CAC model, hence suggesting therapeutic potential[124]. However, most recently, IL-22 signalling was found to be anti-tumorigenic through induction of essential repair cascades in DNA-damaged intestinal stem cells[126]. Full understanding of these contrasting functions is outstanding but temporal dynamics during the tumorigenesis process may be critical in determining the downstream functions of IL-22.[3].

In humans, a role for IL-22 in disease progression is still poorly understood. However, *IL22* mRNA is reportedly increased in CRC tumour sections relative to matched normal tissue[124]. Notably, our lab recently identified a cohort of poor prognosis CRC patients characterised by high *IL22RA1* expression and mutant *KRAS*[127]. More functionally, IL-22 has been shown to promote stemness in the human DLD-1 colon cancer cell line[128] and trigger heightened proliferation of DLD-1 cells specifically in a *KRAS* mutant but not the isogenic wild-type line[127].

## 1.4 A novel CRC hallmark on the horizon: Dysbiosis

Infections are a well-known environmental cause for cancer, with the detrimental role of *Helicobacter pylori* in gastric cancer first identified in 1991 and now a leading example for infection-triggered carcinogenesis[129,130]. Even though evidence for a similar pathogen is lacking in CRC, such observations have led to a growing interest in the role of the microbiota in CRC onset and progression.

Initial evidence for a role of the microbiota in cancer emerged decades ago when it was shown that germ-free rats administered the carcinogen dimethylhydrazine developed less tumours than animals with a microbiota[131]. Years later, mice gavaged with the faecal microbiome of CRC patients developed more intestinal polyps relative to healthy control stool which strongly suggested a functional role for a CRC microbiota[132]. Many questions are outstanding in the field, but due to advances in microbiome sequencing, it is irrefutable that individuals with CRC have a distinct taxonomic composition relative to healthy controls, termed 'dysbiosis'[133]. In general, CRC patients have an increased taxonomic diversity in faecal samples and an outgrowth of certain 'oncogenic' species, including *Fusobacterium nucleatum, Escherichia coli, Bacteroides fragilis, Enterococcus faecalis, Streptococcus gallolyticus and Peptostreptococcus spp.* Simultaneously, proposed protective genera such as *Roseburia, Clostridium, Faecalibacterium* and *Bifidobacterium* are decreased in CRC[1,134-137]. A study on early neoplasia of patients with Lynch Syndrome also highlighted microbial changes and identified a shift of potential immune-stimulating flagellin producers[134,138]. In particular, the levels of the traditionally oral bacterium *Fusobacterium* in CRC have been consistently correlated with poor clinical outcome[137] highlighting a biomarker potential.

Determining whether observed taxonomic changes are a cause or an effect of CRC and whether dysbiosis is a novel cancer hallmark is undoubtedly a work in progress. Various studies have begun to dissect how CRC-associated microorganisms and their metabolites shape disease progression more functionally. Often, these mechanisms underly maladaptations within the host-microbiota interface and include the interplay of many

cellular players within the TME as illustrated in **Figure 1.5**, including the epithelial, immune, and mesenchymal compartments.

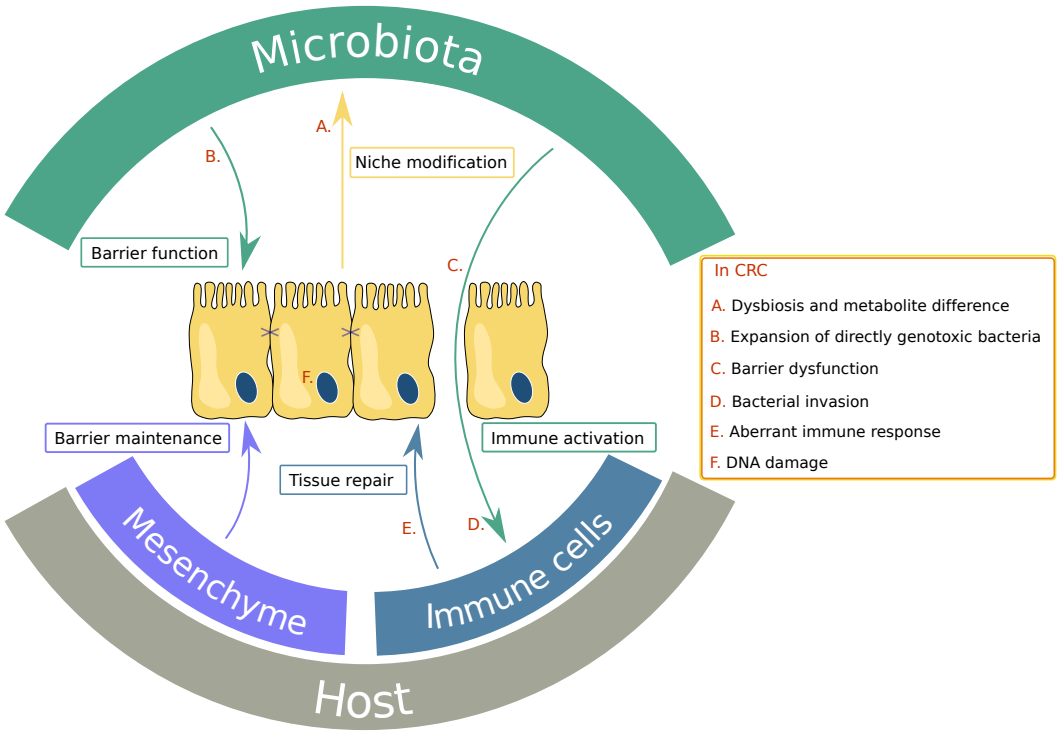



**Figure 1.5: The host microbiota interface in CRC**
The host-microbiota interface consists of a dynamic and complex microbiota which neighbours host mesenchymal and immune cells underlying a single layer of epithelium. In homeostasis, communication between the host and the microbiota facilitate barrier function, maintenance and repair. Also, the genetic make-up of the host can shape the microbiota niche formation. However, in CRC these interactions are maladapted (orange box). Maladaptations include dysbiosis (A), an increase of genotoxic bacteria (B), intestinal barrier defects (C), influx of microbes (D), aberrant inflammation (E) and DNA damage (F).

(Figure prepared for Nature by N.Ilott and A. Janney[3])

Barrier dysfunction and subsequent bacterial-induced inflammation positively correlates with tumour multiplicity[139] and triggers the production of cancer-related proinflammatory cytokines by mesenchymal and immune cells[3]. Mechanistic studies have focussed on a few

candidate microbes to date. One such species is colibactin-producing polyketide synthase (pks)$^{+}$ *E. coli* which has been shown to have genotoxic abilities which induce DNA double strand breaks and improper cell division[140-142]. Further, 'oncomicrobes' such as *F. nucleatum* or enterotoxigenic *B. fragilis* (ETBF) create a permissive TME by recruiting immunosuppressive MDSCs[143-145] or by directly inhibiting the cytotoxic activity of NK-cells[146]. Overall, these and other studies reviewed elsewhere[3] support the idea that the microbiota is a causal contributor to CRC progression.

### 1.4.1 Inflammatory mechanisms underlying microbiota interactions in CRC

Inflammation is a well-known driver of CRC (**1.3.1**) with research on which inflammatory mechanisms and cytokine networks underlie the microbiota and contribute to CRC an ongoing endeavour (**Figure 1.6**).

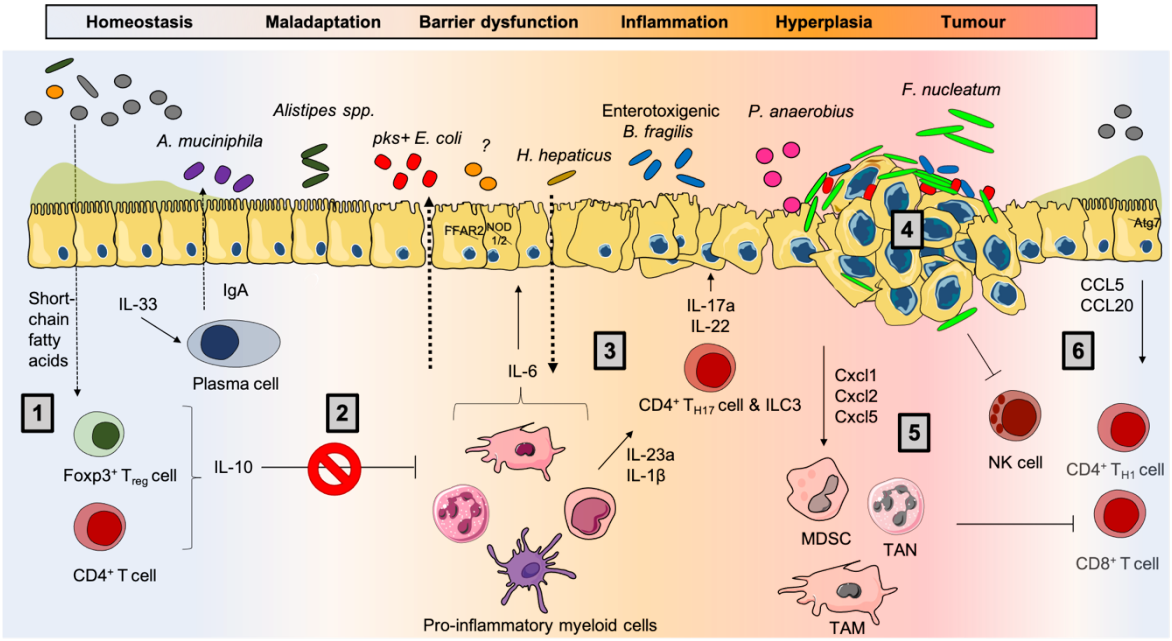



**Figure 1.6 Inflammatory mechanisms underpinning the microbiota in CRC.**
The microbiota is known to contribute to CRC via a variety of mechanisms. (1) Microbiota-derived short-chain fatty acids can induce IL-10 secretion which in turn can dampen undesired inflammation. IL-33 is also protective by inducing IgA production and limiting the activity of mucus degrading microbes such as *A. munciphilia*. (2) IL-10 deficiency triggers inflammation and allows the outgrowth

of harmful (*Alistipes spp*.), genotoxic (pks$^+$ *E. coli*) and other still unknown species which further fuel inflammation and promote genomic instability. (3) Increased epithelial permeability allows the influx of bacteria, including enterotoxigenic *B. fragilis* (ETBF), which stimulates pro-inflammatory cytokine production. IL-6 can drive aberrant epithelial cell proliferation and in combination with IL-1β and IL23A, promote a CD4$^+$ $T_{H17}$ responses. Altered innate immunity permits *H. hepaticus* triggered IL-22 production by ILC3s which leads to epithelial cell proliferation (4) Human tumours can be coated by biofilms which contain ETBF, pks+ *E. coli* and *F. nucleatum*. These coincide with increased levels of IL-6, IL-23A and IL-17A. (5) Some species can alter the immune cell composition of the TME. For example, ETBF and *P. anaerobius* recruit immunosuppressive MDSCs, TAMs and TAN by secreting CXCL1, CXCL2, and CXCL5 whereas *F. nucleatum* can directly inhibit cytotoxic NK cell activity. (5) Some features of the microbiota can be protective in CRC, for example *Clostridia* species, which promote CD4 $T_{H1}$ and CD8 T-cell responses (6).

IL, interleukin; CD, cluster of differentiation; Foxp3, forkhead box P3; Ig, immunoglobulin; TH, T-helper CXCL, C-X-C motif ligand; NK, natural killer; CCL, C-C motif ligand.

(Figure prepared for Nature by A. Janney[3])

In homeostasis, microbiota derived factors such as short chain fatty acids can induce the production of IL-10 by CD4$^+$ and regulatory T cells which dampens undesired inflammation and the outgrowth of genotoxic species such as pks$^+$ *E. coil*[147-150]. Similarly protective, the cytokine IL-33 triggers IgA production by B cells which prevents a dysbiosis characterised by destructive mucus-degrading *Akkermansia* species[151].

However, CRC is often accompanied by a physical breakdown of the IEC barrier which allows the influx of potentially harmful microbes (**1.4**) as well as the formation of invasive bacterial aggregates termed 'biofilms' in human specimens[152]. This increased proximity of the microbiota to the host correlates and is hypothesized to trigger the secretion of many cancer-related pro-inflammatory cytokines including IL-1β, TNFa, IL-6, IL-17A IL-23 and IL-22[3]. In mice and in humans, the expression of IL-6, IL-23 and IL17A have been shown significantly elevated within tumours infiltrated by bacteria[153,154], but further studies are needed to confirm if microbes indeed trigger their elevated expression.

Importantly, cytokines have diverse and pleiotropic effects during CRC. In particular, cytokines of the $T_H17$ pathway such as IL-23, IL-22 and IL-17 are important mediators of mucosal tissue healing and host defence during homeostasis, but can be hijacked to promote colitis and/or CRC in genetically maladapted individuals[74]. For example, *Helicobacter hepaticus* (*Hh*) is a mouse-related pathobiont that is known to trigger inflammation in an IL-23- dependent manner and lead to pro-tumorigenic IL-22 production in genetically susceptible mouse models of colitis and cancer[55,124,155,156].

## 1.5 Mouse models recapitulating CRC

Mouse models are an essential tool to understand CRC biology. However, due to the complexity of the mutational landscape and TME, no known mouse models adequately represent all forms of disease. Thus, consideration is needed when choosing a suitable model for the scientific questions asked. **Table 1.1** summarises the most commonly used mouse models in CRC and CAC to date.

**Table 1.1 Mouse models of colorectal and colitis-associated cancer.**
(Expanded table on the basis of West et al, 2015)[74]

| Mouse model | Primary tumour | Metastasis | CRC/CAC | Driver |
|---|---|---|---|---|
| **129RAG$^{-/-}$**<br>***H. hepaticus***<br>**AOM** | Distal colon dysplasia, aberrant foci and adenocarcinomas | No | CAC | Innate inflammation by pathobiont and AOM mutagenesis |
| **DSS**<br>**AOM** | Colon aberrant foci and adenomas | No | CAC | Intestinal barrier disruption by chemical irritant DSS, inflammation and AOM mutagenesis |
| **IL10$^{-/-}$**<br>**+/- pks$^{+}$ *E. coli*** | Colon adenocarcinomas | No | CAC | Loss of immune regulation, aggressive inflammation[147] |
| **APC$^{MIN/+}$** | Small intestine adenomas | No | CRC | Heterozygous for *APC* loss of function |

| | | | | |
|---|---|---|---|---|
| **APC$^{\Delta 468}$** | Small intestine adenomas | No | CRC | Mutation of APC loss of function |
| **CPC-APC** | Colon & distal small intestine | No | CRC | Deletion of APC allele by Cdx2 cre-lox recombination |
| **APC$^{MIN/+}$ ETBF** | Distal colon | No | CRC/CAC | APC loss of function, $T_H17$ and innate inflammation[144] |
| **APC$^{MIN/+}$ *+F. nucleatum*** | Colon | No | CRC/CAC | APC loss of function, Immune suppressive environment[143] |
| ***Vil1*-Cre *Trp53*$^{fl/fl}$ AOM** | Colon | No | CRC | Epithelial loss of p53 and AOM mutagenesis |
| ***Vil1-Cre*$^{ER}$ *Braf*$^{SL\text{-}V600E}$ *Alk5*$^{fl/fl}$** | Proximal colon | No | CRC | Tamoxifen induced mutation of BRAF and loss of loss of epithelial TGFβ-receptor signalling[157] |
| ***Vil1-Cre*$^{ER}$ *Kras*$^{G12D}$ *Trp53*$^{fl/fl}$ *Rosa26*$^{N1icd}$** | Small intestine and colon | Yes | CRC | Tamoxifen induced Mutation of KRAS, loss of p53 and activated Notch in the epithelium [107] |
| **Intrarectal injections** | Rectum (variable invasiveness) | Variable | CRC | Variable |
| **Xenografts** | Often subcutaneous or orthotopic | No | NA | Variable [158,159] |
| **Experimental metastasis** | Lung (intravenous injection) or liver (intrasplenic injection) | Yes | NA | Variable |

Abbreviations: AOM, azoxymethane; CAC, colitis-associated cancer; CRC, colorectal cancer; DSS, dextran sodium sulfate; Il10, interleukin-10; APC, adenomatous polyposis coli protein; ETBF, enterotoxigenic Bacteroides fragilis; Il10, interleukin-10; fl/fl, homozygous expression of floxed genes; RAG, recombination-activating gene; TH, T helper; NA, not applicable; KRAS, Kirsten rat sarcoma virus; BRAF, B-raf proto-oncogene; TGF, transforming growth factor.

The most commonly used CAC mouse models combine mutagenesis caused by the carcinogen azoxymethane (AOM) with inflammation, either triggered by either dextran sodium sulphate (DSS)[74] barrier disruption or pathobiont administration in genetically susceptible mice (129RAG$^{-/-}$ HhAOM model)[124]. Dysfunction can also be induced through GEMMs in which genes such as *Il10* or *Muc2* are knocked-out. In particular, IL-10$^{-/-}$ mice have been investigated thoroughly due to their development of spontaneous intestinal inflammation and high incidence of colonic adenocarcinomas[147,148].

Further, mouse models for familial adenomatous polyposis (FAP) are widely used in the field. Mice that carry mutations or truncated versions of *APC* develop multiple, non-invasive polyps within the small intestine[160]. However, due to the infrequency of small intestinal tumours in human disease[161], an alternative 'CPC-APC' model was generated in which deletion of *APC* is specific to CDX2$^{+}$ colon IECs thereby driving colonic tumour development[162].

Additionally, a range of other GEMMs have been created in which mutations of oncogenes *Kras* or *Braf* represent an alternative spontaneous route to adenocarcinoma development[157,163]. On this, great progress has been made recently with the development of a late-stage, metastatic CRC model combining *Kras, Tp53* mutations with enhanced Notch signalling[107]. Lastly, to reduce the often long duration of GEMM models, many studies inject tumour cells orthotopically or subcutaneously into host mice[164]. A recently developed and physiological alternative for CRC, is the intrarectal injection of colon cancer cells or tumour-derived organoids. The latter representing a more physiological three-dimensional 'mini-gut' consisting of intestinal epithelial or cancer cells[158,159].

## 1.6 Research hypotheses and experimental aims

An 'oncogenic' microbiota is now considered a key characteristic of CRC but if this is indeed a cause or consequence of tumorigenesis is unknown[3]. However, previous studies revealed that *Helicobacter hepaticus (Hh)* administration was crucial in establishing inflammation-driven cancer in the colon of mutagen treated 129RAG$^{-/-}$ mice[124,155]. Thus, we hypothesised that changes in the microbiome during *Hh*-driven colitis would have functional consequences on

tumour development in this genetically-susceptible model. In this thesis we therefore aimed to identify novel host-microbiota interactions by answering the following questions with unbiased and longitudinal approaches:

1. Which host factors drive colon cancer development in the *Hh*AOM CAC model?
2. Are identified factors driven by colon inflammation, mutagenic insult or the synergistic interplay of both?
3. Does *Helicobacter hepaticus* play a direct or indirect role in disease progression?

# 2 Materials and methods

## 2.1 Mice

Wildtype 129SvEv.RAG$^{-/-}$ mice were either bred and maintained under specific pathogen-free (SPF) or germ-free (GF) conditions in the accredited animal facility of the Kennedy Institute, Oxford University. All mice were negative for *Helicobacter spp.* and all other known pathogens. GF flexible film isolators were further routinely screened for overall bacterial contamination. Only females with at least 6 weeks of age were used for all indicated experiments. All procedures were conducted under the UK Scientific Procedures Act of 1986 by a Personal License holder under a Project License authorised by the UK Home office.

## 2.2 Mouse models

### 2.2.1 *Hh*AOM 129SvEv.RAG$^{-/-}$

Mice were gavaged with 1x10$^{8}$ colony forming units (CFU) of *Helicobacter hepaticus* (Hh) using a 22G curved needle on days 1 and 2. At week 4 (or week 6 during initial experiments) 10mg/kg azoxymethane (AOM) was injected interperitoneally (i.p) weekly for 5 consecutive weeks. Mice were sacrificed at week 15 (or week 17 during initial experiments). Animals were monitored for weight loss and signs of colitis weekly and on a daily basis from week 13 (or week 15 during initial experiments) until experimental end.

### 2.2.2 GSK484 (PAD4 inhibitor)

Mice gavaged with Hh and treated with AOM were injected daily with 20mg/kg of the reversible PAD4 inhibitor GSK484 (Cayman Chemicals) from week 13 until experimental end (week 15). Prior to injections, the crystalline solid stock was resuspended in DMSO (Sigma) and diluted accordingly with filtered MilliQ water. Control mice were injected with diluted DMSO only.

### 2.2.3 Culturing Hh

To induce disease, *Helicobacter hepaticus* (*Hh*) NCI-Frederick isolate 1A (strain 51449) was grown on laked horse blood agar plates containing the antibiotics Trimethoprim, Vancomycin and Polymixin B (TVP, Fischer/Oxoid) at 37° for 48 hours (48h) under microaerophilic conditions (3-5% O2, 10% CO2, 10% H2, balance N2) utilising a CampyPak jar (Oxoid). *Hh* was harvested with cotton swabs and transferred into a vented Erlenmeyer flask (Corning) containing Tryptic Soy Broth (TSB, Sigma-Aldrich) with 10% Fetal Bovine Serum (FBS, Life Technologies) and TVP (Fischer/Oxoid). The liquid culture was kept at 37°, under microaerophilic conditions in a shaking incubator and split 1/10 every 24h (200rpm). Prior to *Hh* gavage, the optical density of Hh was measured at 600nm using a BioPhotometer (Eppendorf). The appropriate amount of Hh was then centrifuged at 3000rpm for 15min and subsequently resuspended in sterile PBS.

## 2.3 Histology

### 2.3.1 Definitions of proximal versus distal colon

The murine colon was harvested by disconnecting the proximal colon from the caecum and collecting the complete colon including the rectum. After sampling of the utmost distal and proximal areas for RNA/DNA/protein/methacarn, the proximal and distal colon was defined by dividing the resulting tissue in two equal halves as illustrated below (**Figure 2.1**).

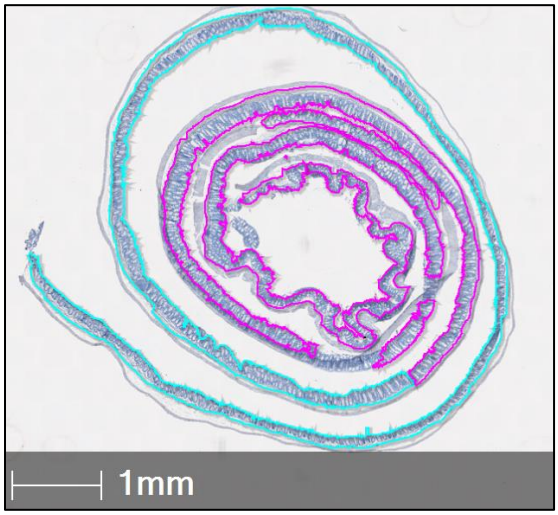



**Figure 2.1 Definition of murine proximal (pink) and distal (blue) colon used throughout this thesis. Colons are embedded as 'swiss rolls'.**

### 2.3.2 Formalin-fixed and paraffin-embedded (FFPE) swiss rolls

Murine colons were harvested and flushed with PBS. After opening longitudinally, colons were pinned on humidifying paper (Biotechne) with luminal side up and immersed into 10% formalin (Sigma) for 24h. Colons were then washed with 70% ethanol and "rolled" from proximal colon to distal colon (Fig 2.1|), pinned and stored in 70% ethanol. The swiss rolls were next embedded into paraffin and sectioned to 5µm thickness (Kennedy Histology Facility).

### 2.3.3 Methacarn fixation

An approximately 1cm piece of not cleaned nor opened distal colon and/or ceacal tissue was immediately fixed in freshly-prepared mucus-preserving methacarn (60% methanol, 30% chloroform and 10% acidic acid, (Sigma)). After 7 days of fixation at room temperature, tissue was washed twice with 100% methanol, 100% ethanol and xylene. Dehydrated tissue was embedded into paraffin and subsequently sectioned to 5µm thickness.

### 2.3.4 Histopathological analysis

Three µm-thick FFPE colon swiss roll tissue sections were stained with haematoxylin & eosin (H&E, Kennedy Institute Histology facility with DRS 2000 autoStainer, Sakur) and assessed in a blinded manner using digital scans and the NDP.view2 Viewing software (Hamamatsu) by a Board-Certified Veterinary Pathologists (ASB), American College of Veterinary Pathologists. Sections were assessed and graded for the presence and type of colitis as well for the presence and type of proliferative epithelial lesions using the INHAND (International Harmonization of Nomenclature and Diagnostic Criteria for Lesions in Rats and Mice) method[165].

Histopathological examination, number of tumours (absolute counts), and areas of non-neoplastic and neoplastic proliferative tissue ($\mu m^2$ - $mm^2$) were measured. Epithelial dysplasia (pre-neoplastic lesion that may progress to a benign/malignant neoplasm) was indicated by disruption of the intestinal crypts, shown as distorted glands. Adenoma was noted as a sessile, polypoid, or papillary focus with histological evidence of cellular atypia that exceeds that observed in epithelial dysplasia, with no evidence of basement membrane invasion. Well-

differentiated adenocarcinoma was noted where areas of increased cellular atypia were observed, with or without basement membrane invasion. Moderately-differentiated adenocarcinoma exhibits increased and more frequent cellular atypia and mitotic activity. Inflammation scores were obtained as per the following grading system outlined below (**Figure 2.2**).

| | | | |
|---|---|---|---|
| **A.** | **EPITHELIUM** | **HYPERPLASIA and/or** | **GOBLET CELL DEPLETION** |
| | **0** | **None** | **None** |
| | **1** | **Mild (1.5x)** | **Mild (25%)** |
| | **2** | **Moderate (2-3x)** | **Marked (25-50%)** |
| | **3** | **Severe (>3x)** | **Substantial (>50%)** |
| **B.** | **INFLAMMATION IN LAMINA PROPRIA** | | |
| | **0** | **None - few leucocytes** | |
| | **1** | **Mild - some increase in leucos at tips of crypts OR many lymphoid follicles** | |
| | **2** | **Moderate - marked infiltrate (notable broadening of crypt)** | |
| | **3** | **Severe - dense infiltrate throughout** | |
| **C.** | **AREA AFFECTED** | **(% of section)** | |
| | **0** | **None** | |
| | **1** | **up to 25%** | |
| | **2** | **25-50%** | |
| | **3** | **>50%** | |
| **D.** | **MARKERS OF SEVERE INFLAMMATION** | | |
| | **0** | **None** | |
| | **1** | **Submucosal inflammation OR Few crypt abscesses (<5)** | |
| | **2** | **Submucosal inflammation AND Few Crypt Absecesses (<5)** | |
| | **2** | **Many crypt abscesses (<5) OR Extensive submucosal inflamm OR Crypt branching (*Hh*)** | |
| | **3** | **Many crypt abscesses (<5) AND Extensive submucosal inflamm / Crypt branching (*Hh*)** | |
| | **3** | **Ulceration OR Extensive fibrosis** | |

**Figure 2.2 Criteria for the histological scoring of inflammation in murine colons**

## 2.4 Cell isolation

### 2.4.1 Cell isolation from the spleen

Murine spleens were harvested, mashed through 100µm nylon mesh with syringes, washed with PBS containing 0.1% BSA and 5mM EDTA (PBS/BSA/EDTA, Sigma) and centrifuged for 5min at 500g at 4°C. To lyse red blood cells, the cell pellet was resuspended in 1ml ACK (Ammonium-Chloride-Potassium) lysis buffer for 3 minutes at room temperature and

subsequently washed and centrifuged as priorly stated. Cells were resuspended in an appropriate volume of PBS/BSA/EDTA and counted with either a CASY Cell Counter (Cambridge BioScience) or CountBright Absolute Counting Beads (Life Technologies) during flow cytometry acquisition.

### 2.4.2 Cell isolation from the colon

Murine colons were harvested as stated in **2.2.3**, opened longitudinally, cleaned with ice cold PBS and cut into ~2cm pieces. After rigorous shaking of tissue pieces with ice cold PBS to remove mucus, colon pieces were minced and transferred to 5ml of RPMI containing 10% FCS, 15mM HEPES, 50U/ml collagenase VIII and 50U/ml DNAse I ("Digestion buffer", Sigma). Samples were then incubated for 20min at 37°C in a shaking incubator at 180rpm. The digested tissue was vortexed and filtered through a 100µM strainer and ice cold 10ml PBS/BSA/EDTA was added. Remaining tissue was collected from the strainer and the digestion step repeated until colon tissue completely digested. Collected cell suspensions were kept on ice and subsequently centrifuged at 500g for 10min. Cells were washed, resuspended in PBS/BSA/EDTA and counted as stated in **2.4.1.**

### 2.4.3 LPL supernatant generation

Cells were processed as in **2.4.2**, counted with the CASY Cell Counter and resuspended in serum-free RPMI and 1% Penicillin- Streptomycin (P/S, Sigma) to an overall concentration of 4 million cells/ml. Cells were seeded into 48-well flat-bottom plates (200 µl/well, SLS) and incubated at 37°C 5% $CO_2$ for 16h. Supernatants were subsequently harvested, filtered and frozen to -80°C.

### 2.4.4 Flow cytometry

Single cell suspensions were plated into 96 well V bottom plates for cell surface staining. All of the following procedures were performed with PBS/BSA/EDTA, all centrifugation steps were at 500g for 5min and all stainings were performed in the dark at 4°C.

### 2.4.5 Surface Staining

To stain for viability, cells were incubated with Fixable Viability dye eFlour® 780 (1:1000, eBioscience) for 15min. Following a wash step, anti-CD16/CD32 (1:400, eBioscience) was added for 15min to minimise non-specific binding. After further washing, appropriate antibody mixes (**Table 2.1**) were added to cells and stained for 30min at 4°C. Cells were then washed and fixed with BD Cytofix/Cytoperm (BD Biosciences) for 15min at RT. Lastly, cells were washed twice, filtered and transferred to round bottom polystyrene tubes (SLS) for flow cytometer acquisition.

**Table 2.1 Antibodies used in flow cytometry experiments**

| Primary antibody | Fluorochrome | Clone | Manufacturer |
|---|---|---|---|
| Ly6G | FITC | 1A8 | Biolegend |
| CD11c | PE | N418 | Biolegend |
| CD101 | PE-Cy7 | Muschi101 | ebioscience |
| CD177 | AF647 | Y127 | BD Pharmingen |
| MHCII | AF700 | M5/114.15.2 | invitrogen |
| Siglec-F | BV421 | E50-2440 | BD Horizon |
| F480 | BV510 | BM8 | Biolegend |
| CD11b | BV605 | M1/70 | Biolegend |
| CD45 | BV711 | A20 | Biolegend |
| Ly6C | BV785 | HK1.4 | Biolegend |

### 2.4.6 Acquisition and analysis

Flow cytometric aquisition was performed on a BD Fortessa X-20 following bead-based (BD Biosciences) compensation. Data was analysed using FlowJo v10 software.

## 2.5 Gene expression analysis

### 2.5.1 Tissue RNA and DNA extraction

2mm$^2$ tissue pieces were stored in RNAlater (Qiagen) at -80°C. To isolate RNA and DNA, tissue was thawed and transferred to RLT lysis buffer (Qiagen) containing 1% β-mercaptoethanol (Sigma-Aldrich) and disrupted using a single steel bead (Qiagen) and the TissueLyserII

(Qiagen) at 30 Hz for 14min. Simultaneous isolation of RNA and DNA was performed used the AllPrep DNA/RNA Mini Kit (Qiagen) as per manufacturer's instructions. Isolated RNA and DNA were stored at -20°C and -80°C, respectively.

### 2.5.2 qRT-PCR

Up to 2µg of RNA was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions and including 5% RNase Inhibitor (Applied Biosystems). Transcription was performed in a C1000 TouchTM Thermal Cycler (Biorad) and cDNA was subsequently diluted 1:5 with Nuclease free water (Life Technologies). Quantitative PCR (qPCR) was performed using Taqman assays (Applied Biosystems, **Table 2.2**) and PrecisionPlus Mastermix (PrimerDesign, UK) on a ViiA7 Real-time PCR detection system (Applied Biosystems). All expression levels were normalised to the internal housekeeping gene *Hprt* and calculated using the ΔCT method: $2^{-\Delta CT}$ where ΔCT = CT $_{\text{(target gene)}}$ – CT $_{\text{(}Hprt\text{ endogenous control)}}$.

**Table 2.2 Taqman probes used for quantitative PCR**

| Gene | Assay ID |
|---|---|
| *Hprt* | Mm01545399_m1 |
| *Padi4* | Mm01341658_m1 |

### 2.5.3 RNA sequencing

Library preparation, RNA-sequencing (HiSeq4000) and quality control were performed on a service basis by the High-Throughput Genomics Group (Wellcome Trust Centre for Human Genetics, Oxford).

### 2.5.4 RNA *in silico* analysis

Raw data was processed using a pipeline by Dr. Nicholas Ilott. Quality control was performed by running *fastQC* (v0.11.9). Gene mapping and counts were produced by *kallisto* (v.0.46.0). All subsequent analysis was performed using R software. Analysis included principal component analysis (PCA) of transformed read counts (log2 counts per million, *prcomp*), differential gene expression analysis (*DESeq*) and gene ontology (GO) enrichment analysis (*fgsea*). The github link for R code created and used for all analysis is included in Appendix 8.1.

## 2.6 Stool sampling

Murine faecal samples were collected into autoclaved 1.5 ml tubes (Eppendorf) during defecation. Samples were immediately snap frozen on dry ice and stored at -80°C.

### 2.6.1 Faecal DNA extraction

Faecal DNA was isolated using the DNeasy PowerSoil kit (Qiagen) following manufacturer's instructions.

### 2.6.2 Faecal supernatant extraction

To prepare faecal supernatants, faecel pellets were transferred to screw-capped tubes filled with 500µl cOmplete ULTRA EDTA-free protease inhibitor (Roche). Samples were homogenized using the TissueLyser II (Qiagen) for 10min at 30 Hz and subsequently centrifuged at 10,000g for 5min at 4°C. Supernatants were collected and immediately used for ELISA (**2.8.2**).

## 2.7 Bacterial DNA analysis

### 2.7.1 Quantification of bacterial colonisation

Universal bacterial colonisation was quantified as described[166] using a real time PCR method based on 16S ribosomal RNA primers below (**Table 2.3**) and performed on the ViiA7 Real-Time PCR detection system (Applied Biosystems) with FastStart Universal SYBR Green with Rox (Roche). Standard curves were constructed using *H.h* DNA that was purified from bacterial cultures using the DNeasy kit (Qiagen).

**Table 2.3 Primers used for 16S ribosomal RNA quantitative PCR**

| Primers | Sequence |
|---|---|
| 515 F | GTG CCA GCM GCC GCG GTA A |
| 805 R | GAC TAC CAG GGT ATC TAA T |

### 2.7.2 Quantification of *Hh* colonisation

*Hh* colonisation was quantified as described[167] using a real-time PCR method based on primers below (**Table 2.4**) and performed on the ViiA7 Real-Time PCR detection system (Applied Biosystems) with PrecisionPlus Mastermix (PrimerDesign, UK). Standard curves were constructed using *H.h* DNA that was purified from bacterial cultures using the DNeasy kit (Qiagen).

**Table 2.4 Primers used for 16S ribosomal RNA quantitative PCR**

| **Primers** | Sequence |
|---|---|
| *cdt*B Forward | CCG CAA ATT GCA GCA ATA CTT |
| *cdt*B Reverse | TCG TCC AAA ATG CAC AGG TG |
| *cdt*B Fluorescence Probe | AAT ATA CGC GCA CAC CTC TCA TCT GAC CAT |

### 2.7.3 16S rRNA Sequencing

Microbial genomic DNA was PCR amplified targeting the V3 and V4 regions and subsequently subjected to quality control and sequencing (MiSeq_300PE) on a service basis by the High-Throughput Genomics Group (Wellcome Trust Centre for Human Genetics, Oxford).

### 2.7.4 16S rRNA *in silico* analysis

16S rRNA sequencing data was quality checked using *fastqc* and processed using the *dada2* pipeline within the NGSKit pipeline (https://github.com/nickilott/NGSKit) by Dr. Nicholas Ilott. This generated an amplicon sequence variant (ASV) table for downstream analysis using R software. Analysis included abundance and ordination analysis (R package *phyloseq*) and

the MaAsLin2 package to identify candidates differentially abundant over time (https://github.com/biobakery/Maaslin2). Analysis was run in collaboration with Dr. Sena Bae (Harvard University).

## 2.8 Analysis of protein expression

### 2.8.1 Luminex assay

Proteins were quantified using a bead-based Luminex assay (R&D) following manufacturer's instructions. 25µl per sample were quantified in duplicate. Fluorescence was measured and subsequently analysed by a Luminex 100/200 instrument (R&D) and xPONENT software (in collaboration with Department of Paediatrics, Oxford).

### 2.8.2 Enzyme-linked immunosorbent assays

Proteins were also quantified using enzyme-linked immunosorbent assays (ELISA) according to the respective manufacturer's instructions (Invitrogen). 50µl of sample was quantified in duplicate and absorbance was measured by a spectrophotometric plate reader (SPECTROstar Nano).

### 2.8.3 Proteomics on patient CRC samples

Patients diagnosed with CRC scheduled for colectomy at Oxford University Hospitals were consented for research use of their resected cancer and matched normal mucosa during a pre-operative appointment under the established GI Illness Ethics through the University of Oxford Translational Gastroenterology Unit (Reference numbers: 11/YH/0020, 16/YH/0247,

HTA 12217). All patients gave informed consent and metadata were fully anonymised by biobankers prior to analysis. Surgical resection samples were processed by a pathologist then placed immediately on ice in PBS until collected for further processing within 2 hours. 1-10mm$^3$ tissue pieces were placed in 300µl RIPA buffer plus 1X HALT protease inhibitor (ThermoFisher). Upon thawing tissue was firstly homogenised using Soft Tissue Homogenizing CK14 Kit vials (Stretton Scientific) and a TissueLyserII (Qiagen) machine. Protein concentrations in bulk tissue were determined using bicinchoninic acid assay (BCA) assay (ThermoFisher) prior to performing mass spectrometry in collaboration with Dr. Andy Howden (University Dundee) and Dr. Elizabeth Mann.

## 2.9 *In situ* analysis

### 2.9.1 Immunohistochemistry and HALO quantification

Following standard de-waxing procedures (2x 5min Xylene 2x 1min 100% Ethanol on DRS 2000 autoStainer, Sakur), FFPE swiss roll sections were exposed to heat-induced antigen retrieval by Target Retrieval solution (DAKO, pH=6) for 15min (microwave, boiling). Subsequently, sections were blocked with Bloxall solution for 15min (Vector Labs), followed by avidin and then biotin blocking for 15min each (respectively) and 1h blocking in Tris Buffered Saline (TBS, Insight Biotechnology) solution containing 10% serum (Sigma) of the secondary antibody species. Sections were washed with TBS between each blocking step and all steps were performed at RT.

Without a subsequent wash step, tissue sections were then incubated with primary antibodies in 10% TBS/serum overnight at 4°C. Sections were subsequently washed and

directly-conjugated horseradish peroxidase (HRP) secondary antibodies (Vector Labs) applied for 1h at RT. Chromogenic staining was then developed by application of a DAB substrate solution (Vector Labs) at manufacturer's instructions and counterstained with Haematoxylin (Sigma) for 5min at RT. Slides were dehydrated and mounted in DPX mounting media (Sigma). Whole section imaging (Akoya Biosciences Vectra Polaris) was performed on chromogenic swiss rolls and subsequently analysed using Indica Labs HALO® image analysis platform. Swiss rolls were annotated manually and scored using the analysis module CytoNuclear which detected DAB positive and negative cells. After training the software, the percentage of positive and negative cells per annotated area was automatically calculated by the HALO® software (**Figure 2.3**) and graphed in Prism 9 (GraphPad).

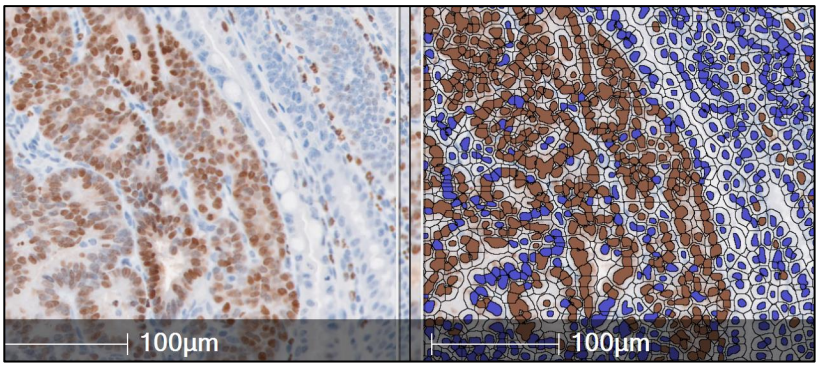



**Figure 2.3 Example of the automatic calculation of DAB positive cells (brown) versus negative cells (blue) by the HALO® software.**

## 2.9.2 Immunofluorescence

Initial steps for imunofluorescence stainings on FFPE sections were identical to those mentioned in **2.9.1**. However, no avidin or biotin blocking was required. Primary antibodies (**Table 2.5**) were directly conjugated with a fluorophore or secondary antibodies (Life technologies) were utilised. Nuclei were stained with Hoechst 28332 (Life technologies) prior

to imaging on a Zeiss confocal LSM 880/980 or a Zeiss Axioscope A1 microscope. All images were analysed with ImageJ software (NIH).

**Table 2.5 Antibodies used in immunofluorescence stainings**

| Primary antibody | Clone | Manufacturer | Secondary antibody |
|---|---|---|---|
| PAD4 | EPR20706 | Abcam | goat anti-rabbit 555 |
| EpCAM 488 | G8.8 | Biolegend | - |
| Ly6G | 1A8 | Abcam | donkey anti-rat 488 |
| MPO | polyclonal | R&D | rabbit anti-goat 555 |

### 2.9.3 BugFISH

Following previously defined de-waxing procedures, methacarn-fixed and paraffin-embedded sections were hybridised with the conjugated Eub338 universal and Hep647 *Hepaticus*-specific DNA probe (**Table 2.6**). Both probes were applied at a concentration of 10ng/µl in 50°C pre-warmed hybridisation buffer (900mM NaCl, 20mM Tris pH 7.5, 5mM EDTA). A DAPI (1:500 in PBS) counterstain was performed at 4°C for 1h and slides mounted using ProLongGold antifade reagent (Life Technologies). All imaging was performed on a Zeiss Axioscope A1 microscope and analysed with ImageJ software (NIH).

**Table 2.6 DNA probes used for BugFISH**

| Primers | Specificity | Sequence | Fluorophore |
|---|---|---|---|
| Eub338 | All bacteria | GCT GCC TCC CGT AGG AGT | Cy3 |
| Hep647 | *Helicobacter hepaticus* | CCC ACA CTC CAG AGT AAC AGT | Cy5 |

## 2.10 Statistical analysis

Statistical analysis and graphical representation were done using Prism 9 (GraphPad Software). Respective statistical and normality tests are indicated in each figure legend in this thesis. Differences were considered statistically significant when $p \leq 0.05$ and shown according to the following: * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

# 3 Investigation of host-microbe interactions in a genetically-susceptible colitis-associated cancer model

## 3.1 Introduction

Patients diagnosed with colitis have an increased risk of developing CRC[70], seemingly due to aberrant inflammation driving tumorigenesis (**1.2**). A variety of murine models have begun to identify mechanisms of transition to adenocarcinoma[168], but have primarily focused on established tumours leaving early on-set mechanisms yet to be determined.

An innate model of colitis-associated cancer established in our laboratory possesses the hallmarks of human disease because it is driven by the interplay of genetic susceptibility and environmental exposures[124,155]. The pathobiont *Hh* triggers chronic colitis in *RAG*-deficient 129SvEv mice in contrast to mild inflammation observed in some other immune-deficient strains[155,169]. Together with the administration of the DNA-damaging agent azoxymethane (AOM), bacteria-induced chronic innate inflammation promotes adenocarcinoma development specifically in the distal colon[124]. The genetic interval 'HICCS' was identified to drive this innate response[155] and more recently, the gene alpha kinase 1 (*Alpk1*) was shown to be an important inflammatory regulator within this locus[169]. In particular, the IL-23/Th17 pathway has been well-established to play a role in driving tumorigenesis[74]. Ablating ILCs or blocking the Th17-derived cytokine IL-22 resulted in a therapeutic reduction in tumours in this *Hh*AOM model[124].

However, alongside the contribution of ILCs to this disease model, the role of the myeloid compartment has yet to investigated in detail. In particular, the overall environment in early cancer on-set and which aspects of disease are driven by inflammation versus mutagenic insult remain unexplored. To address this gap, we performed gene expression profiling of distal and proximal colon tissue harvested at 4 (colitis), 10 (mid-tumourigenesis) and 15 (cancer) weeks and correlated that with histopathology on colon swiss rolls across all groups and time points (**Figure 3.1 A**).

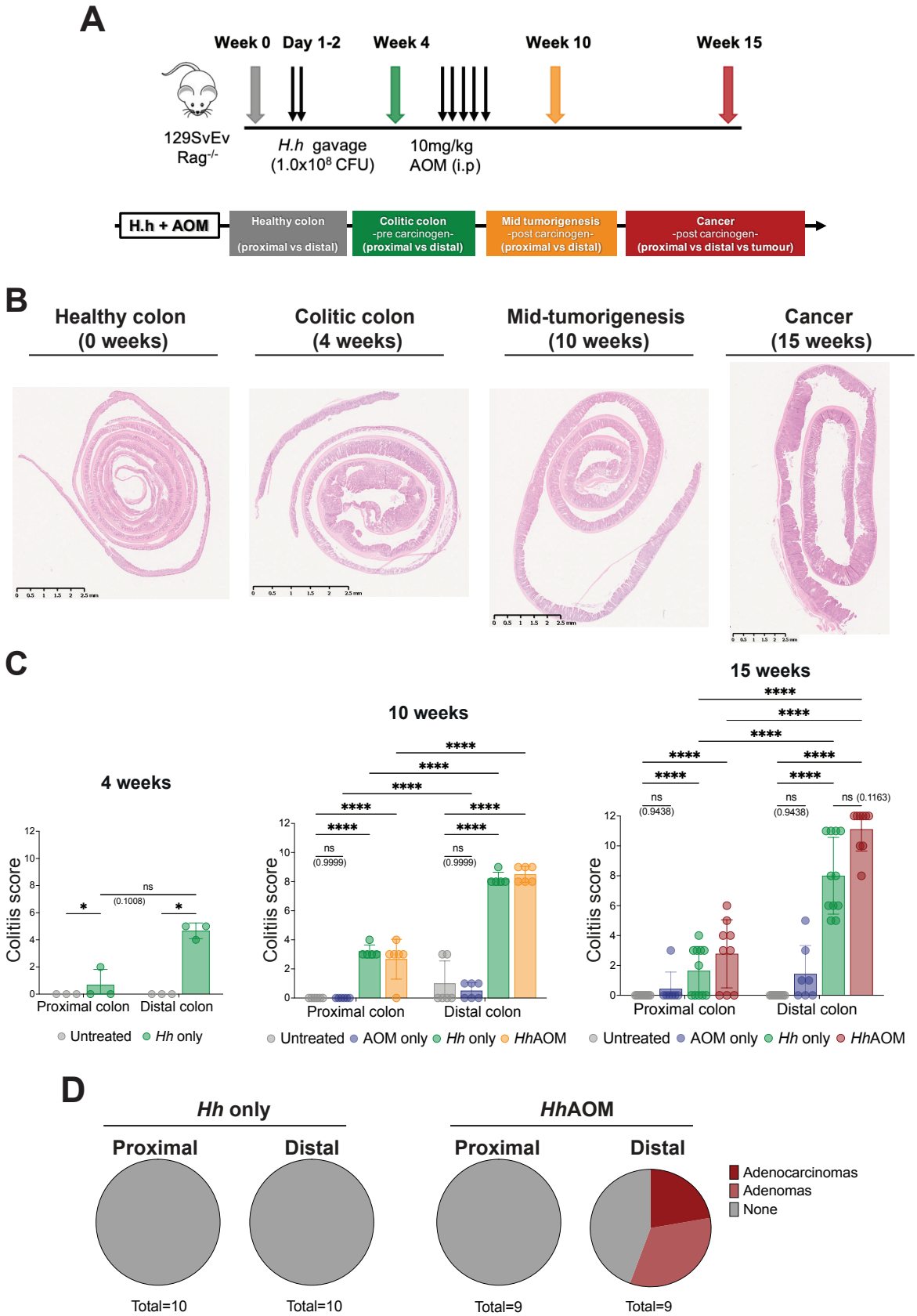
## 3.2 Aim

Utilising RNA sequencing, histopathology and cell profiling we aim to characterise the distal colon environment in early- and late-stage disease to identify novel cancer-promoting mechanisms. We hypothesise that the distal colon harbours a specific microenvironment that promotes tumorigenesis upon challenge.

## 3.3 Results

### 3.3.1 Colitis-associated cancer (CAC) occurs in the distal colon in the *Hh*AOM model

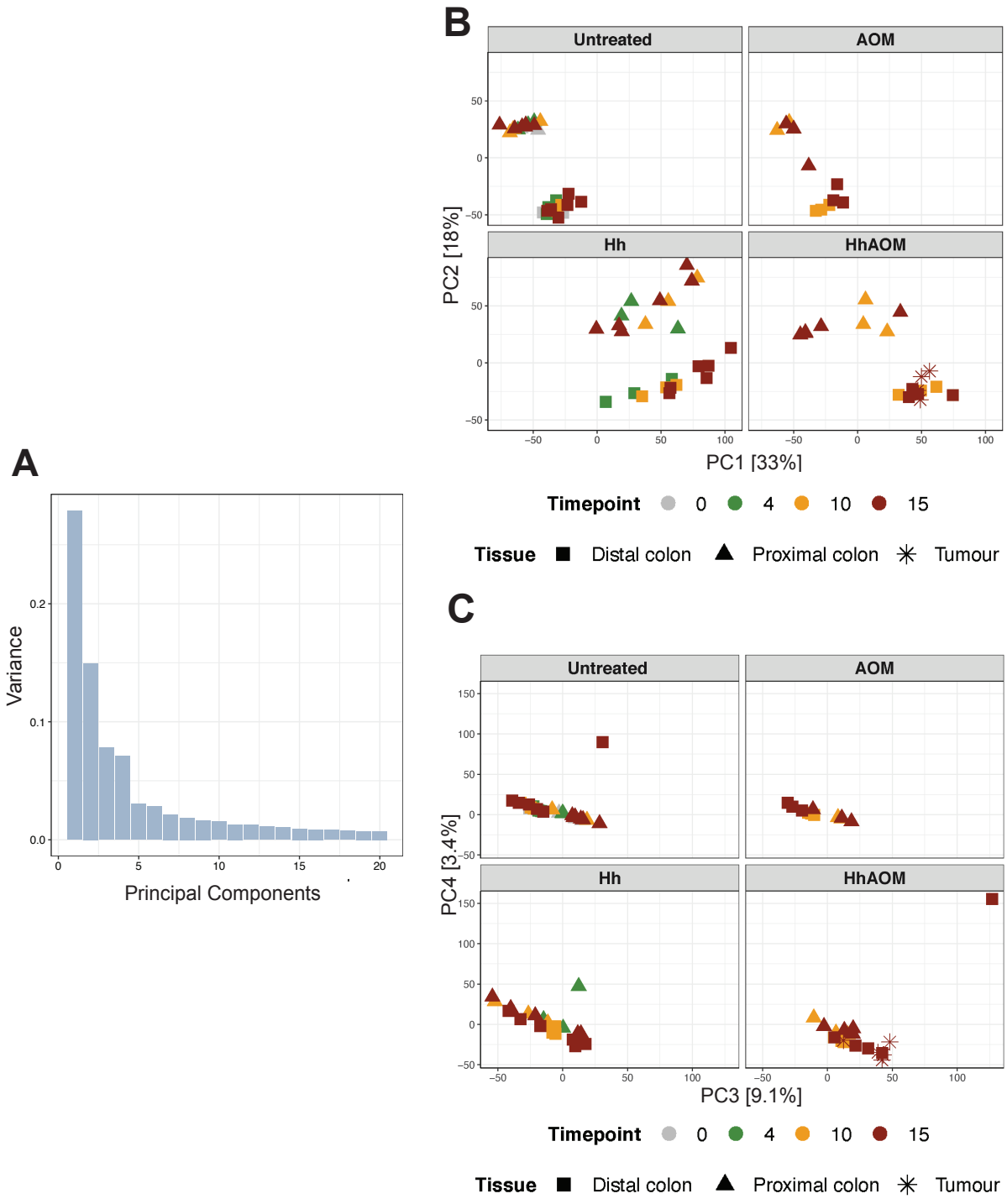
To identify pathological features associated with distinct phases of disease, H&E-stained whole colon swiss rolls **(Figure 3.1 B)** were scored (**2.3.4**) for colitis and cancer (kind collaboration with Dr. Suarez-Bonnet of the Royal Veterinary College (RVC, London)). Proximal and distal colons were separated and scored individually (**Figure 2.1**).


A
Week 0
Day 1-2
Week 4
Week 10
Week 15
129SvEv
Rag-/-
H.h gavage
(1.0x10⁸ CFU)
10mg/kg
AOM (i.p)
H.h + AOM
Healthy colon
(proximal vs distal)
Colitic colon
-pre carcinogen-
(proximal vs distal)
Mid tumorigenesis
-post carcinogen-
(proximal vs distal)
Cancer
-post carcinogen-
(proximal vs distal vs tumour)
B
Healthy colon
(0 weeks)
Colitic colon
(4 weeks)
Mid-tumorigenesis
(10 weeks)
Cancer
(15 weeks)
C
4 weeks
10 weeks
15 weeks
Colitis score
Proximal colon
Distal colon
Untreated
AOM only
Hh only
HhAOM
ns
(0.1008)
(0.9999)
(0.9438)
ns (0.1163)
D
Hh only
HhAOM
Proximal
Distal
Adenocarcinomas
Adenomas
None
Total=10
Total=10
Total=9
Total=9

**<u>Figure 3.1</u> Longitudinal assessment reveals the distal colon as the distinct site of colitis and cancer during disease progression.**
129SvEvRAG$^{-/-}$ mice were left untreated (Week 0 n=3, Week 4 n=3, Week 10 n=6, Week 15 n=6), injected with AOM (Week 10 n=6, Week 15 n=7), gavaged with *Hh* (Week 4 n=3, Week 10 n= 5, Week 15 n=10), or exposed to *Hh* and AOM (Week 10 n=6, Week 15 n=9). At sacrifice, colons were harvested and embedded as swiss rolls.

A. Schematic of experimental set-up.
B. Representative haematoxylin and eosin (H&E) images of whole colon swiss rolls.
C. Colitis score of proximal and distal colons across experimental groups and time points. Each dot represents one mouse and bars represent mean ±SEM. Data from 4 weeks represents one independent experiment, data from 10 and 15 weeks represent two independent experiments. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (*, $p < 0.05$; ****, $p < 0.0001$; ns, no significance).
D. Pie chart showing the distribution of adenocarcinoma stages across groups.

Data showed colitis in the proximal colon post-*Hh* administration, however distal colon had the highest disease score across all time points. Colitis scores were comparable between *Hh*-gavaged mice and *Hh* mice intra-peritoneally injected with AOM, but development of adenomas and invasive adenocarcinomas were restricted to the distal colons of *Hh* and AOM-treated mice after 15 weeks **(Figure 3.1 C).** Administration of the carcinogen AOM alone did not result in severe colitis or cancer **(Figure 3.1 D)**.




**<u>Figure 3.2</u> Principal component analysis shows sample clustering due to inflammation and location.**

129SvEvRAG$^{-/-}$ mice were left untreated (week 0 n=3, week 4 n=3, week 10 n=3, week 15 n=3), injected with AOM (week 10 n=3, week 15 n=3), gavaged with *Hh* (week 4 n=3, week 10 n= 5, week 15 n=5), or exposed to *Hh* and AOM (week 10 n=3, week 15 n=5). At sacrifice, proximal and distal colon was harvested for RNA isolation and subsequent bulk RNA sequencing. Principal component analysis (PCA) was performed on all samples (n=87).

A. Scree plot of all principal components (PC).

B. PC1 versus PC2 on all samples. Symbols represent the origin of sample whereas colors the timepoint of sampling in weeks.
C. PC3 versus PC4 on all samples.

Bulk RNA from proximal and distal colons across treatment groups and timepoints was extracted and subjected to RNA sequencing. Principal component analysis (PCA, *prcomp*) on the complete dataset highlighted *Hh* exposure and colon location to be the predominant modes of variation in this model, with the first four principal components explaining most of the variance within the overall dataset **(Figure 3.2 A).** The first principal component (PC)1 segregated *Hh*-gavaged samples from those left untreated or injected with AOM, whereas proximal and distal samples were separated by PC2 irrespective of treatment or tumorigenesis **(Figure 3.2 B)**. Further segregation of samples was not seen with PC3 versus PC4 showing no clustering of samples **(Figure 3.2 C)**.

### 3.3.2 Tumours are enriched in neutrophil-associated markers.

To understand what is driving tumorigenesis, we compared an established tumour and a healthy distal colon at week 15 (**Figure 3.3 A**). Differential expression analysis (*DESeq2*) identified *Lrg1*, *Ly6d* and *Tacstd2* to be among the top upregulated genes in tumours, whereas *Sult1a1*, *Sycn* and *Dpyd* were among the genes downregulated (**Figure 3.3 B, Appendix 8.2**). To identify the underlying biological processes, we performed pathway enrichment analysis (*enrichGO* from the *clusterProfiler* package) and observed a significant enrichment of neutrophil-associated pathways (**Figure 3.3 C**). Labelling of neutrophil-associated genes on the differential gene volcano plot further highlights their enrichment in tumours **(Figure 3.3 D, Appendix 8.3**).

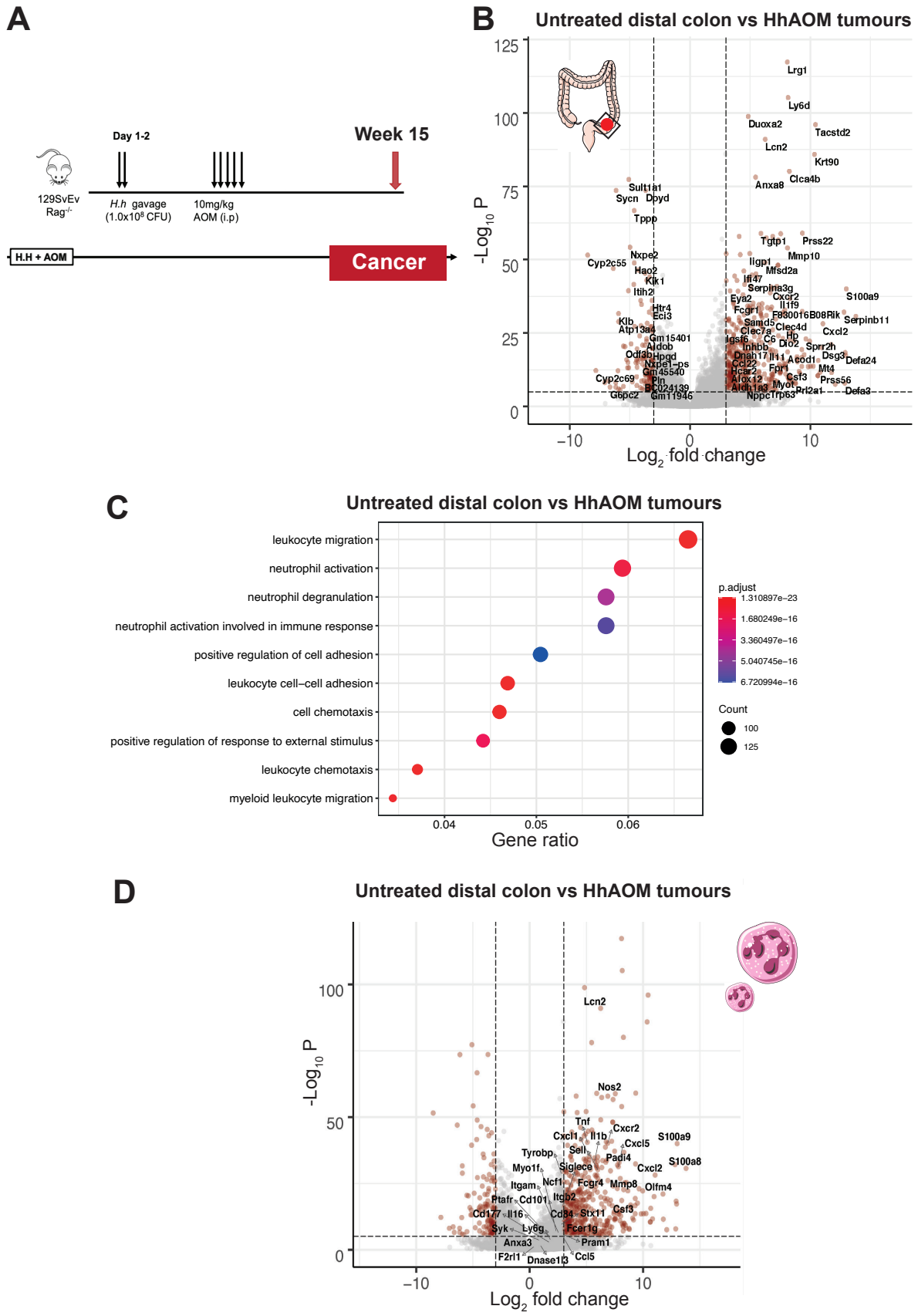

A
Day 1-2
Week 15
129SvEv
Rag-/-
H.h gavage
(1.0x10⁸ CFU)
10mg/kg
AOM (i.p)
H.H + AOM
Cancer
B
Untreated distal colon vs HhAOM tumours
-Log10 P
Log2 fold change
C
Untreated distal colon vs HhAOM tumours
leukocyte migration
neutrophil activation
neutrophil degranulation
neutrophil activation involved in immune response
positive regulation of cell adhesion
leukocyte cell–cell adhesion
cell chemotaxis
positive regulation of response to external stimulus
leukocyte chemotaxis
myeloid leukocyte migration
Gene ratio
p.adjust
Count
D
Untreated distal colon vs HhAOM tumours
-Log10 P
Log2 fold change

**Figure 3.3 Tumours are enriched in neutrophil markers when compared to a healthy distal colon.**
Differential expression analysis was performed comparing healthy distal (n=3) and tumour tissue (n=5) harvested from mice upon sacrifice at week 15.

A. Schematic illustrating the time point of mouse sacrifice.
B. Volcano plot showing the top differentially expressed genes coloured in red ($log_2$Fold > 3, $log_{10}$P > 6). Positive fold changes represent upregulated genes in tumours while negative fold changes correspond to downregulated genes in comparison to healthy distal tissue (*DESeq, ggplot*)
C. Dot plot of the 10 gene ontology (GO) processes with the largest gene ratios. The size of each dot represents the number of significantly differentially expressed genes from the prior DESeq analysis (B) and the dot colour illustrates the p-adjusted values (*enrichGO* from *clusterProfiler*)
D. Volcano plot as in (B) with neutrophil- associated genes labelled.

### 3.3.3 Neutrophil accumulation precedes CAC.

Next, we sought to probe the landscape of distal colons exposed to *Hh* and AOM prior to adenocarcinoma development (**Figure 3.4 A**) by performing *DESeq* on healthy distal colons versus distal colons of *Hh*AOM mice at 10 weeks (**Figure 3.4 B**). The antimicrobial genes *Reg3b* and *Reg3g* were found to be highly upregulated in diseased colons and again, *Sycn* was a top downregulated gene (**Figure 3.4 C**).

As in later stage tumours, pathway analysis revealed neutrophil-related biology to be a dominant component of the tissue environment at 10 weeks (**Figure 3.4 D**). To probe if this observation was distal colon specific, we ran *DESeq* and *clusterGo* on corresponding healthy proximal vs diseased proximal tissue and found that neutrophil-related genes and pathways were indeed enriched in the distal colon. A distal colon neutrophil signature was similarly observed in *Hh*-only controls.

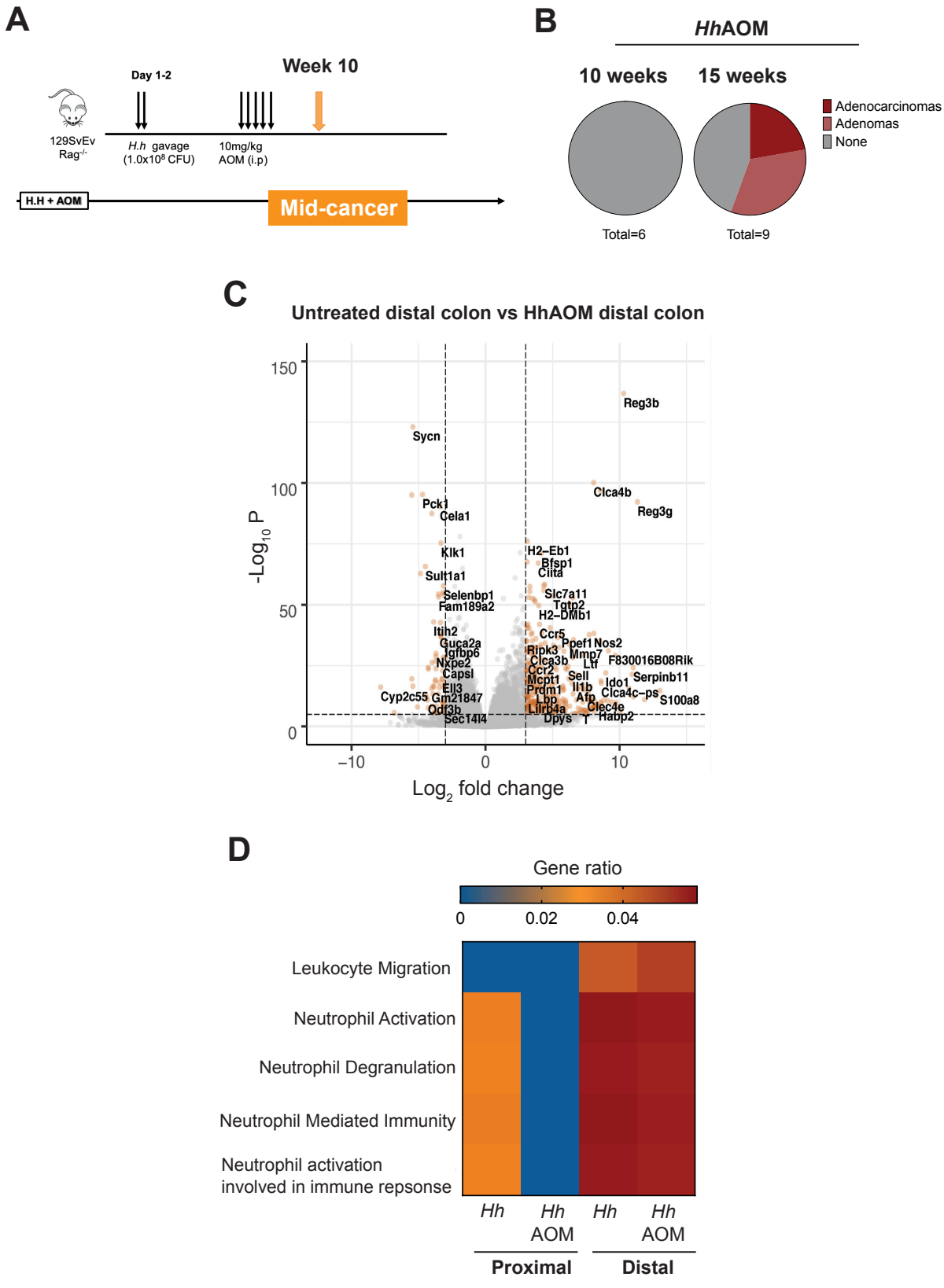
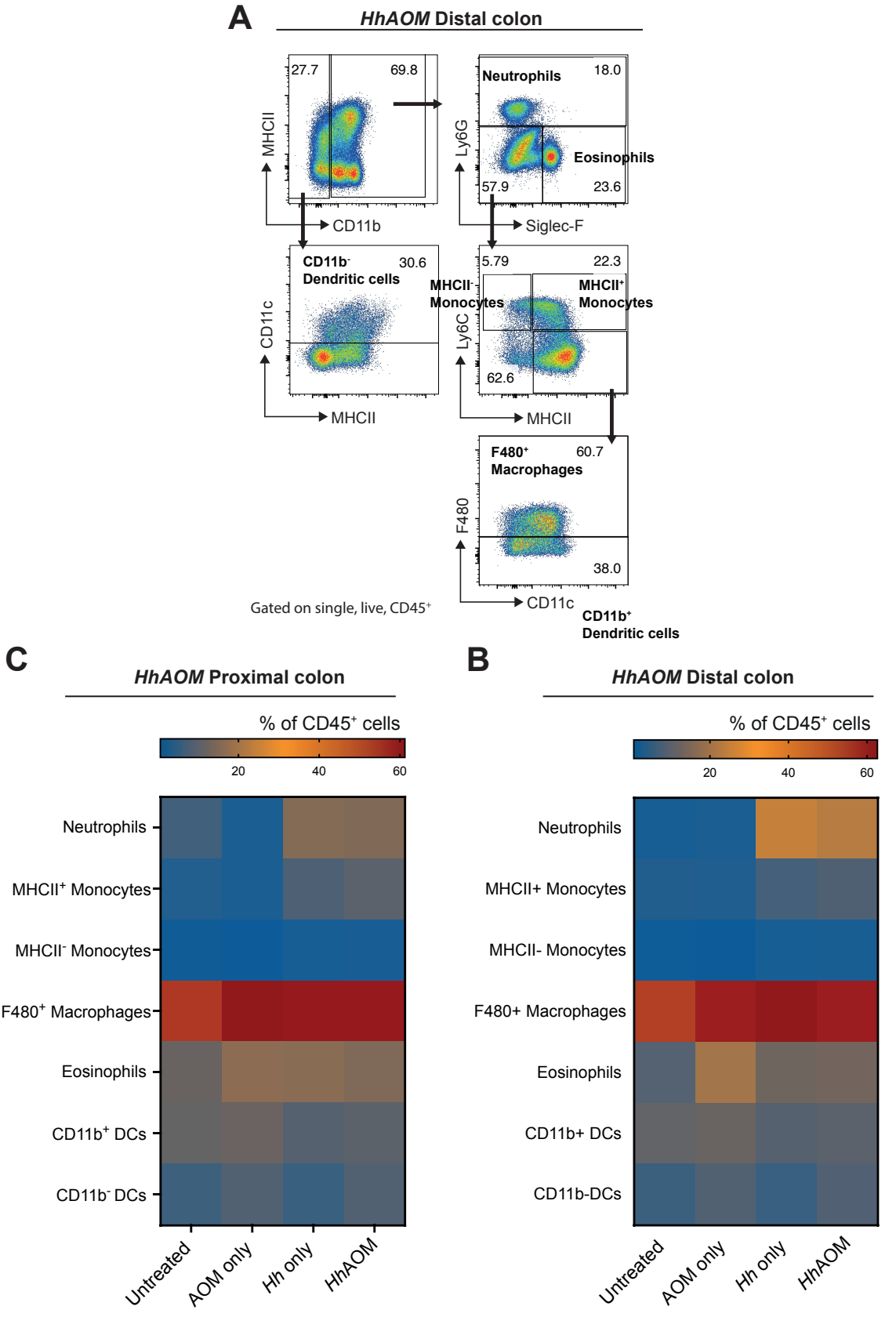

A
Week 10
Day 1-2
129SvEv
Rag-/-
H.h gavage
(1.0x10⁸ CFU)
10mg/kg
AOM (i.p)
H.H + AOM
Mid-cancer
B
HhAOM
10 weeks
15 weeks
Adenocarcinomas
Adenomas
None
Total=6
Total=9
C
Untreated distal colon vs HhAOM distal colon
-Log10 P
Log2 fold change
Reg3b
Sycn
Clca4b
Reg3g
Pck1
Cela1
Klk1
Sult1a1
H2-Eb1
Bfsp1
Clta
Slc7a11
Tgtp2
H2-DMb1
Selenbp1
Fam189a2
Itih2
Guca2a
Igfbp6
Nxpe2
Capsl
Eil3
Cyp2c55
Gm21847
Odf3b
Sec14l4
Ccr5
Ppef1
Nos2
Ripk3
Clca3b
Mmp7
Ltf
F830016B08Rik
Ccr2
Mcpt1
Sell
Prdm1
Il1b
Ido1
Serpinb11
Lbp
Afp
Clca4c-ps
S100a8
Lilrb4a
Clec4e
Dpys
T
Habp2
D
Gene ratio
0
0.02
0.04
Leukocyte Migration
Neutrophil Activation
Neutrophil Degranulation
Neutrophil Mediated Immunity
Neutrophil activation
involved in immune repsonse
Hh
Hh
AOM
Hh
Hh
AOM
Proximal
Distal

**Figure 3.4 Neutrophil signatures observed prior to cancer establishment in *Hh*-colonised mice.**
Differential expression analysis was performed comparing healthy proximal colons (n=3) to *Hh (n=3)* or *Hh*AOM (n=3) treated colons alongside healthy distal colons (n=3) to *Hh* (n=3) and *Hh*AOM (n=3) treated distal collected from mice at week 10.

A. Schematic illustrating the time point of mouse sacrifice.
B. Pie chart showing the distribution of adenocarcinoma stages at 10 and 15 weeks.
C. Volcano plot showing the top differentially expressed genes coloured in orange ($log_2$Fold > 3, -$log_{10}$P > 6). Positive fold changes represent upregulated genes in distal colon tissue of *Hh*AOM mice while negative fold changes correspond to downregulated genes in comparison to healthy distal tissue.
D. Heatmap of neutrophil – associated GO processes showing corresponding gene ratios from prior calculated differential expression analysis run on respective tissue site and treatment as in Figure 1.3C.

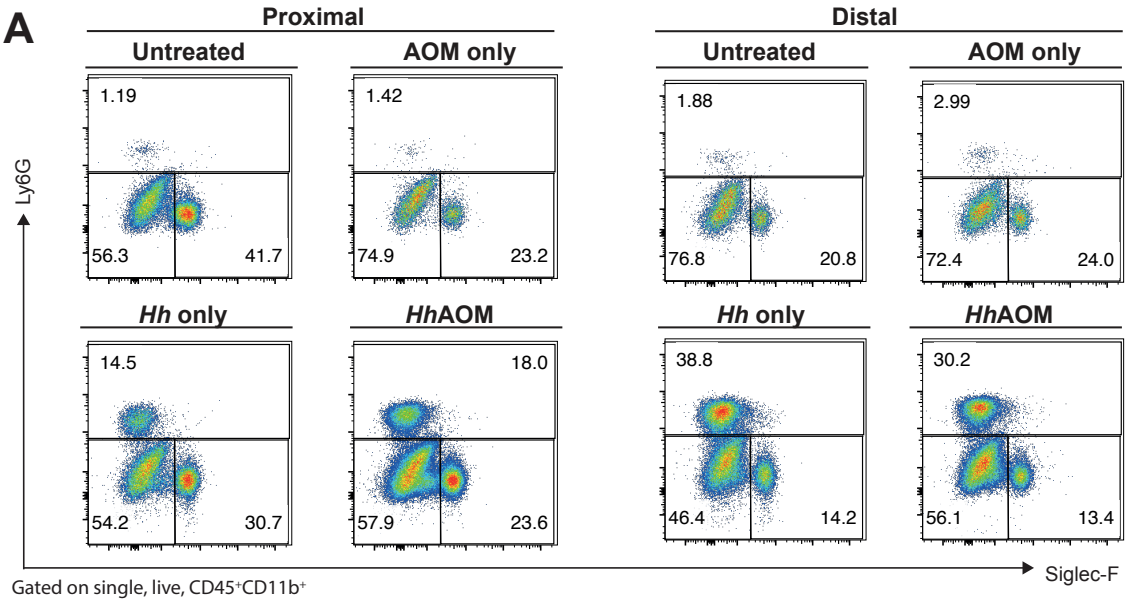
To validate our findings at a cellular level we generated single cell suspensions of proximal and distal colons to analyse myeloid populations by flow cytometry at 10 weeks (**Figure 3.5 A**). We confirmed an increase in $Ly6G^+$ neutrophil abundancy in colitis-driven cancer compared to other myeloid subsets in the proximal and distal colon (**Figure 3.5 B and C**).


A
HhAOM Distal colon
27.7
69.8
MHCII
CD11b
Neutrophils
18.0
Eosinophils
57.9
23.6
Ly6G
Siglec-F
CD11b- Dendritic cells
30.6
CD11c
MHCII
5.79
22.3
MHCII- Monocytes
MHCII+ Monocytes
62.6
Ly6C
MHCII
F480+ Macrophages
60.7
38.0
F480
CD11c
CD11b+ Dendritic cells
Gated on single, live, CD45+
C
HhAOM Proximal colon
% of CD45+ cells
20
40
60
Neutrophils
MHCII+ Monocytes
MHCII- Monocytes
F480+ Macrophages
Eosinophils
CD11b+ DCs
CD11b- DCs
Untreated
AOM only
Hh only
HhAOM
B
HhAOM Distal colon
% of CD45+ cells
20
40
60
Neutrophils
MHCII+ Monocytes
MHCII- Monocytes
F480+ Macrophages
Eosinophils
CD11b+ DCs
CD11b-DCs
Untreated
AOM only
Hh only
HhAOM

**<u>Figure 3.5.</u> Flow cytometry analysis of myeloid populations shows a differential increase of neutrophils in colitis and early cancer.**
129SvEvRAG$^{-/-}$ mice were left untreated (n=7), injected with AOM (n=7), gavaged with *Hh* (n=7) or exposed to both (n=8). Upon sacrifice at week 10, colons were harvested, split into proximal and distal and single cell suspensions analysed by flow cytometry.

A. Representative flow cytometry plots gated on live, CD45$^{+}$ single cells.
B. Heatmap of frequency means of indicated immune cell populations in CD45$^{+}$ leukocytes across experimental groups of the proximal colon.
C. Heatmap of frequency means of indicated immune cell populations in CD45$^{+}$ leukocytes across experimental groups of the distal colon.

A significant increase in neutrophil abundance as a proportion of CD45$^{+}$ leukocytes was observed in *Hh* and *HhAOM* conditions (**Figure 3.6 A-C**), alongside significantly higher neutrophil numbers in both colon locations (**Figure 3.6 D & E**). However, comparing the total number of neutrophils between proximal and distal colons, we found a significant increase in the distal colon (**Figure 3.6 F**). In agreement with previous gene expression analysis (**Figure 3.4 D**), there was no difference observed between the neutrophil numbers in *Hh* only or *Hh*AOM conditions (**Figure 3.6 G**).


A
Proximal
Distal
Untreated
AOM only
Hh only
HhAOM
1.19
56.3
41.7
1.42
74.9
23.2
1.88
76.8
20.8
2.99
72.4
24.0
14.5
54.2
30.7
18.0
57.9
23.6
38.8
46.4
14.2
30.2
56.1
13.4
Ly6G
Siglec-F
Gated on single, live, CD45+CD11b+

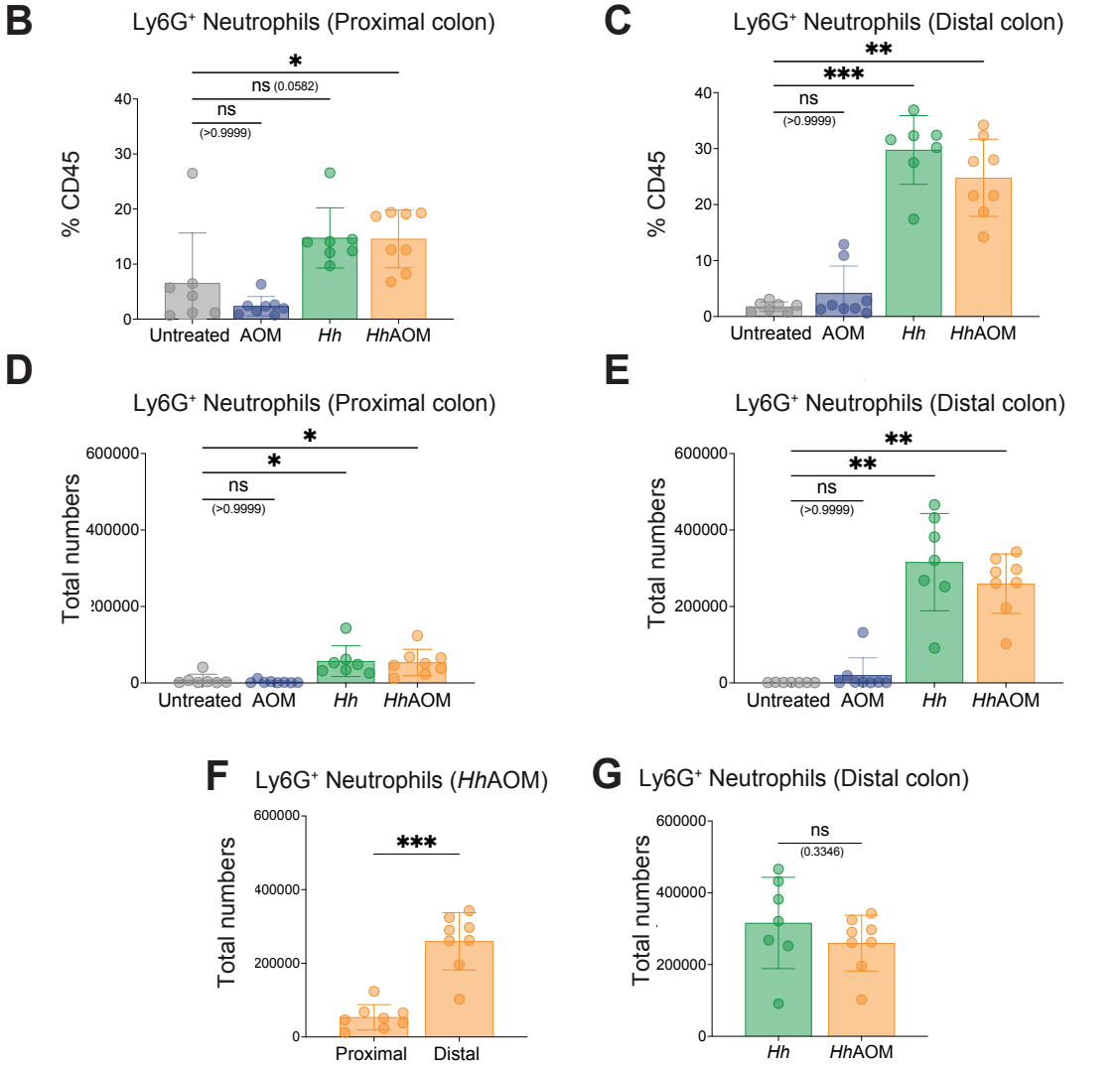

B
Ly6G+ Neutrophils (Proximal colon)
% CD45
Untreated
AOM
Hh
HhAOM
ns
ns (0.0582)
ns (>0.9999)
C
Ly6G+ Neutrophils (Distal colon)
D
Ly6G+ Neutrophils (Proximal colon)
Total numbers
E
Ly6G+ Neutrophils (Distal colon)
F
Ly6G+ Neutrophils (HhAOM)
Proximal
Distal
G
Ly6G+ Neutrophils (Distal colon)
ns (0.3346)

**Figure 3.6 The neutrophil influx in *Hh* and *Hh*AOM mice is distal colon specific.**
129SvEvRAG$^{-/-}$ mice were left untreated (n=7), injected with AOM (n=7), gavaged with *Hh* (n=7) or exposed to both (n=8). Upon sacrifice at week 10, colons were harvested, split into proximal and distal and single cell suspensions analysed by flow cytometry.

A. Representative flow cytometry plots of proximal and distal colons gated on live, CD45$^+$ CD11b$^+$ single cells.
B. Frequencies of Ly6G$^+$ neutrophils in untreated, AOM only, *Hh* only or *Hh*AOM proximal colons. Each dot represents one mouse and bars represent mean ±SEM. Data represents two independent experiments. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (*, $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$; ns, no significance).
C. Frequencies of Ly6G$^+$ neutrophils in distal colons. Figure outlines as in B. (**, $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; ns, no significance).
D. Total number of Ly6G$^+$ neutrophils in proximal colons. Figure outlined as in B (*, $p < 0.05$; ns, no significance).
E. Total number of Ly6G$^+$ neutrophils in distal colons. Figure outlined as in B (**, $p < 0.01$; ns, no significance).
F. Comparison of total number of Ly6G$^+$ neutrophils in proximal and distal colons. Data were assessed by a paired t-test (*** $p < 0.001$).
G. Comparison of total number of Ly6G$^+$ leukocytes in *Hh* and *Hh*AOM distal colons. Data were assessed by a Welch's t-test (ns, no significance).

In line with this observation, the expression of neutrophil-attracting chemokines *Cxcl1*, *Cxcl2,* and *Cxcl5* was upregulated to a greater extent in distal colons during early tumorigenesis (**Figure 3.7A**). Bar plots of normalised *Cxcl1* and *Cxcl2* gene counts showed that this was also the case during late-stage cancer at 15 weeks (**Figure 3.7B**). Further, a downregulation of both chemokines was observed in solely the proximal colons at 15 weeks (**Figure 3.7B**). However, a significant change in neutrophil numbers was not observed in the proximal or distal colon between 10 and 15 weeks **Figure 3.7C**).

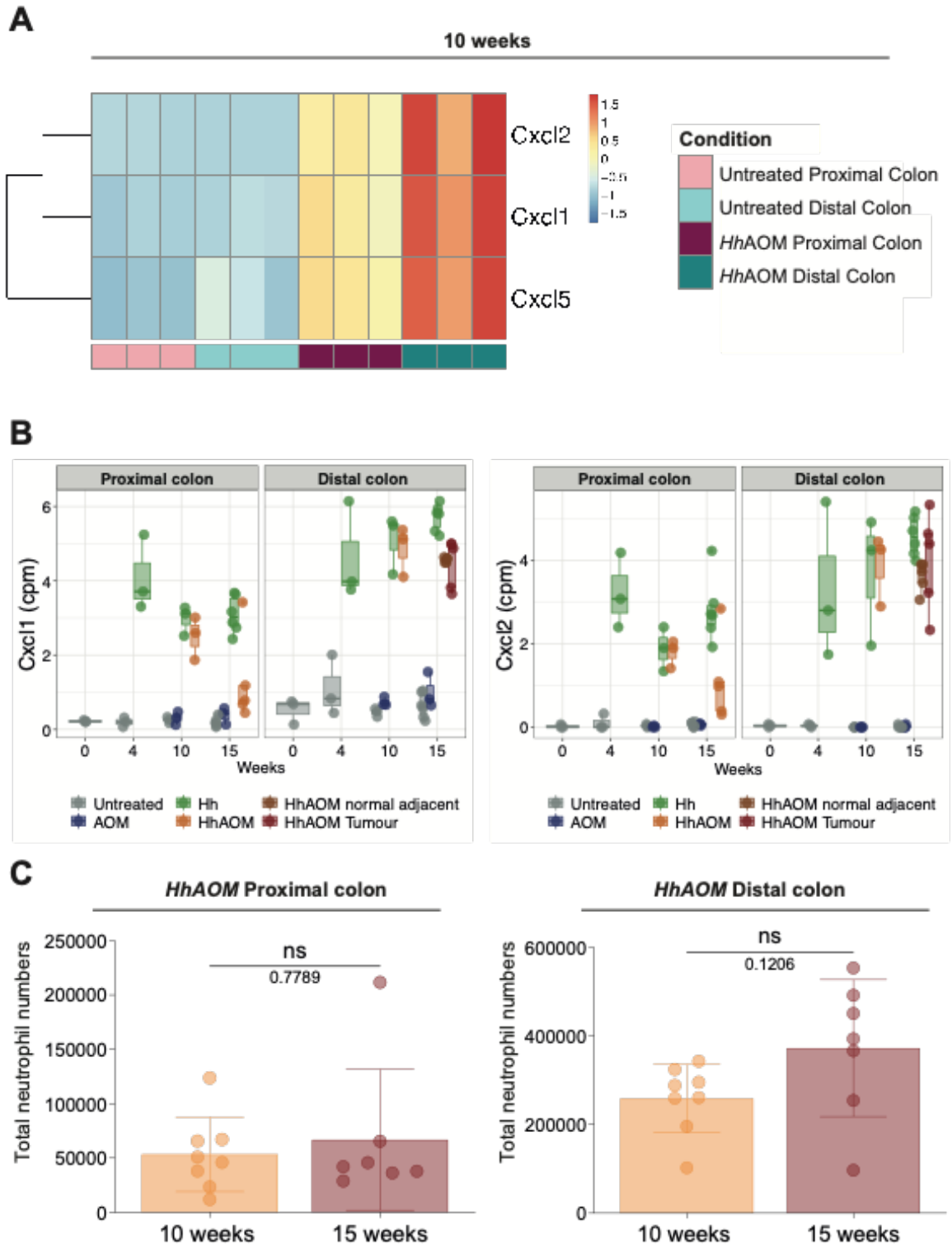



**Figure 3.7 Expression of neutrophil attracting chemokines is increased in the distal colon**
129SvEvRAG$^{-/-}$ mice were left untreated (week 0 n=3, week 4 n=3, week 10 n=3, week 15 n=3), injected with AOM (week 10 n=3, week 15 n=3), gavaged with *Hh* (week 4 n=3, week 10 n= 5, week 15 n=5), or exposed to both (week 10 n=3, week 15 n=5). Upon sacrifice, proximal and distal colon was harvested

for RNA isolation and subsequent bulk RNA sequencing or FACS analysis. Gene counts were normalised and downstream unbiased analysis performed on *Cxcl1*, *Cxcl2* and *Cxcl5.*

A. A heatmap of chemokines (normalised and scaled counts per million, *pheatmap*) was generated with untreated and *Hh*AOM samples harvested at 10 weeks.
B. Bar plot of normalised counts per million of *Cxcl1* and *Cxcl2* in full longitudinal RNA sequencing dataset.
C. Total number of $Ly6G^+$ neutrophils in proximal and distal colons at 10 and 15 weeks of the model. Data were assessed by a Mann-Whitney test after D'Agostino-Pearson omnibus normality test (ns, no significance).

These observations concur with lamina propria inflammation scoring performed by Dr. Suarez-Bonnet (RVC London) which showed significantly higher inflammation in the distal colon at 10 and 15 weeks (**Figure 3.8 A**). Representative immunofluorescence images of *Hh*AOM colonic swiss rolls further illustrated increased $Ly6G^+$ (green) and myeloperoxidase $(MPO)^+$ cells (red) predominantly in the distal colons and tumours of *Hh*AOM mice (**Figure 3.8 B**). In tumours, $Ly6G^+$ $MPO^+$ cells were observed in clusters similar to crypt abscesses. More detailed histopathological analysis is warranted to confirm this.

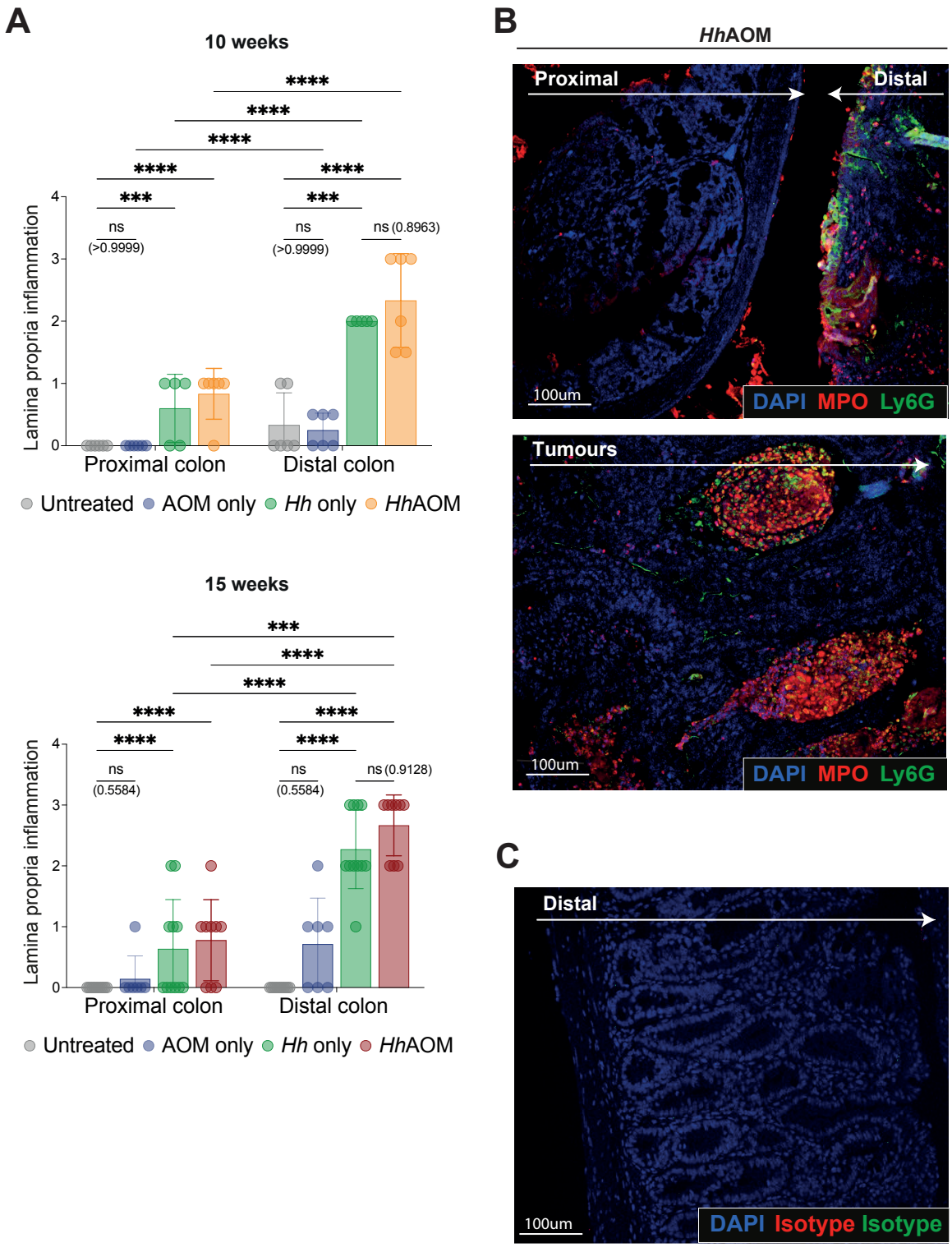



**Figure 3.8 Immunofluorescent imaging illustrates a distal colon neutrophil influx.**
129SvEvRAG$^{-/-}$ mice were gavaged with *Hh* and injected with AOM. Upon sacrifice at week 10 and 15, l colon tissue was harvested, fixed in formalin and swiss rolls processed for immunofluorescence and H&E staining.

A. Lamina propria inflammation score of proximal and distal colons across experimental groups at week 10 and week 15. Each dot represents one mouse and bars represent mean ±SEM. Data represent two independent experiments. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (*, $p < 0.05$; ****, $p < 0.0001$; ns, no significance).
B. Representative, merged images of myeloperoxidase (MPO, red), Ly6G (green) and DAPI (blue) alongside an isotype control (red, green) are shown.
C. Representative, merged images of isotype controls (rabbit IgG, red; rat IgG, green).

### 3.3.4 Immature and mature neutrophils are increased during colitis and cancer

We next to sought to probe $Ly6G^+$ neutrophil subsets in the distal colon using flow cytometry. We utilised surface CD101 to distinguish 'immature' ($CD101^-$) from 'mature' neutrophils ($CD101^+$) (**Figure 3.9 A & B**)[21] and analysed CD177 known to be expressed on a variable proportion of human neutrophils[170].

In untreated distal colons, we found the low number of neutrophils to be largely $CD101^+$. In AOM-only conditions we observed a slight expansion of $CD101^-$ neutrophils which was increased further in *Hh*-only and *Hh*AOM conditions (**Figure 3.9 C**). With regards to total numbers, we registered a trend towards an increase of both $CD101^+$ and $CD101^-$ neutrophils (**Figure 3.9 D**). All $Ly6G^+$ neutrophils isolated from the mice expressed CD177 in health and during inflammation (**Figure 3.9 B-D**).

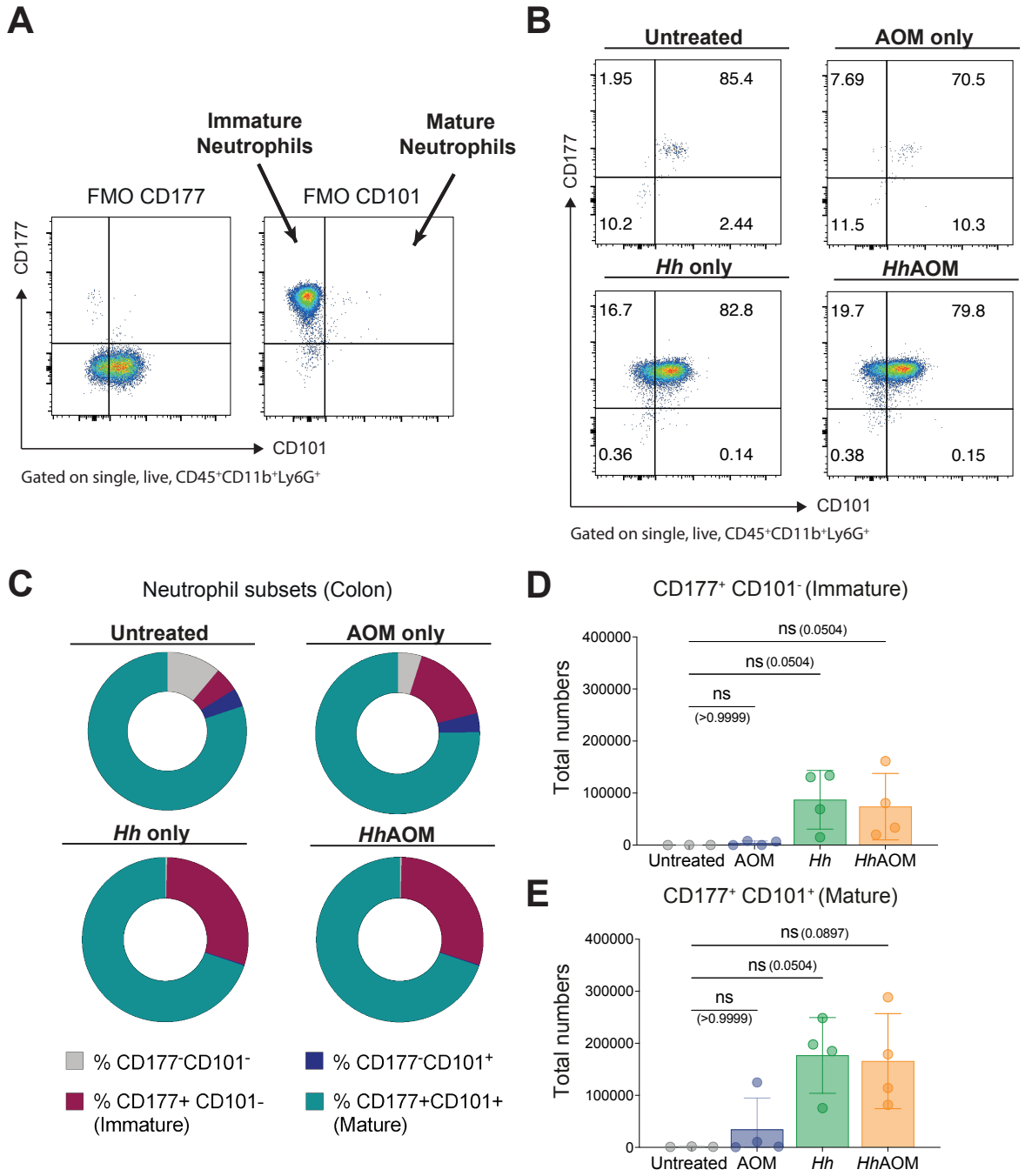



**Figure 3.9 Immature and mature neutrophils are present in colitis and early cancer.**
129SvEvRAG$^{-/-}$ mice were left untreated (n=3), injected with AOM (n=4), gavaged with *Hh* (n=4) or exposed to both (n=4). Upon sacrifice at week 10, distal colons were harvested and single cell suspensions analysed by flow cytometry.

A. Representative flow cytometry plot depicting FMO controls of CD177 and CD101 (gated on live, CD45$^+$ CD11b$^+$ Ly6G$^+$ single cells) and illustrating the gates to define immature and mature neutrophil subsets.
B. Representative flow cytometry plots of distal colons gated on live, CD45$^+$ CD11b$^+$ Ly6G$^+$ single cells.

C. Pie charts displaying the frequencies of indicated populations for each experimental group. Frequencies of defined populations were averaged amongst live, $CD45^+$ $CD11b^+$, $Ly6G^+$ cells.
D. Total number of $CD177^+$ $CD101^-$ ("immature") neutrophils in untreated, AOM only, *Hh* only or *Hh*AOM distal colons. Each dot represents one mouse and bars represent mean ±SEM. Data represents one independent experiment. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (ns, no significance).
E. Total number of $CD177^+$ $CD101^+$("mature") neutrophils in distal colons. Figure outlined as in D.

$CD101^-$ neutrophils were similarly enriched in the spleen of inflamed mice at 10 weeks (**Figure 3.10 A & B**). In line with this, we observed a significant increase of total $CD101^+$ and $CD101^-$ cells in all *Hh* exposed spleens, irrespective of AOM administration. (**Figure 3.10 C**).

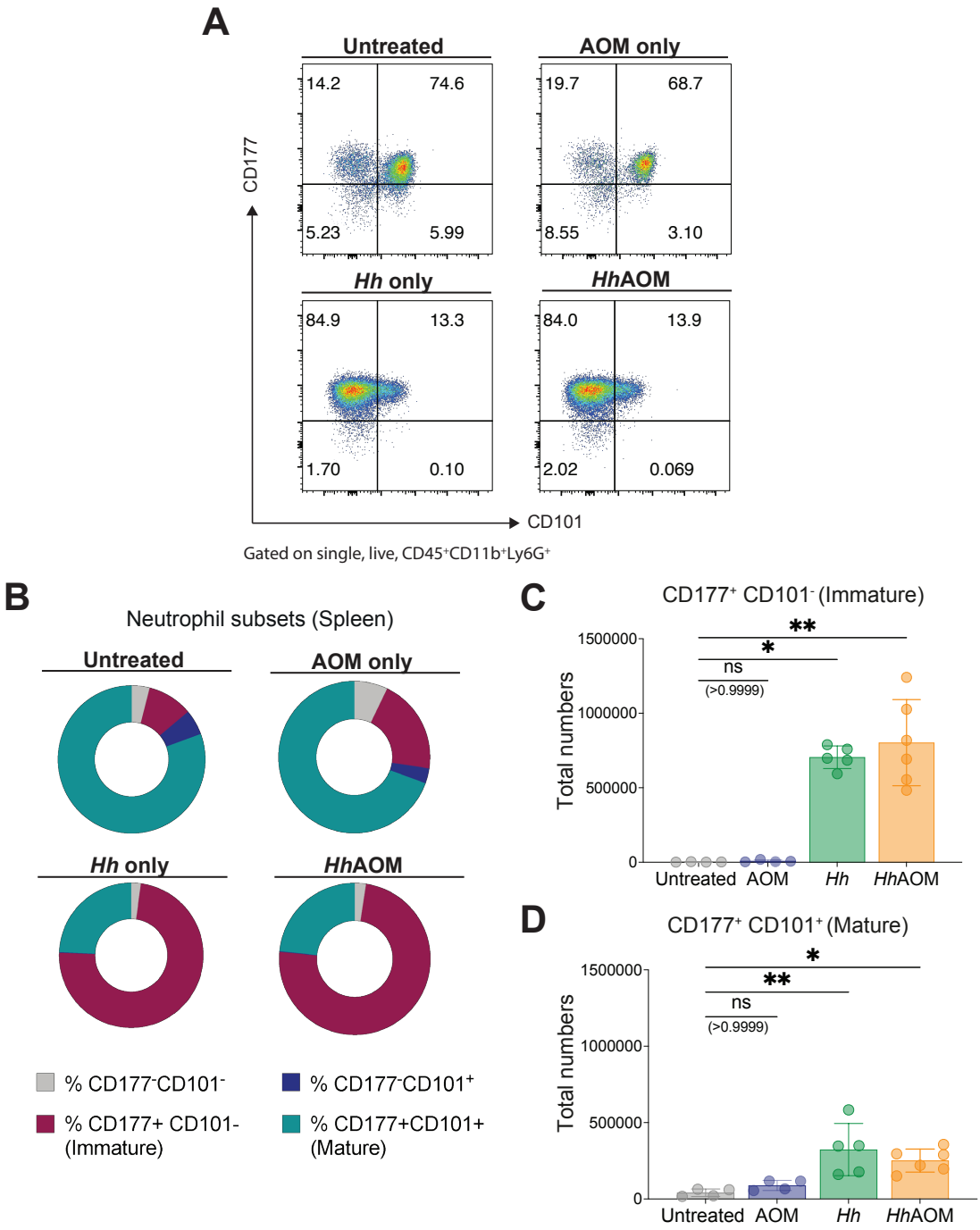



**Figure 3.10 Accumulation of immature neutrophils observed in the spleens of inflamed animals.**

129SvEvRAG$^{-/-}$ mice were left untreated (n=3), injected with AOM (n=3), gavaged with *Hh* (n=4) or exposed to both (n=4). Upon sacrifice at week 10, spleens were harvested and single cell suspensions analysed by flow cytometry.

A. Representative flow cytometry plots of spleens gated on live, CD45$^+$ CD11b$^+$ Ly6G$^+$ single cells.
B. Pie charts displaying the frequencies of indicated populations for each experimental group. Frequencies of defined populations were averaged amongst live, CD45$^+$ CD11b$^+$, Ly6G$^+$ cells.

C. Total number of $CD177^+$ $CD101^-$ ("immature") neutrophils in untreated, AOM only, *Hh* only or *Hh*AOM spleens. Each dot represents one mouse and bars represent mean ±SEM. Data represents one independent experiment. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (*, $p < 0.05$; ** $p < 0.01$; ns, no significance).
D. Total number of $CD177^+$ $CD101^+$("mature") neutrophils in spleens. Figure outlined as in D.

Finally, we examined neutrophil subsets in the colon at early (4 week) and late (15 week) time points to probe changes over time. We could see an expansion of $CD101^-$ neutrophils at 4, 10 and 15 weeks in *Hh* as well as *Hh*AOM mice (**Figure 3.11 A**). We noted higher $CD101^-$ and $CD101^+$ numbers at 15 weeks as compared to 4 weeks, but no significant difference between 10 and 15 weeks in *Hh*AOM mice (**Figure 3.11 B**). As seen previously, there was no significant difference between the numbers of $CD101^-$ or $CD101^+$ in *Hh* only and *Hh*AOM mice at 15 weeks (**Figure 3.11 C**).

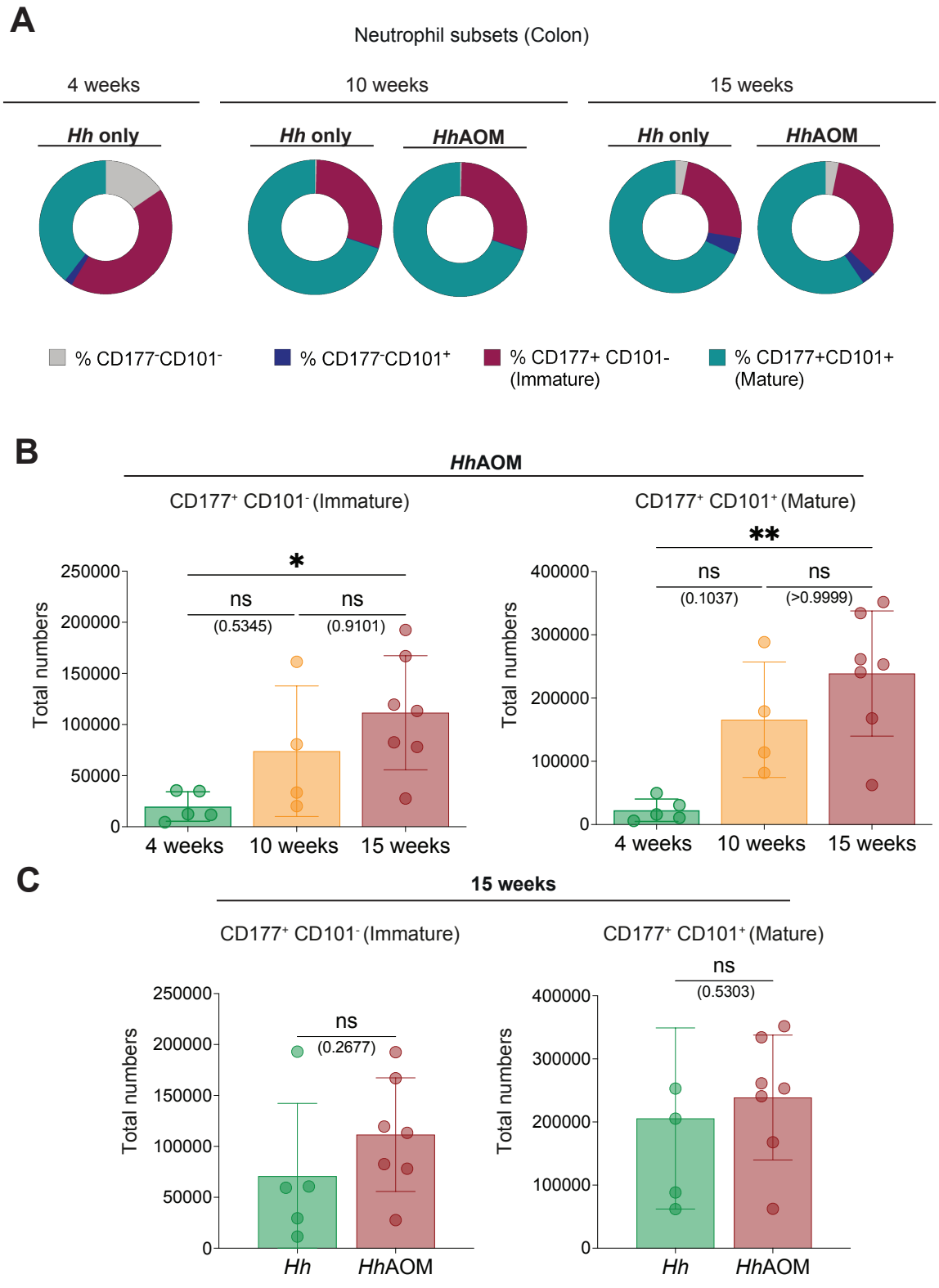



**Figure 3.11 Inflammation and tumorigenesis are associated with an increase in neutrophil numbers.**

129SvEvRAG$^{-/-}$ mice were left untreated (Week 4 n=4, Week 10 n=3, Week 15 n=4), injected with AOM (Week 10 n=4, Week 15 n=4), gavaged with *Hh* (Week 4 n=5, Week 10 n=4, Week 15 n=5) or exposed

to both (Week 10 n=4, Week 15 n=7). Upon sacrifice, distal colons were harvested and single cell suspensions analysed by flow cytometry.

A. Pie charts displaying the frequencies of indicated populations for each time point and experimental group. Frequencies of defined populations were averaged amongst live, $CD45^+$ $CD11b^+$, $Ly6G^+$ cells.
B. Total number of $CD177^+$ $CD101^-$ ("immature) and $CD177^+$ $CD101^-$ ("mature") neutrophils in *Hh*AOM distal colons at 4, 10 and 15 weeks of model. Each dot represents one mouse and bars represent mean ±SEM. Data represents one independent experiment. Data were assessed by a Mann-Whitney test after D'Agostino-Pearson omnibus normality test (ns, no significance).
C. Total number of $CD177^+$ CD101 ("immature") and $CD177^+$ $CD101^-$ ("mature") neutrophils in distal colons of *Hh* and *Hh*AOM mice at 15 weeks. Data was assessed as in B.

Next to $Ly6G^+$ neutrophils, we observed an increased influx of $MHCII^+$ monocytes (**Figure 3.12A, orange box**) within the inflamed distal colon relative to proximal colon of *Hh*AOM mice at 10 and 15 weeks (**Figure 3.12B**). However, in contrast to similar neutrophil numbers between week 10 and week 15, we observed a decrease in $MHCII^+$ monocytes in the proximal and distal colon of *Hh*AOM mice (**Figure 3.12C**). This decrease was not observed in $MHCII^-$ monocytes (**Figure 3.12D**).

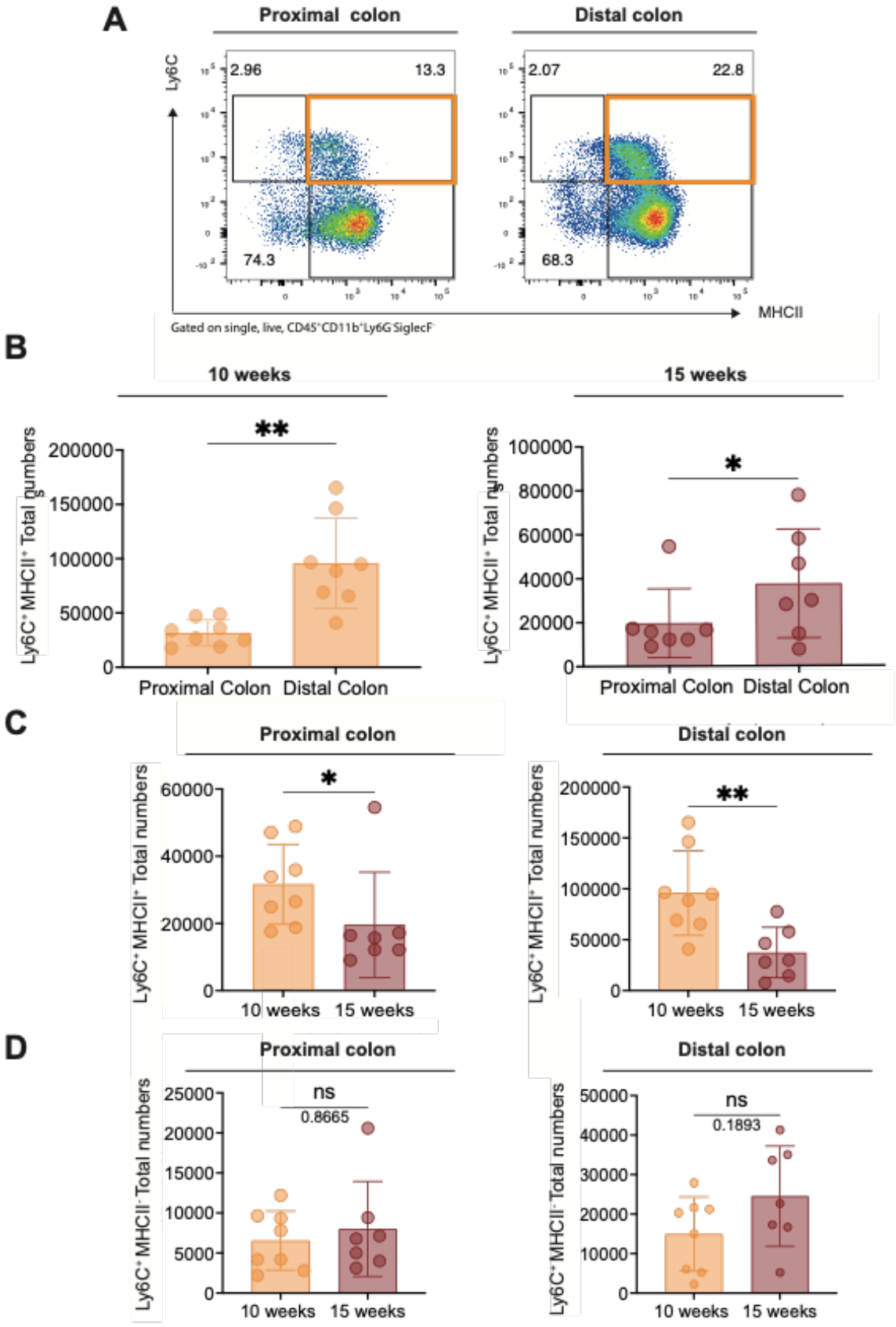

A
Proximal colon
Distal colon
Ly6C
2.96
13.3
74.3
2.07
22.8
68.3
MHCII
Gated on single, live, CD45+CD11b+Ly6G SiglecF-
B
10 weeks
15 weeks
Ly6C+ MHCII+ Total numbers
**
*
Proximal Colon
Distal Colon
C
Proximal colon
Distal colon
*
**
10 weeks
15 weeks
D
Ly6C+ MHCII- Total numbers
ns
0.8665
0.1893

**Figure 3.12 MHCII$^+$ Ly6C$^+$ monocytes are increased in the distal colon but decreased over time.**
129SvEvRAG$^{-/-}$ mice were gavaged with *Hh* and administered AOM (Week 10 n=8, Week 15 n=7). Upon sacrifice, proximal and distal colons were harvested and single cell suspensions analysed by flow cytometry. Each dot represents one mouse and bars represent mean ±SEM. Week 10 represent two independent experiments and week 15 represent one independent experiment.

A. Representative flow cytometry plots of colons at 10 weeks gated on live, CD45$^+$ CD11b$^+$ Ly6G$^-$ SiglecF$^-$ single cells.
B. Comparison of total number of MHCII$^+$ monocytes in proximal and distal colons. Data were assessed by a paired t-test after D'Agostino-Pearson omnibus normality test (** $p < 0.01$, * $p < 0.05$).
C. Comparison of total number of MHCII$^+$ monocytes in proximal and distal colons at 10 and 15 weeks. Data were assessed by a Mann-Whitney test after D'Agostino-Pearson omnibus normality test (** $p < 0.01$).
D. Comparison of total number of MHCII$^-$ monocytes in proximal and distal colons at 10 and 15 weeks. Data were assessed by a Mann-Whitney test after D'Agostino-Pearson omnibus normality test (** $p < 0.01$).

### 3.3.5 S100A8, IL-1β, IL-17A and IL-22 are increased in the distal colon of HhAOM mice

In order to not only probe the cellular composition of the distal colon, but also measure their secreted output, we investigated the expression of cytokines in our dataset. We first generated a heatmap of the normalised gene counts from untreated and *Hh*AOM mice for candidate cytokines at 10 weeks which illustrated a distinct expression pattern amongst proximal and distal colon samples. Among others, the mRNA of *S100a8*, *S100a9 Cxcl1, Il17a and Osm,* was enriched in the distal colon at 10 weeks, whereas *Il33 and Il12a* showed higher expression in the proximal colon. In contrast, the expression of *Ifng*, *Csf2*, *Tnf* and *Il1b* mRNA was similar in early tumorigenic proximal and distal colons. No differences were detected in *Selp* and *Il33* expression between untreated or HhAOM conditions (**Figure 3.13***).*

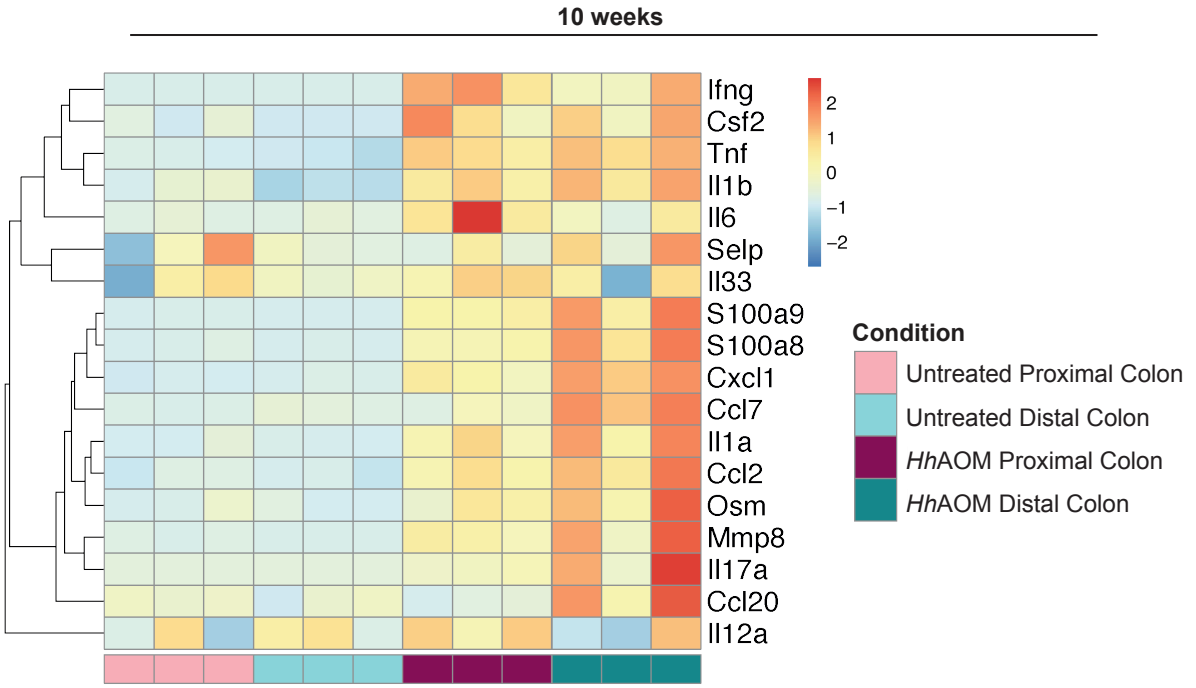



**Figure 3.13 mRNA expression of target genes depicts a location-driven expression pattern in early tumorigenesis.**
Gene counts of the longitudinal RNA sequencing dataset were normalised and downstream unbiased analysis performed on the previously identified genes. A heatmap of genes (normalised counts per million, *pheatmap*) was generated depicting proximal and distal colons of untreated and *Hh*AOM mice at 10 weeks.

In order to examine cytokine expression at a protein level, we created single cell suspensions from the proximal and distal colons of untreated or *Hh*AOM mice at 10 weeks and cultured these in serum-free media (**2.4.3**). After 16 hours cell supernatants were collected and a Luminex assay performed for all cytokines previously analysed within our RNA sequencing data. We generated a heatmap of the average pg/ml measured for each respective cytokine; CXCL1, IL-12A and MMP9 values are excluded due to unreliable standard curves and levels of IFN-γ, IL-33, CCL20 and OSM were below the standard curve. However, in contrast to mRNA expression, we observed no changes in IL-6, GM-CSF, CCL2 or TNFa secretion in untreated relative to *Hh*AOM conditions. (**Figure 3.14**).

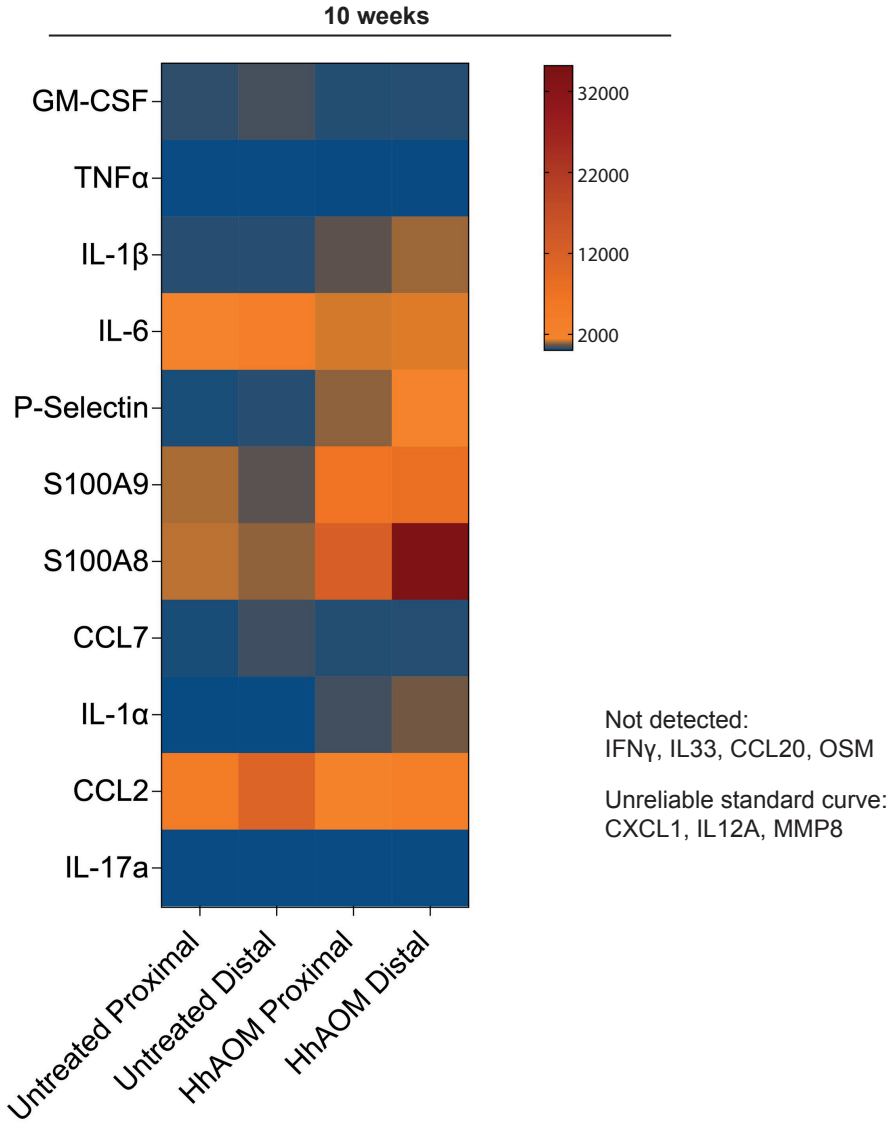
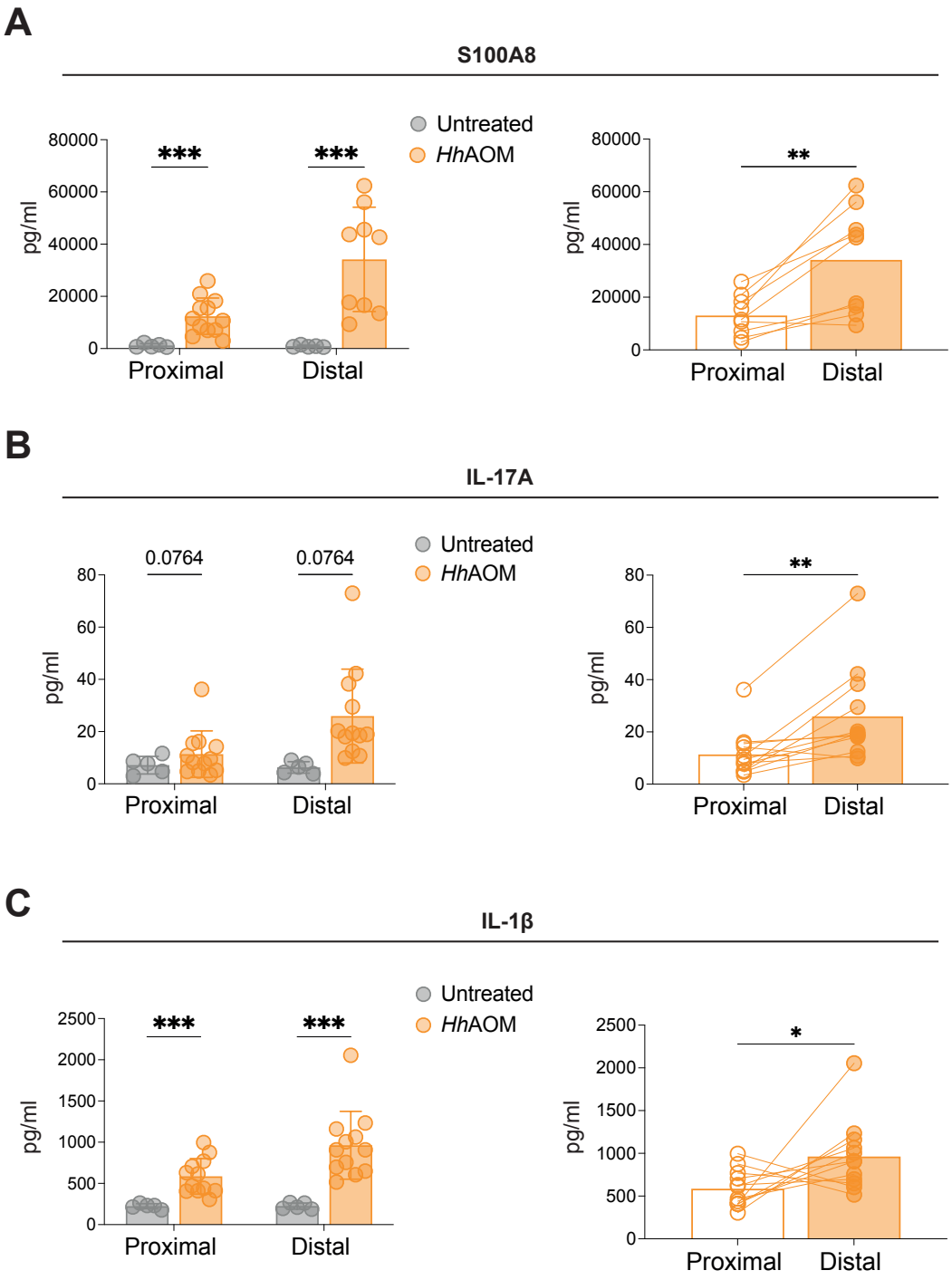



**Figure 3.14 Heatmap of Luminex data shows increase of certain cytokines in cell supernatants.**
129SvEvRAG$^{-/-}$ mice were left untreated (n=5) or exposed to *Hh* and AOM (n=12). Upon sacrifice at week 10, proximal and distal colons were harvested for cell supernatant generation. Single cell suspensions were incubated in media for 16h. Supernatants were harvested and used for cytokine measurements by Luminex. A heatmap was generated using the average pg/ml of each measured cytokine per condition (*GraphPad Prism*). Data represents two independent experiments.

The S100A8 subunit of the heterodimer calprotectin (S100A8/S100A9), IL-17A and IL-1β were secreted to a greater extent in pre-tumorigenic inflammation and significantly increased in the distal colon supernatants compared to proximal colon counterparts (**Figure 3.15**).




**<u>Figure 3.15</u> S100A8, IL-1β and IL-17A production are increased in the distal colon of *Hh*AOM mice**

129SvEvRAG$^{-/-}$ mice were left untreated (n=5) or exposed to *Hh* and AOM (n=12). Upon sacrifice at week 10, proximal and distal colons were harvested for cell supernatant generation. Single cell suspensions were incubated in media for 16h. Supernatants were harvested and used for cytokine measurements by Luminex. Data represents two independent experiments. Each dot represents one mouse and bars represent mean ±SEM. Untreated versus *Hh*AOM analysis was performed by Kruskal-

Wallis test after D'Agostino-Pearson omnibus normality test and proximal versus distal data was assessed by Wilcoxon test.

A. Measurements of S100A8 in cell supernatants (*** $p < 0.001$, ** $p < 0.01$).
B. Measurements of IL-17A in cell supernatants (** $p < 0.01$).
C. Measurements of IL-1β in cell supernatants (*** $p < 0.001$, * $p < 0.05$).

Lastly, analysis of bulk tissue gene expression (**Figure 3.16 A**) and secreted protein levels of IL-22 from colon-derived cell cultures (**Figure 3.16 B**) showed increases in inflamed distal colon samples as compared to proximal colon.

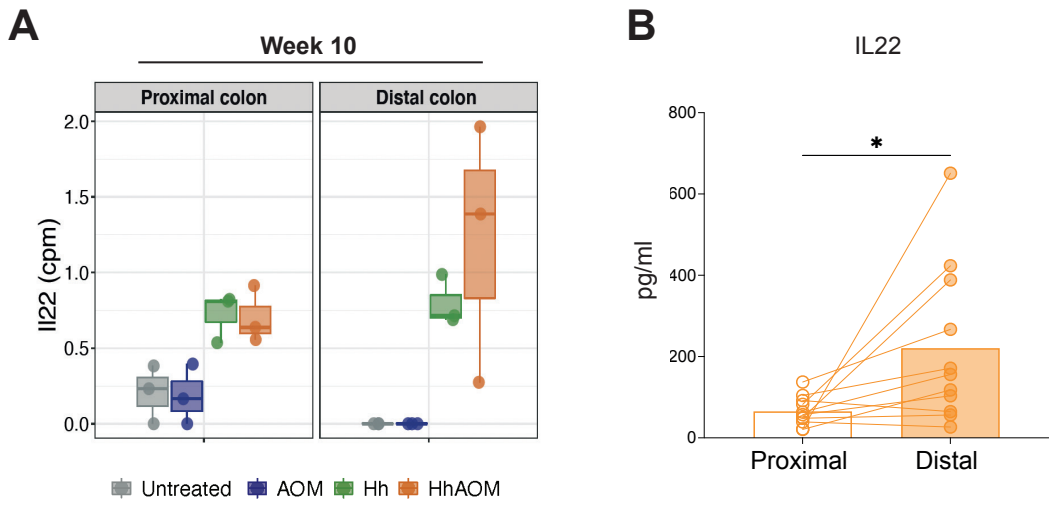



**Figure 3.16 IL-22 protein is greater in the distal colon during early tumorigenesis.**
129SvEvRAG$^{-/-}$ mice were left untreated (n=3), injected with AOM (n=3), gavaged with *Hh* (n=3) or exposed to both (n=3). Upon sacrifice at week 10, proximal and distal colons were harvested for bulk RNA sequencing. Additional *Hh*AOM treated proximal and distal colons were collected for cell supernatant generation (n=11). Single cell suspensions were incubated in media for 16h and supernatants were harvested ELISA measurements.

A. Normalised gene counts per million of *IL22* at 10 weeks in longitudinal RNA sequencing dataset.
B. Measurement of IL-22 in cell supernatants. Each dot represents one mouse and bars represent mean ±SEM. Data represents two independent experiments. Data was assessed with a Welch's t-test after D'Agostino-Pearson omnibus normality test (*, $p < 0.05$).

## 3.4 Discussion

By investigating this pathobiont-driven cancer model in a spatiotemporal manner, we found that colitis and cancer are characterised by a predominant neutrophil mRNA signature which correlates with an increase disease score in the distal colon. We further verified this at a cellular level and observed a higher abundance and number of neutrophils in the distal colon at early and late time points of colitis as well as cancer. In particular, the frequency of neutrophils increased over the course of disease, suggesting a possible role in tumorigenesis within this model.

### 3.4.1 The distal colon niche supports CRC.

Initial clustering analysis revealed that the colon location explained a considerable amount of variation in our longitudinal dataset, much more so than treatment with the mutagen AOM or progressive timepoints. This was in line with histopathological scoring which showed increased inflammation and tumorigenesis in the distal colon relative to the proximal colon in the *Hh*AOM model. We therefore concluded that the nature of the distal colon environment is a key factor in disease. Central to this environment is a significant enrichment in neutrophil-related genes and pathways in early and established tumorigenesis.

Investigating the myeloid compartment by flow cytometry, we observed an increase in $Ly6G^+$ neutrophils as well as $MHCII^+$ monocytes in both colitis and early CAC. Notably, the increase in neutrophils was specific to the distal colon at early and late timepoints. This contrasts data on the inflammation-driven increase in $MHCII^+$ monocytes which was high in early disease but decreased upon tumour establishment, a finding which warrants further investigation. *In situ*

analysis identified clusters of Ly6G$^+$ neutrophils in close proximity to the tumour which we hypothesis to be associated with crypt abscesses. Lastly, the mRNA expression of specific cytokines including *S100A8*, *S100A9, Il22*, *Il17a* and *OSM* were upregulated in the distal colon relative to the proximal colon during early tumorigenesis further showing location-specific exacerbated inflammation. Given evidence for different immune and microbial niches throughout the colon[171], alongside location-specific effects of the microbiota in other CRC models[141], such location dependent studies are crucial.

Distal colon specific pathology was in line with an earlier study which revealed that enterotoxic *B. fragilis* can create a pro-tumorigenic environment in the distal colon by driving locally-restricted polymorphonuclear myeloid-derived suppressor cell (PMN-MDSC) accumulation. This was shown to be due to enhanced NfκB signalling and subsequent *Cxcl1*, *Cxcl2* and *Cxcl5* expression in distal colon IECs[144]. Notably, these chemokines are also expressed to greater extent in the distal colon of the model used here. Further cytokines that shape the distal colon environment in our model are discussed in detail in section **3.4.5.**

### 3.4.2 The ambiguous role of neutrophils in cancer

Neutrophils are polymorphonuclear cells (PMNs) that play a crucial role in the innate immune response and are often associated with inflammatory diseases. In colitis, the infiltration of PMNs is a big contributor to disease severity scores and progression[172], and neutrophils have been recently linked to therapeutic non-response and distinct pathotypes in IBD[173]. Furthermore, PMNs have the highest correlation with adverse outcomes in colon, lung and pancreatic cancer[93]. In the *Hh*AOM model used here, colitis scores positively correlated with

neutrophil amounts which indicates that they may contribute to severe inflammation and subsequent cancer.

However, the precise functional role(s) of neutrophils in cancer remain ambiguous (**1.3.1**). Indeed, a growing body of literature strongly suggests that neutrophils are a crucial component of tumour-promoting inflammation with multiple studies depleting neutrophils and observing decreased tumorigenesis[96,144,174,175]. On the contrary, a DSS-driven cancer model performed with $LysM^{cre}Mcl1^{-/-}$ mice, showed a protective role for neutrophils. In that study neutrophil depletion led to a higher bacterial load in tumours which the authors proposed promoted further deleterious inflammation[176]. In line with this, the cytokine Tgfβ has been reported to drive pro-tumorigenic properties of N2-like TANs and if the microenvironment was void of this cytokine, cytotoxic 'N1' TANs were essential in dampening disease. These publications highlight the importance of the surrounding tumour microenvironment in shaping neutrophil function. Therefore, although we hypothesise that the observed marked neutrophil expansion is playing a deleterious role in cancer progression, neutrophils may also reinforce barrier protection especially in early stages of disease. Overall, the role of neutrophils in the cause or effect of both colitis and cancer is yet to be elucidated fully with more longitudinal and functional data needed.

### 3.4.3 Neutrophil subsets increase over the course of disease.

It is now widely accepted that neutrophils are not a homogenous population of cells. Next to distinct TAN subtypes classified by either cytotoxic or immune suppressive capabilities[108], other studies have discovered the cell surface marker CD101 to segregate immature

neutrophils from mature neutrophils in the circulation and tissue sites[21]. In a pancreatic cancer model, increases of both mature ($CD101^+$) and immature ($CD101^-$) neutrophil numbers were found in blood and tumour sites, in line with our observed increases of both populations in the spleen and distal colon. However, by contrast with the pancreatic cancer data, increases in *Hh*AOM distal tumours were not significant, possibly due to a premature cancer timepoint or low sample numbers.

Evrard and colleagues propose $CD101^-$ immature neutrophils to be associated with tumorigenesis based on a positive correlation with tumour load but those studies lack mechanistic data. However, their study does reveal mature $CD101^+$ neutrophils to harbour a higher migratory capacity and to be more prone to DNA-damaging ROS induction[21,95]. This raises the question of whether $CD101^-$ neutrophils represent an immature reservoir and/or biomarker of disease whereas the $CD101^+$ neutrophils are equipped with tumour promoting functions in the TME. However, definitive data is missing due to a lack of a uniform neutrophil nomenclature and clear definitions of TANs, PMN-MDSC or immature/mature neutrophils [89].

### 3.4.4 The HICCS locus drives neutrophil influx

Transcriptomic analysis performed on colons from $129SvEvRAG^{-/-}$ and control $RAG^{-/-}$ mice, revealed neutrophil chemotaxis to be among the top pathways differentially regulated after *Hh* infection[169]. This highlights the interaction of the microbiome and the previously identified HICCS susceptibility locus as upstream of the observed neutrophil mediated pathology. However, in $ALPK1^{-/-}RAG^{-/-}$ mice, the neutrophil influx observed was not distal colon specific, showing that other genes within this locus may play an important role in determining lesion

site. Studies are needed to assess if similar genetic predispositions are found in cancer patients and if these harbour a similar pathology. Also, due to the high similarity of the immune landscape in colitis and cancer within our model, we predict mechanisms found in this setting could identify an innate-driven mechanism that explains the enlarged risk of cancer development when pre-diagnosed with ulcerative colitis[70]. Interestingly, colleagues have identified a subset of colitic patients with a strong neutrophil signature, however the genetic make-up of these patients is unknown[173].

### 3.4.5 A distinct cytokine network in the distal colon

It has been well established that secreted cytokine networks are critical mediators of inflammation and tumorigenesis[74], and they were therefore probed in the RNA sequencing dataset and quantified at protein levels. The distal colon showed a distinct expression pattern relative to the proximal colon with mRNA of well-known pro-inflammatory molecules such as *S100A8*, *S100A9, Il17a, Osm and Ifng* increased.

The S100A8/S100A9 heterodimer calprotectin is expressed by neutrophils and a well-known biomarker for IBD which correlates with disease severity[177]. The enhanced expression in distal relative to proximal colons further validates previous observations of an exacerbated inflammatory environment present in the distal colon during early colon cancer. Oncostatin M (OSM) is upregulated in IBD patients[178] and suggested to function as an inducer of mesenchymal plasticity in CRC[179]. In contrast to upregulated mRNA levels in inflammation, we failed to detect OSM at a protein level. However, this could be technical and should be verified by other methods such as Luminex and/or ELISAs on whole tissue lysates. Similarly,

the mRNA expression of *Ifng,* an important mediator of anti-tumour immunity[82] was upregulated in *Hh*AOM proximal and distal colons but undetected at the protein level. A lack of IFNγ production could potentially aid early tumour progression in the *Hh*AOM model but further studies are required to confirm this. Other cytokine gene expression increased at 10 weeks was not detected at a protein level, including GM-CSF and CCL2. Further assessment of this observation will be of great interest due to their importance in myeloid cell attraction and maintenance[180] alongside our preliminary data showing downregulation of inflammatory monocytes over time. Lastly, protein levels of known cancer-promoting cytokine IL-6[74] were unchanged which was in line with previous studies in the *Hh*AOM model[124] suggesting a minor role in this disease setting.

Consistent with our RNA sequencing data, cells isolated from the inflamed distal colon of *Hh*AOM mice secreted more IL-1β, IL-17A and IL-22 than those from the proximal colon. Colonic ILC3s are a dominant source of IL-17A and IL-22 and both have been shown to maintain tumorigenesis in this and other colon cancer models[124,153]. We therefore hypothesize that these factors are a crucial part of an early distal-specific tumour-promoting network. In particular IL-17A has been studied in relation to neutrophils, due to its ability to act on epithelial cells and promote neutrophil chemotaxis[124,144]. In the ETBF model of CRC, IL-17A$^{-/-}$ mice showed less mRNA expression of CXCL1, CXCL2 and CXCL5 which resulted in a lower number of PMN-MDSCs and decreased tumorigenesis in the distal colon[144]. Previously in our lab, blocking IL-17A reduced tumour numbers alongside a lower granulocyte influx[124]. Therefore, the increased secretion of IL-17A which we observe at 10 weeks could explain the higher number of neutrophils in the distal colon. Notably, anti-IL-22 treatment prevented a

colonic neutrophil influx[124]. Whether both IL-17A and IL-22 have overlapping or synergistic roles to promote innate recruitment in CAC remains to be elucidated.

IL-1β is a potent inducer of IL-17A and IL-22 in Th17, γδ T-cells and ILC3s[96,181] and its production positively correlates with levels of IL-17A and IL-22. As neutrophils comprise approximately 50% of leukocytes in the *Hh*AOM distal colon, we propose that these cells are the main source of IL-1β in our model and hypothesise that neutrophil-derived IL-1β subsequently triggers IL-22 and IL17A secretion by colonic ILC3s **(Figure 3.17).** By binding to receptors on epithelial cells, IL-22 and IL-17A could upregulate production of neutrophil chemoattractants in a positive feedback loop. Even though it is well-known that neutrophils are not the sole cellular source of IL-1β production, they have previously been shown to be the main reservoir in another microbe-driven model of CAC[174]. Intracellular cytokine staining for IL-1β and/or sorting neutrophil populations could address this hypothesis.

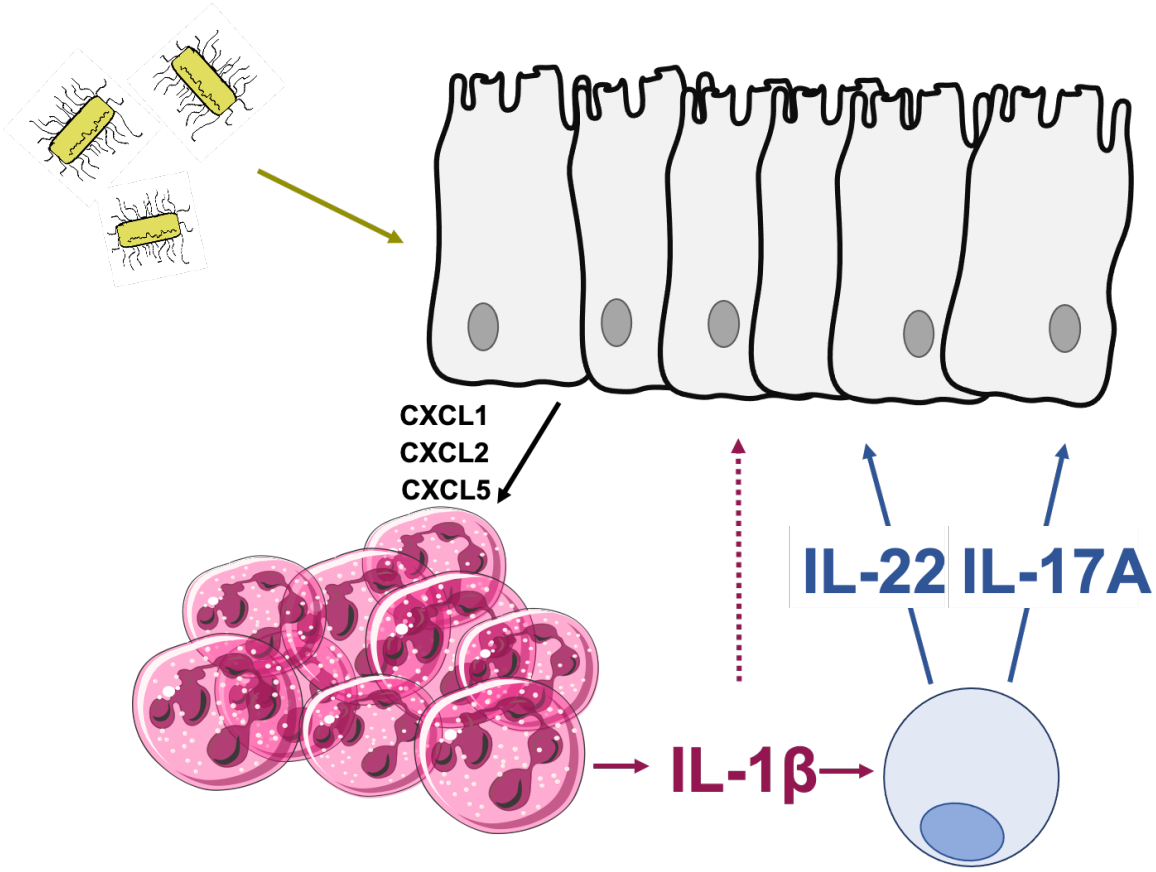



**Figure 3.17 Proposed mechanism driving neutrophil influx into the distal colon of *Hh* and *Hh*AOM mice.**
*Hh* creates a dysbiotic environment which leads to a distal colon influx of neutrophils. Distal colon neutrophils secrete IL-1β which stimulates the secretion of IL-22 and IL-17A by ILC3s. Both cytokines have been shown to in turn trigger chemokine expression of epithelial and cancer cells resulting in a positive neutrophil attracting loop. Also, IL-1β has been shown to promote tumorigenesis directly by stimulating intestinal epithelial cell proliferation.

CXCLs, chemokines; IL, interleukin

Alternatively, IL-1β also could act directly on cancer cells or influence neutrophil function in an autocrine manner[181]. In particular, the antimicrobial function of neutrophils is dependent on IL-1 signalling[176,181] and the AMP calprotectin was shown to be expressed by neutrophils in mammary tumour-mice[96]. Therefore, we propose that IL-1β and the downstream antimicrobial features of TANs could contribute to tumorigenesis.

## 3.5 Conclusion

Results presented in this chapter identify the distal colon as the site of significantly higher inflammation, tumorigenesis, and influx of immature and mature neutrophils during the course of *Hh*AOM disease. These changes are all observed in mice administered *Hh* regardless of AOM administration, highlighting the important role of this pathobiont in shaping the tumour promoting microenvironment. We hypothesise that these effects are driven by the enhanced production of, among other cytokines, calprotectin, IL-17A, IL-22 and IL-1β which were observed to form a distal-colon enriched cytokine network.

## 3.6 Future outlook and limitations

Next steps should focus on dissecting how neutrophils shape colon pathology and how they might set the scene for colitis-associated cancer. One approach to do this would be to deplete neutrophils in our genetically-susceptible model of either colitis or cancer at different stages of disease. This however remains a technical challenge due some depletion strategies being non-specific[42] and others appearing to reduce neutrophil numbers just in the circulation and not in primary tumours[107].

Alternatively, after determining whether neutrophils are the main source of cleaved and active IL-1β by cell sorting with subsequent culture and ELISA, this cytokine could be blocked *in vivo*. It will also be important to understand if all neutrophils secrete IL-1β or if there indeed is a subset in the cancer and/or colitis triggered by specific stimuli. This could be addressed by quantitative spatiotemporal imaging of Ly6G and IL-1β, adding in CD101 to understand if

immature or mature cells are the main source in tumour lesions. In general, such imaging techniques will help us understand the role of cytokines and their cellular sources during disease progression. Spatial techniques can be complemented by measuring cytokine protein levels by intracellular cytokine staining and flow cytometry.

In particular, data on the function of neutrophils in disease settings is limited. By comparing intensity data from proteomics on bulk neutrophils we will be able to unbiasedly assess differences in neutrophils isolated from mice with colitis versus early and late cancer. To this end, initial mass spectrometry of sorted colon neutrophils at week 10 has proven to be technically achievable (data not shown). Any notably differentially expressed proteins could then be analysed *in situ* to validate differences and determine localisation relative to dysplastic crypts, tumours or other pathotypes.

There are of course limitations to be aware of, such as the immunodeficiency of our mouse model and missing information of the mutational status of cancer cells. Both factors are known to be crucial in shaping cancer biology[79]. Nevertheless, this reductionist model harbours great potential to discover novel neutrophil driven mechanisms in colitis and colon cancer.

# 4 Identification of the enzyme PAD4 as a potential cancer promoting enzyme in CRC.

## 4.1 Introduction

CRC is not only driven by the accumulation of genetic mutations[72], but also by epigenetic modifications such as DNA methylation which alters chromatin structure and DNA accessibility[182,183]. Histone citrullination, catalysed by peptidylargine deiminase (Pad) enzymes 1-4, is another epigenetic modification which drives chromatin de-condensation and is involved in autoimmune diseases as well as carcinogenesis[184]. In particular, peptidylargine deiminase 4 (PAD4/Padi4) has been studied extensively for its role in neutrophil extracellular trap (NET) formation[44], but has also been found to promote tumorigenesis in murine models of melanoma[104,105]. Notably in metastasis, NETs reportedly capture circulating tumour cells which aids colonisation at sites such as the lung and liver[185-188]. A recent study of sustained lung inflammation revealed that NET-derived proteases neutrophil elastase and matrix metalloproteinase 9 can cleave laminin and thereby awaken dormant cancer cells[106].

With regards to cancer cells, the role of PAD4 has been less explored and results thus far are ambiguous. In breast and lung cancer cell lines, PAD4 expression reportedly represses epithelial to mesenchymal transition (EMT)[189,190], a crucial hallmark of cancer [79]. Additionally, studies have linked PAD4 expression to the tumour suppressor protein p53, with DNA damage caused by irradiation triggering PAD4 induction, histone citrullination and downstream

apoptosis of cancer cells[191,192]. In contrast, citrullination of the ECM in CRC promoted liver metastasis and PAD4 inhibition decreased subcutaneous tumourigenesis[193]..

Finally, conflicting studies have implicated PAD4 and citrullination in the regulation of two further cancer hallmarks: uncontrolled epithelial cell proliferation and stemness. In multipotent hematopoietic stem cells PAD4 and downstream effects on histone H3 promoted cell proliferation[194]. Conversely, PAD4-driven citrullination of histone H1 resulted in the expression of key genes such as *Pou5f1* which are necessary for stemness in murine pluripotent stem-cells[195].

## 4.2 Aim

Cells of the myeloid compartment characterised in the previous chapter showed no apparent differences between colitis, early and established CAC. However, we and others have established that an interplay of inflammation with a mutagenic insult is needed for cancer development in this genetically susceptible model[124] , as neither *Hh* nor AOM alone promote tumorigenesis. Therefore, this chapter aims to identify differences between *Hh* and *Hh*AOM treatment groups in order to pinpoint a cancer-driving mechanism. Amongst 118 differentially expressed genes, various antimicrobial genes including defensins and lysozyme were upregulated in *Hh*AOM mice. We however decided to focus on the enzyme PAD4 due to 1) its role in tumour-promoting NETosis given the neutrophil-rich environment defined in Chapter 3, and 2) the availability of tools to study this molecule *in situ* and *in vivo*. Despite a previously characterised neutrophil-rich environment in both colitis and cancer, we

hypothesised PAD4-driven NETosis to be a cancer-specific neutrophil function in this CAC model.

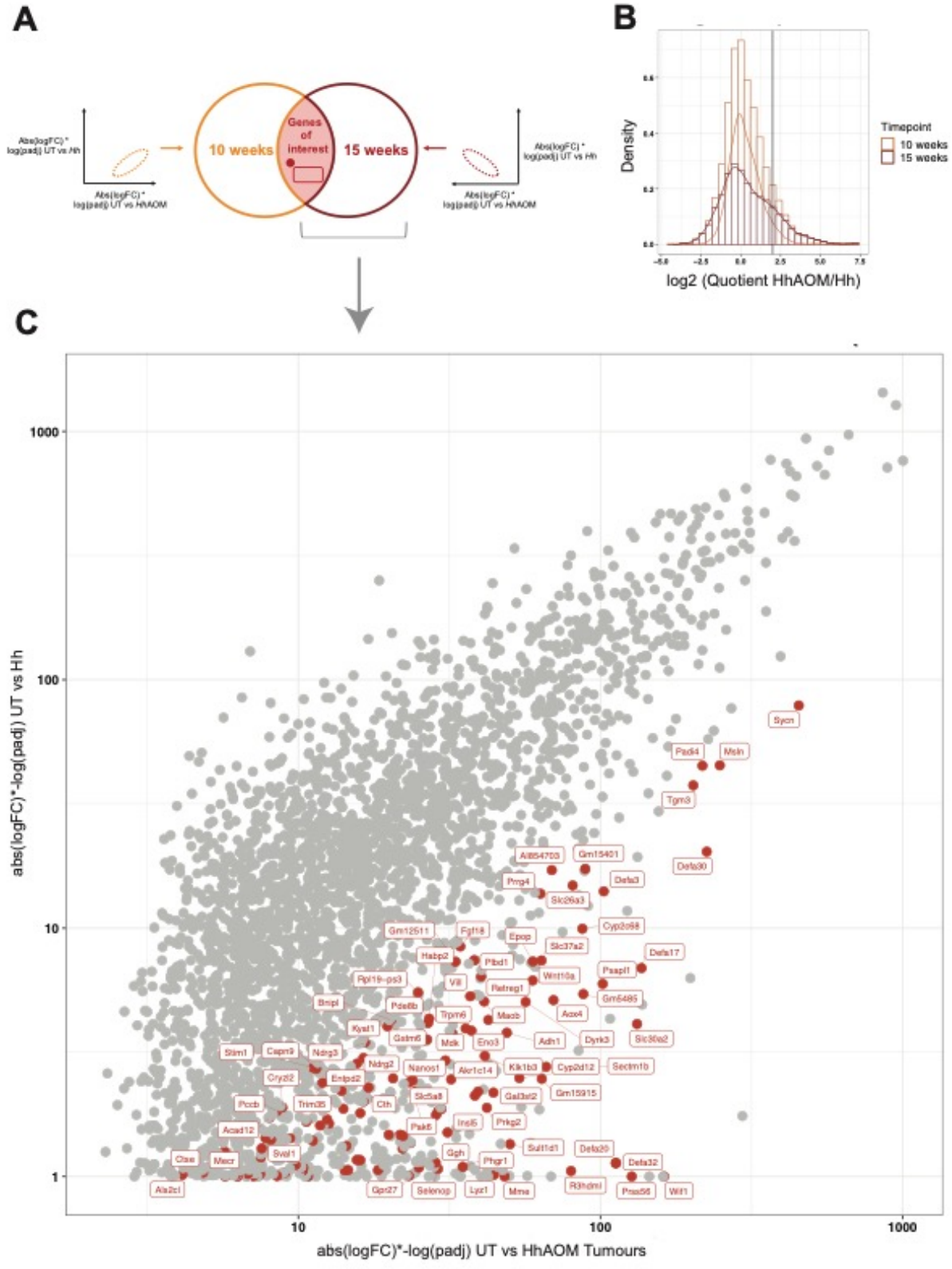
## 4.3 Results

### 4.3.1 118 genes are differentially expressed in tumorigenesis as compared to colitis

To determine differentially expressed genes in cancer versus colitis, we performed *DESeq* on 10- and 15- week distal colon samples. *Hh* and *Hh*AOM samples were compared to their respective untreated distal colon controls, using tumours as the 15-week *Hh*AOM tissue of interest. Post-*DESeq* analysis, we multiplied the absolute $\log_2$ fold change with the adjusted p-value (log(padj)), creating a single value per gene termed the 'effect size' which reflects the magnitude of differential expression. Subsequently we divided *Hh*AOM effect sizes with *Hh*-only effect sizes which resulted in a quotient used for downstream filtering (**Figure 4.1 A**). Plotting *Hh*AOM effect sizes against those of *Hh*-only controls identified genes that were enriched specifically in *Hh*AOM (close to the '*Hh*AOM' x axis).

We hypothesised that a high quotient will pinpoint genes dysregulated to a larger extent in cancer versus colitis. In order to determine an appropriate quotient cut-off for downstream analysis, we plotted the binary logarithm of calculated quotients as a histogram for both time points (*ggplot*) and filtered out all genes with a quotient lower than '4' (**Figure 4.1 B**). Finally, we overlayed the filtered results from our 10-week timepoint with those of our 15-week tumour analysis and identified 118 genes increased or decreased in tumorigenesis compared

to colitis, which we termed ‘genes of interest’ (**Figure 4.1 A**). These 118 genes are labelled in a dot plot depicting all gene effect sizes calculated at week 15 (**Figure 4.1 C**).




**<u>Figure 4.1</u> Mathematical computation identifies 118 genes enriched or decreased in early and late cancer in comparison to colitis only controls.**

Differential expression analysis was performed comparing healthy distal (n=3) and *Hh*AOM treated distal tissue (n=3) harvested from mice upon sacrifice at week 10 and 15 (healthy; n=3 tumours; n=5).

The same analysis was also run comparing healthy distal colons (n=3) and *Hh* gavaged mice at week 10 (n=3) and 15 (n=5).

A. Schematic of the strategy to identify genes expressed to a greater or lesser extent between *Hh* colitic and *Hh*AOM tumorigenic mice. Post differential expression (DESeq) analysis, the absolute $log_2$Fold changes (Abs(logFC)) and adjusted p values (log(padj)) of genes were multiplied to generate a single value ('effect size') representing the differential expression of each gene in a given condition. Subsequently, the quotient of *Hh*AOM and *Hh* effect sizes was calculated for each gene at 10 weeks (orange) as well as 15 weeks (red) of the model. A quotient 'cut-off' of 4.0 was chosen to use for overlap analysis. To illustrate this, the effect sizes of genes in *Hh*AOM and *Hh* conditions were plotted against each other on an x (*Hh*AOM) and y (*Hh*) axis.
B. Histogram illustrating the binary logarithm of calculated quotients (HhAOM/Hh) with a cut-off mark at '2' (corresponding to the original quotient value of '4').
C. Dot plot showing effect sizes calculated at week 15 (grey) and plotted on x (*Hh*AOM) and y axis (*Hh*). Labelled genes were found enhanced or decreased in established tumours as well as early tumorigenesis by using the above-mentioned quotient cut-off.

To probe if our strategy identified altered gene expression in *Hh*AOM versus *Hh* only mice, we generated a heatmap (*pheatmap*) of the 118 gene list across all time points and treatment groups; this resulted in an unbiased clustering of *Hh*AOM samples in comparison to *Hh* (**Figure 4.2**). Pathway analysis of our gene list did not however identify any differentially regulated biological processes, probably due to the restricted list size (data not shown).

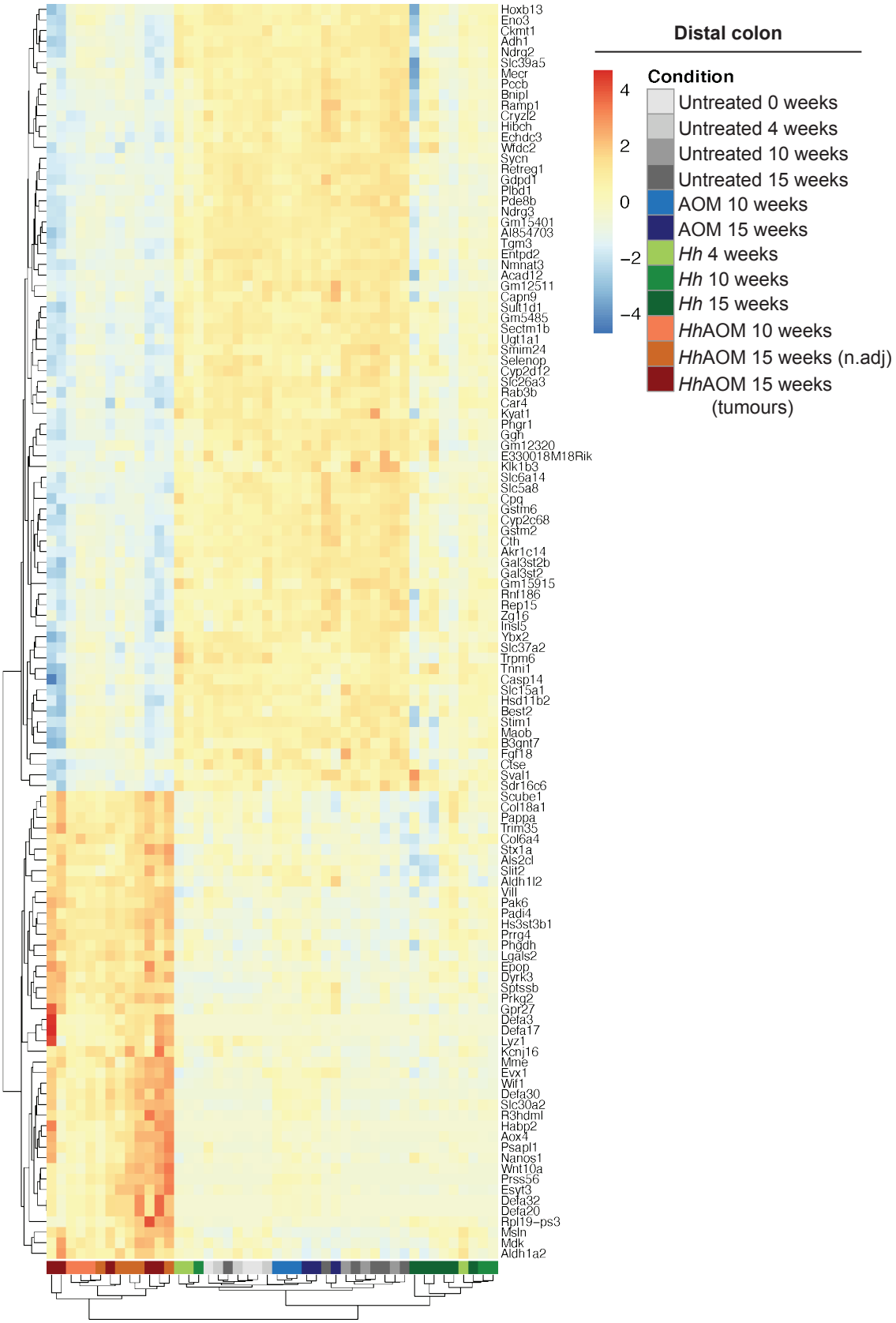

Distal colon
Condition
4
2
0
−2
−4
Untreated 0 weeks
Untreated 4 weeks
Untreated 10 weeks
Untreated 15 weeks
AOM 10 weeks
AOM 15 weeks
Hh 4 weeks
Hh 10 weeks
Hh 15 weeks
HhAOM 10 weeks
HhAOM 15 weeks (n.adj)
HhAOM 15 weeks (tumours)
Hoxb13
Eno3
Ckmt1
Adh1
Ndrg2
Slc39a5
Mecr
Pccb
Bnipl
Ramp1
Cryzl2
Hibch
Echdc3
Wfdc2
Sycn
Retreg1
Gdpd1
Plbd1
Pde8b
Ndrg3
Gm15401
Al854703
Tgm3
Entpd2
Nmnat3
Aca d12
Gm12511
Capn9
Sult1d1
Gm5485
Sectm1b
Ugt1a1
Smim24
Selenop
Cyp2d12
Slc26a3
Rab3b
Car4
Kyat1
Phgr1
Ggh
Gm12320
E330018M18Rik
Klk1b3
Slc6a14
Slc5a8
Cpq
Gstm6
Cyp2c68
Gstm2
Cth
Akr1c14
Gal3st2b
Gal3st2
Gm15915
Rnf186
Rep15
Zg16
Insl5
Ybx2
Slc37a2
Trpm6
Tnni1
Casp14
Slc15a1
Hsd11b2
Best2
Stim1
Maob
B3gnt7
Fgf18
Ctse
Sval1
Sdr16c6
Scube1
Col18a1
Pappa
Trim35
Col6a4
Stx1a
Als2cl
Slit2
Aldh1l2
Vill
Pak6
Padi4
Hs3st3b1
Prrg4
Phgdh
Lgals2
Epop
Dyrk3
Sptssb
Prkg2
Gpr27
Defa3
Defa17
Lyz1
Kcnj16
Mme
Evx1
Wif1
Defa30
Slc30a2
R3hdml
Habp2
Aox4
Psapl1
Nanos1
Wnt10a
Prss56
Esyt3
Defa32
Defa20
Rpl19-ps3
Msln
Mdk
Aldh1a2

**Figure 4.2 Heatmap of identified genes reveals HhAOM specific clustering.**
Gene counts of the longitudinal RNA sequencing dataset were normalised and downstream unbiased analysis performed on the previously identified genes. A heatmap of identified genes (normalised and scaled counts per million, pheatmap) was generated depicting an unbiased clustering of *Hh*AOM samples in comparison to *Hh* only samples.

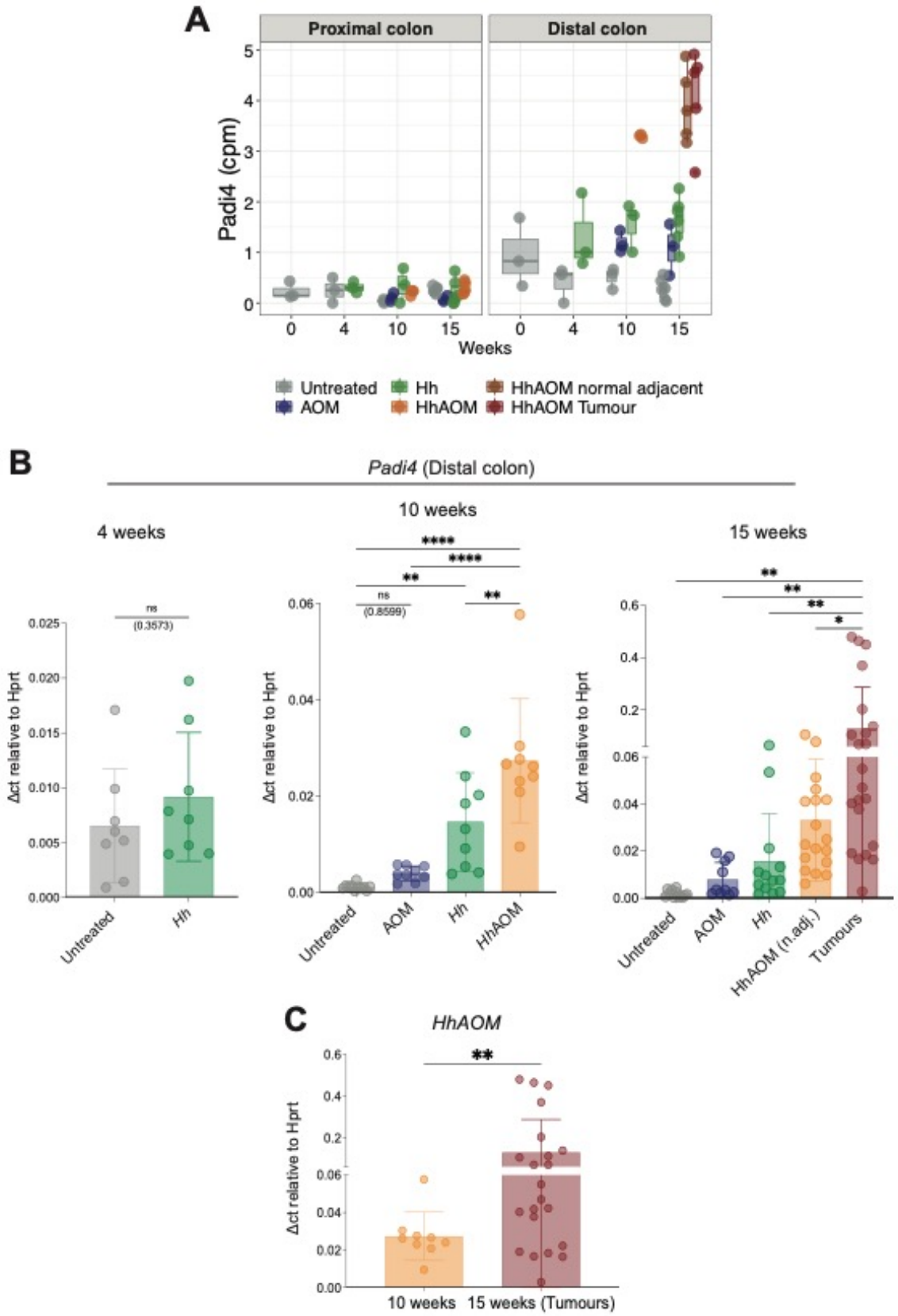
### 4.3.2 *Padi4* mRNA and protein is increased in distal colons of HhAOM mice.

Due to our neutrophil-rich environment in the area of cancer progression, we assumed some of our 118 candidate genes to be involved in neutrophil biology. In line with this, we were particularly interested in investigating the citrullination enzyme *Padi4* in more detail. Not only does PAD4 have a well-known role in the neutrophil function netosis[44], but netosis has further been linked to enhanced tumorigenesis in melanoma models[104,105] and the promotion of CRC metastasis[193]. We therefore hypothesised Padi4 to promote a cancer-specific neutrophil function in our model.

To evaluate the normalised counts of the *Padi4* genes in our dataset, we first plotted the gene counts per million (cpms) of *Padi4* across time points, locations and conditions and verified an increased expression in the distal colon. We detected a subtle upregulation in *Hh*- and AOM-only controls at 4 and 10 weeks respectively, but confirmed that expression was further elevated in *Hh*AOM samples both at 10 weeks and even more so at 15 weeks (**Figure 4.3 A**).

We validated this observation by performing quantitative PCR (qPCR) on samples from three independent experiments. *Padi4* mRNA was increased in inflamed conditions, with a higher magnitude of expression in mice exposed to both *Hh* and AOM (**Figure 4.3 B**). We further

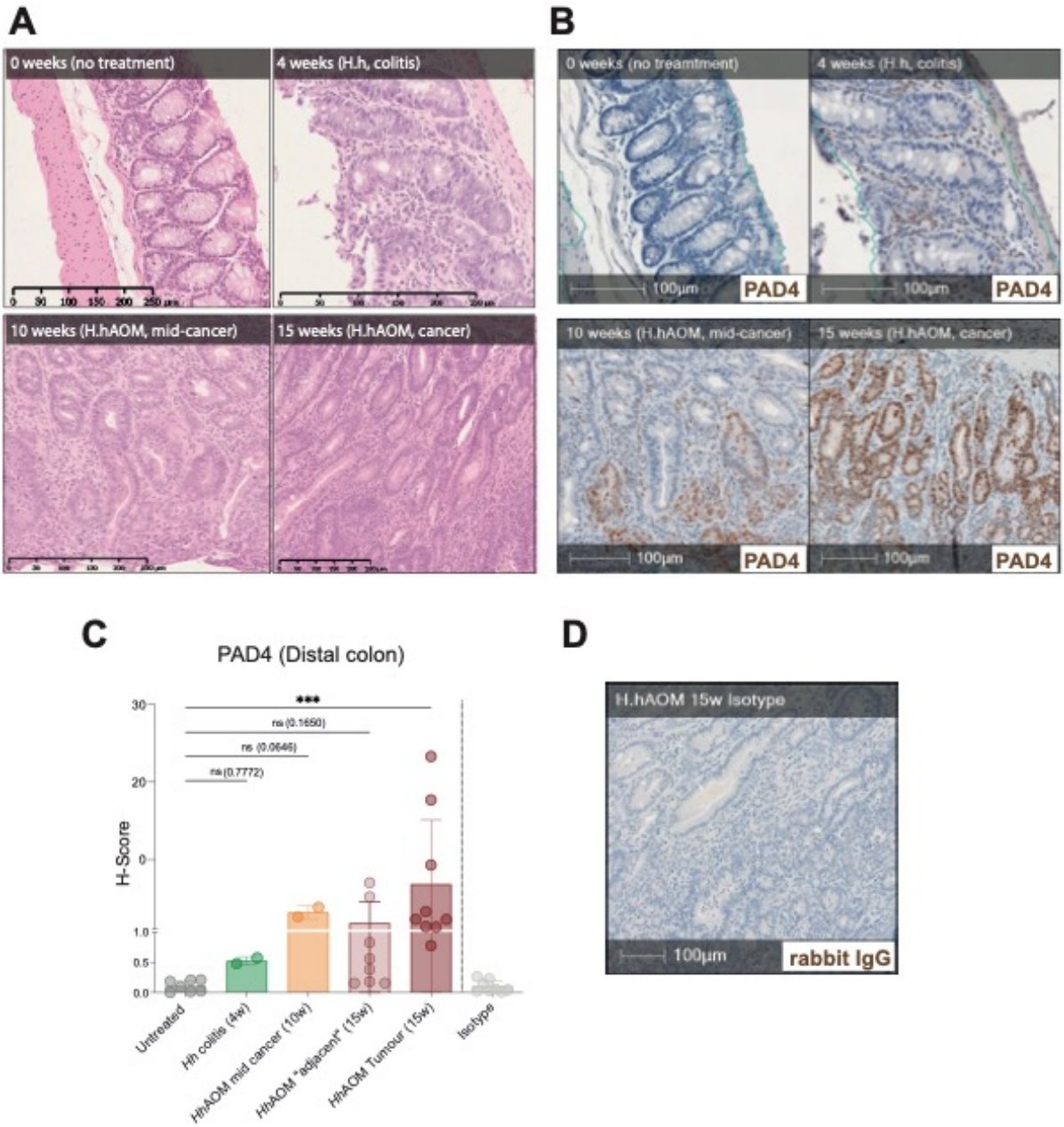
confirmed significantly increased *Padi4* expression in *Hh*AOM distal colons from 15-week as compared to 10-week mice (**Figure 4.3 C**).


A
Proximal colon
Distal colon
Padi4 (cpm)
Weeks
Untreated
AOM
Hh
HhAOM
HhAOM normal adjacent
HhAOM Tumour
B
Padi4 (Distal colon)
4 weeks
10 weeks
15 weeks
Δct relative to Hprt
ns
(0.3573)
ns
(0.8599)
Untreated
Hh
AOM
HhAOM
HhAOM (n.adj.)
Tumours
C
HhAOM
10 weeks
15 weeks (Tumours)

**<u>Figure 4.3</u> Increased *Padi4* expressed measured in distal colons of *Hh*AOM mice by qPCR.**
Gene counts of the longitudinal RNA sequencing dataset were normalised and expression of candidate gene *Padi4* in proximal and distal colon investigated.

A. Bar plot of normalised counts per million of *Padi4* in longitudinal RNA sequencing dataset.
B. *Padi4* qPCR of extracted distal colon mRNA at 4, 10 and 15 weeks of the model. Relative expression to the housekeeping gene *Hprt* is shown for each mouse from three independent experiments, with bars representing mean +/- SEM. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (*, $p < 0.05$; **, $p < 0.01$, ***, $p < 0.001$; ****, $p < 0.0001$; ns, no significance).
C. Comparison of *Padi4* expression in the distal colon of *Hh*AOM mice at 10 weeks and tumour tissue at 15 weeks. Data were assessed by a Welch's t-test (**, $p < 0.01$).

Alongside gene expression, we probed PAD4 protein expression by performing chromogenic stainings on FFPE colonic swiss rolls. PAD4 was observed in colitic, adenoma and adenocarcinoma tissue (**Figure 4.4 A**). Quantification of PAD4 stainings using HALO software revealed a significant upregulation of the PAD4 protein between untreated and *Hh*AOM tumours (**Figure 4.4 C**), with the isotype control staining negative (**Figure 4.4 D**).




**<u>Figure 4.4</u> PAD4 is significantly upregulated on protein level within *Hh*AOM cancer.**
129SvEvRAG$^{-/-}$ mice were left untreated (Week 0 n=3, Week 4 n=3, Week 10 n=6, Week 15 n=6), injected with AOM (Week 10 n=6, Week 15 n=7), gavaged with *Hh* (Week 4 n=3, Week 10 n= 5, Week 15 n=10), or exposed to both (Week 10 n=6, Week 15 n=9). Upon sacrifice, colons were harvested and embedded as swiss rolls. Chromogenic staining of PAD4 was performed on processed swiss rolls.

A. Representative H&E images.
B. Corresponding images of PAD4 expression (brown).
C. Bar plot of PAD4 expression in the distal colon. Area of proximal and distal colons were annotated manually and the amount of protein expression was automatically quantified by HALO® software (H-Score). The H-Score is shown from each mouse from three independent experiments (one experiment for 4- and 10-week timepoint), with bars representing mean +/-

SEM. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (**, $p < 0.01$; ns, no significance).

D. Representative image of the rabbit isotype control.

### 4.3.3 Epithelial PAD4 expression coincides with the loss of surface marker EpCAM

To assess the localisation of PAD4 expression relative to tumours, we performed immunofluorescence staining for PAD4 (magenta) and the epithelial marker EpCAM (green). In *Hh*AOM swiss rolls harvested from mice after 15 weeks, we observed PAD4 expression in multifocal areas and locally extensive adenocarcinomas coinciding with a striking lack of EpCAM expression (**Figure 4.5 A & B, white arrows**).

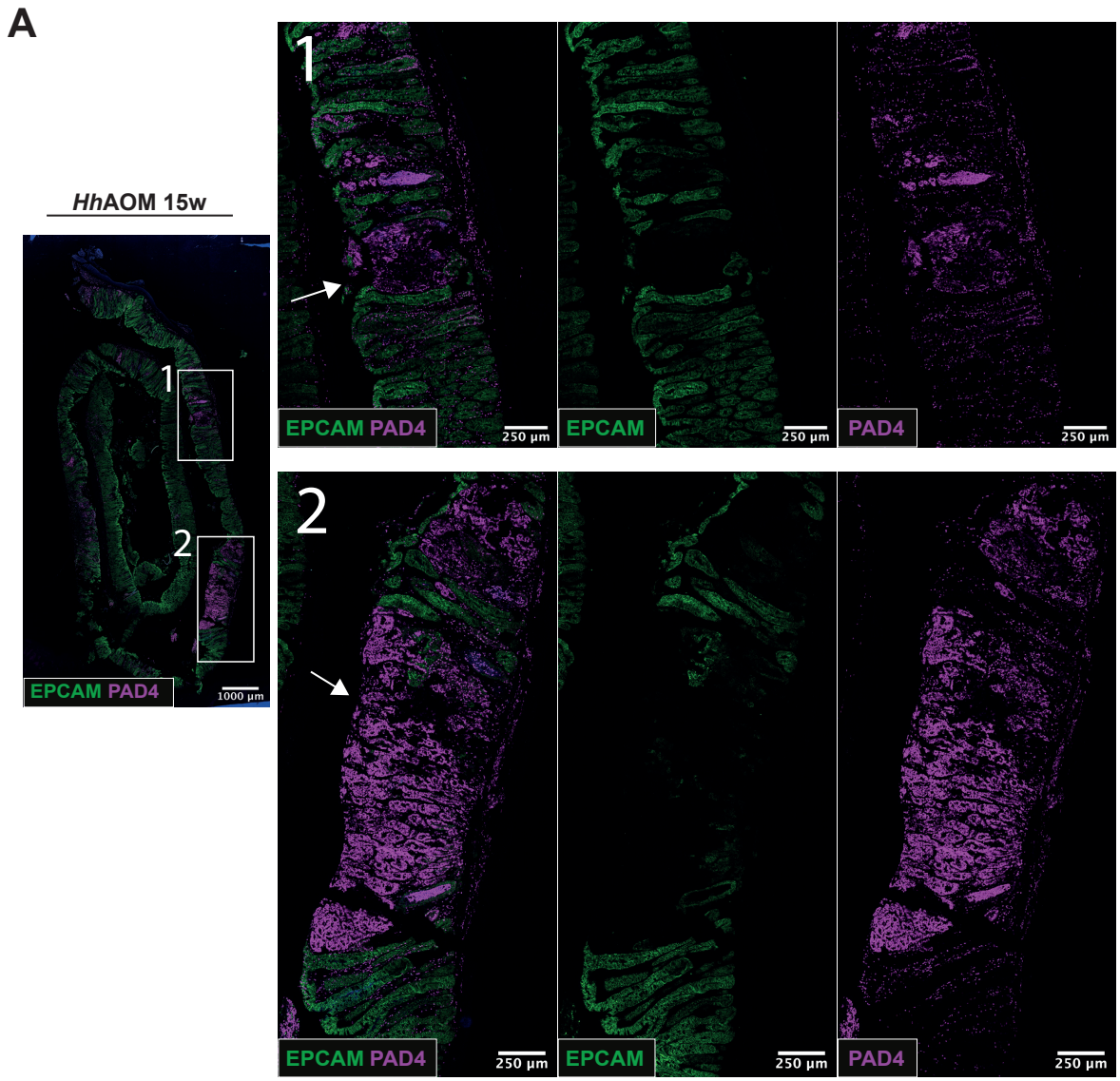
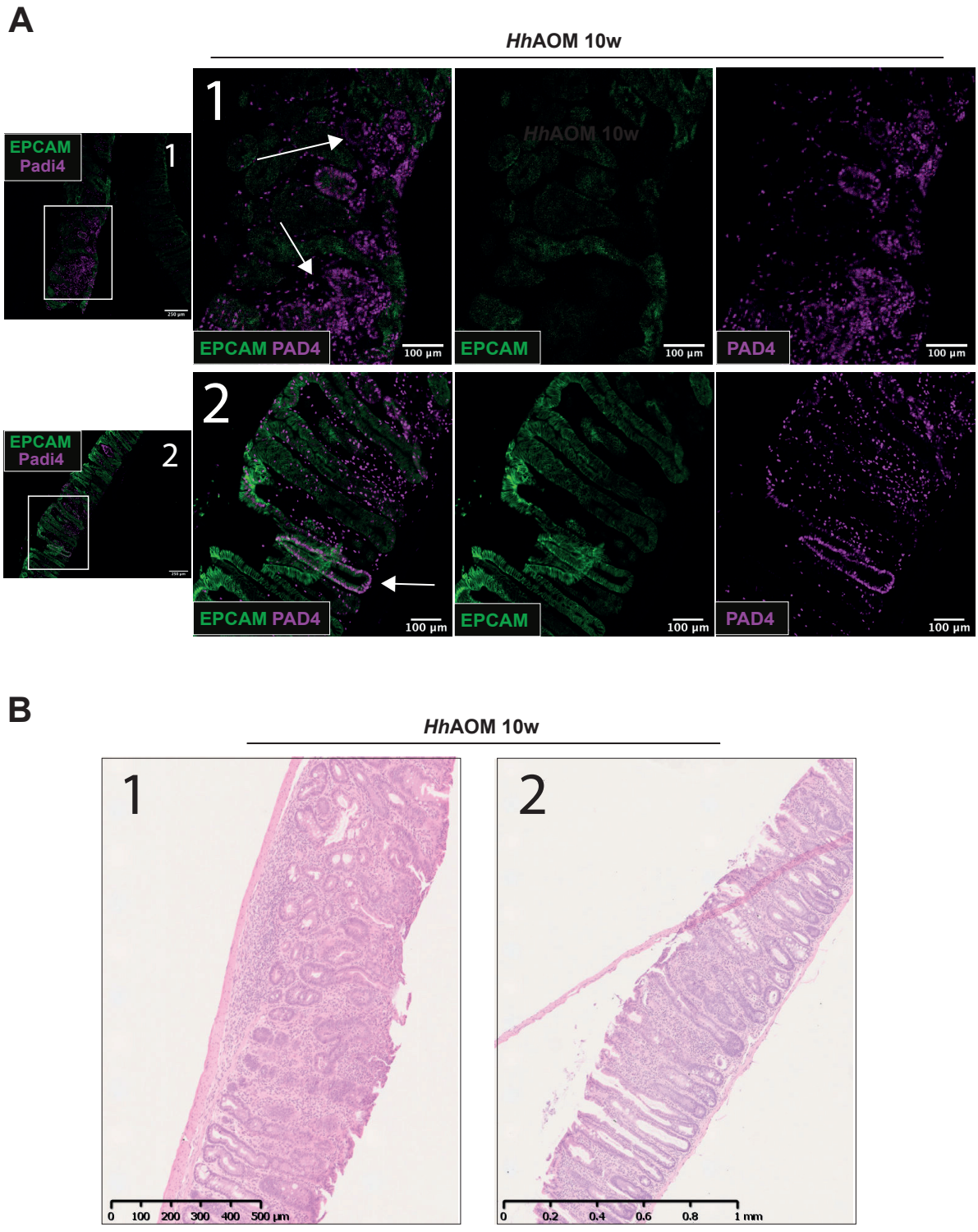

A
HhAOM 15w
1
2
EPCAM PAD4
1000 µm
EPCAM PAD4
250 µm
EPCAM
250 µm
PAD4
250 µm
B
HhAOM 15w
1
2
0 0.2 0.4 0.6 0.8 1 mm

**Figure 4.5 Epithelial PAD4 expression coincides with the loss of surface marker EPCAM.** 129SvEvRAG$^{-/-}$ mice were gavaged with *Hh* and injected with AOM. Upon sacrifice at week 15, distal colon tissue was harvested, fixed in formalin and processed for immunofluorescence staining.

A. Representative merged and single-coloured images of PAD4 (magenta) and Ly6G (green).
B. Corresponding H&E images.

We also noted this pattern of expression during early tumorigenesis at 10 weeks (**Figure 4.6, white arrow**). However, at this earlier time point we simultaneously observed some PAD4 expression in dysplastic but still EpCAM$^{+}$ epithelial crypts (**Figure 4.6 A & B; 2, white arrow**).




**<u>Figure 4.6</u> Crypt PAD4 expression observed during early tumorigenesis indicating oncogenic role.**

129SvEvRAG$^{-/-}$ mice were gavaged with *Hh* and injected with AOM. Upon sacrifice at week 10, distal colon tissue was harvested, fixed in formalin and processed for immunofluorescence staining.

A. Representative merged and single-coloured images of PAD4 (magenta) and Ly6G (green).

B. Corresponding H&E images.

In *Hh*-only controls, PAD4 resided solely in non-crypt tissue, whereas PAD4 expression in AOM-only controls was rarely detected. However, we did discover one crypt expressing PAD4 alongside EpCAM at week 10 (**Figure 4.7, white arrow**).

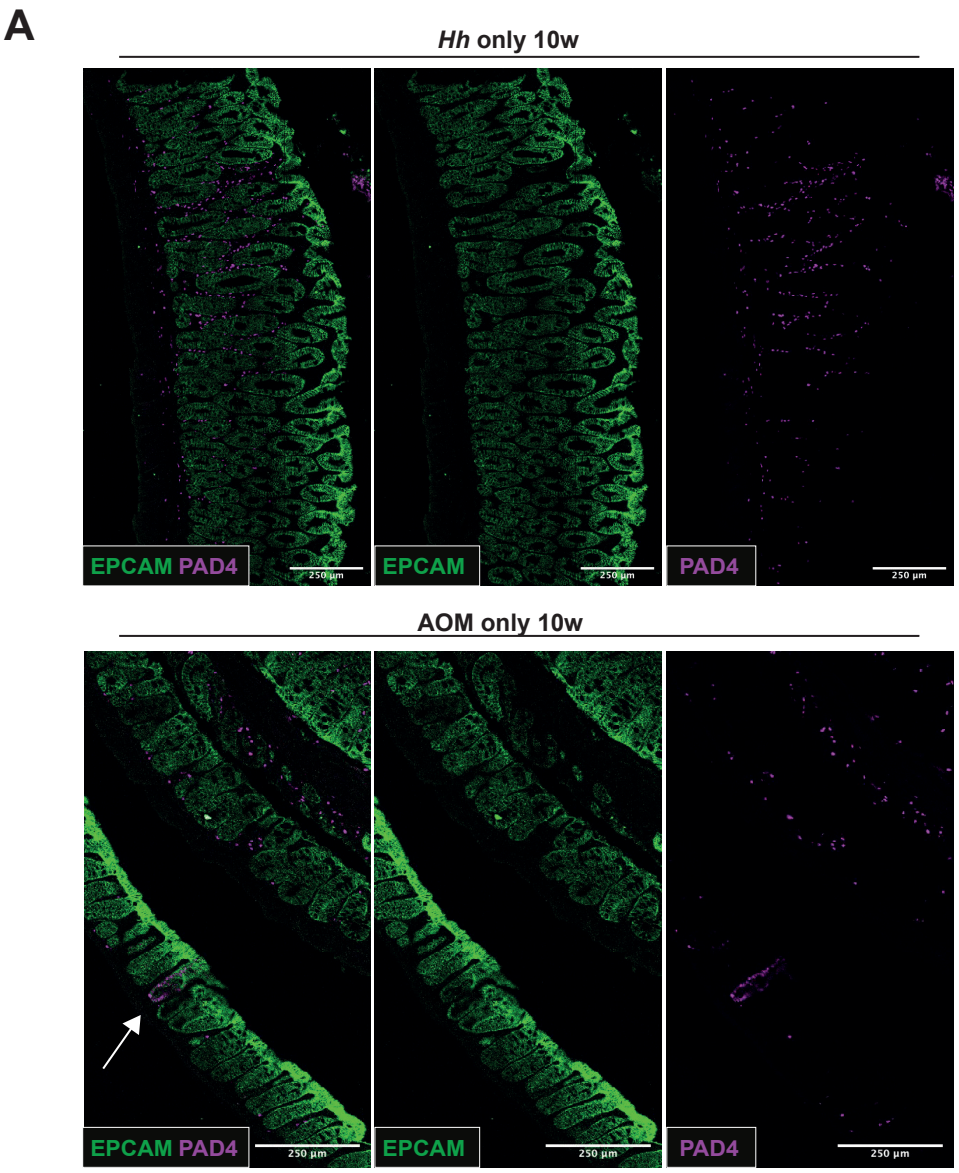

A
*Hh* only 10w
EPCAM PAD4
EPCAM
PAD4
250 µm
AOM only 10w
EPCAM PAD4
EPCAM
PAD4
250 µm
B
*Hh* only 10w
AOM only 10w
0 100 200 300 400 500 µm
0 50 100 150 200 250 µm

**Figure 4.7 PAD4 expression seen solely in EPCAM$^-$ cells in *Hh*-only controls and in a limited number of crypts in AOM-only mice.**
129SvEvRAG$^{-/-}$ mice were gavaged with *Hh* or injected AOM. Upon sacrifice at week 10, distal colon tissue was harvested, fixed in formalin and processed for immunofluorescence staining.

A. Representative merged and single-coloured images of PAD4 (magenta) and Ly6G (green).
B. Corresponding H&E images.

We also performed immunofluorescent stainings with conjugated goat and rabbit immunoglobulins to ensure fluorescent signal was not due to non-specific binding (**Figure 4.8**).

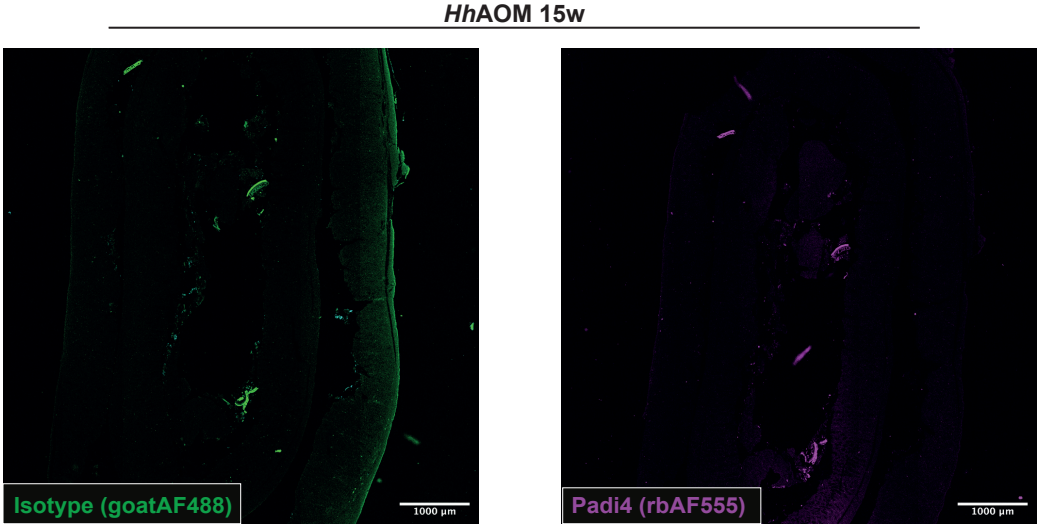



**Figure 4.8 Immunofluorescent stainings with isotype controls verify target specific binding of EPCAM and PAD4 antibodies.**
129SvEvRAG$^{-/-}$ mice were gavaged with *Hh* and injected with AOM. Upon sacrifice at week 15, distal colon tissue was harvested, fixed in formalin and processed for immunofluorescence staining. Representative merged and single-coloured images of goat (green) or rabbit (magenta) isotype controls.

### 4.3.4 Comparable expression pattern of Lgr5 and PAD4.

To better understand which epithelial cells are expressing PAD4, we collaborated with Nadia Nasreddin (Wellcome Centre for Human Genomics, Oxford) and stained tumour-bearing *Hh*AOM swiss rolls with intestinal stem cell markers *Lgr5* and *Ly6a* to probe cancer stemness with corresponding RNA probes *in situ*. By staining consecutive swiss rolls with PAD4, we noticed PAD4-positive tumours also express *Lgr5* (**Figure 4.9 A**). Normalised gene counts of *Lgr5* in our RNA sequencing data further showed a similar expression pattern to *Padi4* with an elevated expression in *Hh*AOM mice over time (**Figure 4.9 B**).

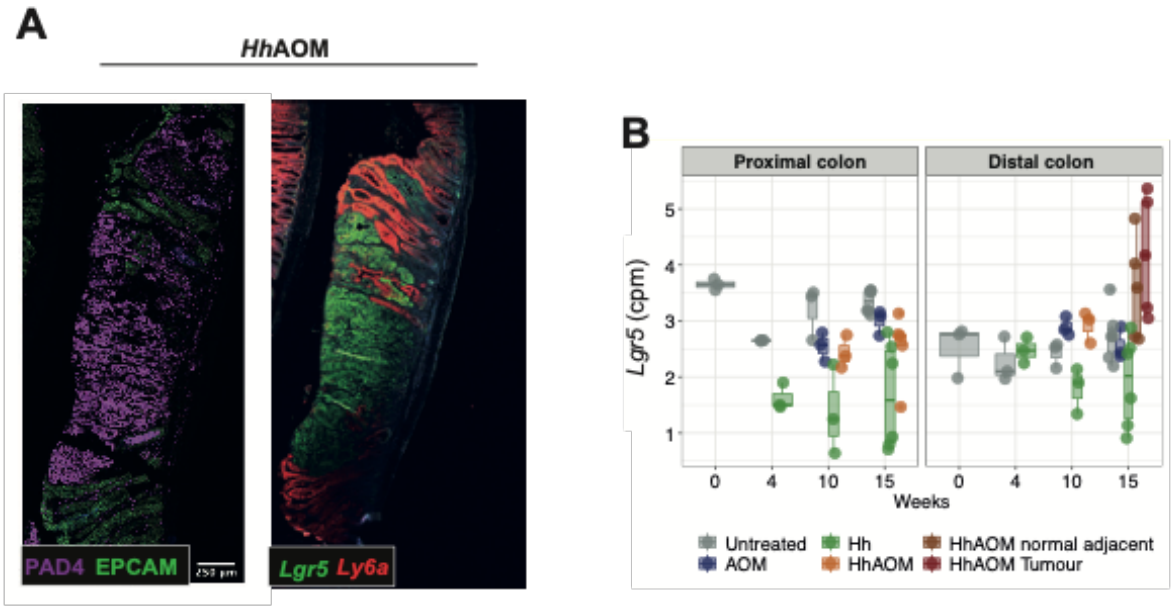



**Figure 4.9 The expression stem cell marker *Lgr5* is comparable to the PAD4 expression pattern.**
129SvEvRAG$^{-/-}$ mice were left untreated (Week 0 n=3, Week 4 n=3, Week 10 n=3, Week 15 n=3), injected with AOM (Week 10 n=3, Week 15 n=3), gavaged with *Hh* (Week 4 n=3, Week 10 n= 5, Week 15 n=5), or exposed to both (Week 10 n=3, Week 15 n=5). Upon sacrifice, proximal and distal colon was harvested for RNA isolation and subsequent bulk RNA sequencing**.** Tissue was further embedded as swiss rolls for downstream sectioning and staining.

A. Representative immunofluorescent images of consecutive *Hh*AOM tissue sections stained with PAD4 and EPCAM (left) and *Lgr5* and *Ly6a* (right; in kind collaboration with Nadia Nasreddin, The Wellcome Centre for Human Genetics, Oxford).

B. Normalised gene counts per million of *Lgr5* in longitudinal RNA sequencing dataset.

### 4.3.5 *In vivo* PAD4 inhibition results in a decreased tumorigenesis.

Having confirmed PAD4 upregulation during tumourigenesis, we used a commercially available small molecule PAD4 inhibitor (GSK-484) to probe function. 20mg/kg GSK-484 or DMSO vehicle control was injected i.p. into *Hh*AOM mice daily, beginning at week 13 (**Figure 4.10 A**). After 14 days of treatment, we embedded colon swiss rolls and performed H&E stainings (**Figure 4.10 B**) which were scored by Dr. Alejandro Suarez-Bonnet (RVC London). Mice treated with GSK-484 had significantly fewer and smaller tumour areas than DMSO controls (**Figure 4.10 C**).

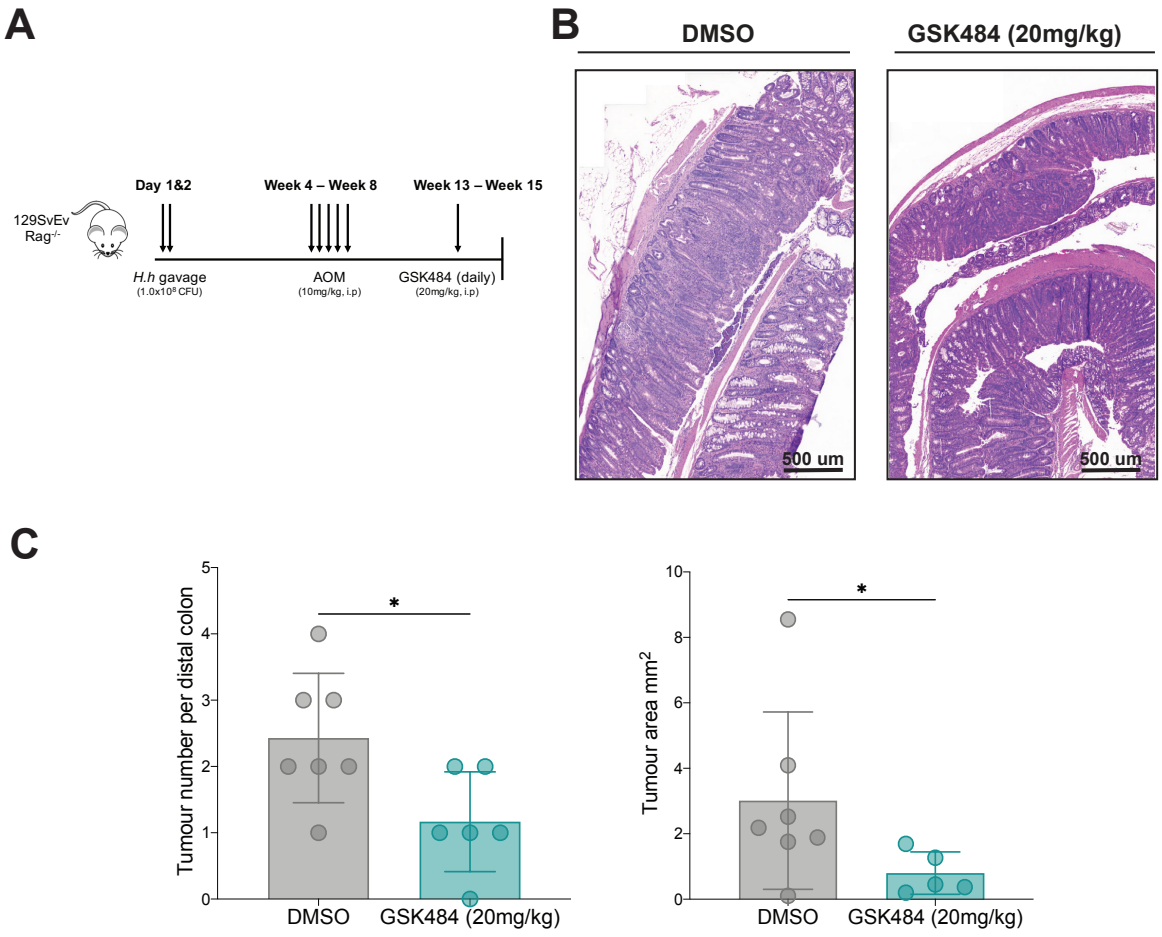
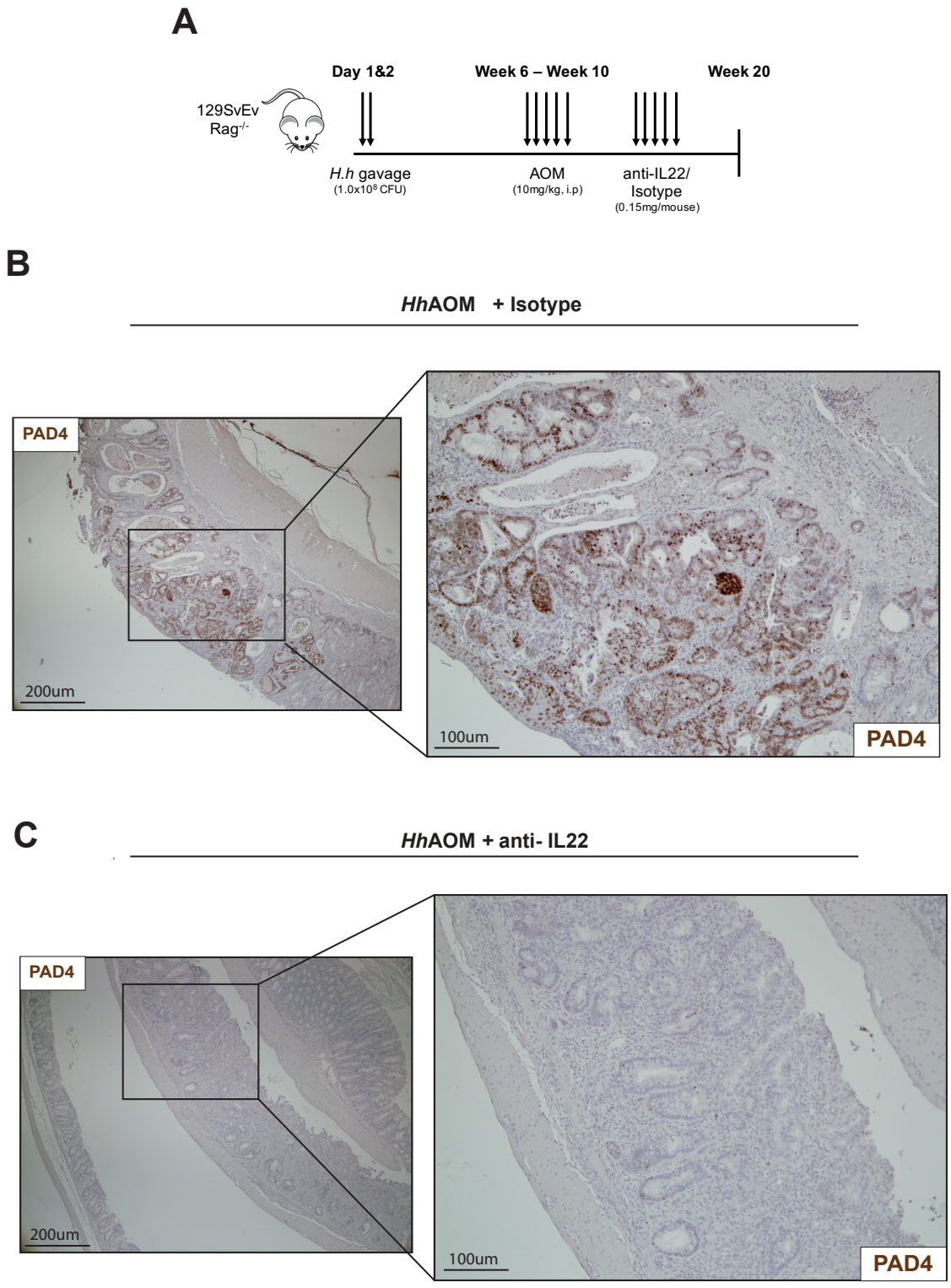


**Figure 4.10** ***In vivo i*nhibition of PAD4 during tumorigenesis results in a decreased tumour number and size.**
129SvEvRAG$^{-/-}$ mice were gavaged with *Hh* and injected with AOM. Daily injections of PAD4 inhibitor (GSK484, 20mg/kg) or DMSO alone were administered intraperitoneally (i.p) from week 13 to week 15. Upon sacrifice, colons were harvested and embedded as swiss rolls. Additionally, distal colon tissue was collected for protein analysis.

A. Schematic of experimental set-up.
B. Representative haematoxylin and eosin (H&E) images of whole colon swiss rolls.
C. Bar plots showing tumour number and tumour area of swiss rolls across experimental groups. Each datapoint depicts one mouse from one independent experiment with bars representing mean +/- SEM. Data was assessed with Mann-Whitney test after D'Agostino-Pearson omnibus normality test by (*, $p < 0.05$).

### 4.3.6 Lack of PAD4 expression in *Hh*AOM colons treated with anti-IL22.

Due to previous work in our lab revealing a therapeutic effect of blocking the cytokine IL-22[124], we investigated PAD4 expression in *Hh*AOM mice administered anti-IL22 or an isotype control thrice weekly (**Figure 4.11 A**, experiment performed by S. Kirchberger). Chromogenic staining revealed a loss of overall PAD4 expression in the colon of anti-IL22-treated mice as compared with the isotype group (**Figure 4.11 B/C**). Further repeats and quantification are required to confirm these results.




**<u>Figure 4.11</u> PAD4 is not expressed in the colons of *Hh*AOM mice treated with anti- IL22.**
*Hh*AOM mice were treated with an anti-IL22 (n=3) or an isotype control monoclonal antibody (n=3). Intraperitoneal injections (i.p) were performed weekly from week 12 to week 19 by Dr. Fanny Franchini (The Kennedy Institute, Oxford) and Dr. Abigail D'Cruz (The Wellcome Centre for Human Genetics, Oxford). Chromogenic staining of PAD4 was performed on processed swiss rolls.

A. Treatment schematic
B. Representative images of PAD4 expression (brown) in isotype treated colons at week 19.
C. Representative images of PAD4 expression (brown) in anti-IL22 treated colons at week 19.

### 4.3.7 Variable *PADI4* expression in human CRC

Finally, we sought to study a potential translational role for *PADI4* and therefore in collaboration with Dr. Elizabeth Mann, probed human CRC datasets for *PADI4* expression. Despite low *PADI4* expression, a microarray of human colorectal adenomas, sessile serrated adenomas (SSA), tumours and adjacent mucosa tissue (GSE117606) showed a significant upregulation in adenomas, but not tumours, relative to normal mucosa (**Figure 4.12 A**). Furthermore, there was a trend towards elevated PAD4 protein in bulk CRC tumour tissue as compared to paired normal adjacent tissue (**Figure 4.12 B**). Finally, to asses which cells express *PADI4* in human CRC we investigated a published single cell CRC dataset[196], which revealed the majority of expression in immune cells. However, *PADI4* expression was also detected in the epithelial compartment (**Figure 4.12 D**).

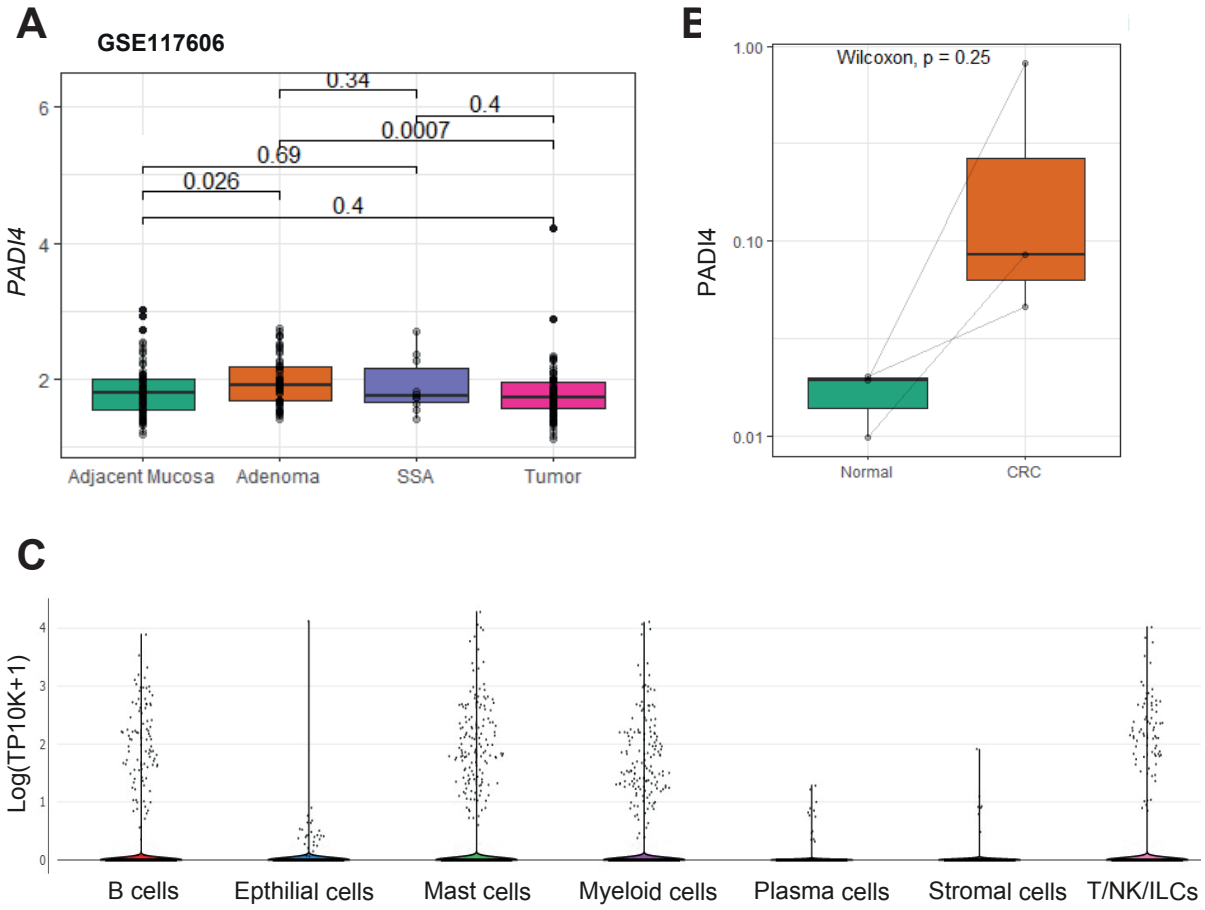



**Figure 4.12** ***PADI4*** **expression in humans is highly variable.**
Publicly available human colon cancer sequencing datasets were probed for *Padi4* expression (A/C), alongside in-house mass spectrometry analysis (B) (in kind collaboration by Elizabeth Mann).

A. *PADI4* expression within microarray of human colorectal adenomas, sessile serrated adenomas (SSA), tumours and adjacent mucosa tissue (GSE117606 cohort; n=70, Data were assessed by a Kruskal-Wallis test).
B. PAD4 expression in bulk CRC (n=3) and normal adjacent (n=3) tissue. Protein intensity values were normalised by the sum of all intensities of each respective sample. Data is shown in log scale and assessed by a Wilcoxon test.
C. *PADI4* expression within the GSE178341 CRC single cell sequenced dataset[196] (n=62).

## 4.4 Discussion

We discovered over 100 genes that are differentially expressed specifically during CAC but not colitis. Among others, we identified enriched expression of a range of AMPs alongside *Padi4* at 10- and 15-weeks during tumorigenesis. Subsequently, we verified elevated tumoral

PAD4 mRNA and protein in *Hh*AOM CAC and observed reduced tumour numbers following therapeutic PAD4 blockade. Of note, we observed similarly high *Padi4* expression in tumours and normal adjacent distal colon tissue of *Hh*AOM mice, but presume this is due to sampling errors discussed further in **5.6.**

### 4.4.1 PAD4 expression in DNA-damaged epithelium of *Hh*AOM mice.

Due to the description of a neutrophil-rich environment in the distal colon (Chapter 3), we initially hypothesised that many of our 118 genes would be linked to neutrophil functions. Various antimicrobial peptides were indeed enriched in the *Hh*AOM cancer setting, including defensins and lysozyme which are known to be expressed by both epithelial cells and neutrophils[197]. AMPs are well-known double-edged swords in tumorigenesis because they are able to promote proliferation and cancer cell migration of cervical cancer cells[198], but also induce tumour cell apoptosis in lung cancer xenograft models[199]. In human CRC, the often neutrophil-derived alpha defensins 1-3 were shown elevated in the tumour tissue and serum of patients providing biomarker potential[200]. Notably, AMPs are also known to be loaded on tumour-promoting NETs[104,105] with the key NETosis mediator *Padi4*[44] a top hit in our gene list. Recently, a study revealed PAD4-dependent NET formation to be induced by neutrophil chemoattractants CXCL1, CXCL2 and CXCL5[105], molecules elevated within the distal colon of this CAC environment as well (**Fig 3.7**). Therefore, we hypothesised that although numbers of neutrophils may be the same in *Hh* and *Hh*AOM (Chapter 3), neutrophils from cancer-developing *Hh*AOM mice were to be more prone to NETosis than colitic *Hh* only controls.

Chromogenic and immunofluorescence stainings of PAD4 revealed a punctate staining pattern, which indeed suggested PAD4 expression in leukocyte populations. However, this pattern was observed in colitic as well as cancer-bearing mice. Importantly, it was PAD4 expression within aberrant crypts and tumours that was observed specifically in the cancer setting. Therefore, we now hypothesise that PAD4 expression by malignant epithelial cells is a cancer-specific feature in this model.

In mice treated with AOM alone, we did observe some $PAD4^{+}$ epithelial crypts, albeit very few, but nonetheless suggesting that DNA damage triggered by AOM administration induces PAD4 expression in epithelial cells. To this end, it was been previously shown that PAD4 and downstream protein citrullination are upregulated after harmful irradiation[191]. However, a similar increase was not seen after exposure to ultraviolet light[191]**.** Irradiation is known to create double-strand DNA breaks[201] but UV exposure induces pyrimidine dimers[202]. Therefore, it may be just specific types of DNA damage that induce PAD4. In this CAC model, it is yet to be understood which mutagenic insult is caused by AOM with only sparse evidence of DNA cross-linking properties of the activated metabolite methylazoxymethanol[203]. Further, preliminary data collected in collaboration with the Leedham Laboratory in Oxford did not detect any known CRC-driver mutations in tumours but there was a trend towards an enrichment of the non-canonical WNT pathway (data now shown).

### 4.4.2 The enzyme PAD4 augments tumorigenesis.

Strikingly, therapeutic inhibition of PAD4 *in vivo* significantly reduced tumour number and tumour area in *Hh*AOM mice. This strongly suggests that PAD4 is not merely a potential DNA damage marker, but possesses tumour-promoting or sustaining functions. This observation is

in line with studies performed to decipher the role of NETosis and in which *in vivo* PAD4 inhibition also resulted in reduced subcutaneous tumorigenesis[104,105]**.** It is worth noting that those studies drew NETosis-specific conclusions despite using universal PAD4 inhibition and do not report on PAD4 expression in their respective cancer cells. Single cell sequencing of human CRC[196] recently revealed PAD4 expression in immune as well as epithelial cells**,** emphasising the importance of disentangling the roles of PAD4-expressing immune cells from those of epithelial origin.

In contrast with our results showing a deleterious role for PAD4, a number of studies indicate PAD4 curtails tumorigenesis[191,192]. This function involves tumour suppressor p53 which has been shown to target PAD4 and initiate apoptosis following DNA damage in epithelial cells[191]. However, these conclusions are drawn from reductionist *in vitro* experiments and $PAD4^{-/-}$ mice in which an absent microenvironment or unknown interplay of various $PAD4^{+}$ cell types might alter results. A recent study which subcutaneously injected $PAD4^{-}$ colon cancer cells into otherwise unaltered mice, observed significantly smaller tumours relative to cancer cells expressing PAD4[193]. That publication contradicts the notion of PAD4 suppressing tumorigenesis particularly when $PAD4^{+}$ cancer cells are exposed to a more complex environment.

In further support of this, studies have shown an upregulation of PAD4 in human CRC[204]. A recent publication also measured a strong upregulation in CRC liver metastasis relative to healthy liver and primary CRC, which points to a role for PAD4 in late-stage disease[193]. Harnessing publicly available datasets, we observed an overall low expression of *PADI4* but a

significant increase just in pre-cancerous adenomas. Also, our preliminary data on protein expression showed a trend towards increased PAD4 in bulk CRC tumour tissue relative to that from healthy controls. Overall, due to a wide variety of cells expressing *PADI4*[196], measurements in human bulk specimens depend on cell types present with further phenotyping and/or CRC stratification needed to understand the underlying clinical significance.

### 4.4.3 Preliminary evidence of PAD4-driven stemness.

By probing the phenotype of PAD4-expressing cells in more detail, we identified downregulation of surface marker EpCAM and a potential co-expression of Leucine-rich repeat-containing G-protein coupled receptor (Lgr5) by RNAscope **(Figure 5.13).**

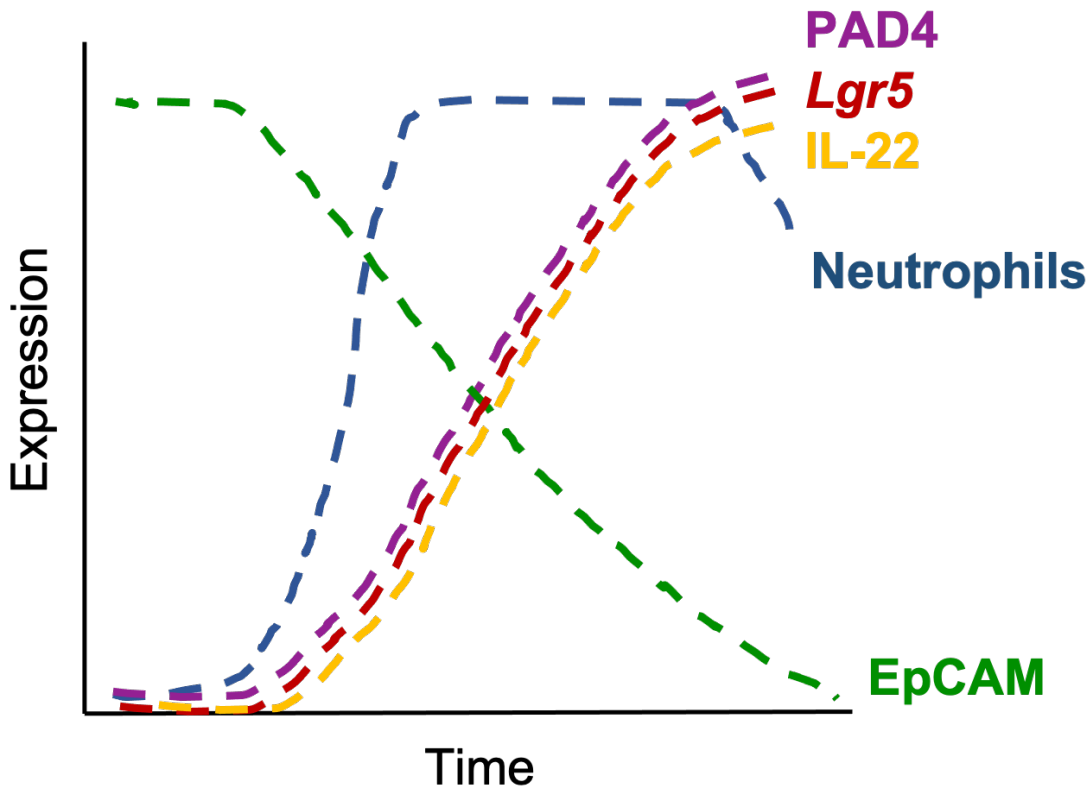



**Figure 5.13 Schematic of the longitudinal expression patterns of key molecules in the distal colon.** This figure illustrates the expression levels of PAD4 (purple), Lgr5 (red), IL-22 (yellow), neutrophils (blue) and EpCAM (green) during *Hh*AOM tumourigenes.

The biological relevance of the observed decrease of cell-adhesion molecule[205] EpCAM remains unclear. The downregulation of adhesion molecules on cancer cells has been attributed to EMT and therefore cancer progression[79]. By contrast others have found an increase of EpCAM expression by immunohistochemistry in a variety of tumours, including adenocarcinomas[206]. Also, PAD4 has been shown to inhibit EMT[189]. Together, this leads us to speculate on either a yet to be identified consequence of EpCAM loss or potential EMT despite PAD4 expression in this setting.

Lgr5 is a well-established stem-cell marker[207]**.** In the normal intestinal tissue, Lgr5 expression on stem cells is lost as they migrate and undergo differentiation[208]. However, in CRC, Lgr5 is reported to act as a cancer stem cell (CSC) marker[209-212]. One CSC alone can drive tumorigenesis in a process that is thought to be cell-autonomous and driven by genetic and epigenetic factors[213]; additional studies have also identified external cues such as cytokines that enhance CSC self-renewal properties[128,214]**.** The potential co-localisation of PAD4 and *Lgr5* in tumours leads us to hypothesise that PAD4 regulates the stemness of cancer cells by opening chromatin which leads to altered gene expression. This would be in line with a recent paper which identified PAD4 to promote stemness in murine embryonic tissue by histone citrullination and chromatin de-condensation[195].

We hypothesise that PAD4 expression functions in a cell-intrinsic manner triggered by DNA-damaging AOM, but is shaped by factors found in the inflamed distal environment induced by Hh infection. Notably, IL-22 has been shown to promote stemness and the upregulation of stem cell markers SOX2 and POU5F1 in a murine model of CRC. Researchers linked this phenomenon to the enzyme DOT1L and an increased downstream epigenetic methylation of histone H3K79[128]**.** In our hands, neutralising IL-22 *in vivo* ablated PAD4 expression in a pilot experiment, which we propose is preliminary evidence for IL-22 directly (or indirectly) influencing PAD4 expression and stemness features. More precisely, we hypothesise IL-22-triggered PAD4 to lead to histone citrullination, chromatin de-condensation and ultimately the expression of stemness-related genes such as *Lgr5*.

Although IL-22 has been shown to control DNA-damage induced apoptosis and therefore play a protective role after genotoxic insults[126], it can sustain expansion of already transformed cells[116] which we propose is the mechanism of action in our studies. Collectively the data highlight contrasting roles for IL-22 in tumorigenesis that are context-dependent, governed by concentration and temporal dynamics in disease. This warrants further investigation in our study which will include measurements of possible PAD4 induction by IL-22 in EC lines and/or organoids.

## 4.5 Conclusion

In this chapter we have identified 118 genes expressed to a lesser or greater extent in developing cancer relative to colitis. Amongst these, we verified a significant upregulation of the citrullination enzyme peptidylargine deiminase 4 (PAD4) during tumorigenesis in aberrant epithelium and demonstrated a functional role in driving disease. Areas of $PAD4^{+}$ epithelial cells overlaid EpCAM loss and *Lgr5* expression which is suggestive that cells harbour stem-like properties. Mechanistically, we hypothesise that PAD4 is an oncogene that is triggered by DNA damage and exacerbated by the inflamed microenvironment in which IL-22 is a crucial player.

## 4.6 Future outlook and limitations

To understand the role of PAD4 in more detail, we plan to further validate our functional findings alongside characterising $PAD4^{+}$ cancer cells and their surrounding environmental niche in more detail. First and foremost, we will repeat the PAD4 inhibition experiment and additionally investigate the disease state prior to our first administration at week 13. This will

also allow us to decipher if PAD4 inhibition merely halts tumorigenesis or can actively push towards cure. Subsequent unbiased mass-spectrometry of citrullinated proteins will provide us with a catalogue of potential downstream targets of PAD4 and help us pinpoint relevant biology. Additionally, by measuring the expression of genes involved in stemness or apoptosis, we will be able to test our hypothesis.

However, interpretation of this experiment is complicated due to PAD4 expression by various cell types making it difficult to know the biologically relevant source or sources. Ideally, utilising conditional murine PAD4 knockouts or bone marrow chimeras could help disentangle the contribution of hematopoietic versus non-hematopoietic PAD4 expression. This would unfortunately be a time-consuming endeavour due to the 129RAG$^{-/-}$ genetic background of this CAC model. A potential alternative to conditional murine models would be rectal injections of PAD4$^{+}$ and PAD4$^{-}$ cancer cells into the inflamed distal colon. This approach would allow us to assess the oncogenic potential of PAD4 when exposed to an identical microenvironment. To date such an experiment has been thwarted by inadequate surface markers to precisely cell sort and expand PAD4$^{+}$ tumour cells *in vitro.* Despite our identification of Lgr5 as a possible candidate, a suitable antibody has not yet been developed and dual stainings on the same slide are required in the future. Murine cell lines or colonic organoids could however be engineered by CRISPR-CAS9 or transgenic PAD4 expression.

Furthermore, unbiased but costly spatial transcriptomics and/or proteomics could be performed in the hope of discovering a suitable epithelial surface marker. These approaches would also help to unravel the downstream biology of PAD4$^{+}$ cells and simultaneously

phenotype the TME. An alternative and less costly approach would lie in multiplex imaging on the basis of acquired knowledge from the previous chapter. Deciphering the immune niche in which $PAD4^{+}$ cancer cells locate will also aid in discovering potential drivers of their expansion. Addition of markers such as IL22RA1 or SOX2 would also help probe our mechanistic hypothesis of IL-22 driven stemness in these cells.

Lastly, a limitation to be aware of is the similarity of HhAOM "normal adjacent" and tumour tissue in our RNA sequencing dataset. This is potentially due to macroscopic sampling of adjacent distal colon tissue that was in a similar cancerous state. We therefore decided to perform *in silico* analysis solely between untreated colons, AOM-only, *Hh* only and/or tumour tissue, hence disregarding normal adjacent samples from our analysis. However, by imaging swiss rolls we could distinguish "normal adjacent" from tumour tissue and used this to verify any tumour specific hypothesis we had post the RNAseq such as the increased expression of PAD4 in tumours.

# 5 The role of *Hh* in shaping dysbiosis-driven cancer

## 5.1 Introduction

In the first two chapters of this thesis, we found that *Helicobacter hepaticus (Hh)* administration triggered a neutrophil influx and subsequent tumorigenesis specifically in the distal colon of the *Hh*AOM mouse model of CAC. We further identified a causative role for the citrullination enzyme PAD4 in carcinogenesis. Here we set out to investigate the spatiotemporal dynamics of *Hh* and other bacterial members within the host during CAC progression.

*Helicobacter* species are well-known triggers of colitis and CAC in a range of susceptible mouse strains[215], but are relatively unexplored in the human setting[216]. The exception is infection with *Helicobacter pylori* which is now an established cause for gastric cancer in humans and a leading example of pathogenic infections driving tumorigenesis[129,130,217].

*Hh* is a gram-negative *Proteobacteria* that colonises the murine colon[218,219]. Although originally isolated from the liver and associated with chronic hepatitis and hepatocellular carcinoma[219-221], *Hh* has been repeatedly implicated in the development of colitis in several immunodeficient mouse strains[218,222]. These deficiencies are often due to defective T-cell function, ablated IL-10 signalling[55,124,155] or linked to the genetic HICCS (*Hh*-induced colitis and associated cancer susceptibility) locus[124,155]**.** Spatial investigations have revealed *Hh* to preferentially reside within the host mucus layer and intestinal crypts[219,223].

More functionally, *Hh* expresses the virulence factor cytolethal distending toxin (CDT) which has DNA damaging capabilities[224]. Indeed, 129SvEVRAG$^{-/-}$ mice administered *Hh* that is deficient in CDT developed fewer pre-neoplastic lesions[225]. In a similar vein, mono-colonisation with the 'oncomicrobe' *C.jejuni* was shown to promote CDT-dependent colon cancer in DSS-APC$^{Min/+}$ mice[226]. Genotoxic damage and subsequent tumourigenesis can also be triggered indirectly as a result of *Hh*-driven innate inflammation leading to the release of

reactive nitrogen and oxygen species [227,228]. Despite this handful of studies, the exact roles of *Hh* over the course of this CAC model are yet to be determined.

It is now well-known that CRC is associated with a dysbiosis which includes shifts in several 'oncomicrobes' that possess genotoxic or TME altering capabilities (**Chapter 1.4**). Furthermore, the unrestrained inflammation in colitis and/or CAC models can reshape the intestinal ecological niche and lead to an outgrowth of better adapted microbes[3]. For example, increased quantities of genotoxic pks$^{+}$ *E. coli* were identified in inflamed IL10$^{-/-}$ intestines[147,148], and the additional inhibition of bacterial scavenger protein lipocalin 2 (Lcn2) led to an outgrowth of harmful *Alistipes* species[229]. However, if inflammation and/or subsequent cancer shapes a niche to which *Hh* is better adapted to is not known.

## 5.2 Aim

In this chapter we generated a 16S rRNA dataset to probe differences in the microbiota composition during CAC. Specifically, we sought to investigate the spatiotemporal location of *Hh* in relation to the host and probe if this microbe alone has the ability to trigger disease using gnotobiotic mice.

## 5.3 Results

To characterise variation within the bacterial community during CAC, we performed 16S rRNA sequencing of DNA isolated from the proximal and distal colon tissue across all treatment groups and timepoints (**Figure 5.1A**). Ordination analysis (*phyloseq*) of this data revealed *Hh* administration to drive the largest differences within our dataset (sample segregation on 'Axis.1'). Simultaneously, 'Axis.1' separated 10- and 15-week *Hh*AOM samples. 'Axis.2' did not reveal any further segregation between samples (**Figure 5.1B**). The analysis of DNA extracted from longitudinally collected faecal samples resulted in the same inter-group patterns of variation (**Figure 5.1C**).

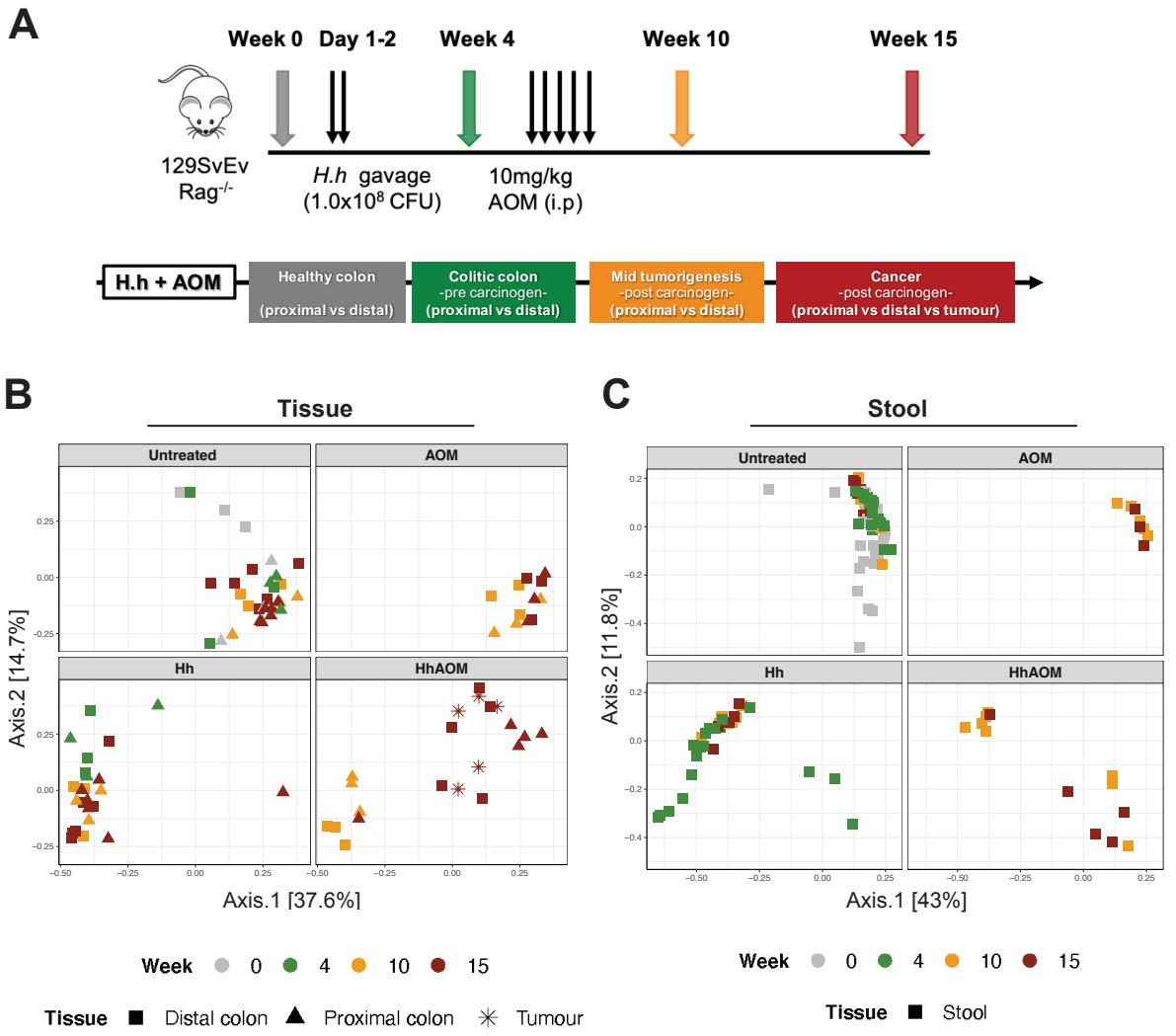
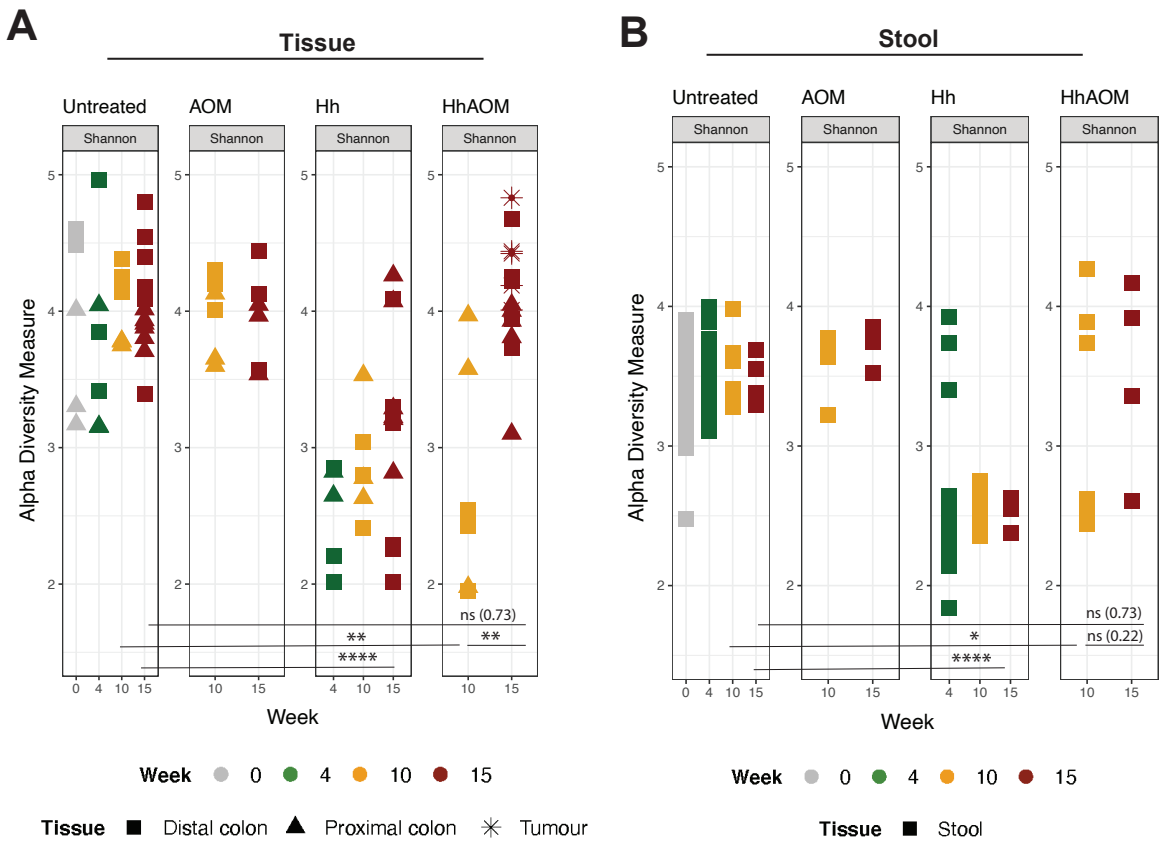



**Figure 5.1 Clustering analysis of extracted microbial DNA reveals longitudinal variability in *Hh*AOM mice.**

129SvEvRAG$^{-/-}$ mice were left untreated (week 0, 4, 10, n=3, respectively week 15 n=5), injected with AOM (week 10, 15 n=3, respectively), gavaged with *Hh* (week 4 and 10 n=3, respectively; week 15 n=6), or exposed to both (week 10 n=3, week 15 n=5). Upon sacrifice proximal, distal and tumour tissue was harvested for DNA isolation. Faecal pellets were collected for DNA isolation at week 0 (n=41, grey), week 4 (n=37, green), week 10 (n=31, orange) and week 15 (n=20, red) from all mice in each experimental group.

A. Schematic of experimental set-up.
B. Axis 1 and Axis 2 of ordination analysis run with all tissue samples (n=87). Treatment groups are shown in separate panels with symbols representing tissue location and colours representing time points of sacrifice (phyloseq, ggplot, R). Data is shown for each mouse and from one independent experiment (week 15 excluding AOM represents two independent experiments).
C. Axis 1 and Axis 2 of ordination analysis run with all stool samples (n=216). Figure outlined as in B.

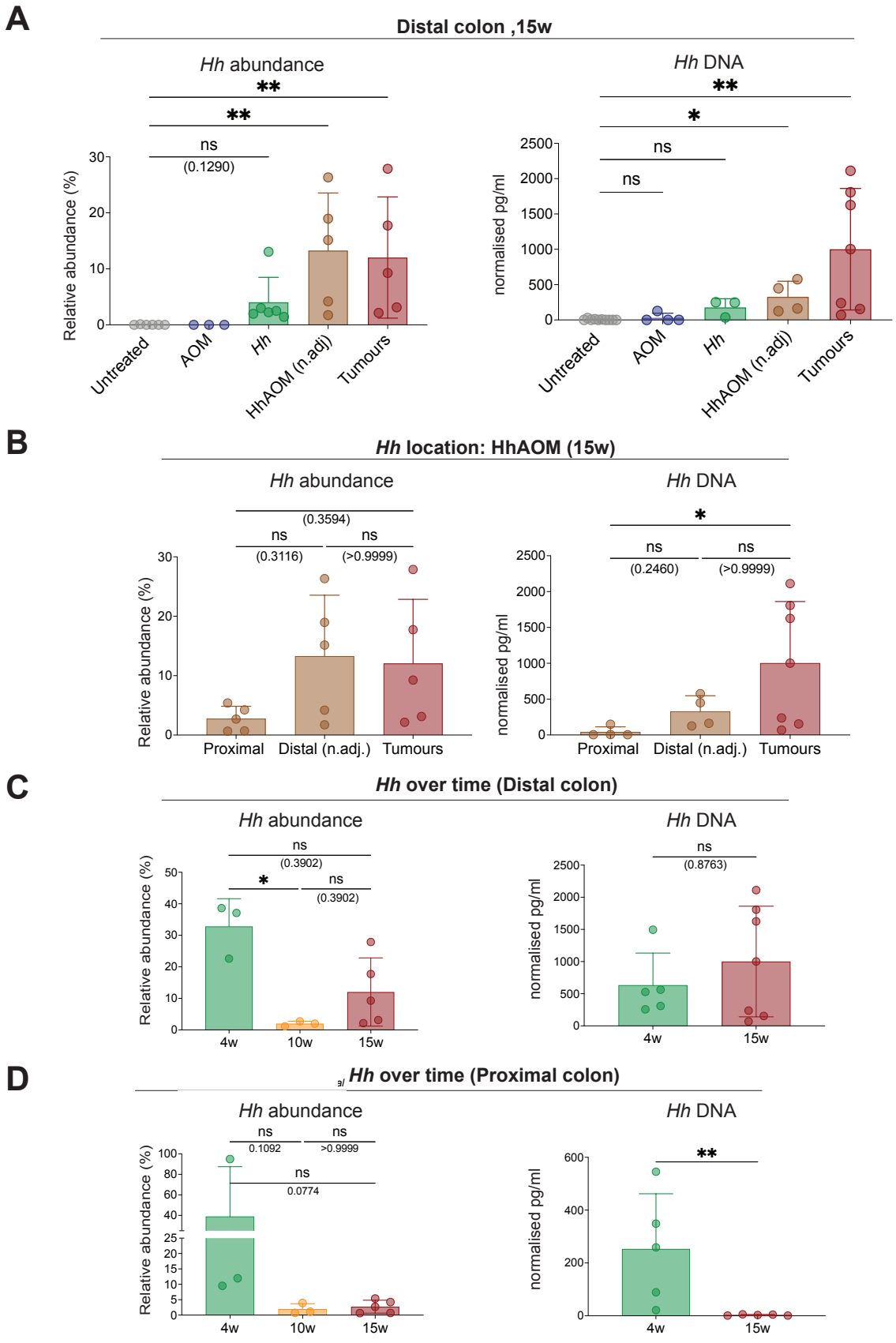
Given that colitis and cancer are associated with an altered flora diversity[230,231], we next sought to measure the alpha diversity of our bacterial communities in kind collaboration with Dr. Sena Bae from Harvard University. We observed decreased diversity within all tissue locations (**Figure 5.2A**) and stool (**Figure 5.2B**) of *Hh*-gavaged mice relative to non-infected controls. This decrease was less evident in *Hh*AOM mice with samples taken at 15 weeks displaying similar alpha diversity measurements to healthy controls (**Figure 5.2 A&B**)




**<u>Figure 5.2.</u> Alpha diversity decreases in colitis but increases during tumorigenesis.**
129SvEvRAG$^{-/-}$ mice were left untreated (week 0, 4, 10 n=3, respectively, week 15 n=5), injected with AOM (week 10, 15 n=3, respectively), gavaged with *Hh* (Week 4 and 10 n=3, respectively; week 15 n=6), or exposed to both (week 10 n=3, week 15 n=5). Upon sacrifice proximal, distal and tumour tissue was harvested for DNA isolation. Faecal pellets were collected for DNA isolation at week 0 (n=41, grey), week 4 (n=37, green), week 10 (n=31, orange) and week 15 (n=20, red) from all viable mice in each experimental group. Measures are shown for each mouse and from one independent experiment (week 15 excluding AOM represents two independent experiments). Data were assessed by an unpaired student's T-test after D'Agostino-Pearson omnibus normality test (*, $p < 0.05$; **, $p < 0.01$; ****, $p < 0.0001$; ns, no significance) in kind collaboration with Dr. Sena Bae.

A. Alpha diversity measurements (Shannon) from all tissue samples (n=87) (*MaAslin2*, ggplot, R)
B. Alpha diversity measurements (Shannon) run with all tissue with all stool samples (n=216) (*MaAslin2*, ggplot, R).

These unbiased analyses suggest a central role for *Hh* in altering the bacterial composition in both the proximal and distal colon. To date, *Hh* is known to reside within the caecum and colon of mice[218] but colonisation levels in the distal colon, which is the major site of disease in our model, are unknown. To assess *Hh* location, we probed 16S rRNA sequencing alongside *Hh*-specific quantitative PCR. With both approaches we detected significant increases of *Hh* in the distal colon of mice relative to their non-gavaged controls (**Figure 5.3A**). At week 15, the most significant increase in absolute *Hh* DNA was within established tumours (**Figure 5.3B**). Further longitudinal analysis revealed a significant decrease of *Hh* abundancy from 4 weeks to 10 weeks and an increased trend in *Hh* colonization from 10 weeks to 15 weeks (**Figure 5.3C**).




**<u>Figure 5.3</u> *Hh* colonises proximal and distal colon tissue with high incidence in tumors.**
For *Hh* abundance analysis 129SvEvRAG$^{-/-}$ mice were left untreated (week 0, 4, 10, n=3, respectively week 15 n=5), injected with AOM (week 10, 15 n=3, respectively), gavaged with *Hh* (week 4 and 10

n=3, respectively; week 15 n=6), or exposed to both (week 10 n=3, week 15 n=5). Upon sacrifice proximal, distal and tumour tissue was harvested for DNA isolation and 16S Sequencing (left). To verify findings, 129SvEvRAG$^{-/-}$ mice were left untreated (n=10, grey), injected with AOM (n=4, blue), gavaged with *Hh* (early, n=5, green; late n=3, yellow) or exposed to both (n=5, maroon; tumours n=7, red). Upon sacrifice proximal and distal colon tissue was harvested for quantitative PCR and the *Hh* DNA concentration of each sample was normalised to its respective 16S DNA concentration. (right)

A. Relative abundance of *Hh* (*dada2*, *MaAslin2)* in the proximal colon shown alongside 16S real-time (RT) PCR. Percentages and normalised pg/ml values are shown for each mouse from one independent experiment (week 15 excluding AOM represents two independent experiments). with bars representing mean +/- SEM. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (*, $p < 0.05$; **, $p < 0.01$; ns, no significance).
B. Comparison of *Hh* DNA between tissue sites. Data were assessed by a Kruskal-Wallis test (*, $p < 0.05$; ns, no significance).
C. Relative abundance of *Hh* (*dada2*, *MaAslin2*) in distal colons of *HhAOM* mice over time. Percentages and normalised pg/ml values are shown for each mouse from one independent experiment (week 15 represents two independent experiments) with bars representing mean +/- SEM. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (*, $p < 0.5$; ns, no significance).
D. Relative abundance of *Hh* (*dada2*, *MaAslin2*) in proximal colons of *HhAOM* mice over time and processed as in C (**, $p < 0.01$; ns, no significance).

To confirm *Hh* colonisation in the distal colon, we performed fluorescence *in situ* hybridization (FISH) imaging on methacarn-preserved tissue of *Hh*AOM mice at 15 weeks. We detected invasive *Hh* DNA (orange, white arrow) alongside other bacterial species (red) within malignant tissue of the distal colon (**Figure 5.4A**). Lack of staining in gnotobiotic controls confirmed that this observation was not due to non-specific binding (**Figure 5.4B**).

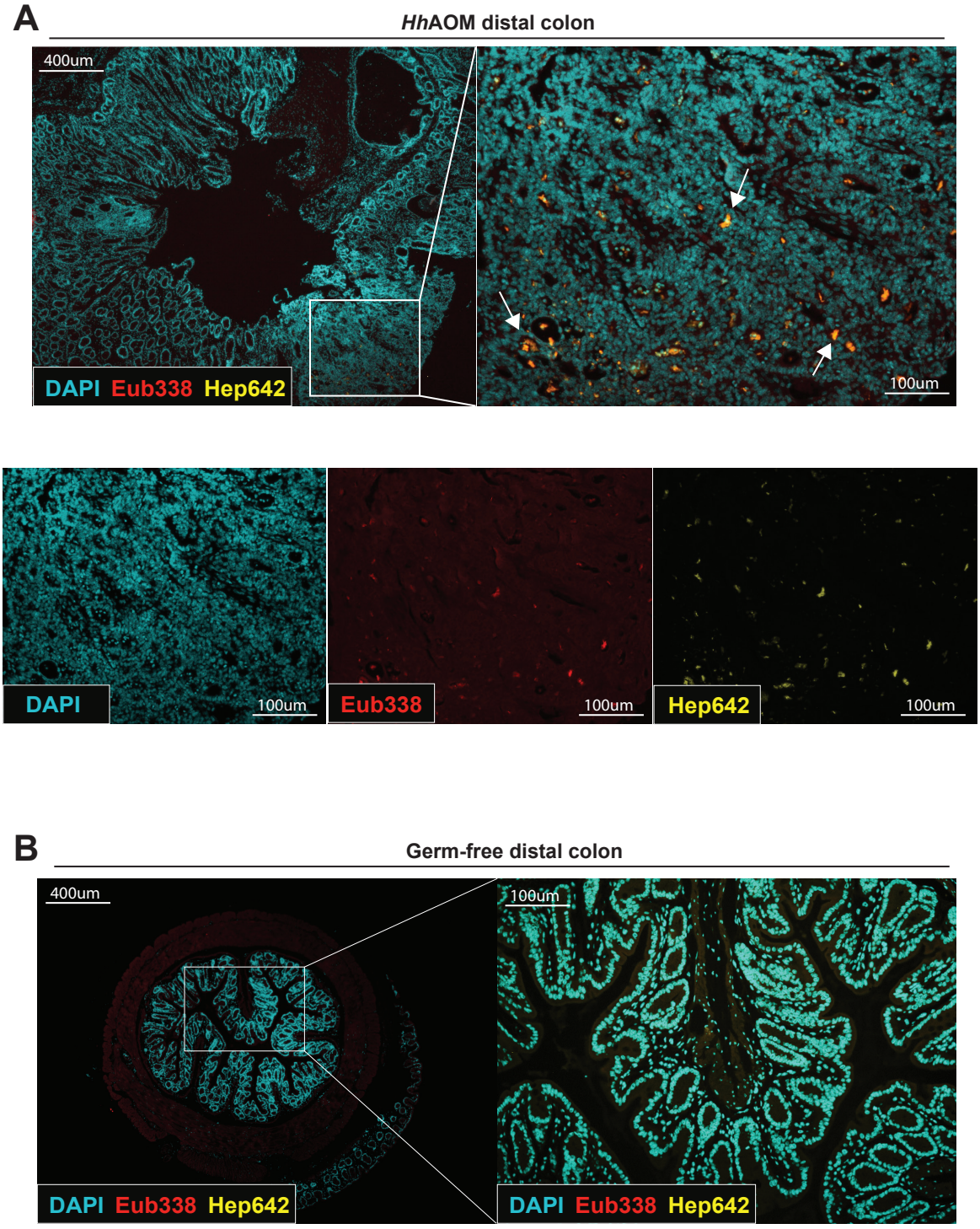
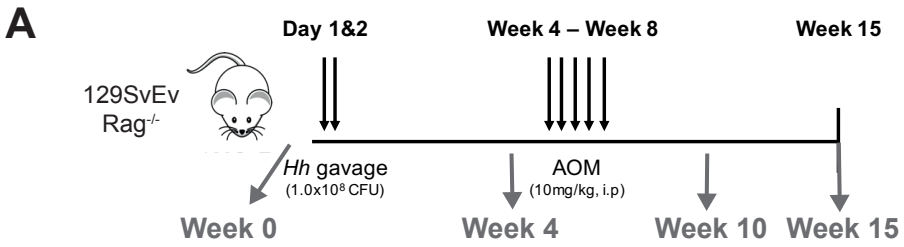
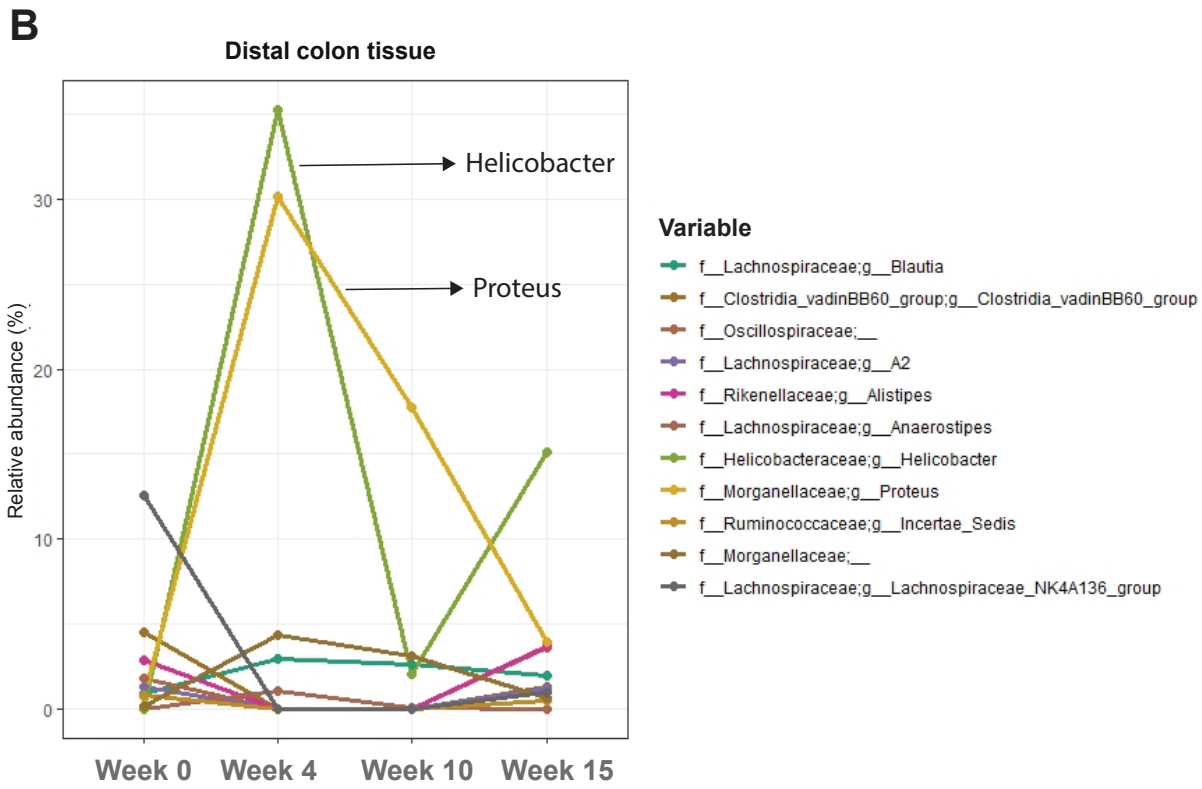



**Figure 5.4 Invasive *Hh* detected within malignant distal colon tissue of *Hh*AOM mice.**
129SvEvRAG$^{-/-}$ mice were gavaged with *Hh* and injected with AOM. Upon sacrifice at week 15, distal colon tissue was harvested, fixed in methacarn and subsequently embedded in paraffin. Alongside, germ-free wild-type (BL6) mouse was sacrificed and the distal colon harvested.

A. Representative fluorescence *in situ* hybridization (FISH) of all bacteria (Eub338, PE, red), *Hh* (Hep642, APC, yellow) and counterstained with the nuclear staining DAPI (cyan). Three single stain images are shown (bottom) as well as the resulting overlayed image (top).
B. Representative FISH image of all bacteria in a germ-free distal colon (panel as in A).

In order to detect abundancy changes in other bacteria, we performed differential sequencing analysis (*MaAslin2*) on the distal colon 16S rRNA dataset (in kind collaboration with Dr. Sena Bae, Harvard University). We identified 11 genera differentially expressed between untreated and *Hh*AOM samples over time and plotted the average abundance of each at week 0 (health), week 4 (colitis/pre-AOM), week 10 (early tumorigenesis) and week 15 (tumours) during the *Hh*AOM time course (**Figure 5.5 A&B**). Similar to the *Helicobacter* genus, *Proteus* abundancy increased in colitis (week 4) and decreased during early tumorigenesis (week 10). However, at week 15, *Proteus* levels decreased further whereas *Helicobacter* increased from week 10 to week 15 (**Fig 5.5B**). *Helicobacter* and *Proteus* abundance was significantly upregulated in tumour tissue relative to untreated controls (week 15), but *Proteus* abundancy was notably lower in *Hh*AOM relative to *Hh*-only controls (**Fig 5.5C**).


A
Day 1&2
Week 4 – Week 8
Week 15
129SvEv
Rag-/-
Hh gavage
(1.0x10⁸ CFU)
AOM
(10mg/kg, i.p)
Week 0
Week 4
Week 10
Week 15


B
Distal colon tissue
Helicobacter
Proteus
Relative abundance (%)
Week 0
Week 4
Week 10
Week 15
Variable
f__Lachnospiraceae;g__Blautia
f__Clostridia_vadinBB60_group;g__Clostridia_vadinBB60_group
f__Oscillospiraceae;__
f__Lachnospiraceae;g__A2
f__Rikenellaceae;g__Alistipes
f__Lachnospiraceae;g__Anaerostipes
f__Helicobacteraceae;g__Helicobacter
f__Morganellaceae;g__Proteus
f__Ruminococcaceae;g__Incertae_Sedis
f__Morganellaceae;__
f__Lachnospiraceae;g__Lachnospiraceae_NK4A136_group

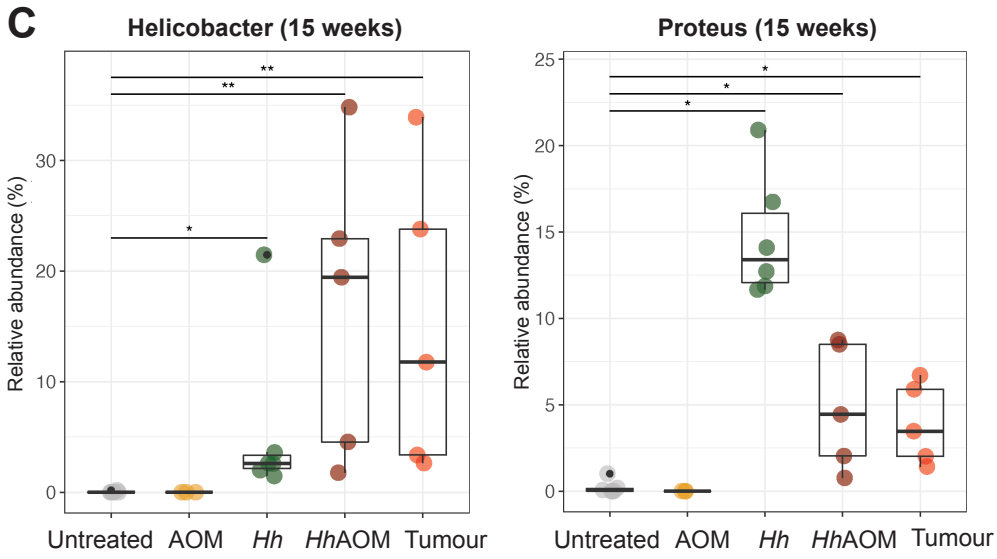
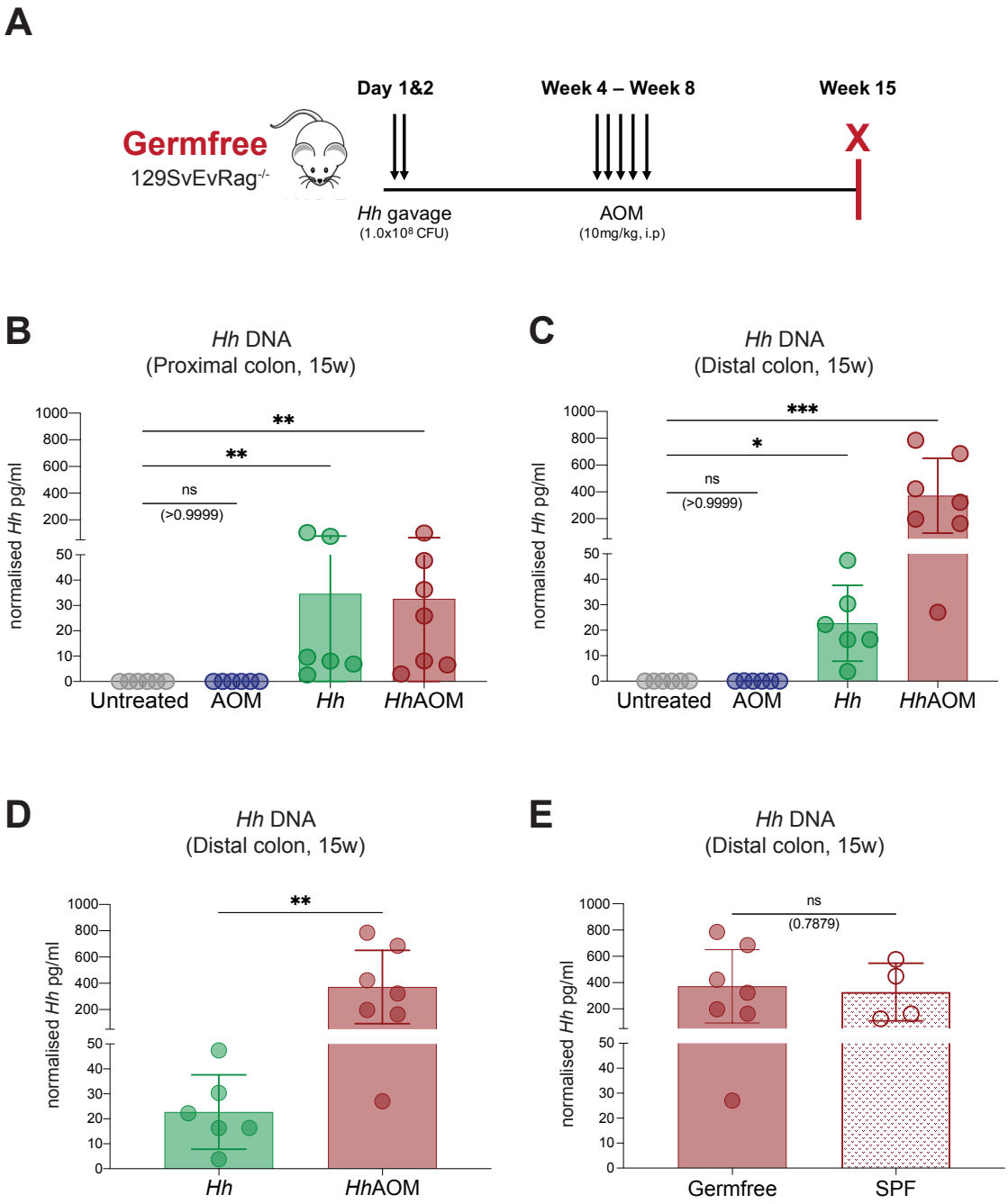

C
Helicobacter (15 weeks)
Proteus (15 weeks)
Relative abundance (%)
Untreated
AOM
Hh
HhAOM
Tumour

**Figure 5.5 The abundance of the Proteus genera changes during the course of disease.**
129SvEvRAG$^{-/-}$ mice were left untreated (Week 0 n=3, Week 15 n=6), administered AOM (Week 15 n=3) gavaged with *Hh* (Week 4 n=3, Week 10, n=3, Week 15 n=6), or exposed to both *Hh* and AOM (Week 10 n=3, Week 15 n=5). Upon sacrifice distal and tumour tissue was harvested for DNA isolation. Data analysis was performed by Dr. Sena Bae

A. Schematic of *Hh*AOM experimental set-up with time points of analysis (week 0, week 4, week 10 and week 15) highlighted in grey.
B. *Dada2* and *MaAslin2* analysis and subsequently plotted the average percentage abundance of genera that differed significantly at time points indicated in A. Data is from one independent experiment (week 15 excluding AOM represents two independent experiments).
C. The relative abundancy of the *Helicobacter* and *Proteus* genera in the distal colon of week 15 as determined by *MaAslin2* analysis. Percentages are shown for each mouse and boxes indicate the first, second (median), and third quartiles of data. Data is from two independent experiment (excluding AOM). The points outside of boxplot whiskers are outliers. Statistical analysis was performed using a Wald's test *p < 0.05, **p < 0.01, ***p < 0.001.

We have therefore established that *Hh* colonization plays a crucial role in triggering disease (Chapter 3), *Hh* can live in the distal colon (**Figures 5.3 and 5.4**) and have identified abundancy changes of other bacterial species during disease (**Figure 5.5**). However, whether inflammation and subsequent CAC is due to *Hh* alone or changes in other microbial-derived factors that we describe remains unclear. To examine this, we mono-colonized gnotobiotic 129SvEvRAG$^{-/-}$ mice with *Hh* and ran the equivalent 15-week *Hh*AOM model (**Figure 5.6A**).

*Hh* colonisation at 15 weeks was significantly increased in the proximal **(Figure 5.6B)** and distal colon **(Figure 5.6C)** of *Hh* and *Hh*AOM mice relative to uninfected controls. Notably, the distal colons of *Hh*AOM mice had significantly higher *Hh* colonisation levels than mice receiving *Hh* but no AOM **(Figure 5.6D).** Overall, comparable levels of distal colon *Hh* DNA were detected in germ-free and SPF conditions **(Figure 5.6E)**. However, despite successful *Hh* colonization, mono-colonized *Hh* and *Hh*AOM mice did not show an upregulation of the scavenger protein and inflammatory marker Lcn2 in fecal samples (**Figure 5.7**). Lcn2, largely expressed by neutrophiles and stored in their secondary granules[232,233], is thereby used as a disease activity marker which is also utilized in the human IBD setting[234]. Preliminary analysis of H&Es also shows a lack of inflammation and tissue in these mono-colonized mice (data not shown).




**<u>Figure 5.6</u> *Hh* amounts in germ-free 129SvEvRAG$^{-/-}$ mice higher when treated with AOM.**
Germ-free 129SvEvRAG$^{-/-}$ mice were left untreated (n=6, grey), injected with AOM (n=6, blue), gavaged with *Hh* (n=6, green) or exposed to both (n=7, red). Upon sacrifice at week 15, proximal and distal colon tissue was harvested for DNA isolation. *Hh* DNA was normalised to the total DNA concentration of each sample.

A. Schematic of experimental set-up.
B. DNA was isolated from proximal tissue and subjected to *Hh* RT-PCR. Normalised pg/ml values are shown for each mouse from one independent experiment with bars representing mean

+/- SEM. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (**, $p < 0.01$; ns, no significance).

C. DNA was isolated from the distal tissue and processed as in B. (*, $p < 0.05$; ***, $p < 0.001$; ns, no significance).
D. Comparison of *Hh* DNA in the distal colon tissue of *Hh* and *Hh*AOM mice. Data were assessed by a Mann-Whitney test (**, $p < 0.01$).
E. Comparison of *Hh* DNA in the distal colon tissue of germ-free and SPF *Hh*AOM mice. Data were assessed by a Mann-Whitney test (ns, no significance)

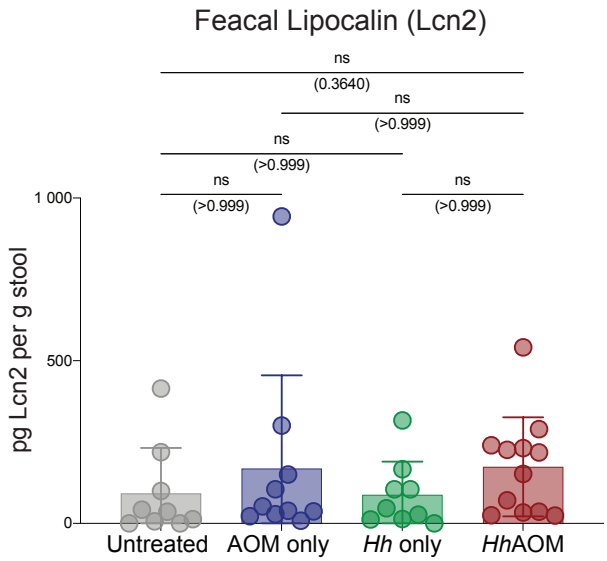
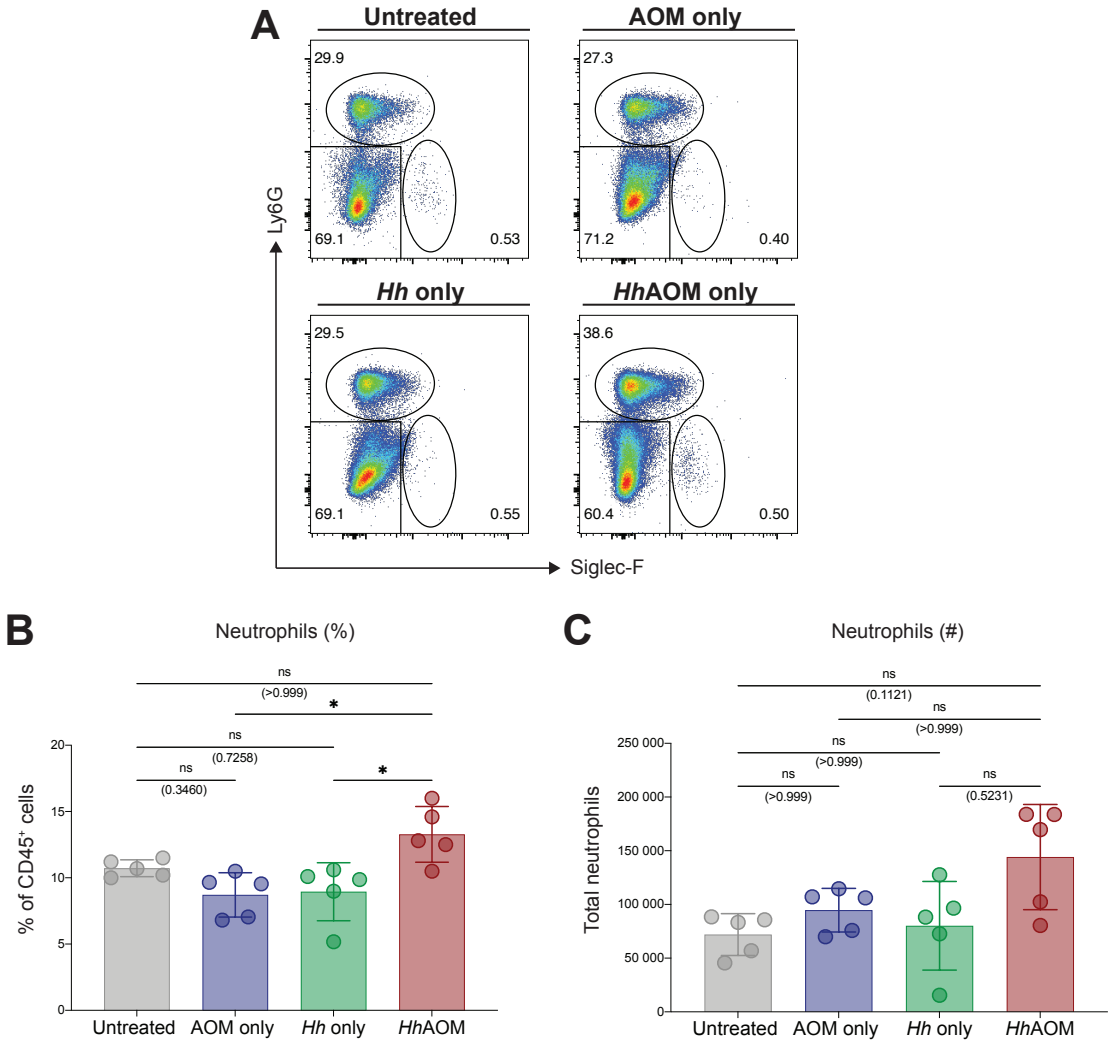



**Figure 5.7 Faecal Lcn2 is not significantly increased in gnotobiotic *Hh*AOM mice.**
Germ-free 129SvEvRAG$^{-/-}$ mice were left untreated (n=9, grey), injected with AOM (n=10, blue), gavaged with *Hh* (n=9, green) or exposed to both (n=12, red). One day prior to sacrifice, faecal pellets were collected for faecal supernatant generation. Lipocalin (Lcn2) ELISA was performed on feacal supernatants and normalised by stool weight. Each dot represents one sample and bars represent mean standard error of mean (±SEM). Data represents two independent experiments and was assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (ns, no significance).

Lastly, we sought to probe systemic inflammation in mono-colonised mice. Similar to our analysis in SPF mice (**Chapter 3**), we generated single cell suspensions of spleens and analysed neutrophil abundancy by flow cytometry (**Figure 5.8 A**). In contrast to our SPF observations (**Figure 1.9**), we did not detect a significant increase in absolute Ly6G$^{+}$ neutrophils in the spleens of mono-colonised mice (**Figure 5.8C**). However, we observed a slight increase in Ly6G$^{+}$ neutrophil abundancy relative to single-treated controls (**Figure 5.8B**). There were no significant differences seen in other myeloid populations (data not shown).




**<u>Figure 5.8</u> A higher frequency of neutrophils observed in *Hh*AOM spleens when compared to *Hh* or AOM only controls.**

Germ-free 129SvEvRAG$^{-/-}$ mice were left untreated (n=6, grey), injected with AOM (n=6, blue), gavaged with *Hh* (n=6, green) or exposed to both (n=7, red). Upon sacrifice at week 15, spleens were harvested and single cell suspensions analysed by flow cytometry.

A. Representative flow cytometry plots of indicated experimental groups, gated on live, singlets, CD45$^+$ and CD11b$^+$ cells. Neutrophils are defined by expression of Ly6G.
B. Frequencies of neutrophils across experimental groups. Each dot represents one mouse and bars represent mean ±SEM. Data represents one independent experiment. Data were assessed by a Kruskal-Wallis test after D'Agostino-Pearson omnibus normality test (*, $p < 0.05$; ns, no significance).
C. Total number of neutrophils across experimental groups. Figure outlined as in B.

## 5.4 Discussion

Changes in the bacterial composition during *Hh*AOM tumourigenesis was driven predominantly by *Hh* administration. Although *Hh* was located within distal disease location,

preliminary data shows that it failed to elicit acute inflammation in mono-colonised gnotobiotic controls either alone or with AOM administration.

### 5.4.1 *Hh*-driven dysbiosis is altered during tumorigenesis.

16S rRNA sequencing analysis revealed decreased diversity in bacterial communities within all mice gavaged with *Hh* relative to their healthy and mutagen-only controls. No variation was detected between AOM-injected mice and healthy controls. Together, these observations strongly suggest a different taxonomic composition in *Hh* administered mice and therefore a *Hh*-triggered 'dysbiosis'.

In colitis (*Hh*-only) controls, we measured a decreased bacterial diversity which remained unchanged over time and in line with observations made within IBD patients[230]. A similar reduced bacterial diversity was seen in an early 10-week timepoint of *Hh*AOM mice, but it then increased to levels of uninfected controls at the 15-week tumour-bearing time point. This observation agreed with the detected shift of 15-week *Hh*AOM samples in the previous PCA. Together, both observations suggested a different community as a consequence of established tumorigenesis relative to colitis or early cancer. In the CRC literature, diversity measurements and subsequent conclusions are still heavily disputed due to high intra-individual heterogeneity and, unlike this dataset, missing longitudinal assesements[3,136]. However, a more diverse microbiota was recently observed in samples from young CRC patients[231]. Conflicting results highlight the importance of measuring a more wide-spanning measure of functional diversity rather than taxonomic diversity, a topic further discussed in Chapter **6.4**.

### 5.4.2 The tumour microenvironment is enriched with *Hh*.

Our results indicated *Hh* induces microbial dysbiosis but how this shapes inflammation and cancer is still unclear. We therefore analysed *Hh* location across treatment groups with a particular interest in colonization levels within the distal colon due to the site-specific neutrophil signature and tumour development observed in **Chapter 3**.

Results mapped *Hh* to the distal colon of gavaged mice both in terms of abundance and absolute amounts. Imaging of *Hh*AOM colons further revealed invasive *Hh* alongside other microbes within malignant areas of the distal colon tissue. Although preliminary, this suggested that the TME harbours an ecological niche favourable for *Hh* and/or other bacteria. Indeed, it has been well-recorded that *Hh* resides close to the host epithelium[219,223] and previous studies have demonstrated inflammation-driven CRC to establish a niche in which certain species thrive[147,148,229].

In support of this hypothesis, we observed a trend towards an increase of *Hh* in *Hh*AOM mice over time. Even though non-significant to date, sequencing analysis and quantitative PCR showed a trend towards higher *Hh* colonisation in tumours relative to 10-week pre-tumorigenic distal colons. In particular, the comparison of *Helicobacter* and *Proteus* dynamics supports the idea of a hospitable TME for *Hh*, as both genera are highly abundant at 4 weeks and decrease after mutagenic insult. However, only *Hh* seemingly re-emerges in established tumours which suggests a possible host-microbe synergy, potentially due to a changing metabolic environment[136,235].

Nevertheless, whether the developing tumour truly harbours advantages for *Hh* colonisation and how this shapes tumour progression remains to be confirmed in future experiments with higher experimental numbers. Similar to the cytokine IL-22, *Hh* might have different roles depending on disease progression and context.

### 5.4.3 *Hh* does not trigger disease in a gnotobiotic setting.

Although previous publications have shown a crucial role for DNA-damaging CDT expressed by *Hh*, we could not detect inflammation nor verify tumorigenesis in our mono-colonised *Hh*AOM experiments. Mice mono-colonised with *Hh* had comparable levels of faecal Lcn2 and splenic neutrophils relative to their germ-free counterparts suggesting a lack of local and systemic inflammation. These observations were not due to failed colonisation as *Hh* was shown to be comparably abundant between SPF and gnotobiotic conditions. Strikingly, we

observed a higher amount of colonisation in mono-colonised *Hh*AOM mice relative to *Hh*-only controls which supported the notion that *Hh* is better adapted to a colon environment exposed to a mutagen.

Based on those results we speculate that other members of the microbiome must be present to elicit the inflammatory response characteristic of this model. It is tempting to speculate a role for the *Proteus* genus given its increase in colitis controls. To this end, a recent publication discovered two distinct *Proteus mirabilis* species more prevalent in patients with Crohn's disease[236].

## 5.5 Conclusion

Our results indicate that *Hh* administration triggers dysbiosis and a decreased bacterial diversity in the colon tissue and stool of 129SvEvRAG$^{-/-}$ mice. We further observed a compositional shift in bacteria during tumour development and detected *Hh* within established tumours, which we hypothesise to be a consequence of tumorigenesis. Lastly, we found that *Hh* alone does not trigger disease suggesting that other species are needed to initiate the underlying inflammation which is essential to perpetuate tumorigenesis.

## 5.6 Future outlook and limitations

The next steps in investigating the role of *Hh* in inflammation and cancer will focus upon utilising the reductionist Oligo-Mouse-Microbiota (OMM$^{12}$). This established microbiota model consists of 12 sequenced and publicly available strains which represent five bacterial phyla that naturally and stably colonise the murine intestine[237]. This will not only allow us to probe host-microbiota interactions of known OMM$^{12}$ strains during CRC, but also function as a tool to understand human CRC-identified microbes, such as *F.nucleatum*, in a more physiological manner.

The main limitation of the results presented in this chapter, in addition to low mouse numbers, was the suboptimal sequencing depth of our 16s RNA dataset. This did not allow us

to detect species or strain level differences and constrained us to focus solely on high-level analysis. In addition, the analysis presented here was restricted to bacterial DNA. Extending this to meta-transcriptomic or metabolomic datasets will provide further insight into wide-spanning functions rather than only taxonomic shifts underlying disease progression.
By introducing the OMM[12] into germ-free 129SvEvRAG$^{-/-}$ mice and performing *Hh*AOM experiments, we can probe longitudinal and functional differences in a more precise manner. Strain specificity will grant the potential to detect more granular differences between tissue locations which, in our SPF analysis, are seemingly similar. Recent publications on intra-strain variability of the 'oncomicrobe' *F. nucleatum* have highlighted the importance of strain specificity[238,239].

The knowledge of all strains present will also aid future multiplex imaging. Not only can specific probes be generated, but spatiotemporal changes of one candidate strain can be assessed within the context of all other strains using samples that have been banked throughout these studies. Lastly, with an OMM[12] approach, single strains can be experimentally depleted or introduced. This provides a valuable and physiological toolbox to assess downstream host mechanisms triggered by candidate strains within a complex community.

Lastly, it will be of great importance to probe functions of *Hh* and/or other identified microbes in immunocompetent mouse models. Most recently, *Hh* was shown to have a protective effect in DSS-AOM mice by inducing CD4$^{+}$ T cell and B-cell dependent anti-cancer immunity[240]. This strikingly highlights the crucial role of host genetics in shaping host-microbiota interactions in CRC.

# 6 General discussion

The incidence of CRC is increasing world-wide and at an alarmingly high rate in younger individuals[69,241]. Despite screening techniques, CRC remains the second leading cause of cancer-related death[241]. Therefore, the development of novel diagnostics, biomarkers and therapeutics is needed. Such initiatives rely on studies that mechanistically investigate the factors underlying cancer progression and, in particular in CRC, probe the role of the host-microbiota interface.

## 6.1 Summary

Given that microbial dysbiosis and mutagenic insult are both associated with CRC in many individuals, this thesis illustrates an interplay of environmental factors and host maladaptation in promoting tumorigenesis. Based on our data to date, we propose that *Hh*-driven dysbiosis triggers a local and neutrophil-rich inflammation which fuels the tumorigenic properties of DNA-damaged cells. More specifically, we hypothesise that the accumulated neutrophils are a crucial source of IL-1β which in turn stimulates the production of ILC3-derived IL-22. We further suggest that IL-22 shapes IEC biology, for example, through the observed PAD4 amplification and subsequent alteration of stem properties characterised by *Lgr5* expression and EpCAM loss. This thesis describes many of these changes during tumourigenesis which, using published literature, we have connected into a proposed model **(Figure 6.1)**. Verifying that these factors do indeed influence each other *in vivo* is an ongoing body of work.

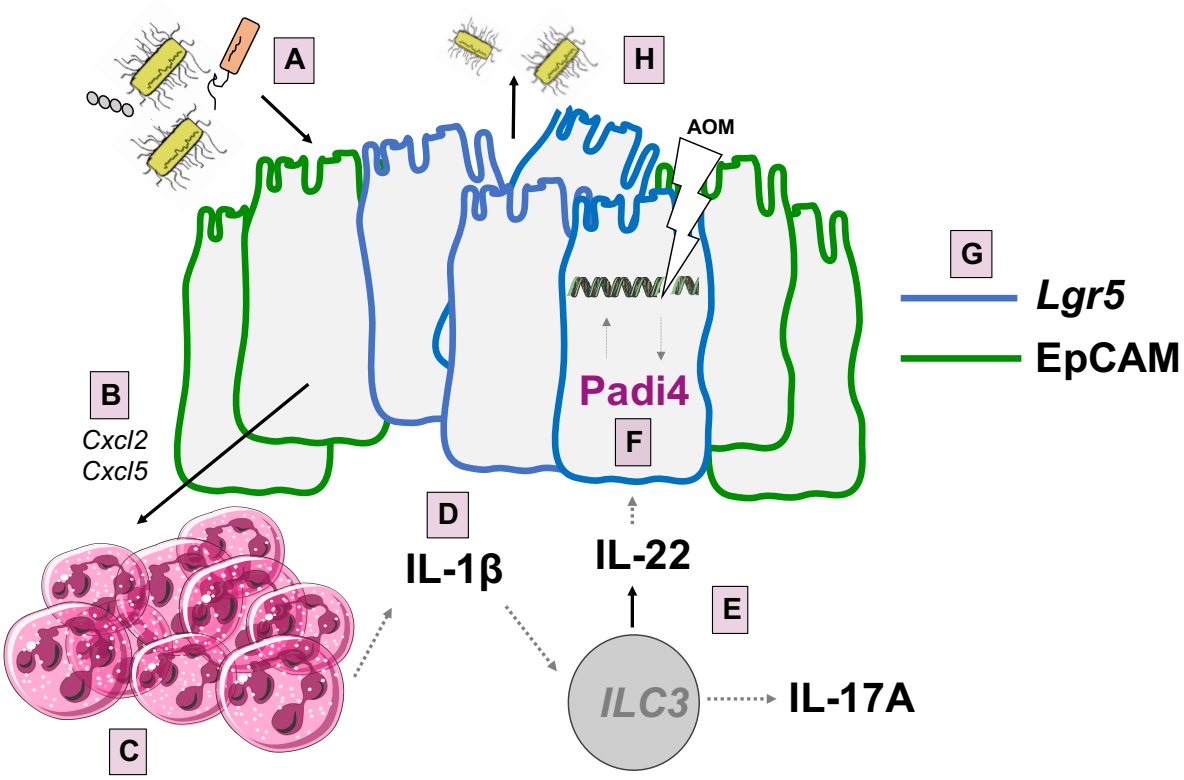



**Figure 6.1: Overview of our proposed tumorigenesis model in genetically susceptible *Hh*AOM mice.**
We propose *Hh* to trigger a dysbiosis (A) leading to an enriched neutrophil chemokine expression (B) and neutrophil influx within the distal colon (C). Further, we hypothesise that accumulated neutrophils secrete IL-1β (D) and stimulate the production of IL17A and IL-22 by ILC3s (E). In turn, IL-22 either directly or indirectly influences tumour progression by amplifying cell intrinsic processes linked to the citrullination enzyme PAD4 (F) and cancer stemness characterised by *Lgr5* expression and EpCAM loss (G). Lastly, we suggest that established tumours harbour an ecological niche preferred by *Hh* (H).

CXCL, chemokine; IL, interleukin; ILC, innate lymphoid cell; Padi, peptidyl arginine deiminase; AOM, azoxymethane; Lgr5, Leucine Rich Repeat Containing G Protein-Coupled Receptor; EpCAM, epithelial cell adhesion molecule

## 6.2 New insights into how inflammation and epigenetics impact CRC progression

Inflammation is a well-known driver of tumorigenesis, with bacteria-induced inflammation correlating with tumour multiplicity[139]. In this study, we sought to identify novel mechanisms driving inflammation-induced cancer by utilising a genetically susceptible mouse model. This allowed us to study the functions of innate cells, and in particular neutrophils, in CRC

progression for which literature is lacking despite a high correlation of neutrophils with adverse outcomes[93].

Notwithstanding the immunocompromised nature of the *Hh*AOM model, our data provides insight into how neutrophil-rich inflammation may shape the epigenetic landscape of damaged epithelial cells. Indeed, an emerging area of CRC research is posttranslational modifications and epigenetics[242]. For example, studies have shown that microbiota-derived butyrate can function as a histone de-acetylase (HDAC) inhibitor to dampen colitis and colon cancer progression[243,244]. HDAC inhibitors are now being tested in a large range of clinical trials for solid tumours including CRC[245]. Notably, this example provides strong evidence that an epigenetic function detected under specific circumstances can be broadly applicable in the cancer field and therapeutically tangible in human disease, irrespective of the specific protein mediating the effect.

It remains to be elucidated if the oncogenic PAD4 observation in our mouse model (**Figure 4.10**) can be similarly utilised. PAD4 is a well-known epigenetic regulator due its ability to citrullinate histones and consequently open chromatin[195]. Therefore, a broad assessment of downstream citrullination bears therapeutic potential. Citrullination has been found to play a role in different disease pathologies including multiple sclerosis, rheumatoid arthritis (RA) and cardiovascular disease[246]. Although citrullination data is limited in the cancer field, Yuzhalin and colleagues have shown extensive ECM protein citrullination in bulk CRC liver metastasis by proteomics[193].

Despite observing elevated *Padi4* mRNA at week 10, the precise trigger is unknown. Independently, *Padi4* expression and citrullination have however been shown to be triggered by environmental factors including DNA damage[191,192] and cytokines such as IL-6[247]. These factors are known to be crucial in tumorigenesis[74,79] , but whether they upregulate *Padi4* is unknown. Irrespective of the trigger, knowing that PAD4 is upregulated early in disease (**Figure 4.6**) has intriguing clinical potential. To this end, measurements of citrullinated protein fragments are widely used as a diagnostic approach in RA[248].

## 6.3 The importance of TME characterisation and patient stratification

Undoubtedly, more data must be accumulated to prove the roles of PAD4 and cytokines in any crosstalk outlined in our proposed model. Prospective studies could greatly benefit from assessing dynamic expression of targets at a protein level and simultaneously bringing them into a spatial context. Temporal assessment will be critical because enzymes and/or cytokines can play pleiotropic or even contrasting roles at different stages of disease[3,74]. In particular IL-22, which we suggest is vital in promoting tumorigenesis in this model, is shown to be protective at early stages but detrimental in late disease[116]. Thus, only by assessing key factors such as PAD4 and IL-22 in a greater framework, i.e by longitudinal imaging and spatial proteomics, can downstream functions be clarified and therapeutic intervention be considered. Along these lines, only imaging approaches allowed us to identify that cancerous epithelial cells were key expressors of PAD4 (**Figure 4.5**).

Beyond tumour cells, the TME consists of many cell types including vascular cells, cancer-associated fibroblasts (CAFs) and leukocytes[78]. Unbiased analysis of the *Hh*AOM model revealed a dominant neutrophil signature within tumours leading us to hypothesise that neutrophils may play a crucial role in disease. In previous studies, neutrophils and NETs, notably triggered by PAD4, have been shown to play tumour-promoting roles ranging from enhanced proliferation to barricading tumour cells from cytotoxic cell attack[104,105]. These conclusions are based in part on utilising the GSK-484 PAD4 inhibitor as a tool to block netosis; they do not however mention what contribution blocking PAD4 activity in other cells of the TME, including cancer cells, may have. Therefore, despite therapeutic potential, it will be crucial to define the cellular composition of a patient TME and disentangle cell-specific from common mechanisms targeted by any drug to ensure accuracy and beneficial outcome.

CAF research is rapidly expanding with multiple subsets having been described with functions ranging from producing ECM components to secreting exosomes, cytokines and chemokines[249]. Of particular relevance here, CAFs were recently shown to trigger PAD4-driven netosis and tumorigenesis in melanoma models[104]. This, taken together with their crucial ability to remodel the ECM during cancer[249] suggests that CAFs also warrant further investigation in our model. Fortunately, we have banked tissue from longitudinal studies, enabling assessment of the frequency, location and potential functional output of neutrophils, CAFs and many other TME constituents over time.

Mapping the cellular composition of the *Hh*AOM TME will also be highly beneficial in assessing the overall translational potential of this study. Segregating heterogenous CRC into distinct subtypes with biological interpretability has provided a basis for more targeted analysis and

interventions[75]. Most recently, distinct multi-cellular inflammatory hubs were characterised in mismatch repair deficient versus proficient tumours. Therefore, a more targeted CRC patient stratification on the basis the *Hh*AOM murine TME, could help explain variable *PADI4* expression in bulk tissue from CRC patients (**Figure 4.12**).

## 6.4 Deciphering the role of the microbiome in CRC

The intricacy of the dialogue between the microbiota and cells of the TME can be seen by comparing the results presented in this thesis with a recently published study probing the downstream effects of *Hh* in CRC. In this *Hh*AOM model, mice lack an adaptive immune system and are therefore susceptible to the pathobiont *Hh* driving inflammation and subsequent tumorigenesis. In contrast, *Hh* colonisation leads to $CD4^+$ T cell and B cell-dependent anti-tumour immunity in immunocompetent mice[240]. In the future, gnotobiotic studies in different genetically maladapted mouse models with *Hh* and/or other identified microbes will aid our understanding of this complex crosstalk in different contexts.

Notably, to uncover further mechanisms between the microbiome and cancer, the CRC field must continue to distinguish between causation and consequence. This thesis has noted causal observations, such as the dependency of *Hh* colonisation to create dysbiosis and trigger inflammation needed for cancer onset, but also consequential observations including a potential tumour-specific niche for better adapted bacteria. This niche could be the result of a potential metabolic shift during tumour establishment, such as hypoxia[250], which could favour anaerobic species. However, this awaits future verification by longitudinal BugFISH imaging.

Lastly this, and many other studies before, solely focus on bacteria and neglect other important members of the microbiome such as fungi, viruses or archaea, all of which might shape CRC[3]. Therefore, future analysis must take all microbiome members into consideration with a potential focus on their collective functional output. With improved mechanistic insight it may be possible to stratify patients using metabolomic readouts.

## 6.5 Conclusion

In conclusion, functional studies on the host-microbiota interface in intestinal diseases, such CRC and IBD, will be crucial going forward. Ideally, these will be on the basis of stratified patient cohorts which identify targets with translational potential alongside GEMMs and other reductionist approaches to probe and validate any mechanisms. Overall, only by unravelling the wide-spanning functions which underlie the host-microbiota interface in health and disease will we be able to access the key to the sought-after door to remedy.

# 7 References


1 Garrett, W. S. The gut microbiota and colon cancer. *Science* **364**, 1133-1135, doi:10.1126/science.aaw2367 (2019).

2 Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121-141, doi:10.1016/j.cell.2014.03.011 (2014).

3 Janney, A., Powrie, F. & Mann, E. H. Host-microbiota maladaptation in colorectal cancer. *Nature* **585**, 509-517, doi:10.1038/s41586-020-2729-3 (2020).

4 Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. *Nat Rev Immunol* **14**, 667-685, doi:10.1038/nri3738 (2014).

5 Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C. The healthy human microbiome. *Genome Med* **8**, 51, doi:10.1186/s13073-016-0307-y (2016).

6 Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat Rev Gastroenterol Hepatol* **16**, 19-34, doi:10.1038/s41575-018-0081-y (2019).

7 Kaiko, G. E. *et al.* The Colonic Crypt Protects Stem Cells from Microbiota-Derived Metabolites. *Cell* **165**, 1708-1720, doi:10.1016/j.cell.2016.05.018 (2016).

8 Gerbe, F., Legraverend, C. & Jay, P. The intestinal epithelium tuft cells: specification and function. *Cell Mol Life Sci* **69**, 2907-2917, doi:10.1007/s00018-012-0984-7 (2012).

9 Sprangers, J., Zaalberg, I. C. & Maurice, M. M. Organoid-based modeling of intestinal development, regeneration, and repair. *Cell Death Differ* **28**, 95-107, doi:10.1038/s41418-020-00665-z (2021).

10 Johansson, M. E., Larsson, J. M. & Hansson, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4659-4665, doi:10.1073/pnas.1006451107 (2011).

11 Van der Sluis, M. *et al.* Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* **131**, 117-129, doi:10.1053/j.gastro.2006.04.020 (2006).

12 Bevins, C. L. & Salzman, N. H. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nat Rev Microbiol* **9**, 356-368, doi:10.1038/nrmicro2546 (2011).

13 Peterson, L. W. & Artis, D. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat Rev Immunol* **14**, 141-153, doi:10.1038/nri3608 (2014).

14 Shalapour, S. & Karin, M. Immunity, inflammation, and cancer: an eternal fight between good and evil. *J Clin Invest* **125**, 3347-3355, doi:10.1172/JCI80007 (2015).

15 Koh, T. J. & DiPietro, L. A. Inflammation and wound healing: the role of the macrophage. *Expert Rev Mol Med* **13**, e23, doi:10.1017/S1462399411001943 (2011).

16 Medzhitov, R. Inflammation 2010: new adventures of an old flame. *Cell* **140**, 771-776, doi:10.1016/j.cell.2010.03.006 (2010).

17 Gasteiger, G. *et al.* Cellular Innate Immunity: An Old Game with New Players. *J Innate Immun* **9**, 111-125, doi:10.1159/000453397 (2017).

18 Casanova-Acebes, M. *et al.* Rhythmic modulation of the hematopoietic niche through neutrophil clearance. *Cell* **153**, 1025-1035, doi:10.1016/j.cell.2013.04.040 (2013).

19 Ng, L. G., Ostuni, R. & Hidalgo, A. Heterogeneity of neutrophils. *Nat Rev Immunol* **19**, 255-265, doi:10.1038/s41577-019-0141-8 (2019).

20 Borregaard, N. Neutrophils, from marrow to microbes. *Immunity* **33**, 657-670, doi:10.1016/j.immuni.2010.11.011 (2010).

21 Evrard, M. *et al.* Developmental Analysis of Bone Marrow Neutrophils Reveals Populations Specialized in Expansion, Trafficking, and Effector Functions. *Immunity* **48**, 364-379 e368, doi:10.1016/j.immuni.2018.02.002 (2018).

22 Zhu, Y. P. *et al.* Identification of an Early Unipotent Neutrophil Progenitor with Pro-tumoral Activity in Mouse and Human Bone Marrow. *Cell Rep* **24**, 2329-2341 e2328, doi:10.1016/j.celrep.2018.07.097 (2018).

23 Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271-281, doi:10.1038/ncb3493 (2017).

24 Hedrick, C. C. & Malanchi, I. Neutrophils in cancer: heterogeneous and multifaceted. *Nat Rev Immunol* **22**, 173-187, doi:10.1038/s41577-021-00571-6 (2022).

25 Liu, F., Wu, H. Y., Wesselschmidt, R., Kornaga, T. & Link, D. C. Impaired production and increased apoptosis of neutrophils in granulocyte colony-stimulating factor receptor-deficient mice. *Immunity* **5**, 491-501, doi:10.1016/s1074-7613(00)80504-x (1996).

26 Lieschke, G. J. *et al.* Mice lacking granulocyte colony-stimulating factor have chronic neutropenia, granulocyte and macrophage progenitor cell deficiency, and impaired neutrophil mobilization. *Blood* **84**, 1737-1746 (1994).

27 Walker, F. *et al.* IL6/sIL6R complex contributes to emergency granulopoietic responses in G-CSF- and GM-CSF-deficient mice. *Blood* **111**, 3978-3985, doi:10.1182/blood-2007-10-119636 (2008).

28 Romani, L. *et al.* Impaired neutrophil response and CD4+ T helper cell 1 development in interleukin 6-deficient mice infected with Candida albicans. *J Exp Med* **183**, 1345-1355, doi:10.1084/jem.183.4.1345 (1996).

29 Eash, K. J., Greenbaum, A. M., Gopalan, P. K. & Link, D. C. CXCR2 and CXCR4 antagonistically regulate neutrophil trafficking from murine bone marrow. *J Clin Invest* **120**, 2423-2431, doi:10.1172/JCI41649 (2010).

30 Adrover, J. M. *et al.* Programmed 'disarming' of the neutrophil proteome reduces the magnitude of inflammation. *Nat Immunol* **21**, 135-144, doi:10.1038/s41590-019-0571-2 (2020).

31 Stark, M. A. *et al.* Phagocytosis of apoptotic neutrophils regulates granulopoiesis via IL-23 and IL-17. *Immunity* **22**, 285-294, doi:10.1016/j.immuni.2005.01.011 (2005).

32 Cua, D. J. & Tato, C. M. Innate IL-17-producing cells: the sentinels of the immune system. *Nat Rev Immunol* **10**, 479-489, doi:10.1038/nri2800 (2010).

33 Schwarzenberger, P. *et al.* Requirement of endogenous stem cell factor and granulocyte-colony-stimulating factor for IL-17-mediated granulopoiesis. *J Immunol* **164**, 4783-4789, doi:10.4049/jimmunol.164.9.4783 (2000).

34 Deniset, J. F., Surewaard, B. G., Lee, W. Y. & Kubes, P. Splenic Ly6G(high) mature and Ly6G(int) immature neutrophils contribute to eradication of S. pneumoniae. *J Exp Med* **214**, 1333-1350, doi:10.1084/jem.20161621 (2017).

35 De Filippo, K. & Rankin, S. M. The Secretive Life of Neutrophils Revealed by Intravital Microscopy. *Front Cell Dev Biol* **8**, 603230, doi:10.3389/fcell.2020.603230 (2020).

36 Martin, C. *et al.* Chemokines acting via CXCR2 and CXCR4 control the release of neutrophils from the bone marrow and their return following senescence. *Immunity* **19**, 583-593, doi:10.1016/s1074-7613(03)00263-2 (2003).

37 Gordy, C., Pua, H., Sempowski, G. D. & He, Y. W. Regulation of steady-state neutrophil homeostasis by macrophages. *Blood* **117**, 618-629, doi:10.1182/blood-2010-01-265959 (2011).

38 Furze, R. C. & Rankin, S. M. The role of the bone marrow in neutrophil clearance under homeostatic conditions in the mouse. *FASEB J* **22**, 3111-3119, doi:10.1096/fj.08-109876 (2008).

39 Suratt, B. T. *et al.* Neutrophil maturation and activation determine anatomic site of clearance from circulation. *Am J Physiol Lung Cell Mol Physiol* **281**, L913-921, doi:10.1152/ajplung.2001.281.4.L913 (2001).

40 Capucetti, A., Albano, F. & Bonecchi, R. Multiple Roles for Chemokines in Neutrophil Biology. *Front Immunol* **11**, 1259, doi:10.3389/fimmu.2020.01259 (2020).

41 Ley, K., Laudanna, C., Cybulsky, M. I. & Nourshargh, S. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nat Rev Immunol* **7**, 678-689, doi:10.1038/nri2156 (2007).

42 Jaillon, S. *et al.* Neutrophil diversity and plasticity in tumour progression and therapy. *Nat Rev Cancer* **20**, 485-503, doi:10.1038/s41568-020-0281-y (2020).

43 Zhang, D. *et al.* Neutrophil ageing is regulated by the microbiome. *Nature* **525**, 528-532, doi:10.1038/nature15367 (2015).

44 Papayannopoulos, V. Neutrophil extracellular traps in immunity and disease. *Nat Rev Immunol* **18**, 134-147, doi:10.1038/nri.2017.105 (2018).

45 Xie, X. *et al.* Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. *Nat Immunol* **21**, 1119-1133, doi:10.1038/s41590-020-0736-z (2020).

46 Ballesteros, I. *et al.* Co-option of Neutrophil Fates by Tissue Environments. *Cell* **183**, 1282-1297 e1218, doi:10.1016/j.cell.2020.10.003 (2020).

47 Khoyratty, T. E. *et al.* Distinct transcription factor networks control neutrophil-driven inflammation. *Nat Immunol* **22**, 1093-1106, doi:10.1038/s41590-021-00968-4 (2021).

48 Spits, H. *et al.* Innate lymphoid cells--a proposal for uniform nomenclature. *Nat Rev Immunol* **13**, 145-149, doi:10.1038/nri3365 (2013).

49 Klose, C. S. & Artis, D. Innate lymphoid cells as regulators of immunity, inflammation and tissue homeostasis. *Nat Immunol* **17**, 765-774, doi:10.1038/ni.3489 (2016).

50 Kiessling, R., Klein, E., Pross, H. & Wigzell, H. "Natural" killer cells in the mouse. II. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Characteristics of the killer cell. *Eur J Immunol* **5**, 117-121, doi:10.1002/eji.1830050209 (1975).

51 Diefenbach, A., Colonna, M. & Koyasu, S. Development, differentiation, and diversity of innate lymphoid cells. *Immunity* **41**, 354-365, doi:10.1016/j.immuni.2014.09.005 (2014).

52 Sonnenberg, G. F. *et al.* Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. *Science* **336**, 1321-1325, doi:10.1126/science.1222551 (2012).

53 Hernandez, P. P. *et al.* Interferon-lambda and interleukin 22 act synergistically for the induction of interferon-stimulated genes and control of rotavirus infection. *Nat Immunol* **16**, 698-707, doi:10.1038/ni.3180 (2015).

54 Pearson, C. *et al.* ILC3 GM-CSF production and mobilisation orchestrate acute intestinal inflammation. *Elife* **5**, e10066, doi:10.7554/eLife.10066 (2016).

55 Buonocore, S. *et al.* Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. *Nature* **464**, 1371-1375, doi:10.1038/nature08949 (2010).

56 Chaplin, D. D. Overview of the immune response. *J Allergy Clin Immunol* **125**, S3-23, doi:10.1016/j.jaci.2009.12.980 (2010).

57 Bain, C. C. *et al.* Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. *Nat Immunol* **15**, 929-937, doi:10.1038/ni.2967 (2014).

58 Shaw, T. N. *et al.* Tissue-resident macrophages in the intestine are long lived and defined by Tim-4 and CD4 expression. *J Exp Med* **215**, 1507-1518, doi:10.1084/jem.20180019 (2018).

59 Nielsen, S. C. A. & Boyd, S. D. Human adaptive immune receptor repertoire analysis-Past, present, and future. *Immunol Rev* **284**, 9-23, doi:10.1111/imr.12667 (2018).

60 Kuo, T. C. & Schlissel, M. S. Mechanisms controlling expression of the RAG locus during lymphocyte development. *Curr Opin Immunol* **21**, 173-178, doi:10.1016/j.coi.2009.03.008 (2009).

61 Farber, D. L., Netea, M. G., Radbruch, A., Rajewsky, K. & Zinkernagel, R. M. Immunological memory: lessons from the past and a look to the future. *Nat Rev Immunol* **16**, 124-128, doi:10.1038/nri.2016.13 (2016).

62 Janeway, C. A., Jr. The T cell receptor as a multicomponent signalling machine: CD4/CD8 coreceptors and CD45 in T cell activation. *Annu Rev Immunol* **10**, 645-674, doi:10.1146/annurev.iy.10.040192.003241 (1992).

63 Skapenko, A., Leipe, J., Lipsky, P. E. & Schulze-Koops, H. The role of the T cell in autoimmune inflammation. *Arthritis Res Ther* **7 Suppl 2**, S4-14, doi:10.1186/ar1703 (2005).

64 Vignali, D. A., Collison, L. W. & Workman, C. J. How regulatory T cells work. *Nat Rev Immunol* **8**, 523-532, doi:10.1038/nri2343 (2008).

65 Mestecky, J. & Russell, M. W. Specific antibody activity, glycan heterogeneity and polyreactivity contribute to the protective activity of S-IgA at mucosal surfaces. *Immunol Lett* **124**, 57-62, doi:10.1016/j.imlet.2009.03.013 (2009).

66 Jasperson, K. W., Tuohy, T. M., Neklason, D. W. & Burt, R. W. Hereditary and familial colon cancer. *Gastroenterology* **138**, 2044-2058, doi:10.1053/j.gastro.2010.01.054 (2010).

67 Bogaert, J. & Prenen, H. Molecular genetics of colorectal cancer. *Ann Gastroenterol* **27**, 9-14 (2014).

68 Wu, S., Powers, S., Zhu, W. & Hannun, Y. A. Substantial contribution of extrinsic risk factors to cancer development. *Nature* **529**, 43-47, doi:10.1038/nature16166 (2016).

69 Collaborators, G. B. D. C. C. The global, regional, and national burden of colorectal cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* **4**, 913-933, doi:10.1016/S2468-1253(19)30345-0 (2019).

70 Jess, T., Rungoe, C. & Peyrin-Biroulet, L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. *Clin Gastroenterol Hepatol* **10**, 639-645, doi:10.1016/j.cgh.2012.01.010 (2012).

71 Ulcerative colitis. *Nat Rev Dis Primers* **6**, 73, doi:10.1038/s41572-020-00215-4 (2020).

72 Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767, doi:10.1016/0092-8674(90)90186-i (1990).

73 Punt, C. J., Koopman, M. & Vermeulen, L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nat Rev Clin Oncol* **14**, 235-246, doi:10.1038/nrclinonc.2016.171 (2017).

74 West, N. R., McCuaig, S., Franchini, F. & Powrie, F. Emerging cytokine networks in colorectal cancer. *Nat Rev Immunol* **15**, 615-629, doi:10.1038/nri3896 (2015).

75 Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350-1356, doi:10.1038/nm.3967 (2015).

76 Isella, C. *et al.* Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nat Commun* **8**, 15107, doi:10.1038/ncomms15107 (2017).

77 Paget, S. The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev* **8**, 98-101 (1989).

78 Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**, 309-322, doi:10.1016/j.ccr.2012.02.022 (2012).

79 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

80 Bhome, R. *et al.* The Colorectal Cancer Microenvironment: Strategies for Studying the Role of Cancer-Associated Fibroblasts. *Methods Mol Biol* **1765**, 87-98, doi:10.1007/978-1-4939-7765-9_6 (2018).

81 Burnet, F. M. The concept of immunological surveillance. *Prog Exp Tumor Res* **13**, 1-27, doi:10.1159/000386035 (1970).

82 Medler, T. R., Cotechini, T. & Coussens, L. M. Immune response to cancer therapy: mounting an effective antitumor response and mechanisms of resistance. *Trends Cancer* **1**, 66-75, doi:10.1016/j.trecan.2015.07.008 (2015).

83 Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960-1964, doi:10.1126/science.1129139 (2006).

84 Doshi, A. S. & Asrani, K. H. in *Cancer Immunology and Immunotherapy* 19-61 (2022).

85 Kabelitz, D., Serrano, R., Kouakanou, L., Peters, C. & Kalyan, S. Cancer immunotherapy with gammadelta T cells: many paths ahead of us. *Cell Mol Immunol* **17**, 925-939, doi:10.1038/s41423-020-0504-x (2020).

86 Topalian, S. L. *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* **366**, 2443-2454, doi:10.1056/NEJMoa1200690 (2012).

87 Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509-2520, doi:10.1056/NEJMoa1500596 (2015).

88 Galon, J. & Bruni, D. Tumor Immunology and Tumor Evolution: Intertwined Histories. *Immunity* **52**, 55-81, doi:10.1016/j.immuni.2019.12.018 (2020).

89 Veglia, F., Sanseviero, E. & Gabrilovich, D. I. Myeloid-derived suppressor cells in the era of increasing myeloid cell diversity. *Nat Rev Immunol* **21**, 485-498, doi:10.1038/s41577-020-00490-y (2021).

90 Noy, R. & Pollard, J. W. Tumor-associated macrophages: from mechanisms to therapy. *Immunity* **41**, 49-61, doi:10.1016/j.immuni.2014.06.010 (2014).

91 Brockmann, L., Giannou, A. D., Gagliani, N. & Huber, S. Regulation of TH17 Cells and Associated Cytokines in Wound Healing, Tissue Regeneration, and Carcinogenesis. *Int J Mol Sci* **18**, doi:10.3390/ijms18051033 (2017).

92 Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883-899, doi:10.1016/j.cell.2010.01.025 (2010).

93 Gentles, A. J. *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* **21**, 938-945, doi:10.1038/nm.3909 (2015).

94 Mantovani, A., Cassatella, M. A., Costantini, C. & Jaillon, S. Neutrophils in the activation and regulation of innate and adaptive immunity. *Nat Rev Immunol* **11**, 519-531, doi:10.1038/nri3024 (2011).

95 Canli, O. *et al.* Myeloid Cell-Derived Reactive Oxygen Species Induce Epithelial Mutagenesis. *Cancer Cell* **32**, 869-883 e865, doi:10.1016/j.ccell.2017.11.004 (2017).

96 Coffelt, S. B. *et al.* IL-17-producing gammadelta T cells and neutrophils conspire to promote breast cancer metastasis. *Nature* **522**, 345-348, doi:10.1038/nature14282 (2015).

97 Rice, C. M. *et al.* Tumour-elicited neutrophils engage mitochondrial metabolism to circumvent nutrient limitations and maintain immune suppression. *Nat Commun* **9**, 5099, doi:10.1038/s41467-018-07505-2 (2018).

98 Wislez, M. *et al.* Hepatocyte growth factor production by neutrophils infiltrating bronchioloalveolar subtype pulmonary adenocarcinoma: role in tumor progression and death. *Cancer Res* **63**, 1405-1412 (2003).

99 Owusu, B. Y., Galemmo, R., Janetka, J. & Klampfer, L. Hepatocyte Growth Factor, a Key Tumor-Promoting Factor in the Tumor Microenvironment. *Cancers (Basel)* **9**, doi:10.3390/cancers9040035 (2017).

100 Scapini, P., Marini, O., Tecchio, C. & Cassatella, M. A. Human neutrophils in the saga of cellular heterogeneity: insights and open questions. *Immunol Rev* **273**, 48-60, doi:10.1111/imr.12448 (2016).

101 Nozawa, H., Chiu, C. & Hanahan, D. Infiltrating neutrophils mediate the initial angiogenic switch in a mouse model of multistage carcinogenesis. *Proc Natl Acad Sci U S A* **103**, 12493-12498, doi:10.1073/pnas.0601807103 (2006).

102 van der Windt, D. J. *et al.* Neutrophil extracellular traps promote inflammation and development of hepatocellular carcinoma in nonalcoholic steatohepatitis. *Hepatology* **68**, 1347-1360, doi:10.1002/hep.29914 (2018).

103 Guglietta, S. *et al.* Coagulation induced by C3aR-dependent NETosis drives protumorigenic neutrophils during small intestinal tumorigenesis. *Nat Commun* **7**, 11037, doi:10.1038/ncomms11037 (2016).

104 Munir, H. *et al.* Stromal-driven and Amyloid beta-dependent induction of neutrophil extracellular traps modulates tumor growth. *Nat Commun* **12**, 683, doi:10.1038/s41467-021-20982-2 (2021).

105 Teijeira, A. *et al.* CXCR1 and CXCR2 Chemokine Receptor Agonists Produced by Tumors Induce Neutrophil Extracellular Traps that Interfere with Immune Cytotoxicity. *Immunity* **52**, 856-871 e858, doi:10.1016/j.immuni.2020.03.001 (2020).

106 Albrengues, J. *et al.* Neutrophil extracellular traps produced during inflammation awaken dormant cancer cells in mice. *Science* **361**, doi:10.1126/science.aao4227 (2018).

107 Jackstadt, R. *et al.* Epithelial NOTCH Signaling Rewires the Tumor Microenvironment of Colorectal Cancer to Drive Poor-Prognosis Subtypes and Metastasis. *Cancer Cell* **36**, 319-336 e317, doi:10.1016/j.ccell.2019.08.003 (2019).

108 Fridlender, Z. G. *et al.* Polarization of tumor-associated neutrophil phenotype by TGF-beta: "N1" versus "N2" TAN. *Cancer Cell* **16**, 183-194, doi:10.1016/j.ccr.2009.06.017 (2009).

109 Ponzetta, A. *et al.* Neutrophils Driving Unconventional T Cells Mediate Resistance against Murine Sarcomas and Selected Human Tumors. *Cell* **178**, 346-360 e324, doi:10.1016/j.cell.2019.05.047 (2019).

110 Singhal, S. *et al.* Origin and Role of a Subset of Tumor-Associated Neutrophils with Antigen-Presenting Cell Features in Early-Stage Human Lung Cancer. *Cancer Cell* **30**, 120-135, doi:10.1016/j.ccell.2016.06.001 (2016).

111 Eruslanov, E. B. *et al.* Tumor-associated neutrophils stimulate T cell responses in early-stage human lung cancer. *J Clin Invest* **124**, 5466-5480, doi:10.1172/JCI77053 (2014).

112 Andzinski, L. *et al.* Type I IFNs induce anti-tumor polarization of tumor associated neutrophils in mice and human. *Int J Cancer* **138**, 1982-1993, doi:10.1002/ijc.29945 (2016).

113 Governa, V. *et al.* The Interplay Between Neutrophils and CD8(+) T Cells Improves Survival in Human Colorectal Cancer. *Clin Cancer Res* **23**, 3847-3858, doi:10.1158/1078-0432.CCR-16-2047 (2017).

114 Zilionis, R. *et al.* Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity* **50**, 1317-1334 e1310, doi:10.1016/j.immuni.2019.03.009 (2019).

115 Dudakov, J. A., Hanash, A. M. & van den Brink, M. R. Interleukin-22: immunobiology and pathology. *Annu Rev Immunol* **33**, 747-785, doi:10.1146/annurev-immunol-032414-112123 (2015).

116 Hernandez, P., Gronke, K. & Diefenbach, A. A catch-22: Interleukin-22 and cancer. *Eur J Immunol* **48**, 15-31, doi:10.1002/eji.201747183 (2018).

117 Zindl, C. L. *et al.* IL-22-producing neutrophils contribute to antimicrobial defense and restitution of colonic epithelial integrity during colitis. *Proc Natl Acad Sci U S A* **110**, 12768-12773, doi:10.1073/pnas.1300318110 (2013).

118 Wolk, K. *et al.* IL-22 increases the innate immunity of tissues. *Immunity* **21**, 241-254, doi:10.1016/j.immuni.2004.07.007 (2004).

119 Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nat Med* **14**, 282-289, doi:10.1038/nm1720 (2008).

120 Kamanaka, M. *et al.* Memory/effector (CD45RB(lo)) CD4 T cells are controlled directly by IL-10 and cause IL-22-dependent intestinal pathology. *J Exp Med* **208**, 1027-1040, doi:10.1084/jem.20102149 (2011).

121 Wolk, K. *et al.* IL-22 induces lipopolysaccharide-binding protein in hepatocytes: a potential systemic role of IL-22 in Crohn's disease. *J Immunol* **178**, 5973-5981, doi:10.4049/jimmunol.178.9.5973 (2007).

122 Brand, S. *et al.* IL-22 is increased in active Crohn's disease and promotes proinflammatory gene expression and intestinal epithelial cell migration. *Am J Physiol Gastrointest Liver Physiol* **290**, G827-838, doi:10.1152/ajpgi.00513.2005 (2006).

123 Geremia, A. *et al.* IL-23-responsive innate lymphoid cells are increased in inflammatory bowel disease. *J Exp Med* **208**, 1127-1133, doi:10.1084/jem.20101712 (2011).

124 Kirchberger, S. *et al.* Innate lymphoid cells sustain colon cancer through production of interleukin-22 in a mouse model. *J Exp Med* **210**, 917-931, doi:10.1084/jem.20122308 (2013).

125 Huber, S. *et al.* IL-22BP is regulated by the inflammasome and modulates tumorigenesis in the intestine. *Nature* **491**, 259-263, doi:10.1038/nature11535 (2012).

126 Gronke, K. *et al.* Interleukin-22 protects intestinal stem cells against genotoxic stress. *Nature* **566**, 249-253, doi:10.1038/s41586-019-0899-7 (2019).

127 McCuaig, S. *et al.* The interleukin 22 pathway interacts with mutant KRAS to promote poor prognosis in colon cancer. *Clin Cancer Res*, doi:10.1158/1078-0432.CCR-19-1086 (2020).

128 Kryczek, I. *et al.* IL-22(+)CD4(+) T cells promote colorectal cancer stemness via STAT3 transcription factor activation and induction of the methyltransferase DOT1L. *Immunity* **40**, 772-784, doi:10.1016/j.immuni.2014.03.010 (2014).

129 Nomura, A. *et al.* Helicobacter pylori infection and gastric carcinoma among Japanese Americans in Hawaii. *N Engl J Med* **325**, 1132-1136, doi:10.1056/NEJM199110173251604 (1991).

130 Parsonnet, J. *et al.* Helicobacter pylori infection and the risk of gastric carcinoma. *N Engl J Med* **325**, 1127-1131, doi:10.1056/NEJM199110173251603 (1991).

131 Reddy, B. S., Weisburger, J. H., Narisawa, T. & Wynder, E. L. Colon carcinogenesis in germ-free rats with 1,2-dimethylhydrazine and N-methyl-n'-nitro-N-nitrosoguanidine. *Cancer Res* **34**, 2368-2372 (1974).

132 Wong, S. H. *et al.* Gavage of Fecal Samples From Patients With Colorectal Cancer Promotes Intestinal Carcinogenesis in Germ-Free and Conventional Mice. *Gastroenterology* **153**, 1621-1633 e1626, doi:10.1053/j.gastro.2017.08.022 (2017).

133 Olesen, S. W. & Alm, E. J. Dysbiosis is not an answer. *Nat Microbiol* **1**, 16228, doi:10.1038/nmicrobiol.2016.228 (2016).

134 Wirbel, J. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* **25**, 679-689, doi:10.1038/s41591-019-0406-6 (2019).

135 Saus, E., Iraola-Guzman, S., Willis, J. R., Brunet-Vega, A. & Gabaldon, T. Microbiome and colorectal cancer: Roles in carcinogenesis and clinical potential. *Mol Aspects Med* **69**, 93-106, doi:10.1016/j.mam.2019.05.001 (2019).

136 Tilg, H., Adolph, T. E., Gerner, R. R. & Moschen, A. R. The Intestinal Microbiota in Colorectal Cancer. *Cancer Cell* **33**, 954-964, doi:10.1016/j.ccell.2018.03.004 (2018).

137 Ternes, D. *et al.* Microbiome in Colorectal Cancer: How to Get from Meta-omics to Mechanism? *Trends Microbiol* **28**, 401-423, doi:10.1016/j.tim.2020.01.001 (2020).

138 Yan, Y. *et al.* Structure of the Mucosal and Stool Microbiome in Lynch Syndrome. *Cell Host Microbe* **27**, 585-600 e584, doi:10.1016/j.chom.2020.03.005 (2020).

139 Uronis, J. M. *et al.* Modulation of the intestinal microbiota alters colitis-associated colorectal cancer susceptibility. *PLoS One* **4**, e6026, doi:10.1371/journal.pone.0006026 (2009).

140 Cuevas-Ramos, G. *et al.* Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci U S A* **107**, 11537-11542, doi:10.1073/pnas.1001261107 (2010).

141 Tomkovich, S. *et al.* Locoregional Effects of Microbiota in a Preclinical Model of Colon Carcinogenesis. *Cancer Res* **77**, 2620-2632, doi:10.1158/0008-5472.CAN-16-3472 (2017).

142 Cougnoux, A. *et al.* Small-molecule inhibitors prevent the genotoxic and protumoural effects induced by colibactin-producing bacteria. *Gut* **65**, 278-285, doi:10.1136/gutjnl-2014-307241 (2016).

143 Kostic, A. D. *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207-215, doi:10.1016/j.chom.2013.07.007 (2013).

144 Chung, L. *et al.* Bacteroides fragilis Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host Microbe* **23**, 203-214 e205, doi:10.1016/j.chom.2018.01.007 (2018).

145 Thiele Orberg, E. *et al.* The myeloid immune signature of enterotoxigenic Bacteroides fragilis-induced murine colon tumorigenesis. *Mucosal Immunol* **10**, 421-433, doi:10.1038/mi.2016.53 (2017).

146 Gur, C. *et al.* Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity* **42**, 344-355, doi:10.1016/j.immuni.2015.01.010 (2015).

147 Arthur, J. C. *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120-123, doi:10.1126/science.1224820 (2012).

148 Arthur, J. C. *et al.* Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. *Nat Commun* **5**, 4724, doi:10.1038/ncomms5724 (2014).

149 Dennis, K. L. *et al.* Adenomatous polyps are driven by microbe-instigated focal inflammation and are controlled by IL-10-producing T cells. *Cancer Res* **73**, 5905-5913, doi:10.1158/0008-5472.CAN-13-1511 (2013).

150 Singh, N. *et al.* Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* **40**, 128-139, doi:10.1016/j.immuni.2013.12.007 (2014).

151 Malik, A. *et al.* IL-33 regulates the IgA-microbiota axis to restrain IL-1alpha-dependent colitis and tumorigenesis. *J Clin Invest* **126**, 4469-4481, doi:10.1172/JCI88625 (2016).

152 Dejea, C. M. *et al.* Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc Natl Acad Sci U S A* **111**, 18321-18326, doi:10.1073/pnas.1406199111 (2014).

153 Grivennikov, S. I. *et al.* Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. *Nature* **491**, 254-258, doi:10.1038/nature11465 (2012).

154 Wu, P. *et al.* gammadeltaT17 cells promote the accumulation and expansion of myeloid-derived suppressor cells in human colorectal cancer. *Immunity* **40**, 785-800, doi:10.1016/j.immuni.2014.03.013 (2014).

155 Boulard, O., Kirchberger, S., Royston, D. J., Maloy, K. J. & Powrie, F. M. Identification of a genetic locus controlling bacteria-driven colitis and associated cancer through effects on innate inflammation. *J Exp Med* **209**, 1309-1324, doi:10.1084/jem.20120239 (2012).

156 Hue, S. *et al.* Interleukin-23 drives innate and T cell-mediated intestinal inflammation. *J Exp Med* **203**, 2473-2483, doi:10.1084/jem.20061099 (2006).

157 Leach, J. D. G. *et al.* Oncogenic BRAF, unrestrained by TGFbeta-receptor signalling, drives right-sided colonic tumorigenesis. *Nat Commun* **12**, 3464, doi:10.1038/s41467-021-23717-5 (2021).

158 Fumagalli, A. *et al.* A surgical orthotopic organoid transplantation approach in mice to visualize and study colorectal cancer progression. *Nat Protoc* **13**, 235-247, doi:10.1038/nprot.2017.137 (2018).

159 Tuveson, D. & Clevers, H. Cancer modeling meets human organoid technology. *Science* **364**, 952-955, doi:10.1126/science.aaw6985 (2019).

160 Kwong, L. N. & Dove, W. F. APC and its modifiers in colon cancer. *Adv Exp Med Biol* **656**, 85-106, doi:10.1007/978-1-4419-1145-2_8 (2009).

161 Overman, M. J. *et al.* A population-based comparison of adenocarcinoma of the large and small intestine: insights into a rare disease. *Ann Surg Oncol* **19**, 1439-1445, doi:10.1245/s10434-011-2173-6 (2012).

162 Xue, Y., Johnson, R., Desmet, M., Snyder, P. W. & Fleet, J. C. Generation of a transgenic mouse for colorectal cancer research with intestinal cre expression limited to the large intestine. *Mol Cancer Res* **8**, 1095-1104, doi:10.1158/1541-7786.MCR-10-0195 (2010).

163 Jackstadt, R. & Sansom, O. J. Mouse models of intestinal cancer. *J Pathol* **238**, 141-151, doi:10.1002/path.4645 (2016).

164 Oliveira, R. C., Abrantes, A. M., Tralhao, J. G. & Botelho, M. F. The role of mouse models in colorectal cancer research-The need and the importance of the orthotopic models. *Animal Model Exp Med* **3**, 1-8, doi:10.1002/ame2.12102 (2020).

165 Nolte, T. *et al.* Nonproliferative and Proliferative Lesions of the Gastrointestinal Tract, Pancreas and Salivary Glands of the Rat and Mouse. *J Toxicol Pathol* **29**, 1S-125S, doi:10.1293/tox.29.1S (2016).

166 Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* **104**, 13780-13785, doi:10.1073/pnas.0706625104 (2007).

167 Ge, Z., White, D. A., Whary, M. T. & Fox, J. G. Fluorogenic PCR-based quantitative detection of a murine pathogen, Helicobacter hepaticus. *J Clin Microbiol* **39**, 2598-2602, doi:10.1128/JCM.39.7.2598-2602.2001 (2001).

168 Romano, G., Chagani, S. & Kwong, L. N. The path to metastatic mouse models of colorectal cancer. *Oncogene* **37**, 2481-2489, doi:10.1038/s41388-018-0155-x (2018).

169 Ryzhakov, G. *et al.* Alpha kinase 1 controls intestinal inflammation by suppressing the IL-12/Th1 axis. *Nat Commun* **9**, 3797, doi:10.1038/s41467-018-06085-5 (2018).

170 Bai, M. *et al.* CD177 modulates human neutrophil migration through activation-mediated integrin and chemoreceptor regulation. *Blood* **130**, 2092-2100, doi:10.1182/blood-2017-03-768507 (2017).

171 James, K. R. *et al.* Distinct microbial and immune niches of the human colon. *Nat Immunol* **21**, 343-353, doi:10.1038/s41590-020-0602-z (2020).

172 Wera, O., Lancellotti, P. & Oury, C. The Dual Role of Neutrophils in Inflammatory Bowel Diseases. *J Clin Med* **5**, doi:10.3390/jcm5120118 (2016).

173 Friedrich, M. *et al.* IL-1-driven stromal-neutrophil interactions define a subset of patients with inflammatory bowel disease that does not respond to therapies. *Nat Med*, doi:10.1038/s41591-021-01520-5 (2021).

174 Wang, Y. *et al.* Neutrophil infiltration favors colitis-associated tumorigenesis by activating the interleukin-1 (IL-1)/IL-6 axis. *Mucosal Immunol* **7**, 1106-1115, doi:10.1038/mi.2013.126 (2014).

175 Katoh, H. *et al.* CXCR2-expressing myeloid-derived suppressor cells are essential to promote colitis-associated tumorigenesis. *Cancer Cell* **24**, 631-644, doi:10.1016/j.ccr.2013.10.009 (2013).

176 Triner, D. *et al.* Neutrophils Restrict Tumor-Associated Microbiota to Reduce Growth and Invasion of Colon Tumors in Mice. *Gastroenterology* **156**, 1467-1482, doi:10.1053/j.gastro.2018.12.003 (2019).

177 Jukic, A., Bakiri, L., Wagner, E. F., Tilg, H. & Adolph, T. E. Calprotectin: from biomarker to biological function. *Gut* **70**, 1978-1988, doi:10.1136/gutjnl-2021-324855 (2021).

178 West, N. R. *et al.* Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease. *Nat Med* **23**, 579-589, doi:10.1038/nm.4307 (2017).
179 Junk, D. J. *et al.* Oncostatin M promotes cancer cell plasticity through cooperative STAT3-SMAD3 signaling. *Oncogene* **36**, 4001-4013, doi:10.1038/onc.2017.33 (2017).
180 Gschwandtner, M., Derler, R. & Midwood, K. S. More Than Just Attractive: How CCL2 Influences Myeloid Cell Behavior Beyond Chemotaxis. *Front Immunol* **10**, 2759, doi:10.3389/fimmu.2019.02759 (2019).
181 Dmitrieva-Posocco, O. *et al.* Cell-Type-Specific Responses to Interleukin-1 Control Microbial Invasion and Tumor-Elicited Inflammation in Colorectal Cancer. *Immunity* **50**, 166-180 e167, doi:10.1016/j.immuni.2018.11.015 (2019).
182 Goel, A. & Boland, C. R. Epigenetics of colorectal cancer. *Gastroenterology* **143**, 1442-1460 e1441, doi:10.1053/j.gastro.2012.09.032 (2012).
183 Jung, G., Hernandez-Illan, E., Moreira, L., Balaguer, F. & Goel, A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nat Rev Gastroenterol Hepatol* **17**, 111-130, doi:10.1038/s41575-019-0230-y (2020).
184 Zhu, D., Zhang, Y. & Wang, S. Histone citrullination: a new target for tumors. *Mol Cancer* **20**, 90, doi:10.1186/s12943-021-01373-z (2021).
185 Park, J. *et al.* Cancer cells induce metastasis-supporting neutrophil extracellular DNA traps. *Sci Transl Med* **8**, 361ra138, doi:10.1126/scitranslmed.aag1711 (2016).
186 Najmeh, S. *et al.* Neutrophil extracellular traps sequester circulating tumor cells via beta1-integrin mediated interactions. *Int J Cancer* **140**, 2321-2330, doi:10.1002/ijc.30635 (2017).
187 Cools-Lartigue, J. *et al.* Neutrophil extracellular traps sequester circulating tumor cells and promote metastasis. *J Clin Invest*, doi:10.1172/JCI67484 (2013).
188 Kolaczkowska, E. *et al.* Molecular mechanisms of NET formation and degradation revealed by intravital imaging in the liver vasculature. *Nat Commun* **6**, 6673, doi:10.1038/ncomms7673 (2015).
189 Stadler, S. C. *et al.* Dysregulation of PAD4-mediated citrullination of nuclear GSK3beta activates TGF-beta signaling and induces epithelial-to-mesenchymal transition in breast cancer cells. *Proc Natl Acad Sci U S A* **110**, 11851-11856, doi:10.1073/pnas.1308362110 (2013).
190 Duan, Q., Pang, C., Chang, N., Zhang, J. & Liu, W. Overexpression of PAD4 suppresses drug resistance of NSCLC cell lines to gefitinib through inhibiting Elk1-mediated epithelial-mesenchymal transition. *Oncol Rep* **36**, 551-558, doi:10.3892/or.2016.4780 (2016).
191 Tanikawa, C. *et al.* Regulation of histone modification and chromatin structure by the p53-PADI4 pathway. *Nat Commun* **3**, 676, doi:10.1038/ncomms1676 (2012).
192 Tanikawa, C. *et al.* Regulation of protein Citrullination through p53/PADI4 network in DNA damage response. *Cancer Res* **69**, 8761-8769, doi:10.1158/0008-5472.CAN-09-2280 (2009).
193 Yuzhalin, A. E. *et al.* Colorectal cancer liver metastatic growth depends on PAD4-driven citrullination of the extracellular matrix. *Nat Commun* **9**, 4783, doi:10.1038/s41467-018-07306-7 (2018).
194 Nakashima, K. *et al.* PAD4 regulates proliferation of multipotent haematopoietic cells by controlling c-myc expression. *Nat Commun* **4**, 1836, doi:10.1038/ncomms2862 (2013).

195 Christophorou, M. A. *et al.* Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507**, 104-108, doi:10.1038/nature12942 (2014).

196 Pelka, K. *et al.* Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* **184**, 4734-4752 e4720, doi:10.1016/j.cell.2021.08.003 (2021).

197 Xu, D. & Lu, W. Defensins: A Double-Edged Sword in Host Immunity. *Front Immunol* **11**, 764, doi:10.3389/fimmu.2020.00764 (2020).

198 Xu, D. *et al.* Human beta-defensin 3 contributes to the carcinogenesis of cervical cancer via activation of NF-kappaB signaling. *Oncotarget* **7**, 75902-75913, doi:10.18632/oncotarget.12426 (2016).

199 Xu, N. *et al.* Human alpha-defensin-1 inhibits growth of human lung adenocarcinoma xenograft in nude mice. *Mol Cancer Ther* **7**, 1588-1597, doi:10.1158/1535-7163.MCT-08-0010 (2008).

200 Melle, C. *et al.* Discovery and identification of alpha-defensins as low abundant, tumor-derived serum markers in colorectal cancer. *Gastroenterology* **129**, 66-73, doi:10.1053/j.gastro.2005.05.014 (2005).

201 Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432**, 316-323, doi:10.1038/nature03097 (2004).

202 Branzei, D. & Foiani, M. Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol* **9**, 297-308, doi:10.1038/nrm2351 (2008).

203 Chen, J. & Huang, X. F. The signal pathways in azoxymethane-induced colon cancer and preventive implications. *Cancer Biol Ther* **8**, 1313-1317, doi:10.4161/cbt.8.14.8983 (2009).

204 Chang, X. *et al.* Increased PADI4 expression in blood and tissues of patients with malignant tumors. *BMC Cancer* **9**, 40, doi:10.1186/1471-2407-9-40 (2009).

205 Litvinov, S. V. *et al.* Epithelial cell adhesion molecule (Ep-CAM) modulates cell-cell interactions mediated by classic cadherins. *J Cell Biol* **139**, 1337-1348, doi:10.1083/jcb.139.5.1337 (1997).

206 Went, P. T. *et al.* Frequent EpCam protein expression in human carcinomas. *Hum Pathol* **35**, 122-128, doi:10.1016/j.humpath.2003.08.026 (2004).

207 Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003-1007, doi:10.1038/nature06196 (2007).

208 Morgan, R. G., Mortensson, E. & Williams, A. C. Targeting LGR5 in Colorectal Cancer: therapeutic gold or too plastic? *Br J Cancer* **118**, 1410-1418, doi:10.1038/s41416-018-0118-6 (2018).

209 Merlos-Suarez, A. *et al.* The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell* **8**, 511-524, doi:10.1016/j.stem.2011.02.020 (2011).

210 Schepers, A. G. *et al.* Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. *Science* **337**, 730-735, doi:10.1126/science.1224676 (2012).

211 Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608-611, doi:10.1038/nature07602 (2009).

212 Shimokawa, M. *et al.* Visualization and targeting of LGR5(+) human colon cancer stem cells. *Nature* **545**, 187-192, doi:10.1038/nature22081 (2017).

213 Wainwright, E. N. & Scaffidi, P. Epigenetics and Cancer Stem Cells: Unleashing, Hijacking, and Restricting Cellular Plasticity. *Trends Cancer* **3**, 372-386, doi:10.1016/j.trecan.2017.04.004 (2017).

214 Cui, T. X. *et al.* Myeloid-derived suppressor cells enhance stemness of cancer cells by inducing microRNA101 and suppressing the corepressor CtBP2. *Immunity* **39**, 611-621, doi:10.1016/j.immuni.2013.08.025 (2013).

215 Rogers, A. B. & Houghton, J. Helicobacter-based mouse models of digestive system carcinogenesis. *Methods Mol Biol* **511**, 267-295, doi:10.1007/978-1-59745-447-6_11 (2009).

216 Huang, Y. *et al.* Identification of helicobacter species in human liver samples from patients with primary hepatocellular carcinoma. *J Clin Pathol* **57**, 1273-1277, doi:10.1136/jcp.2004.018556 (2004).

217 Touati, E. *et al.* Chronic Helicobacter pylori infections induce gastric mutations in mice. *Gastroenterology* **124**, 1408-1419, doi:10.1016/s0016-5085(03)00266-x (2003).

218 Ward, J. M. *et al.* Inflammatory large bowel disease in immunodeficient mice naturally infected with Helicobacter hepaticus. *Lab Anim Sci* **46**, 15-20 (1996).

219 Fox, J. G. *et al.* Helicobacter hepaticus sp. nov., a microaerophilic bacterium isolated from livers and intestinal mucosal scrapings from mice. *J Clin Microbiol* **32**, 1238-1245, doi:10.1128/jcm.32.5.1238-1245.1994 (1994).

220 Ward, J. M., Anver, M. R., Haines, D. C. & Benveniste, R. E. Chronic active hepatitis in mice caused by Helicobacter hepaticus. *Am J Pathol* **145**, 959-968 (1994).

221 Ward, J. M. *et al.* Chronic active hepatitis and associated liver tumors in mice caused by a persistent bacterial infection with a novel Helicobacter species. *J Natl Cancer Inst* **86**, 1222-1227, doi:10.1093/jnci/86.16.1222 (1994).

222 Foltz, C. J. *et al.* Spontaneous inflammatory bowel disease in multiple mutant mouse lines: association with colonization by Helicobacter hepaticus. *Helicobacter* **3**, 69-78, doi:10.1046/j.1523-5378.1998.08006.x (1998).

223 Chai, J. N. *et al.* Helicobacter species are potent drivers of colonic T cell responses in homeostasis and inflammation. *Sci Immunol* **2**, doi:10.1126/sciimmunol.aal5068 (2017).

224 Cortes-Bratti, X., Frisan, T. & Thelestam, M. The cytolethal distending toxins induce DNA damage and cell cycle arrest. *Toxicon* **39**, 1729-1736, doi:10.1016/s0041-0101(01)00159-3 (2001).

225 Ge, Z. *et al.* Helicobacter hepaticus cytolethal distending toxin promotes intestinal carcinogenesis in 129Rag2-deficient mice. *Cell Microbiol* **19**, doi:10.1111/cmi.12728 (2017).

226 He, Z. *et al.* Campylobacter jejuni promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut* **68**, 289-300, doi:10.1136/gutjnl-2018-317200 (2019).

227 Erdman, S. E. *et al.* Nitric oxide and TNF-alpha trigger colonic inflammation and carcinogenesis in Helicobacter hepaticus-infected, Rag2-deficient mice. *Proc Natl Acad Sci U S A* **106**, 1027-1032, doi:10.1073/pnas.0812347106 (2009).

228 Mangerich, A. *et al.* Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer. *Proc Natl Acad Sci U S A* **109**, E1820-1829, doi:10.1073/pnas.1207829109 (2012).

229 Moschen, A. R. *et al.* Lipocalin 2 Protects from Inflammation and Tumorigenesis Associated with Gut Microbiota Alterations. *Cell Host Microbe* **19**, 455-469, doi:10.1016/j.chom.2016.03.007 (2016).

230 Ott, S. J. *et al.* Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* **53**, 685-693, doi:10.1136/gut.2003.025403 (2004).

231 Yang, Y. *et al.* Dysbiosis of human gut microbiome in young-onset colorectal cancer. *Nat Commun* **12**, 6757, doi:10.1038/s41467-021-27112-y (2021).

232 Kjeldsen, L., Johnsen, A. H., Sengelov, H. & Borregaard, N. Isolation and primary structure of NGAL, a novel protein associated with human neutrophil gelatinase. *J Biol Chem* **268**, 10425-10432 (1993).

233 Kjeldsen, L., Bainton, D. F., Sengelov, H. & Borregaard, N. Identification of neutrophil gelatinase-associated lipocalin as a novel matrix protein of specific granules in human neutrophils. *Blood* **83**, 799-807 (1994).

234 Stallhofer, J. *et al.* Lipocalin-2 Is a Disease Activity Marker in Inflammatory Bowel Disease Regulated by IL-17A, IL-22, and TNF-alpha and Modulated by IL23R Genotype Status. *Inflamm Bowel Dis* **21**, 2327-2340, doi:10.1097/MIB.0000000000000515 (2015).

235 La Vecchia, S. & Sebastian, C. Metabolic pathways regulating colorectal cancer initiation and progression. *Semin Cell Dev Biol* **98**, 63-70, doi:10.1016/j.semcdb.2019.05.018 (2020).

236 Zhang, J. *et al.* Elucidation of Proteus mirabilis as a Key Bacterium in Crohn's Disease Inflammation. *Gastroenterology* **160**, 317-330 e311, doi:10.1053/j.gastro.2020.09.036 (2021).

237 Brugiroux, S. *et al.* Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. *Nat Microbiol* **2**, 16215, doi:10.1038/nmicrobiol.2016.215 (2016).

238 Allen-Vercoe, E., Strauss, J. & Chadee, K. Fusobacterium nucleatum: an emerging gut pathogen? *Gut Microbes* **2**, 294-298, doi:10.4161/gmic.2.5.18603 (2011).

239 Manson McGuire, A. *et al.* Evolution of invasion in a diverse set of Fusobacterium species. *mBio* **5**, e01864, doi:10.1128/mBio.01864-14 (2014).

240 Overacre-Delgoffe, A. E. *et al.* Microbiota-specific T follicular helper cells drive tertiary lymphoid structures and anti-tumor immunity against colorectal cancer. *Immunity*, doi:10.1016/j.immuni.2021.11.003 (2021).

241 Keum, N. & Giovannucci, E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol* **16**, 713-732, doi:10.1038/s41575-019-0189-8 (2019).

242 Sobhani, I. *et al.* Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures. *Proc Natl Acad Sci U S A* **116**, 24285-24295, doi:10.1073/pnas.1912129116 (2019).

243 Chang, P. V., Hao, L., Offermanns, S. & Medzhitov, R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc Natl Acad Sci U S A* **111**, 2247-2252, doi:10.1073/pnas.1322269111 (2014).

244 Zagato, E. *et al.* Endogenous murine microbiota member Faecalibaculum rodentium and its human homologue protect from intestinal tumour growth. *Nat Microbiol* **5**, 511-524, doi:10.1038/s41564-019-0649-5 (2020).

245 Jenke, R., Ressing, N., Hansen, F. K., Aigner, A. & Buch, T. Anticancer Therapy with HDAC Inhibitors: Mechanism-Based Combination Strategies and Future Perspectives. *Cancers (Basel)* **13**, doi:10.3390/cancers13040634 (2021).

246 Vitorino, R. *et al.* Elucidating Citrullination by Mass Spectrometry and Its Role in Disease Pathogenesis. *J Proteome Res* **20**, 38-48, doi:10.1021/acs.jproteome.0c00474 (2021).

247 Yahagi, A. *et al.* IL-6-PAD4 axis in the earliest phase of arthritis in knock-in gp130F759 mice, a model for rheumatoid arthritis. *RMD Open* **5**, e000853, doi:10.1136/rmdopen-2018-000853 (2019).

248 Trouw, L. A., Rispens, T. & Toes, R. E. M. Beyond citrullination: other post-translational protein modifications in rheumatoid arthritis. *Nat Rev Rheumatol* **13**, 331-339, doi:10.1038/nrrheum.2017.15 (2017).

249 Sahai, E. *et al.* A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer* **20**, 174-186, doi:10.1038/s41568-019-0238-1 (2020).

250 Han, N., Pan, Z., Liu, G., Yang, R. & Yujing, B. Hypoxia: The "Invisible Pusher" of Gut Microbiota. *Front Microbiol* **12**, 690600, doi:10.3389/fmicb.2021.690600 (2021).

-

# 8 Appendix

## 8.1 Link to R Code used for analysis

https://github.com/Ajanney/Thesis-AJanney

## 8.2 Table of top 50 genes* DESeq (healthy distal colons vs tumours)

| | **log2FoldChange** | **padj** | **Effectsize** (log2FoldChange x padj) |
|---|---|---|---|
| ***Tacstd2*** | 10.43661296 | 9.37375E-97 | 1002.207972 |
| ***Lrg1*** | 8.095703234 | 4.5451E-118 | 949.9697005 |
| ***Krt90*** | 10.34999346 | 1.23778E-86 | 889.1405711 |
| ***Ly6d*** | 8.157899879 | 5.6798E-106 | 858.5835844 |
| ***Clca4b*** | 8.261957956 | 7.78536E-81 | 661.8548852 |
| ***Lcn2*** | 6.254939481 | 9.3334E-92 | 569.3868926 |
| ***Prss22*** | 9.345613649 | 9.06315E-60 | 551.7904595 |
| ***S100a9*** | 12.99503052 | 9.30293E-41 | 520.2090058 |
| ***Duoxa2*** | 4.837247708 | 1.4796E-99 | 478.0645636 |
| ***Sycn*** | -6.148477511 | 2.68885E-74 | 452.3461577 |
| ***Nos2*** | 7.545899205 | 1.41941E-59 | 444.0602732 |
| ***Cyp2c55*** | -8.505274564 | 2.76446E-52 | 438.5182576 |
| ***Mmp10*** | 8.109974426 | 1.01062E-54 | 437.9014032 |
| ***Anxa8*** | 5.460023547 | 7.70612E-79 | 426.4997165 |
| ***S100a8*** | 13.78471075 | 2.06591E-31 | 422.9823027 |

| | | | |
|---|---|---|---|
| ***Mmp7*** | 7.348228759 | 2.18737E-57 | 416.3512179 |
| ***Serpinb11*** | 12.80871591 | 7.75878E-33 | 411.2905145 |
| ***Mmp9*** | 6.887587253 | 1.2517E-58 | 398.8085114 |
| ***Sult1a1*** | -5.089472756 | 4.87793E-78 | 393.4761202 |
| ***Ltf*** | 6.348713641 | 3.93671E-58 | 364.4470571 |
| ***Spp1*** | 7.316278936 | 6.95475E-49 | 352.3353031 |
| ***Chil1*** | 7.323006202 | 8.83562E-49 | 351.8980051 |
| ***Tgtp1*** | 5.904585495 | 1.06935E-59 | 348.1986058 |
| ***Cxcl2*** | 11.0539186 | 6.78931E-29 | 311.3687054 |
| ***Tppp*** | -4.638393322 | 1.9113E-67 | 309.4674356 |
| ***Mfsd2a*** | 6.26176523 | 1.00937E-49 | 306.8011443 |
| ***Trim10*** | 8.308902107 | 1.70886E-37 | 305.4958389 |
| ***Reg3g*** | 9.327144696 | 4.29505E-33 | 301.8919915 |
| ***Slc30a10*** | -6.382175246 | 1.01613E-47 | 299.9178761 |
| ***Serpina3f*** | 7.231490749 | 1.24755E-41 | 295.7964815 |
| ***Gm6565*** | -20.68047865 | 6.02725E-15 | 294.0739389 |
| ***Il1f9*** | 7.485816238 | 6.71448E-38 | 278.2701558 |
| ***Cxcr2*** | 6.881764884 | 6.46476E-41 | 276.574327 |
| ***Arg1*** | 6.741848806 | 1.92831E-41 | 274.4931751 |
| ***Dpyd*** | -3.675963372 | 2.48468E-74 | 270.5682894 |
| ***Nxpe2*** | -4.980411265 | 5.69538E-55 | 270.1598047 |
| ***Krt6a*** | 7.889827347 | 1.2271E-33 | 259.6630494 |
| ***Iigp1*** | 4.974998416 | 7.8092E-53 | 259.2342 |

| | | | |
|---|---|---|---|
| ***Bcl2l15*** | 6.673739015 | 2.06693E-39 | 258.1714146 |
| ***Cd74*** | 2.96422293 | 9.11777E-87 | 255.0420706 |
| ***Msln*** | 5.483414343 | 7.63306E-46 | 247.3968571 |
| ***Chil3*** | 7.735148225 | 1.15214E-32 | 247.0489929 |
| ***Olfm4*** | 11.74714549 | 9.49555E-22 | 246.9541321 |
| ***Mab21l3*** | 5.587634374 | 1.04127E-44 | 245.7577753 |
| ***Cxcl5*** | 7.818390771 | 4.68792E-32 | 244.9425196 |
| ***Slc25a48*** | 4.096470175 | 1.14339E-58 | 237.3568728 |
| ***Defa24*** | 12.90833115 | 4.92636E-19 | 236.318938 |
| ***Tnfrsf11b*** | 5.15448785 | 1.95104E-45 | 230.4557833 |
| ***F830016B08Rik*** | 6.797964161 | 1.53032E-34 | 229.8746464 |
| ***Hao2*** | -4.626573536 | 1.27675E-49 | 226.2111965 |

*Calculated effectsizes used to sort top 50 genes after DESeq analysis

## 8.3 Table of neutrophil genes* DESeq (untreated distal colons vs tumours)

| | **log2FoldChange** | **padj** | **Effectsize** (log2FoldChange x padj) |
|---|---|---|---|
| ***Lcn2*** | 6.254939481 | 9.3334E-92 | 569.3868926 |
| ***S100a9*** | 12.99503052 | 9.30293E-41 | 520.2090058 |
| ***Nos2*** | 7.545899205 | 1.41941E-59 | 444.0602732 |
| ***S100a8*** | 13.78471075 | 2.06591E-31 | 422.9823027 |
| ***Cxcl2*** | 11.0539186 | 6.78931E-29 | 311.3687054 |
| ***Cxcr2*** | 6.881764884 | 6.46476E-41 | 276.574327 |
| ***Olfm4*** | 11.74714549 | 9.49555E-22 | 246.9541321 |
| ***Cxcl5*** | 7.818390771 | 4.68792E-32 | 244.9425196 |
| ***Padi4*** | 6.434899429 | 2.27023E-34 | 216.4953021 |
| ***Tnf*** | 5.081985014 | 2.97112E-41 | 205.9580119 |
| ***Mmp8*** | 8.652149337 | 6.03302E-23 | 192.2461311 |
| ***Il1b*** | 5.723778166 | 4.01554E-34 | 191.1527586 |
| ***Cxcl1*** | 5.258525502 | 6.66067E-36 | 184.9764273 |
| ***Sell*** | 6.001667428 | 6.2562E-31 | 181.2724996 |
| ***Csf3*** | 8.029208596 | 1.82814E-13 | 102.2759927 |
| ***Fcgr4*** | 4.258923573 | 8.33411E-23 | 94.03337328 |
| ***Tyrobp*** | 2.786349158 | 1.69506E-29 | 80.16553538 |
| ***Ncf1*** | 2.560271521 | 9.48147E-24 | 58.94544902 |
| ***Siglece*** | 2.924006849 | 1.61047E-17 | 49.10298222 |

| | | | |
|---|---|---|---|
| ***Itgb2*** | 2.791503822 | 1.0149E-17 | 47.43763408 |
| ***Fcer1g*** | 2.933818635 | 4.19324E-10 | 27.51173979 |
| ***Stx11*** | 2.805750219 | 2.32136E-10 | 27.03131895 |
| ***Cd84*** | 2.516448836 | 7.43085E-11 | 25.48901271 |
| ***Myo1f*** | 2.179221487 | 1.35672E-10 | 21.50348943 |
| ***Pram1*** | 3.046534618 | 2.45705E-06 | 17.08979581 |
| ***Itgam*** | 2.522456174 | 2.85183E-07 | 16.50916234 |
| ***Cd101*** | 2.322865616 | 7.87119E-08 | 16.50154369 |
| ***Ccl5*** | 3.078954724 | 0.00012466 | 12.02107625 |
| ***Ptafr*** | 1.649371566 | 6.4267E-06 | 8.563557127 |
| ***Il16*** | 1.306970694 | 8.92103E-07 | 7.906630165 |
| ***Ly6g*** | 1.803989912 | 0.000124829 | 7.042208197 |
| ***Cd177*** | 1.672482794 | 0.001123765 | 4.932694381 |
| ***Syk*** | 0.815457416 | 0.000199763 | 3.016772343 |
| ***Anxa3*** | 0.846735039 | 0.000876223 | 2.588795147 |
| ***Dnase1l3*** | 0.973210108 | 0.031965366 | 1.455260847 |
| ***F2rl1*** | 0.344940027 | 0.084059373 | 0.370953283 |

*Gene list was created on the basis of Gene Ontology terms containing “neutrophils”

## 8.4 Table of 118 candidate genes* (Tumour versus Hh 15 weeks)

| | **Quotient**<br>*(Effectsize Tumours/Effectsize Hh 15w)* | **log2**<br>*(Quotient)* |
|---|---|---|
| ***Wif1*** | 162.4021553 | 7.343426969 |
| ***Prss56*** | 126.8854596 | 6.987382943 |
| ***Defa20*** | 99.3293041 | 6.634147499 |
| ***Defa32*** | 99.3293041 | 6.634147499 |
| ***R3hdml*** | 75.91689739 | 6.246349128 |
| ***Mme*** | 48.10661502 | 5.588163384 |
| ***Lyz1*** | 43.70803311 | 5.449826552 |
| ***Sult1d1*** | 37.17611439 | 5.216304084 |
| ***Slc30a2*** | 32.15073126 | 5.006779647 |
| ***Phgr1*** | 31.99963999 | 4.999983769 |
| ***Selenop*** | 26.99713256 | 4.754734278 |
| ***Gm15915*** | 25.85175062 | 4.692190075 |
| ***Ggh*** | 25.03817081 | 4.646057263 |
| ***Cyp2d12*** | 23.89839555 | 4.578841859 |
| ***Esyt3*** | 23.08715421 | 4.529018448 |
| ***Gpr27*** | 23.07264329 | 4.528111389 |
| ***Prkg2*** | 22.24543724 | 4.47543755 |
| ***Sectm1b*** | 21.62333128 | 4.434516896 |
| ***Insl5*** | 20.71173984 | 4.372376844 |
| ***Gal3st2*** | 20.30380596 | 4.343678281 |

| | | |
|---|---|---|
| ***Defa17*** | 19.8031081 | 4.307654974 |
| ***Slc15a1*** | 18.15721369 | 4.182470926 |
| ***Echdc3*** | 17.89905632 | 4.161811622 |
| ***Hsd11b2*** | 17.28657809 | 4.111580408 |
| ***Sptssb*** | 17.23450276 | 4.10722777 |
| ***Ybx2*** | 17.09959556 | 4.095890297 |
| ***Psapl1*** | 17.05707221 | 4.092298129 |
| ***Gm5485*** | 16.16736751 | 4.015012883 |
| ***Rab3b*** | 16.1437156 | 4.012900759 |
| ***Slc5a8*** | 15.89968904 | 3.990926645 |
| ***Smim24*** | 15.8282315 | 3.984428166 |
| ***Car4*** | 15.24080174 | 3.929866892 |
| ***Gstm2*** | 14.74608942 | 3.882260505 |
| ***E330018M18Rik*** | 13.76786651 | 3.78323311 |
| ***Evx1*** | 13.75467082 | 3.781849708 |
| ***Aox4*** | 13.62715384 | 3.768412368 |
| ***Klk1b3*** | 13.5360228 | 3.758731998 |
| ***Gal3st2b*** | 13.52124413 | 3.757156 |
| ***Zg16*** | 13.42944611 | 3.747327898 |
| ***Tnni1*** | 13.37972071 | 3.741976096 |
| ***Akr1c14*** | 13.02091896 | 3.702759366 |
| ***Adh1*** | 12.92402764 | 3.691983836 |
| ***Dyrk3*** | 11.20143304 | 3.485611409 |

| | | |
|---|---|---|
| ***Slc6a14*** | 11.1209221 | 3.47520451 |
| ***Defa30*** | 11.06275371 | 3.467638637 |
| ***Col6a4*** | 10.8413687 | 3.438475 |
| ***Casp14*** | 10.72638626 | 3.423092207 |
| ***Mdk*** | 10.45816883 | 3.386558361 |
| ***Hs3st3b1*** | 9.96320417 | 3.316609788 |
| ***Sdr16c6*** | 9.799014444 | 3.292636654 |
| ***Nanos1*** | 9.789069901 | 3.29117179 |
| ***Pak6*** | 9.736300888 | 3.283373753 |
| ***Wnt10a*** | 9.688352098 | 3.276251297 |
| ***Eno3*** | 9.681976257 | 3.275301556 |
| ***Rep15*** | 9.565555188 | 3.257848706 |
| ***Trpm6*** | 9.071715127 | 3.181375337 |
| ***Scube1*** | 8.905871917 | 3.154756863 |
| ***Rnf186*** | 8.794972501 | 3.136679067 |
| ***Cyp2c68*** | 8.764312204 | 3.131640877 |
| ***B3gnt7*** | 8.728069667 | 3.125662617 |
| ***Slc37a2*** | 8.608777805 | 3.105808432 |
| ***Ugt1a1*** | 8.567160821 | 3.098817171 |
| ***Aldh1l2*** | 8.463349048 | 3.081228668 |
| ***Ckmt1*** | 8.282766944 | 3.050112795 |
| ***Kcnj16*** | 8.207862377 | 3.037006541 |
| ***Maob*** | 8.137465679 | 3.024579554 |

| | | |
|---|---|---|
| ***Epop*** | 8.136956884 | 3.024489346 |
| ***Stx1a*** | 8.01462901 | 3.002635741 |
| ***Nmnat3*** | 7.740791245 | 2.952481043 |
| ***Phgdh*** | 7.520431482 | 2.910815438 |
| ***Gstm6*** | 7.492560685 | 2.905458864 |
| ***Cth*** | 7.463284662 | 2.899810714 |
| ***Lgals2*** | 7.354113493 | 2.878551441 |
| ***Col18a1*** | 7.328623567 | 2.873542263 |
| ***Best2*** | 7.302121454 | 2.868315665 |
| ***Defa3*** | 7.293374826 | 2.86658654 |
| ***Slit2*** | 7.054509186 | 2.818545711 |
| ***Vill*** | 6.965966228 | 2.800323478 |
| ***Hibch*** | 6.638142444 | 2.730779588 |
| ***Ramp1*** | 6.585734826 | 2.719344423 |
| ***Pde8b*** | 6.478857757 | 2.695739484 |
| ***Gdpd1*** | 6.406726767 | 2.679587464 |
| ***Retreg1*** | 6.304386339 | 2.656355947 |
| ***Aldh1a2*** | 6.280466184 | 2.650871651 |
| ***Cpq*** | 6.279090364 | 2.650555575 |
| ***Slc39a5*** | 6.277112765 | 2.650101126 |
| ***Habp2*** | 6.271772175 | 2.648873154 |
| ***Wfdc2*** | 6.097786087 | 2.608285541 |
| ***Hoxb13*** | 5.825351285 | 2.542345051 |

| | | |
|---|---|---|
| ***Pappa*** | 5.766945395 | 2.527807362 |
| ***Sycn*** | 5.75406474 | 2.524581453 |
| ***Gm12320*** | 5.706078203 | 2.51249952 |
| ***Msln*** | 5.490493175 | 2.456935743 |
| ***Entpd2*** | 5.488516715 | 2.45641631 |
| ***Ndrg2*** | 5.470652274 | 2.451712858 |
| ***Slc26a3*** | 5.447851155 | 2.445687287 |
| ***Sval1*** | 5.435999693 | 2.442545375 |
| ***Tgm3*** | 5.398843599 | 2.432650424 |
| ***Plbd1*** | 5.144534038 | 2.363040412 |
| ***Gm15401*** | 5.137545093 | 2.361079152 |
| ***Trim35*** | 5.038235923 | 2.33291868 |
| ***Kyat1*** | 4.898774754 | 2.292420958 |
| ***Padi4*** | 4.823600273 | 2.270110357 |
| ***Ndrg3*** | 4.792746749 | 2.260852709 |
| ***Bnipl*** | 4.751075184 | 2.248254037 |
| ***Cryzl2*** | 4.666301974 | 2.222279673 |
| ***Ctse*** | 4.664763592 | 2.221803968 |
| ***Acad12*** | 4.658006239 | 2.219712572 |
| ***Pccb*** | 4.629872439 | 2.210972445 |
| ***Prrg4*** | 4.600864934 | 2.201905104 |
| ***Gm12511*** | 4.528672789 | 2.179088304 |
| ***Rpl19-ps3*** | 4.51811347 | 2.175720504 |

| | | |
|---|---|---|
| ***Mecr*** | 4.500132691 | 2.169967541 |
| ***Capn9*** | 4.221005034 | 2.07758655 |
| ***Fgf18*** | 4.07135284 | 2.025508257 |
| ***Als2cl*** | 4.070420668 | 2.025177901 |
| ***Stim1*** | 4.054567755 | 2.019548121 |
| ***AI854703*** | 4.029420244 | 2.010572278 |

*Candidate genes were sorted by calculated quotients

## 8.5 Table of 118 candidate genes* (HhAOM distal colon 10 weeks versus Hh 10 weeks)

| | **Quotient**<br>*(Effectsize HhAOM 10w/Effectsize Hh 10w)* | **log2**<br>(Quotient) |
|---|---|---|
| *Defa30* | 168.6492537 | 7.397882124 |
| *Prss56* | 97.32248757 | 6.604701291 |
| *Sptssb* | 78.35724252 | 6.291994723 |
| *Defa20* | 74.08245453 | 6.211059994 |
| *Defa32* | 74.08245453 | 6.211059994 |
| *Habp2* | 69.29016565 | 6.1145787 |
| *Defa3* | 40.09164459 | 5.325229694 |
| *Defa17* | 34.52463115 | 5.109554096 |

| | | |
|---|---|---|
| *Wnt10a* | 32.06412394 | 5.002888085 |
| *Car4* | 29.93880195 | 4.903944586 |
| *Wif1* | 27.36712786 | 4.774372129 |
| *Tnni1* | 25.03277662 | 4.645746418 |
| *Ctse* | 23.67393768 | 4.565227784 |
| *Adh1* | 22.29492146 | 4.478643212 |
| *Psapl1* | 18.58745896 | 4.216257652 |
| *Aox4* | 15.34576729 | 3.939768877 |
| *Lyz1* | 14.89558524 | 3.896812902 |
| *R3hdml* | 14.52771668 | 3.860736068 |
| *Gm5485* | 13.07163131 | 3.708367292 |
| *Lgals2* | 12.57477105 | 3.652460228 |
| *Gm15915* | 12.30022941 | 3.620613318 |
| *Mme* | 12.13646893 | 3.60127683 |
| *Gpr27* | 10.5364261 | 3.397313691 |
| *Ugt1a1* | 10.4589304 | 3.386663414 |
| *Zg16* | 10.10162515 | 3.336515508 |
| *Sycn* | 9.885469265 | 3.305309453 |
| *Cryzl2* | 9.280790358 | 3.214247671 |
| *Esyt3* | 9.016750872 | 3.172607661 |
| *Hoxb13* | 8.959182949 | 3.163367169 |
| *Echdc3* | 8.804057057 | 3.138168495 |
| *Slc5a8* | 8.638524555 | 3.110784924 |

| | | |
|---|---|---|
| *Hs3st3b1* | 8.466630786 | 3.081787977 |
| *Pde8b* | 8.341566501 | 3.060318339 |
| *Sdr16c6* | 8.336587491 | 3.05945695 |
| *Bnipl* | 8.272729901 | 3.04836348 |
| *Ckmt1* | 8.215394886 | 3.038329922 |
| *Sval1* | 8.061199093 | 3.010994454 |
| *B3gnt7* | 7.937933634 | 2.988763501 |
| *Gdpd1* | 7.805723083 | 2.964532282 |
| *Maob* | 7.644674522 | 2.934455079 |
| *Trim35* | 7.621147728 | 2.930008281 |
| *Pappa* | 7.528735671 | 2.912407608 |
| *Casp14* | 7.225675079 | 2.853132382 |
| *Prkg2* | 7.186057324 | 2.845200444 |
| *Stx1a* | 6.960502476 | 2.799191457 |
| *Sectm1b* | 6.899529528 | 2.786497989 |
| *Sult1d1* | 6.854802277 | 2.777115053 |
| *Smim24* | 6.669257712 | 2.737526198 |
| *Kyat1* | 6.619664762 | 2.726758157 |
| *Selenop* | 6.452640322 | 2.68988961 |
| *Cyp2d12* | 6.286316389 | 2.652214884 |
| *Dyrk3* | 6.242506098 | 2.642125326 |
| *Insl5* | 6.204817429 | 2.633388761 |
| *Als2cl* | 6.129193851 | 2.615697334 |

| | | |
|---|---|---|
| *Aldh1l2* | 6.099502648 | 2.60869161 |
| *Rab3b* | 6.076922142 | 2.603340808 |
| *Pccb* | 5.985683256 | 2.581515938 |
| *Evx1* | 5.97154627 | 2.578104551 |
| *Tgm3* | 5.921490425 | 2.565960345 |
| *Hsd11b2* | 5.89476967 | 2.559435443 |
| *Phgdh* | 5.849505305 | 2.548314621 |
| *Rnf186* | 5.822227808 | 2.541571289 |
| *Slc37a2* | 5.752945128 | 2.52430071 |
| *Scube1* | 5.72925467 | 2.518347468 |
| *Kcnj16* | 5.710953533 | 2.513731646 |
| *Gm12320* | 5.702120551 | 2.51149854 |
| *Gal3st2* | 5.684060059 | 2.506921798 |
| *Slc15a1* | 5.672244536 | 2.50391973 |
| *Slc26a3* | 5.652909897 | 2.498993702 |
| *Aldh1a2* | 5.543890642 | 2.470898799 |
| *Stim1* | 5.538978122 | 2.46961984 |
| *Gstm2* | 5.522354837 | 2.465283591 |
| *Padi4* | 5.517952888 | 2.464133139 |
| *Msln* | 5.372376642 | 2.425560451 |
| *Mdk* | 5.252911013 | 2.393117145 |
| *Col6a4* | 5.232101093 | 2.387390416 |
| *Nanos1* | 5.178746446 | 2.372602925 |

| | | |
|---|---|---|
| *Gstm6* | 5.160640062 | 2.367550011 |
| *Cth* | 5.142230862 | 2.362394382 |
| *E330018M18Rik* | 5.111392203 | 2.353716295 |
| *Klk1b3* | 5.09745236 | 2.349776387 |
| *Ggh* | 5.057932146 | 2.338547683 |
| *Gal3st2b* | 4.985562438 | 2.317756269 |
| *Entpd2* | 4.982952179 | 2.317000729 |
| *Mecr* | 4.962044108 | 2.31093456 |
| *Slc6a14* | 4.929657124 | 2.301487305 |
| *Vill* | 4.924091967 | 2.299857707 |
| *Ybx2* | 4.848587181 | 2.277564425 |
| *Rpl19-ps3* | 4.827389572 | 2.271243258 |
| *Gm15401* | 4.812125056 | 2.266674135 |
| *Phgr1* | 4.78365268 | 2.258112646 |
| *Gm12511* | 4.757653917 | 2.25025033 |
| *AI854703* | 4.756126545 | 2.249787101 |
| *Best2* | 4.738215558 | 2.244343833 |
| *Slc39a5* | 4.731872051 | 2.242411064 |
| *Prrg4* | 4.730107327 | 2.241872919 |
| *Wfdc2* | 4.637185246 | 2.213249361 |
| *Eno3* | 4.635089057 | 2.212597058 |
| *Slit2* | 4.626656301 | 2.20996993 |
| *Ndrg2* | 4.520912271 | 2.176613922 |

| | | |
|---|---|---|
| *Ramp1* | 4.486004236 | 2.165430983 |
| *Hibch* | 4.412368933 | 2.141553425 |
| *Acad12* | 4.396729861 | 2.136430895 |
| *Rep15* | 4.377461276 | 2.130094416 |
| *Akr1c14* | 4.362641909 | 2.12520206 |
| *Nmnat3* | 4.298872467 | 2.103958311 |
| *Col18a1* | 4.28964727 | 2.100859022 |
| *Slc30a2* | 4.269245845 | 2.093981243 |
| *Trpm6* | 4.254901413 | 2.089125705 |
| *Cyp2c68* | 4.246177869 | 2.086164806 |
| *Capn9* | 4.170227706 | 2.060126161 |
| *Retreg1* | 4.147197071 | 2.052136604 |
| *Cpq* | 4.10048026 | 2.035792892 |
| *Plbd1* | 4.094553227 | 2.033706042 |
| *Pak6* | 4.091155258 | 2.032508288 |
| *Ndrg3* | 4.047806091 | 2.01714018 |
| *Fgf18* | 4.041864897 | 2.015021099 |
| *Epop* | 4.002073712 | 2.00074774 |

*Candidate genes were sorted by calculated quotients