



TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Life Sciences

Activating transcription factor 6-altered fatty acid metabolism shapes the tumour-associated microbiota

Adam D. A. Sorbie

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Lindsay Hall

Prüfer der Dissertation: 1. Prof. Dr. Dirk Haller

2. apl. Prof. Dr. Klaus-Peter Janssen

Die Dissertation wurde am 09.02.23 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 04.06.23 angenommen.

Abstract

Colorectal cancer (CRC) is a leading cause of cancer-related death worldwide and is heavily associated with changes in microbiota. We recently developed a microbiota-dependent mouse model of CRC, based on intestinal epithelial cell-specific activation of the nuclear p50 fragment of Activating Transcription Factor 6 (nATF6). Alterations in various taxa were observed, however, their exact involvement and what underlies these changes remained unclear. We hypothesised that transgenic nATF6 activation and downstream signalling integrate with the microbiota to promote tumour formation. The aim of this work was to perform an in-depth characterisation of microbial changes to assess whether nATF6 activation shapes a pro-tumorigenic microbiota and how, as well as evaluate the contribution of nATF6 activation and microbial signalling to chronic inflammation-driven tumour formation.

We examined microbial changes in *nAtf6*^{IEC} mice across tumour development, from pre-tumour to tumour onset and progression, using 16S rRNA sequencing to differentiate between taxa which putatively initiate tumorigenesis and those merely associated with onset, and validated these findings using a newly developed method to spatially map colonic mucosal bacteria. To address the role of chronic inflammation in nATF6-driven intestinal tumour formation, we generated a novel mouse model crossing *nAtf6*^{IEC} mice with *Il10*^{-/-} mice – a model of colitis - and performed phenotypic characterisation as well as the same extensive microbiota profiling. Finally, we performed untargeted metabolomics of luminal content and employed data-integration methods in both mouse models to identify associations between microbial changes and altered metabolites.

Here we delineated changes in microbial composition occurring with tumour development in biallelic *nAtf6*^{IEC} mice (tg/tg), identifying several important mucosa-associated taxa. Further, we validated these taxa in a second spatially resolved cohort, mapping them to tumour sites. In *nAtf6*^{tg/wt;-/-} mice, nATF6 activation combined with *Il10* knockout, led to tumour formation with an incidence of approximately 75%, and exacerbated colitis, post-tumour. This was accompanied by a loss of goblet cells, increased microbial penetration into the mucus layer as well as shifts in mucosa-associated microbiota. Finally, we identified a role for nATF6 activation in modulating the intestinal microbiota via the alteration of lipid metabolism. Untargeted metabolomics of luminal contents in *nAtf6*^{IEC} and *nAtf6*^{IEC};*Il10*^{-/-} mice demonstrated a clear enrichment of lipids and long-chain fatty acid (LCFA) species, which were associated with tumour-associated changes in microbiota composition. Moreover, LCFA were enriched pre-tumour and correlated with changes in driver taxa abundance. Predicted metagenomics revealed an enrichment of genes related to fatty acid metabolism, particularly oleate hydratase (*ohyA*), an enzyme involved in LCFA detoxification. Accordingly, the major product of *ohyA*, 10-hydroxystearic acid, was enriched in tumour samples, further linking the modified microbiota to nATF6-altered fatty acid metabolism. Linking the altered fatty acid milieu to microbiota changes, *ex vivo* incubation with enriched LCFA, led to a marked alteration in microbiota composition, increasing the predicted abundance of *ohyA*. These results demonstrate a role for nATF6 activation in indirectly modifying the gut microbiota and supporting the presence of putatively pro-tumorigenic microbes. In susceptible hosts, chronic nATF6 activation may therefore promote colonisation or expansion of potentially harmful taxa which could promote disease.

Table of Contents

Abstract	
Table of Contents	
1. Introduction	1
1.1 The Microbiota	1
1.1.1 Gut Microbiota	1
1.1.2 Microbial environments within the Gastrointestinal tract	1
1.1.3 Mucosa-associated microbiota	2
1.1.4 Microbe-host interactions in health and disease	5
1.2 Colorectal cancer and the Microbiota	6
1.2.1 Colorectal cancer – epidemiology, risk factors and mechanisms	6
1.2.2 Role of the microbiota in spontaneous and colitis-associated colorectal cancer	9
1.2.3 Microbial succession in multistep colorectal carcinogenesis	9
1.2.4 Microbial-driven models of CRC	11
1.3 The unfolded protein response and ATF6 in health and disease	12
1.3.1 Unfolded protein response and ATF6	12
1.3.2 ATF6 beyond the unfolded-protein response	13
1.3.3 ATF6 in intestinal disease	14
2. Aims & Study Objectives	15
3. Materials & Methods	16
3.1 Phenotypic characterisation	16
3.1.1 Ethics statement, mouse breeding and housing	16
3.1.2 Gnotobiotic Mice	16
3.1.3 Genotyping	16
3.1.4 Histology and tissue staining	17
3.1.5 Histopathological analysis	18
3.1.6 Immunohistochemical/-fluorescent staining	18
3.1.7 Fluorescence <i>in situ</i> hybridisation (FISH)	18
3.1.8 Gene expression analysis	19
3.1.9 ELISA measurement of Lipocalin-2	20
3.2 Microbiota analysis	21
3.2.1 Sampling of mouse colonic mucosa-associated microbiota	21
3.2.2 <i>Ex vivo</i> cultivation of caecal content and long-chain fatty acid co-culture	22
3.2.3 Microbial DNA extraction from mouse caecal content	23
3.2.4 Microbial DNA extraction from mouse colonic tissue	23
3.2.4 Microbial DNA extraction from <i>ex vivo</i> cultivated caecal content	24
3.2.5 16S rRNA amplicon sequencing	24

3.3 Metabolomic analysis	25
3.3.1 Untargeted metabolomics	25
3.4 Bioinformatics & Statistical Analysis	26
3.4.1 Functional potential prediction.....	26
3.4.2 Statistical analysis.....	26
3.4.3 Multi-omic data integration	27
4. Results.....	28
4.1 Mucosa-associated microbiota dysbiosis in <i>nAtf6</i> ^{IEC} mice	28
4.1.1 IEC-specific activation of nATF6 alters mucosa-associated microbiota in.....	28
the absence of tumour formation	28
4.1.2 Mucosa-associated microbiota improves classification of phenotype	31
4.1.3 Analysis of fold change within and between phenotypes differentiates taxa	
associated with tumour initiation and onset in <i>nAtf6</i> ^{tg/tg} mice	36
4.2 Spatial organisation of mucosa-associated microbiota in <i>nAtf6</i> ^{IEC} mice.....	38
4.2.1 Microbiota profiles in <i>nAtf6</i> ^{IEC} mice vary along the colon and are altered by tumour	
presence.....	38
4.2.1 ATF6 activation drives spatial microbial alterations	42
4.2.2 Putative microbial drivers are reproducible across cohorts.....	45
4.3 Characterisation of the <i>nAtf6</i> ^{IEC} ; <i>Il10</i> ^{-/-} Phenotype.....	49
4.3.1 Basic characterisation of a novel nATF6-driven model of colitis-associated	
cancer	49
4.3.2 nATF6 activation enhances colitis in <i>nAtf6</i> ^{tg/wt;-/-} mice.....	53
4.3.3 Tumorigenesis in <i>nAtf6</i> ^{tg/wt;-/-} mice is microbiota-driven	56
4.4 Mucosa-associated dysbiosis in <i>nAtf6</i> ^{IEC} ; <i>Il10</i> ^{-/-} mice	61
4.4.1 Mucosa-associated microbiota profiles are altered regardless of tumour status in	
<i>nAtf6</i> ^{tg/wt;-/-} mice.....	61
4.4.2 Mucosa-associated microbiota enables discrimination of microbial profiles between	
<i>nAtf6</i> ^{IEC} ; <i>Il10</i> ^{-/-} mice.....	62
4.4.4 Mucosa-associated populations do not shift significantly from pre- to post-tumour in	
<i>nAtf6</i> ^{tg/wt;-/-} mice.	67
4.5 Spatial organisation of mucosa-associated microbiota in <i>nAtf6</i> ^{IEC} ; <i>Il10</i> ^{-/-} mice	69
4.5.1 Spatial community structure of microbial communities does not significantly differ	
between <i>nAtf6</i> ^{fl/fl;-/-} and <i>nAtf6</i> ^{tg/wt;-/-} mice	69
4.5.2 Mucosa-associated microbiota alterations are restricted to tumour-susceptible sites	
.....	73
4.5.3 Driver taxa are robust across cohorts and localise to tumour-susceptible sites ...	75
4.6 Metabolomics links tumour-associated taxa to the nATF6 metabolite environment....	77
4.6.1 nATF6 activation and associated tumour formation alters the intestinal	77
lipid environment.....	77
4.6.2 Multi-omic integration links host-altered fatty-acid environment to tumour-	
associated microbiota.....	83

4.6.3 Long-chain fatty acids modulate microbial composition.....	90
5. Discussion	95
5.1 Identification of microbial shifts between pre-tumour and tumour onset in an nATF6-dependent model of sporadic colorectal tumorigenesis	95
5.2 Generation and microbial characterisation of a novel model for colitis-associated cancer	97
5.3 Linking shifts in metabolite environment to tumour-associated microbiota	99
Concluding Remarks	101
List of Figures	103
List of Tables.....	104
Glossary	105
Abbreviations.....	107
References	110
Acknowledgements.....	135
Curriculum Vitae	136

1. Introduction

1.1 The Microbiota

1.1.1 Gut Microbiota

The human body is colonised by trillions of microorganisms encompassing bacteria, archaea, single-celled eukaryotes, and viruses. Collectively, this group, their genomes, metabolic activity, as well as their habitat and environmental conditions is termed the microbiome, while microbiota refers to the assemblage of these microorganisms themselves¹. In recent years, the microbiome has been the focus of intense research and has been shown to play a crucial role in health and disease². The advent of culture-independent methods, such as 16S rRNA sequencing and shotgun metagenomics have allowed microbial community surveys to become routine and enabled large-scale international projects such as the human microbiome project (HMP), and MetaHIT and more recently, region-specific projects such as the Dutch Microbiome project and the Flemish gut flora project, which have catalogued the microbiome across various populations, body sites and disease states³⁻⁸. Moreover, metabolomics, representing the functional readout of the microbiome as well as gnotobiotic mouse models, where the microbiome composition is known or absent, have further enabled causal links between microbiome function and disease⁹. The “gut microbiota” refers to the microbiota associated with the gastrointestinal (GI) tract. Due to the non-invasive method of sample collection, faeces is most often used as a proxy for the gut microbiota and assumed to represent an amalgamation of the entire GI tract¹⁰. The total microbial load differs along the GI tract however, with the colon harbouring the greatest diversity and load ($\sim 10^{11}$ CFU/ml)¹¹. In healthy adults, gut microbiota is dominated by taxa belonging to the phyla Firmicutes and Bacteroidetes⁵. At lower taxonomic ranks however, microbiota composition is highly variable between individuals but exhibits a considerable degree of functional redundancy, indicative of strong environmental selection¹²⁻¹⁴. Indeed, gut microbes and their hosts have evolved commensal and mutualistic relationships over millions of years, resulting in a tightly intertwined biology¹⁵.

1.1.2 Microbial environments within the Gastrointestinal tract

A variety of discrete microbial environments exist within the GI tract, from the oral cavity to the rectum, shaped by the changing environmental conditions at each site¹¹. The oral cavity is the start site of digestion, and functions to enable breathing, secretion of saliva and enzymatic digestion of food. Collectively these functions shape the microbial assemblage, existing in this habitat. Salivary flow, food intake and high oxygen levels restrict which species can reside there and by extension their functions. Exemplifying this, the oral microbiota is dominated by aerobes or facultative anaerobes, mainly from the phyla Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria and Proteobacteria⁵. Despite these restrictive conditions, over 1000 species have been identified in the mouth and any one site can contain approximately 50 different species^{16,17}. Indeed, the oral microbiota is the second most diverse environment after the gut microbiota, defined by species such as *Corynebacterium*, *Streptococcus*, *Veillonella* and *Porphyromonas*¹⁸.

Moving down the GI tract, the stomach, presents an unforgiving environment for microbial colonisation, functioning to digest food by acid breakdown and peristaltic action. For this reason, microbial biomass is at its lowest in the entire GI tract in the stomach, estimated to contain around 10^1 - 10^3 CFU/ml¹¹. Indeed, the stomach was once thought to be sterile or only transiently colonized, however a resident microbiota does exist here, mostly comprising taxa from the genera *Prevotella*, *Streptococcus*, *Veillonella*, *Rothia*, *Haemophilus*, *Lactobacillus* and *Helicobacter*¹⁹⁻²¹. The murine stomach is similar; however, it also contains significant quantities of Lachnospiraceae and Bacteroidales taxa²².

The small, and in particular, the large intestine (LI) are the most densely populated and well-characterised environments in the entire GI tract¹³. The small intestine (SI) mainly functions to absorb nutrients, aid digestion and plays a key role in mucosal immunity²³. Microbial load in the SI is lower than that of the LI, likely due to the low pH and abundance of antimicrobial molecules such as bile acids and digestive enzymes, while transit time is also significantly faster (~5 hours in the SI, compared to ~30 in the colon), hindering establishment of microbial communities, and resulting in highly dynamic microbial communities^{24,25}. Several physical and chemical gradients along the SI also licence increased microbial colonisation, with the concentration of antimicrobial compounds as well as oxygenation decreasing along the length of the SI, while pH and mucus layer thickness increase¹¹. In concordance with increasingly permissive environments along the SI, microbial biomass increases from approximately 10^3 - 10^5 CFU/ml in the Duodenum to 10^7 - 10^8 in the distal Ileum, where bacterial load approaches that found in the caecum or colon²⁶. In mice and humans, the proximal SI is dominated by facultative anaerobes or microaerophilic taxa such as *Streptococcus*, *Escherichia* and *Clostridium* species in humans and members of Enterobacteriaceae and Lactobacillaceae in mice respectively^{22,27-30}. In both cases however, with increasing distance and reduced oxygen, these taxa are gradually replaced by obligate anaerobes^{27,30}.

The caecum and proximal colon are the major sites of microbial fermentation. In humans and other omnivorous animals, the caecum is small and believed to be similar to the SI microbiota, however, due to difficulties in sampling, few studies have profiled the microbiota from this region. In herbivores such as mice, however, the caecum is large and populated by a diverse and abundant population of microbes, mostly from the families Muribaculaceae and Lachnospiraceae, which function to breakdown indigestible plant fibres^{22,31,32}. In the colon, the slow transit time, anaerobic environment and lack of simple carbon sources present a significant selective force, favouring the growth of fermentative bacteria such as Bacteroidaceae and Clostridiaceae in humans and Bacteroidaceae and Prevotellaceae in mice^{11,22,33}.

Microbiota not only differ along the longitudinal axis however but also along the lateral axis, with mucosal microbial communities representing distinct populations³³⁻³⁵. This lateral axis, however, is relatively underexplored in comparison to luminal communities, in part due to sampling difficulties, particularly in humans. Current methods require invasive sampling, and as such most of what is currently known regarding mucosa-associated bacteria comes from animal studies. Nevertheless, due to their proximity to the epithelium, mucosa-associated bacteria are attracting increasing interest as potential key modulators of host health.

1.1.3 Mucosa-associated microbiota

The entire SI and Colon is either completely or partially coated with a mucus layer consisting of gel-forming mucins, glycoproteins, lipids, and salts that function to lubricate luminal contents and act as a physical and chemical barrier against potentially immunogenic resident

commensals^{36,37}. Goblet cells, found throughout the GI tract, continuously secrete mucins, forming the mucus hydrogel. As well as presenting a physical and chemical barrier to microbes, the mucosal environment also acts as an immune barrier, enriched in antimicrobial peptides and immunological factors, such as secretory IgA, which binds microbes and their products to prevent translocation, and secreted proteins such as Ly6/PLAUR domain containing 8, zymogen granulae protein 16 and Resistin-like molecule β , which segregate the inner mucus layer and the microbiota^{38–42}. Together these factors further restrict microbial colonisation^{40,43–45}. Moreover, recent work has identified an enrichment of bacteriophages in mucosal layers of various organisms, via binding to mucus glycoproteins, providing an additional layer to limit mucosal bacteria⁴⁶. Together, these function to restrict microbial contact with the epithelium. Breakdown of this crucial barrier, on the other hand, is often implicated in inflammatory and malignant intestinal pathologies. In a healthy colon, the mucus layer is divided into an outer densely colonised layer and an inner layer, mostly devoid of bacteria, however patients with ulcerative colitis (UC) as well as animal models of UC, show defective mucus, with increased bacterial penetration, which may precipitate disease^{47–49}. Similarly, in mice deficient in *Muc2*, which encodes the major component of intestinal mucus, bacteria penetrate into colonic crypts, ultimately leading to colitis and tumour formation⁵⁰. Thus, tight regulation of the mucosal barrier is necessary to maintain intestinal homeostasis and prevent disease.

In spite of the tightly regulated host control and harsh environment, the mucosal layer in all segments of the intestine harbours a complex microbial community, colonised by highly specialised microbes (**Figure 1**)⁵¹. The mucosal microbiota has long been known to differ from the luminal environment. Histological examinations of stained tissue sections in the 1960s and 1970s reported the presence of distinct “fusiform-shaped” bacteria in the mucosal layer^{52,53}. More recent studies in mice have profiled the mucosal microbiota using 16S rRNA sequencing. Laser-capture microdissection has been used to delineate differences in mouse colonic mucosal and luminal microbiota, with an increased abundance of Lachnospiraceae and Ruminococcaceae observed in interfold regions compared to the lumen³³. At the species level, *Akkermansia muciniphila*, several *Bacteroides* and *Eubacterium* spp. as well as *Faecalibacterium prausnitzii* have been identified in humans, while in mice *Lactobacillus*, *A. muciniphila* and *Mucispirillum* are enriched^{54,55}. Additionally, using aspirates of colonic mucosa and colonic content, Li *et al* identified both taxonomic and functional differences between these sites. Here the authors utilised a combination of gnotobiotic mouse models and RNAseq to associate bacterial transcription with environment, showing that, in monocolonised mice, the same species could have a different transcriptional profile depending on the compartment they were residing in⁵⁶.

Besides the densely populated outer layer, a specific core-microbiota populating the crypts of the murine caecum and proximal colon has also been identified. Pedron *et al* found an enrichment of Firmicutes and Proteobacteria in crypts, including aerobic species suggesting the presence of a highly specialised niche³⁵. Rarely, some microbial taxa may also directly associate with the epithelium. Although this is often associated with pathogenic microorganisms or inflammatory/malignant disorders disrupting the mucosal barrier, there are examples of symbiotic epithelial-associated bacteria, such as Segmented Filamentous Bacteria, which adhere to the ileal epithelium stimulating a strong Th17 response⁵⁷. Commensal epithelial-adherent taxa are also likely present in the colon; some *Lactobacillus* species for example, possess genetic loci associated with adherence, and various members of this species are often found in close association with the colonic epithelial surface^{58–61}. The various layers of microbiota structure, from lumen to mucosa to epithelium, are driven by the selective forces present in the local microenvironment. Thus, under homeostatic conditions,

the host environment selects for benign or beneficial taxa, with the spatial structure and segregation of bacteria from host, key to avoiding potentially hazardous interactions. As of yet, no fine-scale investigations into how microbiota spatial organisation is altered in disease have been conducted, however disruption in microbial niches and spatial structuring on a broader scale is a common event in the pathogenesis of several intestinal diseases, thus deeper investigation of this is key to fully disentangle how local spatial structure impacts microbe-host interactions to develop a more complete picture of microbial ecology in the GI tract⁶².

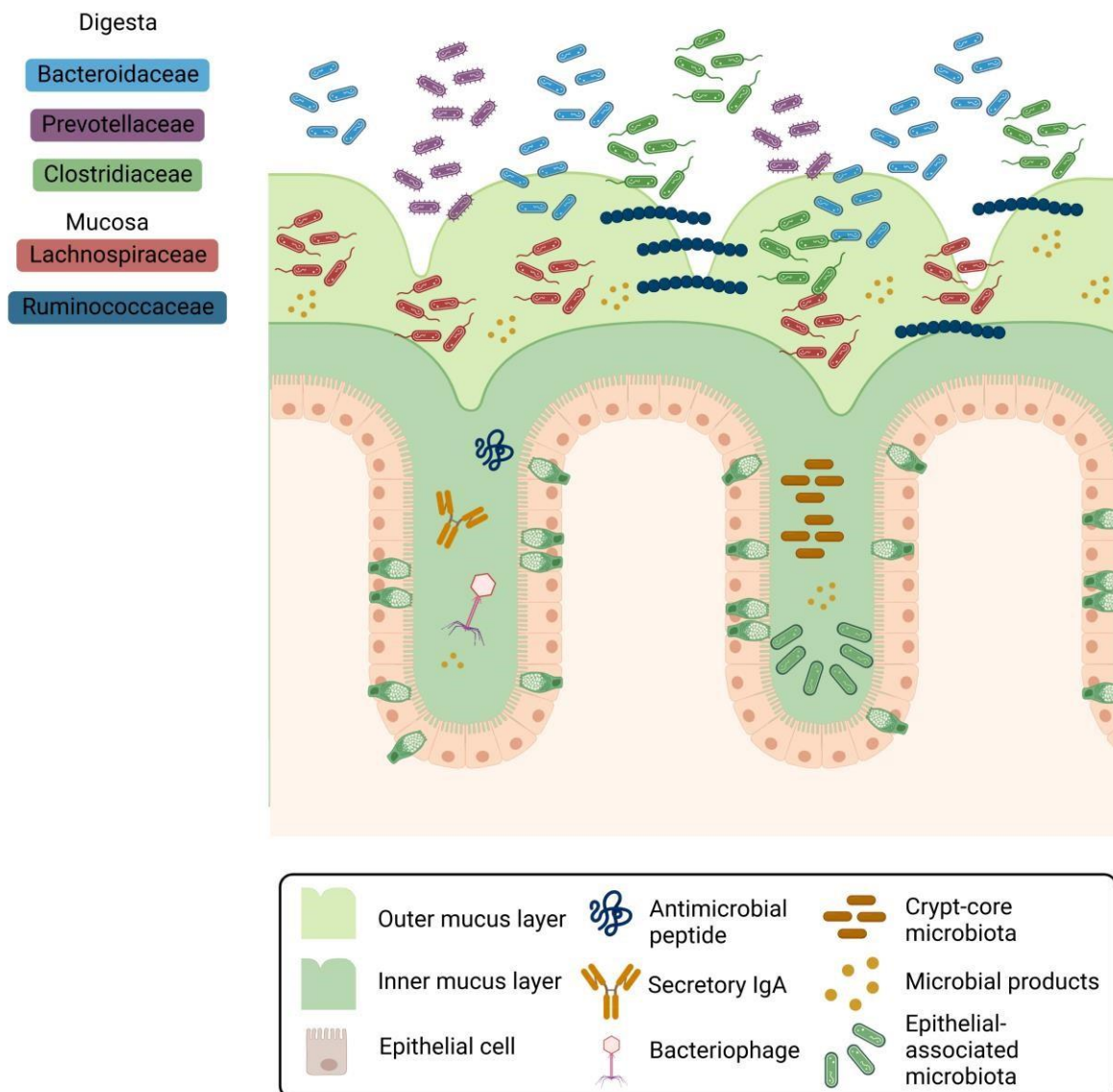


Figure 1. Mucosa-associated bacteria and mucosal defence. The colonic digesta in mice mainly comprises taxa from the families Bacteroidaceae, Prevotellaceae and Clostridiaceae, whereas the mucosal layer is dominated by Lachnospiraceae and Ruminococcaceae. The host restricts microbial colonisation in the inner mucus layer and crypts by secreting antimicrobial peptides, IgA, and anchoring Bacteriophages, however a crypt-core microbiota and even epithelial-associated bacteria are still present. Created with BioRender.com

1.1.4 Microbe-host interactions in health and disease

Lying at the interface between the luminal microbial milieu and the mucosal immune system, the intestinal epithelium functions to absorb nutrients and provide a physical barrier separating host from microbiota. Composed of a single layer of rapidly renewing intestinal epithelial cells (IECs), the epithelium can be broadly categorised into two major lineages: absorptive, comprising enterocytes and M cells, and secretory, comprising Paneth cells, Tuft cells, goblet cells and enteroendocrine cells⁶³. The high turnover rate of the epithelium, renewing approximately every 3-5 days, is maintained by intestinal stem cells (ISC) residing at the bottom of crypts, termed crypt-based columnar cells^{64,65}. These continuously dividing cells, give rise to all the differentiated cell types present in the epithelium. Considering the constantly dividing nature and unlimited expansion potential of ISCs, adequate protection from the luminal environment and microbial products is necessary to restrict the development of malignancies⁶³.

Under homeostatic conditions, the intestinal epithelium, and the mucosal immune system coordinate to keep the microbiota in check. Here, IEC function and metabolism is key, acting as the “middleman” between microbiota and the immune system, relaying stimuli to immune cells in the lamina propria^{66,67}. This is partially governed by the action of Pattern-Recognition receptors (PRR), like Toll-like receptors (TLRs), which are expressed on the apical and basolateral surface of IECs, as well as intracellular variants, like Nod-like receptors (NLRs)^{68,69}. These receptors sense microbial ligands such as lipopolysaccharide, lipoteichoic acid or flagellin and activate NF- κ B, leading to secretion of cytokines and chemokines, priming mucosal immune cells and leading to increased secretion of antimicrobial peptides and mucus^{67,70,71}. Although these receptors play an important role in mediating host-microbe signalling, there are a limited repertoire of PRRs, indicating other means of communication between the microbiota and host may be more important in the maintenance of homeostasis⁶⁹.

Metabolic cues from the microbiota also impact host function. The microbiota produces a broad variety of metabolites, small molecules derived from intermediates or end products of microbial metabolism, which exert a wide array of biological effects. Perhaps the best studied of these are short-chain fatty acids, produced by bacterial fermentation of dietary fibre. Acetate, propionate, and butyrate comprise 95% of the intestinal SCFA pool, occurring in an approximately 60:20:20 ratio respectively^{72,73}. IECs can recognise SCFA, via G-protein coupled (GPCR) or free fatty acid receptors (FFAR), through which SCFA are involved in the regulation of a myriad of host functions^{74,75}. As well as providing a major source of fuel for IECs, butyrate has been shown to regulate the production of antimicrobial peptides and additionally modulate IEC transcription through Histone deacetylase activity⁷⁶⁻⁷⁸. Moreover, both propionate and butyrate activate peroxisome proliferator receptor gamma (PPAR γ) which can activate beta-oxidation in IECs, consuming oxygen and facilitating the maintenance of an anaerobic environment in the gut^{79,80}. Bile acids (BA) are another well-studied class of microbial metabolites, produced via microbial transformation of primary bile acids in the colon. BAs interact with IECs via the Farnesoid-X receptor (FXR) and G-protein-coupled bile acid receptor 1 (GBPAR1, also known as TGR5) and have been shown to play an important role in the regulation of mucosal immunity, for example, by promoting immune tolerance, via modulation of Treg differentiation⁸¹⁻⁸³. More recently, certain tryptophan metabolites, acting via the Aryl-hydrocarbon receptor (AhR) have also been shown to play an important role in host-microbe interactions in the gut, by stimulating colonic motility through the activation of the serotonin type 4 receptor (5-HT₄R)⁸⁴.

Host function alters microbiota composition but at the same time, microbiota metabolites alter host function. Hence, regulation of microbe-host interactions is bi-directional. Given the nature of this relationship, shifts in the state of one constituent can affect the other. Although many microbial products and metabolites, play crucial roles in maintaining intestinal homeostasis, others can also impact the host negatively under certain conditions. Some microbial metabolites, such as hydrogen sulphide (H₂S) or indoleamines can disrupt IEC function or be genotoxic in large quantities, thus an increase in the abundance of microbes producing these metabolites may perturb IEC function⁸⁵⁻⁸⁷. On the host side, mice lacking intestinal TLRs or NLRs are highly susceptible to colitis and cancer, mediated by microbial encroachment into the inner mucus layer, further stressing the importance of correct IEC function in regulating microbe-host interactions⁸⁸⁻⁹⁰. Indeed, disruption of the mucosal barrier is a common feature of multiple GI diseases, including inflammatory bowel diseases (IBD) and colorectal cancer (CRC)^{49,91}. Interestingly, in inflammatory conditions such as IBD, an enrichment of mucolytic taxa such as *Ruminococcus* has been found, implying that the microbiota themselves can affect the mucus layer⁹². Regardless of the source of dysregulation, a disrupted mucosal barrier disturbs compartmentalisation of the microbiota, allowing potentially harmful microbes and metabolites to access the epithelial layer.

1.2 Colorectal cancer and the Microbiota

1.2.1 Colorectal cancer – epidemiology, risk factors and mechanisms

CRC is one of the most common cancers in men and women and a leading cause of cancer-related death⁹³. CRC is most common in individuals over 50 however incidence is rising in younger age groups⁹⁴. Hereditary diseases such as lynch syndrome and Familial adenomatous polyposis present an increased risk for developing CRC, but cases with a genetic predisposition such as these only account for a small number of total CRCs⁹⁴. Sporadic cases account for 70-80% of all CRC incidence, suggesting that environmental factors play a central role in the development of disease^{95,96}. CRC incidence is often associated with increased industrialization, with a sedentary lifestyle and diet high in saturated fat and sugars believed to be associated with the increased incidence observed in Western developed countries^{97,98}. Indeed, the highest rates of CRC incidence are found in Europe, North America, Oceania and East Asia, and in countries with a high or very high human development index (HDI), rates are ~3.5 times higher in men and ~4 times higher in women compared to those with a low or medium HDI⁹⁹. Furthermore, as many developing countries transition from a low HDI to a high HDI, their incidence of CRC increases¹⁰⁰. Troublingly, 10% of all new cases in Western countries occur in individuals under 50 and, although incidence and mortality rates are declining overall, both are increasing in this demographic^{94,101}.

The lack of a strong genetic component and the increasing incidence in younger populations further support a role for environmental factors in driving disease. Several modifiable lifestyle factors appear to increase disease risk. For example, diet appears to play a crucial role, with diets high in red and processed meats strongly associated with increased risk^{102,103}. Certain compounds present within meat, such as haem or choline have been associated with CRC and chemical products of the cooking process such as heterocyclic amines and polycyclic aromatic hydrocarbons are directly carcinogenic¹⁰⁴⁻¹⁰⁶. Obesity is another risk factor for CRC development; each increase of 5kg/m² in BMI constitutes a 5% increase in risk, with comorbid conditions like metabolic disorders, insulin resistance and low-grade systemic inflammation

additionally presenting increased CRC risk^{103,107,108}. Beyond diet, other risk factors include alcohol consumption and smoking, both of which are strongly associated with increased incidence in younger populations^{103,109}. IBD, particularly UC and colonic Crohn's disease (CD) also present a greatly increased risk of developing CRC, also termed colitis-associated cancer (CAC). Epidemiological meta-analyses have demonstrated that 30 years of active disease results in an increased risk of developing CRC of 8% and 18% for CD and UC patients, respectively¹¹⁰.

Regardless of the initiating factor, tumour development occurs via the progressive accumulation of mutations, activating oncogenes and inactivating tumour suppressor genes, ultimately leading to the "hallmarks of cancer": sustained proliferation, inflammation, resistance to cell death, genome instability and immune evasiveness amongst others^{111,112}. This evolution of a healthy epithelium to CRC generally follows a predictable pattern of histopathological and genetic and epigenetic alterations occurring in parallel. In the majority of CRC cases (70-90%), progression of normal colon cells to cancer arises according to the adenoma-carcinoma sequence, where aberrant crypts evolve into adenomas and then further progress into advanced adenoma before finally leading to cancer^{111,113,114}. A small number of CRCs arise according to the serrated neoplasia pathway, which occurs with an initial hyperplastic polyp, advancing to a sessile serrated polyp before giving rise to adenocarcinoma^{111,115}. These pathways involve distinct genetic and epigenetic changes occurring in a specific order¹¹⁵. The adenoma-carcinoma sequence is typically initiated by loss of Adenomatous polyposis coli (*APC*) function, leading to increased Wnt/ β -catenin signalling, promoting increased cell proliferation and survival. This is followed by activation of *KRAS* proto-oncogene (*KRAS*), initiating Mitogen-activated protein kinase (MAPK) and Phosphoinositide-3-kinase (PI3K) signalling which further encourage proliferation and resistance to cell death. Finally, subsequent loss of function of Tumour protein P53 (*TP53*) occurs, leading to the loss of genome stability^{111,113}. Recent work has challenged this dogma however, suggesting loss of *TP53* function may occur before *KRAS*, as well as implicating *KRAS* activation in invasion and metastasis^{112,116}. The serrated neoplasia pathway, on the other hand, occurs via *KRAS* and B-Raf proto-oncogene (*BRAF*) mutations and subsequent CpG island hypermethylation¹¹¹. As well as having a different aetiology, CAC also has a different pathogenesis than sporadic CRC¹¹⁷. Mechanistically, the chronic inflammation which occurs in IBD promotes the development of genetic aberrations, many of which also occur in sporadic CRC, however, the order and frequency in which these mutations appear, often differ. *TP53* loss occurs as an early event in CAC pathogenesis, whereas in sporadic CRC this usually occurs later¹¹⁸. Moreover, *APC* function is lost late in the pathogenesis of CAC and only in a minority of cases (<50%). Similar to sporadic CRC, the Wnt/ β -catenin pathway is upregulated in the majority of early dysplastic lesions; however, this occurs via a different mechanism - likely inflammation - and independently of *APC* loss^{117,119}. *KRAS* mutations also occur late in the pathogenesis of CAC, however they are rare and only present in ~20% of cancers¹²⁰. CAC similarly follows a predictable sequence of histopathological and genetic changes; however, this too differs from sporadic CRC (**Figure 2**). In contrast to the adenoma-carcinoma sequence, disease progression in CAC is defined by the "inflammation-dysplasia-carcinoma" sequence¹¹⁷. Nevertheless, both diseases are defined by the progressive accumulation of mutations which encourage the development of the hallmarks of cancer.

Lying at the interface between environmental risk factors, IBD and CRC/CAC, is the gut microbiota, with each risk factor separately associated with shifts in microbiota composition and function, potentially inducing dysbiosis or expansion of potentially pathogenic species which may promote tumorigenesis^{3,105,121,122}. Moreover, many of these risk factors lead to disruption in intestinal homeostasis or barrier function, enabling the encroachment of resident

microbes towards the epithelium, promoting inflammation and oxidative stress which may encourage tumour development. Considering this, microbiota changes may therefore be responsible for driving increased CRC incidence.

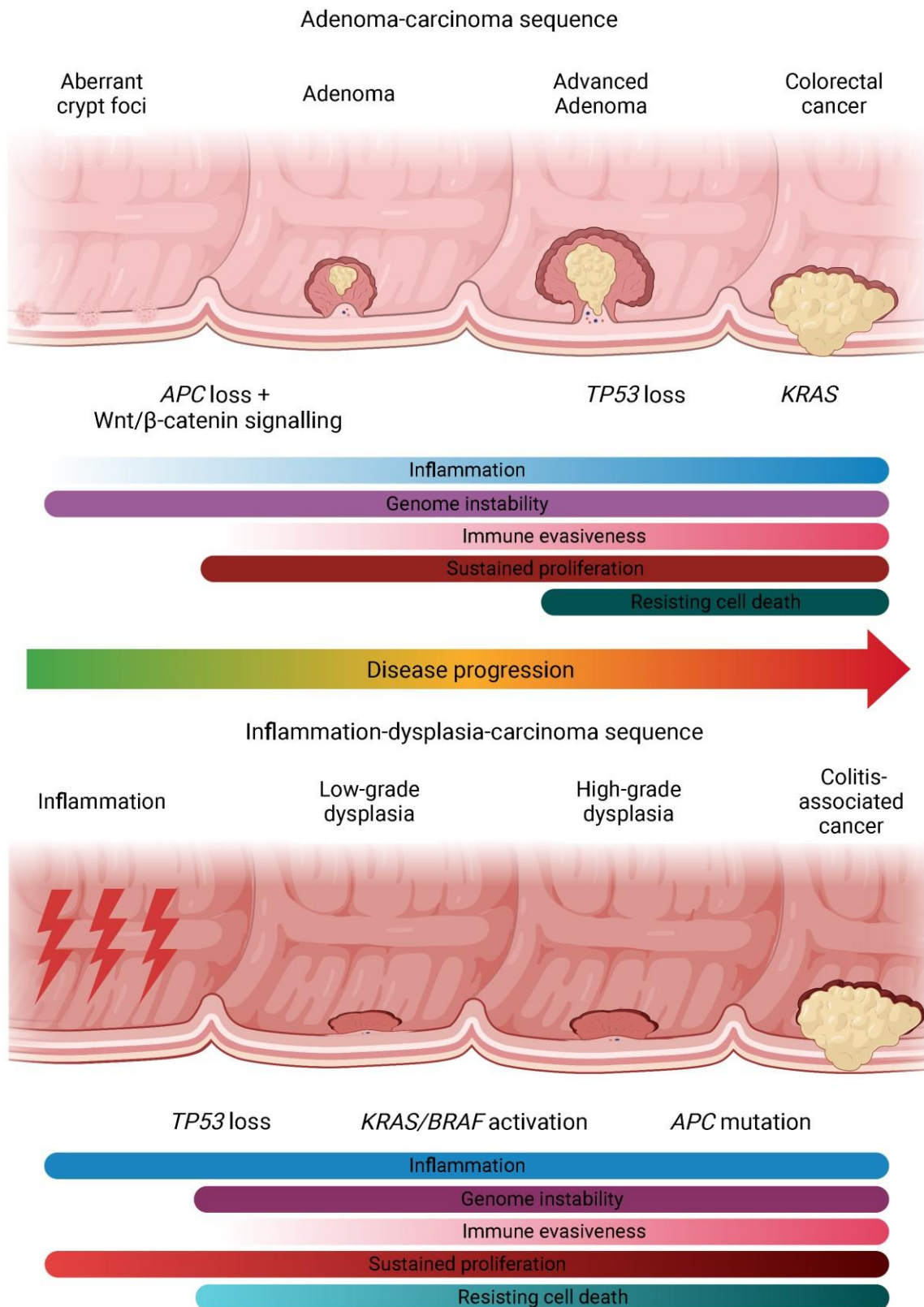


Figure 2. Pathogenesis of Colorectal and colitis-associated cancer. The adenoma-carcinoma sequence (top) and the inflammation-dysplasia-carcinoma sequence (bottom), annotated with typical

mutations in tumour suppressor genes and oncogenes and the stage at which they occur. The hallmarks of cancer and where they occur in the pathogenesis of CRC and CAC are shown below the respective pathway. Created with BioRender.com

1.2.2 Role of the microbiota in spontaneous and colitis-associated colorectal cancer

Accumulating evidence links changes in the microbiota to CRC. Recent large-scale meta-analyses of shotgun metagenomic data have identified distinct cross-cohort taxonomic and functional signatures of CRC in faecal samples, providing potential population-level biomarkers^{105,123–125}. However, whether the presence of those organisms is a cause or consequence of tumour formation remains unclear and mechanistic data linking many of these microbes to CRC pathogenesis is also lacking. Nevertheless, initial surveys of faecal and biopsy samples from CRC patients identified *Fusobacterium nucleatum* sequences as enriched in CRC patients and recently, several more bacterial strains such as *Bacteroides fragilis* and *Escherichia coli*, as well as *F. nucleatum*, have been mechanistically linked to CRC pathogenesis^{126–129}. *F. nucleatum*, for example, has been shown to bind E-cadherin via the *fadD* adhesin, activating the Wnt/ β -catenin signalling pathway¹³⁰. Furthermore, *F. nucleatum* can also elicit immune suppression by binding to immunoinhibitory receptors such as Carcinoembryonic Antigen-Related Cell Adhesion Molecule 1 (CEACAM1) and T Cell Immunoreceptor With Ig And ITIM Domains (TIGIT)¹³¹. *B. fragilis* is a common gram-negative gut bacterium and normally commensal, however a specific subtype termed enterotoxigenic *B. fragilis* (ETBF), expressing the zinc-dependent metalloprotease *Bacteroides fragilis* toxin (BFT) has been identified in biofilm-positive tumours in CRC patients and pre-cancerous lesions^{132,133}. ETBF can induce tumour formation by inducing colitis via the production of BFT, which cleaves E-cadherin, triggering Wnt/ β -catenin signalling and reducing epithelial barrier integrity¹³³. *E. coli* is another typical inhabitant of the healthy human gut, yet some strains encode the polyketide synthase (*pks*) genomic island enabling them to produce colibactin, a potent genotoxin. Importantly, these *pks*+ *E. coli* have also been found in colonic lesions from CRC patients^{129,134}. Microbial metabolites may also play a role in tumour formation. Elevated levels of secondary bile acids in CRC have been identified by multiple studies and have been shown to promote DNA damage^{124,135–137}. Furthermore, an increased abundance of sulphate-reducing bacteria like *Desulfovibrio* and *Bilophila* – which produce the genotoxic metabolite H₂S, is also associated with an increased risk of CRC^{138,139}. In spite of clear evidence that the microbiota and their metabolites are involved in CRC pathogenesis in animal models, questions remain over which taxa actually initiate disease and which merely promote further progression.

The majority of CRC-associated taxa have been identified from stool samples, which may not accurately reflect microbial communities at the tumour site. Indeed, tumour biopsies differ significantly from paired stool samples, and faecal microbiota has been shown to poorly predict adenoma, suggesting early changes in microbiota composition might only be evident at the mucosa^{140–142}. In this regard, understanding of the microbial changes which occur before and during tumour onset at the mucosa is important to delineate those which initiate disease and those which promote progression.

1.2.3 Microbial succession in multistep colorectal carcinogenesis

A handful of theoretical models exist to explain microbial succession in CRC, however one of the models which has gained the most traction is the “driver-passenger” hypothesis. In this

model, drivers – resident species with the potential to promote disease in susceptible hosts – induce tumour formation, leading to irreversible changes in the intestinal environment. Unable to properly adapt to the altered conditions, driver microbes decrease in abundance or die out and are overtaken by so-called passengers – microbes which are able to exploit the newly established niche – that may be benign or may promote cancer progression¹⁴³ (**Figure 3**). Driver bacteria are believed to promote tumorigenesis by inducing DNA damage or a hyperproliferative phenotype in IECs. Suggested examples include *pks+* *E. coli*, which can induce double-strand breaks, aneuploidy, and improper cell division¹³⁴. Short-term exposure of organoids to this *pks+* *E. coli* has also been shown to induce similar genomic aberrations to those seen in CRC, which may implicate this species as a potential driver^{144,145}. Nevertheless *in vivo* evidence to support the role of *pks+* *E. coli* and indeed other CRC-associated bacteria, as drivers of disease is currently lacking, in part due to the limited number of studies utilising longitudinal or time-series study designs. A recent study examining different CRC patients across the adenoma-carcinoma sequence however identified microbial and metabolite differences between different tumour stages in CRC, finding *Bacteroides* and branched-chain amino acids enriched in adenomas, and *Fusobacterium* and amino acid derivatives enriched in later stages¹²⁴. Furthermore, *in silico* analyses using metabolic modelling, have demonstrated that many CRC-enriched taxa likely depend on CRC-enriched metabolites, implying cross-sectional studies may mostly identify passenger rather than driver bacteria¹⁴⁶.

Passenger bacteria may still have some capacity to promote tumour formation though, as colonisation with microbiota from CRC patients in genetically susceptible controls has been observed to induce a higher tumour multiplicity than colonisation with that from healthy controls¹⁴⁷. Similarly, *F. nucleatum* has been identified as a passenger bacterium and observed to increase in abundance with increasing cancer stage, yet its colonisation in certain genetically susceptible mice generally leads to an enhanced tumour burden¹²⁴. This may suggest that, although the driver-passenger model may represent a useful framework, it is likely incomplete. It is certainly possible for example, that taxa which promote disease are also able to adapt to the change in environment and thus do not appear to change in abundance post-tumour-onset. Questions also remain over what would enable driver taxa to promote tumour formation in the first place. Very few studies to date have examined microbial changes, along the adenoma-carcinoma sequence, while the analysis of microbial changes prior to polyp formation would require large, longitudinal cohorts followed over many years. Considering the difficulty in determining microbial changes occurring with CRC initiation and progression in clinical cohorts, genetically susceptible mouse models represent a more practical approach for identifying causal microbes and potential triggers for the switch from a commensal to a pro-tumorigenic microbe, as well as shifts in community composition.

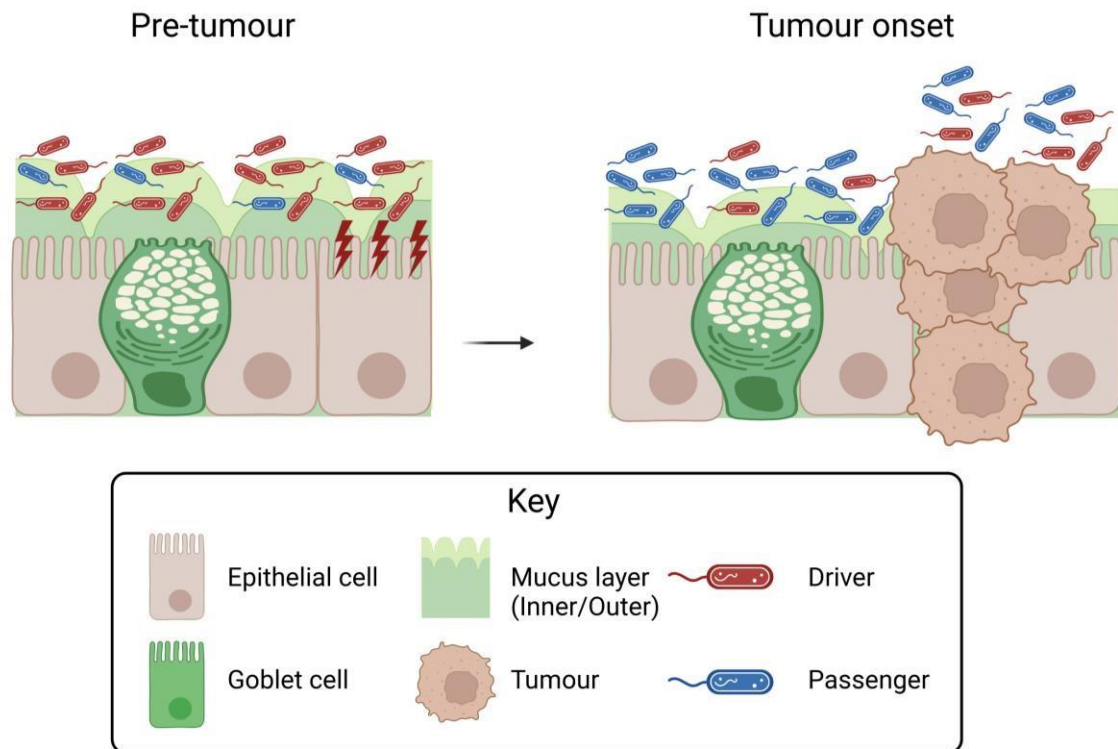


Figure 3. The driver-passenger model of colorectal tumorigenesis. Before tumour onset, environmental or genetic susceptibility allows colonisation/expansion of driver taxa which initiate tumorigenesis by damaging the epithelium. Upon tumour formation, the local milieu is altered, leading to expansion of so-called passenger taxa at the expense of drivers. Created with BioRender.com

1.2.4 Microbial-driven models of CRC

Although a variety of mouse CRC models exist including chemically induced models such as azoxymethane/dextran sodium sulphate (AOM/DSS) and patient-derived xenograft models, there are currently only a handful of spontaneous models¹⁴⁸. Within the spontaneous models, the most commonly used is the *Apc*^{min/+} mouse, driven by a truncation mutation in the murine homolog of the tumour suppressor gene adenomatous polyposis coli (*Apc*). *Apc*^{min/+} mice develop spontaneous tumours and are often combined with other genetic alterations to determine modifiers of tumorigenesis^{149,150}. In addition to investigating other genetic loci, many researchers have focused on the microbiota as a modulator of tumour formation, with germfree (GF) *Apc*^{min/+} mice developing fewer tumours in the SI compared to conventional counterparts¹⁵¹. Despite the common use of this model in CRC research, the human relevance is questionable. Unlike human CRC, *Apc*^{min/+} mice primarily develop tumours in the SI and the hereditary condition forming the basis for the model, only accounts for a small proportion of human cancers⁹⁶. Moreover, although the microbiota is certainly involved in promoting tumour formation, they are not necessary, thus limiting its use in investigating microbes actively driving disease^{151,152}.

In recent years, a small number of purely microbiota-driven models of CRC and CAC have been discovered. Overexpression (OE) of *Zeb2*, a transcription factor involved in the epithelial-to-mesenchymal transition was recently shown to induce colonic tumour formation, dependent on the microbiota¹⁵³. Zinc finger protein 90 homolog (*Zfp90*) has also been demonstrated to

promote inflammation-driven tumour formation in mice, an effect which is abrogated by microbiota depletion¹⁵⁴. Similarly, work from our laboratory demonstrated that the OE of the active nuclear fragment of activating transcription factor 6 (nATF6) induced colonic tumour formation independent of colonic inflammation and dependent on microbiota presence¹⁵⁵. Importantly, microbiota-dependent models such as these will facilitate the identification of driver taxa or communities, while simultaneously allowing the identification of novel host processes governing microbe-host interactions in CRC.

1.3 The unfolded protein response and ATF6 in health and disease

1.3.1 Unfolded protein response and ATF6

ATF6 is a transmembrane endoplasmic reticulum resident protein in the bZIP family of transcription factors. Along with Inositol-requiring enzyme 1-alpha (IRE1 α) and protein kinase R-like endoplasmic reticulum kinase (PERK), ATF6 comprises the endoplasmic reticulum (ER) unfolded protein response (UPR^{ER}), an evolutionary conserved mechanism to alleviate intracellular stress induced by the accumulation of misfolded or unfolded proteins¹⁵⁶. Under normal conditions, ATF6 and the other arms of the UPR^{ER} are kept in an inactive state by Glucose-regulated protein 78 (GRP78), however, upon accumulation of unfolded proteins, GRP78 binds to the exposed hydrophobic domains and dissociates from the three UPR^{ER} transmembrane proteins. Each arm of the UPR^{ER} initiates a separate program, aiming to restore proteostasis or, in the case this cannot be achieved, terminal apoptosis (**Figure 4**). PERK acts by attenuating global translation via phosphorylation of Eukaryotic translation initiation factor 2A (eIF2 α). IRE1 α , on the other hand, dimerises and autophosphorylates upon activation, exposing its RNase domain, resulting in splicing of X-box binding protein 1 (XBP1) mRNA. XBP1 upregulates genes involved in folding, secretion, ER-associated protein degradation (ERAD) and also lipid synthesis^{157,158}. Additionally, via regulated IRE1 α dependent decay (RIDD), IRE1 α itself is involved in the degradation of mRNA and miRNA¹⁵⁸. ATF6 acts in concert with XBP1 and upon UPR^{ER} induction, translocates to the Golgi where it is cleaved by S1P and S2P proteases into the active p50 fragment of ATF6 (nATF6). This fragment then translocates to the nucleus where it induces transcription of chaperones and other proteins involved in folding and maturation. By limiting newly synthesised proteins and increasing the activity of protein folding and secretion machinery, these three pathways cooperate to resolve ER stress. Proper functioning of the UPR^{ER} is therefore crucial to maintain proteostasis and normal cellular function.

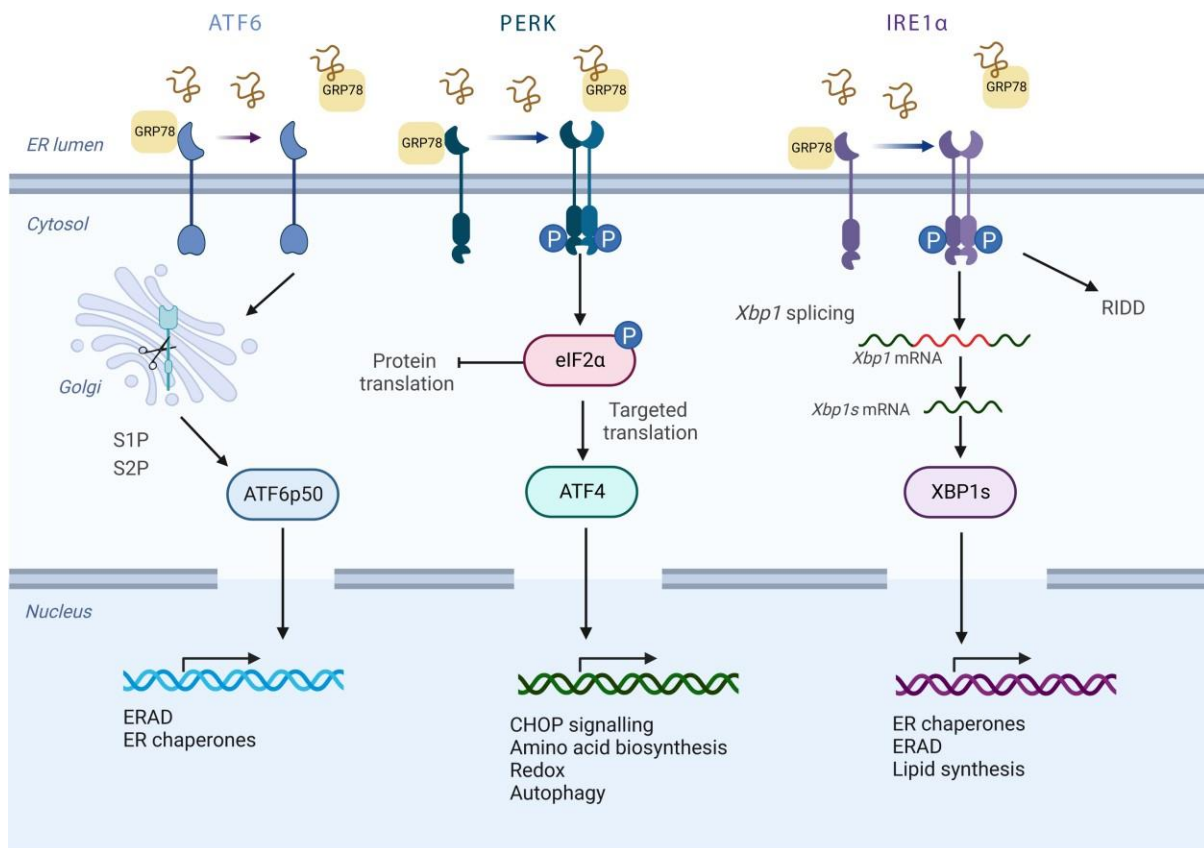


Figure 4. Activation of the endoplasmic reticulum unfolded protein response. Upon sensing of unfolded proteins GRP78 (BiP) dissociates from the three signal transducers, ATF6, PERK and IRE1 α , activating them. ATF6 and IRE1 α mediated transcription of XBP1, function to increase degradation capacity and synthesis of chaperones, while PERK functions mainly to attenuate translation, reducing the folding burden on the cell. Created with BioRender.com

1.3.2 ATF6 beyond the unfolded-protein response

The ER is a multifaceted organelle with diverse functions. Amongst these, the ER plays an important role in the synthesis of secreted and membrane-bound proteins and the synthesis of cellular lipids¹⁵⁹. The ER is thus intimately involved in metabolism and perturbations to ER function can therefore have knock-on effects on cellular metabolism. Accordingly, each of the three arms of the UPR^{ER} has also been shown to play an important role in metabolic regulation, particularly regarding lipids. IRE1 α /XBP1, for example, regulates genes involved in lipid synthesis^{160,161}. With respect to ATF6, the Site 1 and Site (S1P/S2P) proteases which cleave ATF6 into its active fragment, also act on Sterol regulatory element binding protein 2 (SREBP2), a primary regulator of cholesterol metabolism, in a similar manner. OE of *nAtf6* inhibits SREBP2 target gene expression, implying that ATF6 activation can indirectly regulate lipid metabolism¹⁶². Moreover, various studies have demonstrated a role for hepatic ATF6 in modulating lipid metabolism. Here *nAtf6* OE, upregulated genes involved in fatty acid synthesis and elongation, ultimately leading to steatosis in mouse and zebrafish models of disease^{163,164}. Importantly, lipotoxic stress can also activate ATF6 directly, mediated by certain sphingosine and ceramide lipids. This non-canonical activation initiates an alternative transcriptional program, preferentially activating genes related to lipid metabolism as well as ER chaperones¹⁶⁵.

As well as playing a role in the modulation of metabolism, ATF6 also appears to play a role in immunity. In a murine model of liver ischemia, prolonged activation of ATF6 in Kupffer cells alters the response to TLR stimulation, simultaneously increasing pro-inflammatory and decreasing anti-inflammatory cytokines¹⁶⁶. *In vitro* evidence also supports the involvement of ATF6 in NF- κ B activation via phosphorylation of Akt¹⁶⁷. The exact extent of ATF6 involvement in both metabolism and immune regulation remains unclear, and it is unknown whether this involvement is restricted to certain cell types or conditions. The UPR^{ER} is particularly important in cells with a high secretory requirement, such as B-cells and many cells of the intestinal epithelium, thus ATF6 signalling likely plays a more central role in these cell types^{168–171}. Indeed, IECs are highly secretory and additionally have high energy requirements, while facing constant immunological onslaught from the resident microbiota. Linking ATF6 to immune regulation and metabolism, Acyl-CoA synthetase long chain family member 1, an enzyme converting free fatty acids into fatty acyl-CoA esters, was recently found to induce ATF6 signalling in IECs, activating NF- κ B and leading to expression of *Tnfa*¹⁷². Certain aspects of metabolic and immune signalling in IECs may therefore converge on ATF6, indicating disruptions to this pathway could have knock-on effects beyond its role in the UPR^{ER} and restoring proteostasis.

1.3.3 ATF6 in intestinal disease

Activation of the UPR is normally short-lived or terminates in apoptosis of the cell, however unresolved or chronic ER stress is involved in various metabolic and inflammatory disorders, particularly in the intestine^{157,173}. Deletion of *Ire1a* or *Xbp1* in chemically induced models of colitis, for example, leads to enhanced ER stress, breakdown of barrier function, concomitant loss of goblet cells and Paneth cells, aggravating colitis^{174,175}. Additionally, aberrations in ATF6 signalling have also been identified in IBD^{172,176}. Recently, we identified a link between transgenic activated ATF6 and CRC. Biallelic activation of nATF6 in murine IECs, resulted in spontaneous intestinal tumorigenesis, with a loss of goblet cells and increased microbial penetration close to the epithelium preceding tumour onset. Importantly, tumours in this model occurred in the proximal colon, unlike many existing animal CRC models, and a subset of patients in The Cancer Genome Atlas (TCGA) demonstrated genomic and transcriptomic aberrations in ATF6 which was associated with reduced disease-free survival, suggesting this model is human relevant¹⁵⁵. Several other lines of evidence further support the involvement of ATF6 in human intestinal cancers. Hanaoka *et al* identified an increase in ATF6 in precancerous lesions in CAC and CRC, finding high expression could predict neoplastic transformation with high accuracy¹⁷⁷. Additionally, ATF6-mediated upregulation of the oncogene Cancerous inhibitor of protein phosphatase 2A (*CIP2A*) has been found to contribute to poor prognosis by maintaining cancer cell survival¹⁷⁸.

These findings hint at an important yet relatively unexplored role for ATF6 in intestinal tumorigenesis, with potential mechanisms underpinning its involvement remaining elusive. As nATF6-driven tumour formation was dependent on the gut microbiota, with microbial changes occurring upon nATF6 activation prior to tumour formation, a finding further supported by transfer recapitulating disease state, the missing link between ATF6 and CRC may centre on its as-yet-unknown ability to modulate gut microbiota composition.

2. Aims & Study Objectives

The role of the microbiota in CRC is well established, with various microbial taxa associated with tumour formation. Despite this, whether associated microbes are there as a cause or consequence and whether those directly involved instigate tumorigenesis or merely promote progression remains to be seen. Moreover, patterns of microbial succession occurring alongside tumour development are also poorly understood.

To delineate putatively causal from bystander microbes in CRC, a greater understanding of patterns of microbial succession during tumorigenesis is necessary. Given the difficulty in longitudinally sampling human patients over the necessary period to characterise pre-and post-tumour microbiota, suitable mouse models are needed to test this hypothesis. In most current models however, tumorigenesis, although often exacerbated by the microbiota, generally occurs independently of microbial involvement, presenting difficulties in dissecting their exact contribution to disease. Previous work from our laboratory identified a role for biallelic transgenic activation of activating transcription factor 6 (nATF6) in instigating microbiota-dependent tumorigenesis¹⁵⁵. Notably, transfer from transgenic mice as well as control mice at pre-tumour timepoints into susceptible mice led to tumour formation, suggesting nATF6 activation might shape a pro-tumorigenic microbiota, which is already established pre-tumour.

Here we aimed to utilise nATF6-driven models as a platform to examine microbial changes prior to, at onset and with tumour progression in both CRC and CAC. To this end, we utilised 16S rRNA profiling of both luminal and mucosa-associated microbiota, developed an innovative method of describing the spatial organisation of microbial communities and performed untargeted metabolomics. We additionally characterised a novel model for CAC and implemented the same in-depth characterisation of the microbiota and metabolome.

Herein, this thesis is broadly split into six parts. The first section describes the characterisation and contrast of microbial communities from different environments in *nAtf6*^{IEC} mice. We utilised 16S rRNA sequencing, to determine the utility of mucosa-associated microbiota in identifying bacteria driving tumour formation over that of luminal communities. Secondly, we set out to understand the spatial structure of mucosal microbial communities at millimetre resolution and explore how this changes under tumorigenic conditions. Here, we designed a sampling scheme sectioning the mouse colon into 0.5 x 0.5cm plots, to probe alterations in spatial organisation. Third, to examine microbial contribution to disease in the context of chronic inflammation, we crossed *nAtf6*^{IEC} mice with Interleukin-10 deficient (*Il10*^{-/-}) mice generating an nATF6-driven model of CAC and assessed the impact of combined monoallelic nATF6 activation and *Il10* knockout (KO) on tumour formation and IEC function. In the fourth and fifth parts of this work, we additionally carried out the same microbiota characterisation as with the *nAtf6*^{IEC} model, to determine if similar patterns of microbial succession hold in a model of CAC. Finally, to understand how nATF6 activation and tumour presence alter the metabolite environment and link microbial shifts to host changes, we performed an integrated analysis of the mucosal microbiota and host/microbial metabolome.

3. Materials & Methods

3.1 Phenotypic characterisation

3.1.1 Ethics statement, mouse breeding and housing

Mouse experiments were approved and carried out in accordance with the regulations of the relevant local authority (Regierung Oberbayern: Breeding proposals: 55.2-1-54-2532- 217-14 / 55.2-2532.Vet_02-20-58). All experimental animals were housed in the Specific-pathogen free (SPF) mouse facility at the Technical University of Munich. (School of Life Sciences, Weihenstephan). To mitigate potential cage effects, mice of different genotypes were cohoused where possible¹⁷⁹. The *nAtf6*^{IEC};*Il10*^{-/-} mouse model was first generated within the thesis work of Elena Lobner¹⁸⁰. Here *nAtf6*^{IEC};Villin-Cre mice were crossed with Interleukin 10 (*Il10*) KO mice on the same genetic background (C57BL/6). The different genotypes used in this study are described in **Table 1**. Mice received a standard chow diet (Ssniff, Soest, Germany), and autoclaved water *ad libitum* and were maintained under constant 12-hour light/dark cycles at 24-26°C. Due to the severe phenotype of transgenic mice, experimental animals were carefully monitored and scored based on changes in stool consistency, weight loss, rectal prolapse or other signs of poor health, and euthanised if a combined score >20 was reached.

Table 1. Mouse lines and genotype descriptions.

Mouse Model	Genotypes	Characteristics
<i>nAtf6</i> ^{IEC}	<i>nAtf6</i> ^{fl/fl}	Floxed controls.
	<i>nAtf6</i> ^{tg/wt}	Monoallelic Villin-Cre driven constitutive expression of <i>nAtf6</i> in IECs
	<i>nAtf6</i> ^{tg/tg}	Biallelic Villin-Cre driven constitutive expression of <i>nAtf6</i> in IECs
<i>nAtf6</i> ^{IEC} ; <i>Il10</i> ^{-/-}	<i>nAtf6</i> ^{fl/fl} ; <i>-/-</i>	Floxed controls + Global <i>Il10</i> knockout
	<i>nAtf6</i> ^{tg/wt} ; <i>-/-</i>	Monoallelic Villin-Cre driven constitutive expression of <i>nAtf6</i> in IECs + Global <i>Il10</i> knockout

3.1.2 Gnotobiotic Mice

GF mice were housed at the Technical University of Munich (School of Life Sciences, Weihenstephan). GF status was maintained by housing in sterile isolators and checked by cultivating faecal samples on Wilkens-Chalgren Anaerobe (WCA) agar (Oxoid, Basingstoke, UK) as well as periodic gram-staining of faecal smears.

3.1.3 Genotyping

Tail biopsies or ear punches from *nAtf6*^{IEC} and *nAtf6*^{IEC};*Il10*^{-/-} mice were placed in lysis buffer containing 10 mM Tris-HCl, 50 mM KCl, 0.45 % Nonidet P40, 0.45 % Tween 20. 0.5 mg/mL Proteinase K was added to this mixture and samples were incubated overnight at 65°C.

Proteinase K was then inactivated by incubation at 95 °C for 10 min. 1 µl of tail DNA was added to a solution containing 10µl 2x Onetaq Master mix (NEB, Ipswich, MA, USA), and the respective primer combinations. Primer sequences and PCR cycling conditions are given in

Table 2. Genotyping primers used in this study.

Target	Sequence (Concentration [nM])	Product Size	Program	Temperature (°C)/Time (S)	
<i>nAtf6</i> (Rosa 26 locus)	5-TCCCTCGTGATCTGCAACT-3 [125] 5-ATCAGAGCAGCCGATTGTC-3 [125] 5-GGCGGATCACAAGCAATAAT-3 [125]	WT: 262 Mod: 300	Initial Denaturation	94/60	
			Denaturation	94/20	x30
			Annealing	58/20	
			Extension	68/20	
			Final Extension	68/60	
<i>I110^{-/-}</i>	5-GTGGGTGCAGTTATTGTCTTCCCG-3 [200] 5-GCCTTCAGTATAAAAGGGGGACC-3 [200] 5-CCTGCGTGCAATCCATCTTG-3 [200]	WT: 200 KO: 450	Initial Denaturation	94/60	
			Denaturation	94/20	x35
			Annealing	58/20	
			Extension	68/30	
			Final Extension	68/60	
Villin-Cre	5-GACCATATCCACCGAGTCC-3 [300] 5-AGGAATGCGATGAAGTAGAGC-3 [300] 5-CCTTCAGCAAGAGCTGGG-3 [400] 5-GAGACTCTGGCTACTCATCCAGC-3 [400]	DNA: 585 TG: 300	Initial Denaturation	94/60	
			Denaturation	94/15	x30
			Annealing	58/15	
			Extension	68/20	
			Final Extension	68/60	

3.1.4 Histology and tissue staining

Intestinal sections were excised and opened longitudinally immediately after euthanasia. Sections were rolled using the “Swiss roll” method and subsequently fixed in 4% phosphate-buffered formaldehyde for 48 hours¹⁸¹. Samples were dehydrated automatically using a Leica TP1020 (Soest, Germany), before embedding in paraffin (VWR, Ismaning, Germany; Leica EG1150C). Unless otherwise specified, tissue sections were cut at 5µm (Leica RM2255).

3.1.4.1 Haematoxylin & Eosin staining

Sections were incubated for 15 minutes at 65°C. Deparaffinisation, rehydration and subsequent Haematoxylin (Meditate, Burgdorf, Germany) and 0.2% Eosin (Meditate, Burgdorf, Germany) staining, including a bluing step (bluing reagent; Fisher, Dreieich, Germany), was automated using a Leica ST5020 (Soest, Germany). Slides were mounted using xylene-based mounting medium (DPX new, Merck-Millipore, Darmstadt, Germany).

3.1.4.2 Periodic acid-Schiff/alcian blue (PAS/AB) staining

Formalin-fixed paraffin-embedded (FFPE) sections were deparaffinised, before staining with alcian blue solution (Fisher Scientific, Dreieich, Germany) for 1 minute to stain acidic mucins

(0.5 % volume/volume in 3% acetic acid, pH 2.5). The sections were treated with periodic acid solution (0.5%) for 8 minutes before staining for neutral mucins with Schiff's reagent. Colour was developed with two washes in warm water for 2.5 minutes before counterstaining with haematoxylin, dehydration, and subsequent mounting with xylene-based mounting medium (Merck KGaA, Darmstadt, Germany). Goblet cell numbers were counted in five separate regions of approximately ten crypts, averaged, and expressed as the mean number of goblet cells per 100µm². **3.1.4.3 Carnoy fixation**

To preserve the mucus layer for Fluorescent *in situ* hybridisation, unopened and rolled intestinal sections were pinned using a needle and fixed overnight in Carnoy's solution (60% Anhydrous absolute methanol, 30% Chloroform and 10% Glacial acetic acid v/v). Sections were dehydrated by subsequent washes in anhydrous methanol (2 x 30 minutes), absolute ethanol (2 x 15 minutes), xylene/ethanol 1:1 (5 minutes) and finally xylene (2 x 5 minutes) before paraffin embedding.

3.1.5 Histopathological analysis

Combined tumour and inflammation scoring of intestinal sections was carried out in collaboration with Dr. Katja Steiger and Dr. Marianne Reiser, Institute of Pathology, Klinikum Rechts der Isar. Haematoxylin and Eosin (H&E)-stained tissue sections were scored blinded, based on immune cell infiltration into the mucosa/submucosa and muscle layer, epithelial damage, and atypia/dysplasia as previously described^{152,182}. Additionally, dysplasia associated with inflammation was also considered in the resulting combined score.

3.1.6 Immunohistochemical/-fluorescent staining

Immunohistochemical (IHC) and immunofluorescence (IF) staining of tissue sections was carried out as previously described¹⁸³. In brief, slides were deparaffinised and rehydrated before antigen unmasking by boiling in 10Mm citrate buffer at pH 6.0 for 23 minutes. To inhibit nonspecific binding, sections were blocked for 1 hour according to the species the secondary antibody was raised in (Blocking buffer: PBS, 5% serum according to secondary species, 0.3% Triton X-100). Respective antibodies were diluted in antibody dilution buffer (PBS, 1% BSA, 0.3% Triton X-100) to working concentration (**Table 3**), and incubated with the sections overnight at 4°C. In the case of IHC, the signal was developed with DAB (10x; Fisher Scientific, Dreieich, Germany) before counterstaining with haematoxylin and subsequent mounting. IF sections were incubated with the secondary antibody for 1 hour at room temperature and mounted using a water-based mounting medium (Aquatex; Merck KGaA, Darmstadt, Germany). Image acquisition was performed using a PreciPoint M8 microscope (Precipoint, Freising, Germany) and a Fluoview FV10i microscope (Olympus, Shinjuku, Japan) for IHC and IF respectively.

3.1.7 Fluorescence *in situ* hybridisation (FISH)

Fluorescent *in situ* hybridisation staining was carried out as previously described¹⁵⁵. Briefly, sections were deparaffinised, rehydrated, and fixed in 4% PBS-buffered formaldehyde for 15 minutes. Tissue sections were then washed before permeabilisation in a lysis buffer (1.2% Triton-X-100 solution, 20Mm Tris, 2Mm EDTA), supplemented with 40mg/ml lysozyme, for 45 minutes at 37°C. For hybridisation, sections were incubated with Cy5-conjugated EUB338 (5'-GCTGCCTCCCGTAGGAGT-3') or scrambled FITC labelled nonEUB (5'-

ACATCCTACGGGAGGC-3') as a negative control in 100µl sterile hybridisation buffer (20 mmol/L Tris/0.9 mol/L NaCl and 0.01% volume/volume sodium dodecyl sulphate solution, pH 7.3) overnight at 46°C. Sections were co-stained with Rhodamine-conjugated anti-Ulex europaeus agglutinin I (UEA1) (**Table 3**) and counterstained with 4,6-diamidino-2phenylindole (DAPI; Sigma-Aldrich, Taufkirchen, Germany), before mounting in water-based mounting medium (Aquatex; Merck KGaA, Darmstadt, Germany). Bacterial distance to the epithelium was quantified using Volocity software (Quorum technologies, Guelph, ON, Canada). In each colonic compartment (Proximal and Distal) five separate random measurements were taken within five separate regions and averaged.

Table 3. Antibodies used for immunofluorescence and immunohistochemistry and their respective concentrations.

Primary Antibody	Manufacturer	Dilution
IHC		
GRP78	Abcam, Cambridge (UK)	1:200
Ki67	Cell Signaling Technology, (Danvers, MA, USA)	1:400
IF		
HA-tag	Abcam, Cambridge (UK)	1:100
Secondary Antibody	Manufacturer	Dilution
Donkey anti-rabbit secondary Antibody Alexa Fluor 546	Life Technologies, Carlsbad, CA	1:200
Donkey anti-rabbit, Horseradish peroxidase	Dianova, Hamburg, Germany	1:300
Other	Manufacturer	Dilution
UEA1-Rhodamine	Novus Biologicals (Littleton, CO, USA)	1:1000
DAPI	Sigma- Aldrich, Taufkirchen, Germany	1:1000

3.1.8 Gene expression analysis

RNA was isolated from whole colonic Swiss roll sections stored in Optimal cutting temperature compound (Richard-Allan Scientific™ Neg-50™; Thermo Fisher, Kalamazoo, MI, USA) at 80°C. Sections were resuspended in RA1 buffer (Macherey Nagel, Düren, Germany) before homogenisation using a 23-gauge sterile needle and 10µm/ml Dithiothreitol (DTT; Roth, Karlsruhe, Germany) was added as a reducing agent. RNA was then extracted using the Nucleospin RNA mini kit (Macherey Nagel, Düren, Germany I), according to the manufacturer's instructions. Extracted RNA was measured using a Nanodrop ND-1000 Spectrophotometer before synthesis of cDNA from a total of 500ng RNA. RNA was converted to cDNA using the MMLV point mutant kit (Promega, Mannheim, Germany), using random hexamers to prime synthesis and Moloney Murine Leukaemia virus (MMLV) reverse transcriptase (all from Promega, Mannheim, Germany). Cycling conditions and components of the reaction mixture are detailed in **Table 4**. Gene expression was measured using a Roche Lightcycler 480 (Roche, Mannheim, Germany), utilising a probe-based method (universal probe library, Roche, Mannheim, Germany). Expression was calculated using the $2^{-\Delta\Delta Ct}$ method and normalised to *Gapdh*. Primer sequences used in this study and their respective UPL probes are shown in **Table 5**.

Table 4. Protocol for cDNA synthesis.

cDNA synthesis Components	Volume per sample (µl)
Random-Hexamer [200ng/µl]	1
	5 minutes 70°C
	5 minutes 4°C
5x Strand Buffer (Promega)	5
Molecular grade H ₂ O	3.1
rRNasin® (Promega)	0.65µl
dNTP Mix 10mM	1,25µl
MMLV (200U/µl) (Promega)	1µl
	10 minutes 25°C
	50 minutes 48°C

Table 5. qPCR primers used in this study.

Primer Name	Sequence	Probe Number
<i>Atf6</i> endogenous	F: 5-GGACGAGGTGGTGTTCAGAG-3 R: 5-GACAGCTCTTCGCTTTGGAC-3	110
<i>Atf6</i> -total	F: 5-CCACCAGAAGTATGGGTTTCG-3 R: 5-GGTTCTTTATCATCCGCTGCT-3	73
<i>CXCL10 (IP-10)</i>	F: 5-AATGAAAGCGTTTAGCCAAAAA-3 R: 5-AGGGGAGTGATGGAGAGAGG-3	56
<i>Gapdh</i>	F: 5-TCCACTCATGGCAAATTCAA-3 R: 5-TTTGATGTTAGTGGGGTCTCG-3	9
<i>Grp78</i>	F: 5-CTGAGGCGTATTTGGGAAAG-3 R: 5-TCATGACATTCAGTCCAGCAA-3	105
<i>Muc2</i>	F: 5-GGCAGTACAAGAACCGGAGT-3 R: 5-GGTCTGGCAGTCCTCGAA-3	66
<i>Oasis</i>	F: 5-GATGGAGGACCACTCAAGA-3 R: 5-CCATGATGGAGCACAGCTT-3	81
<i>Tff3</i>	F: 5-GTAACAACCGTGGCTGCTG-3 R: 5-GAGCCTGGACAGCTTCAAAA-3	109
<i>Tnfa</i>	F: 5-TGCCTATGTCTCAGCCTCTTC-3 R: 5-GAGGCCATTTGGGAATTCT-3	49

3.1.9 ELISA measurement of Lipocalin-2

To generate faecal water for ELISA, frozen faecal samples were thawed, and a small portion of sample removed using a sterile spatula. This was then placed into a sterile pre-weighed Eppendorf tube containing 0.5mm glass beads. Based on the weight, samples were diluted 1:10 in sterile PBS and vortexed for 1 minute before centrifugation at 1000×g for 5 minutes. The resultant supernatant was transferred to a new Eppendorf tube and centrifuged again at 6000×g for 5 minutes. Samples were stored at -20°C before use. Lipocalin-2 was measured using the LEGEND MAX™ Mouse NGAL (Lipocalin-2) ELISA Kit (Biolegend, San Diego, CA, USA), according to the manufacturer's instructions. Concentration was determined using the standard curve method and expressed as nanograms per gram of faecal content.

3.2 Microbiota analysis

3.2.1 Sampling of mouse colonic mucosa-associated microbiota

To sample colonic tissue for the analysis of mucosa-associated microbiota, mice were euthanised, the colon removed and then opened longitudinally before clearing of colonic content using a sterile needle. The tissue was then washed with sterile PBS until residual content was removed. Tissue sections were then excised using a sterile scalpel, pre-treated with DNA Away™ (Fisher Scientific, Dreieich, Germany) to destroy contaminating DNA fragments. For the analysis of spatial microbiota, the entire colon was laid on a 0.5cm x 0.5cm grid and individually sectioned using a scalpel and cleaned with 80% ethanol between each cut. Each separate 0.5cm x 0.5cm section was considered an individual sample and frozen at -80 °C before DNA extraction and sequencing (**Figure 5**). All steps were carried out in a laminar flow cabinet.

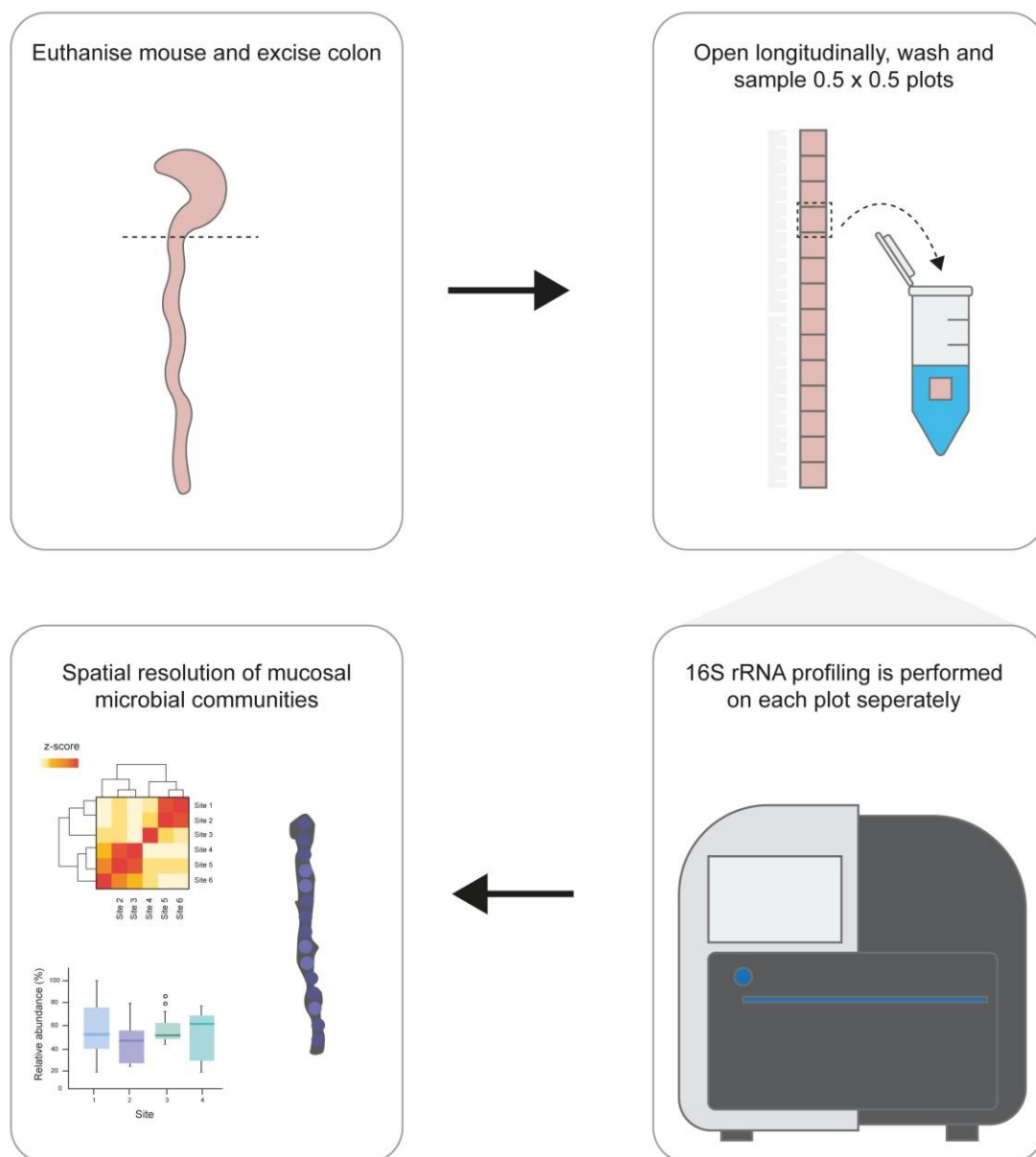


Figure 5. Schematic of spatial organisation experimental design.

3.2.2 *Ex vivo* cultivation of caecal content and long-chain fatty acid co-culture

Caecal content was harvested, immediately after euthanasia, from three *nAtf6^{fl/fl}* and *nAtf6^{fl/fl}; -/-* control mice respectively and diluted 1:10 in 40% glycerol before storage at -80°C. On the day of the experiment, caecal content was thawed under anaerobic conditions for 5 minutes, then left for a further 10 minutes to allow particulates to settle. The supernatant was then removed, and samples were diluted 1:25 in BBL™ Chopped Meat Carbohydrate Broth (BD Diagnostics, Sparks, MD, USA) and then pooled according to genotype. Three technical replicates of 200µl were pipetted per timepoint and genotype, to which 2µl of a long-chain fatty acid solution containing 0.5µmol Nervonic acid (24:1, n-9) and 2µmol each of Behenic (22:0) and Eicosanoic acid (20:0) was added. 0-hour samples were immediately harvested, while the remaining samples were incubated for 7 hours, in an anaerobic cabinet at 37°C.

3.2.3 Microbial DNA extraction from mouse caecal content

DNA extraction from frozen caecal content was carried out using a modified version of the protocol described by Godon *et al*¹⁸⁴. Briefly, caecal samples were homogenised by vortexing and transferred to autoclaved screw-cap Eppendorf tubes containing 500mg silica beads (0.1mm Carl Roth) and kept on ice. To this, 600µl of DNA stabiliser (Macherey Nagel, Düren, Germany), 250µl 4M Guanidine thiocyanate and 500µl N-lauroyl-sarcosine was added. Samples were incubated under moderate shaking (700rpm) for 1 hour at 70°C. Lysis was achieved via mechanical disruption with a FastPrep ®-24 bead beater (MP Biomedicals, Eschwege, Germany) using three 40-second cycles at a speed of 6.5 m/s. 15mg Polyvinylpyrrolidone (PVPP; Sigma-Aldrich, Taufkirchen, Germany) was added to homogenate to remove phenol contamination, before centrifugation at 15,000×g for 3 minutes at 4°C. The resulting supernatant was recovered, and centrifuged again under the same conditions, resulting in a clear supernatant containing lysed bacterial cells. To this mixture, 10mg/ml RNase was added to degrade bacterial RNA and incubated at 37°C for 30 minutes under constant shaking (700rpm). The resulting genomic DNA was then purified using the NucleoSpin ® gDNA clean-up kit (Macherey Nagel, Düren, Germany), according to the manufacturer's instructions. The concentration and purity of extracted genomic DNA was determined using a Nanodrop ND-1000 Spectrophotometer. Samples were stored at -20°C before sequencing.

3.2.4 Microbial DNA extraction from mouse colonic tissue

3.2.3.1 DNA extraction and clean-up

DNA extraction from colonic tissue was performed using enzymatic digestion. Where possible samples were extracted using the same kit batch, to limit differences in inherent kit contamination between different batches¹⁸⁵. Tissues were placed in 180µl sterile lysis buffer (20Mm Tris/HCL, 2Mm EDTA, 1% Triton-X100; pH 8 supplemented with 20mg/ml lysozyme) and incubated for 1 hour in a shaking incubator (950rpm) at 37°C. 10mg/ml Proteinase K (Macherey Nagel, Düren, Germany) was then added, and samples were incubated for 1-3 hours (until complete lysis of the tissue was obtained) at 56°C under moderate shaking (950rpm). Downstream lysis and clean-up were performed using the Nucleospin tissue kit (Macherey Nagel, Düren, Germany) according to the manufacturer's instructions. The concentration and purity of extracted genomic DNA was determined using a Nanodrop ND1000 Spectrophotometer. Extracted DNA was stored at -20°C before sequencing.

3.2.3.1 Spike-in

In order to quantitatively determine the number of 16S rRNA gene copies in a given sample a known amount of artificially constructed 16S rRNA sequences were added. Spiked samples were processed as above, except mucosal samples were weighed before and after the addition of lysis buffer to which 1µl solution containing artificially constructed 16S rRNA sequences at a concentration of 150pg/ml prior to extraction. Spike-in sequences were adapted from Tourlousse *et al.* and are listed in **Table 6**¹⁸⁶.

Table 6. Spike-in DNA sequences.

Name	Length
Ec5001	1525
Ec5002	1525
Ec5003	1525
Ec5004	1525
Ec5005	1525
Ec5501	1525
Ec5502	1525
Ec6001	1525
Bv5501	1520
Ca5501	1495
Ga5501	1508
Tb5501	1554

3.2.4 Microbial DNA extraction from *ex vivo* cultivated caecal content

Microbial suspensions were centrifuged at 8000×g for 5 minutes to pellet bacteria. DNA was subsequently extracted using the Nucleospin Microbial DNA kit (Macherey Nagel, Düren, Germany), following the manufacturer's instructions except DNA was eluted in 25µl, instead of 100µl to increase concentration. Lysis was achieved using a FastPrep ®-24 bead beater (MP Biomedicals, Eschwege, Germany) using a single 40-second cycle at a speed of 6.0 m/s.

3.2.5 16S rRNA amplicon sequencing

High throughput amplicon sequencing of the 16S rRNA gene was carried out as previously described¹⁸⁷. In brief, the V3 and V4 hypervariable regions were amplified via a 2-step protocol, using the 341f and 785r primer pair (341f 5'-CCTACGGGNGGCWGCAG-3', 785r 5'-GACTACHVGGGTATCTAATCC-3'). With 10x15 cycles for caecal content samples and 15x15 for mucosal samples. Samples were barcoded with a double index, according to Kozich *et al*¹⁸⁸. The resulting amplicons were purified using AGENCOURT AMPure XP Beads (Beckman Coulter, Krefeld, Germany), pooled in equimolar ratios and then sequenced on an Illumina MiSeq system (Illumina, San Diego, CA, USA), in paired-end mode (2 x 275bp). Raw reads were processed using IMNGS, which wraps the UPARSE/USEARCH software pipeline^{189,190}. Sequences were demultiplexed, trimmed to first base with a quality score <3 and merged. To remove spurious sequences, a length filter was applied based on the expected amplicon size of 444bp, removing those smaller than 300 and larger than 600bp. The resulting paired sequences were then dereplicated and denoised using UNOISE3, to generate Amplicon sequence variants (ASVs; also known as zero-radius OTUs or zOTUs), reflecting true biological sequences¹⁹¹. Artificial 16S rRNA sequences in spiked samples were first enumerated and then removed using a custom python script, running BLAST to align ASV reference sequences against the spike sequences¹⁹². Taxonomy was assigned using the RDP classifier version 2.11¹⁹³. Samples with less than 5000 total reads were excluded from further analysis. Since low biomass samples are susceptible to contamination which can lead to spurious conclusions, all mucosal samples were further processed with the R package decontam to remove putative contaminants¹⁹⁴. This package utilises the relationship between pooled DNA concentration prior to sequencing and the prevalence and abundance of taxa in negative controls to identify putative contaminants. Identification of contaminants was

performed with the *isContaminant* function using default parameters. Downstream analyses were performed using Rhea and phyloseq^{195,196}. Briefly, ASV tables were normalised using minimum sum scaling or relative abundance. Alpha diversity was measured using Richness, Shannon Effective and for phylogenetic diversity, Faith's phylogenetic diversity (Faith's PD)¹⁹⁷. To assess differences between groups, beta diversity was calculated based on generalized UniFrac distance (GUniFrac)¹⁹⁸. Mean GUniFrac dissimilarity for a given sample was calculated as the dissimilarity to the respective control mean. Differentially abundant taxa were identified using the Linear Discriminant analysis Effect size (LEfSe) algorithm, with a Linear discriminant analysis threshold (LDA) threshold of 3.0, to limit false positives^{199,200}. The total 16S copy number based on spike-in measurement was calculated using the formula:

$$\text{Absolute abundance} = \frac{\left(\frac{s}{r}\right) * n}{\left(\frac{b}{w}\right)}$$

where *s* is the spike-in reads per sample, *r*, a randomly chosen sample to adjust to, *n* spike DNA concentration in ng, *b* bacterial reads per sample and *w*, the weight of the respective sample.

3.3 Metabolomic analysis

3.3.1 Untargeted metabolomics

Untargeted metabolomics measurement was performed in collaboration with Sinah Reiter, Miriam Wimmer (Chair of Food Chemistry and Molecular Sensory Science, Technical University of Munich) and Andreas Dunkel (Leibniz-Institute for Food Systems Biology at the Technical University of Munich).

3.3.1.1 Sample preparation

Sample preparation was performed as previously described¹⁸⁷. Briefly, mouse caecal content (~20mg) or tissue (~25mg) was mixed with 1ml methanol-based extraction solvent in a 2ml bead beater tube (CKMix 2ml, Bertin Technologies, Montigny-le-Bretonneux, France) filled with ceramic beads (1.4 mm and 2.8 mm ceramic beads). Samples were then homogenised using a bead beater (Precellys Evolution, Bertin Technologies) supplied with a Cryolys cooling module (Bertin Technologies, cooled with liquid nitrogen) three times for a duration of 20 seconds, at a speed of 8000rpm. Subsequently, the resulting suspension was centrifuged for 10 minutes, at 6000rpm. Finally, 100µl of supernatant was mixed with 20µl internal standard solution (7 µmol/l) and injected into the Liquid chromatography time-of-flight mass spectrometer (LC-TOF-MS) system for untargeted analysis.

3.3.1.3 Liquid Chromatography-Time of Flight-Mass Spectrometry (LC-TOF-MS)

Untargeted LC-TOF-MS analysis was carried out as described in Metwaly *et al* 2020¹⁸⁷. Briefly, untargeted analysis was performed on an ExionLC Ultra-high performance liquid chromatography (UHPLC) system (Sciex, Darmstadt, Germany), connected to a 6600 TripleTOF instrument (Sciex, Darmstadt, Germany) operating in positive and negative electrospray mode and calibrated using ESI positive and negative calibration solutions (Sciex, Darmstadt, Germany). UHPLC phase separation was performed in reverse as well as hydrophilic interaction stationary phase (HILIC), on a Kinetex C18 column (Phenomenex,

Aschaffenburg, Germany) and ACQUITY BEH Amide column (Water, Eschborn, Germany) respectively. Mass spectrometry was performed using SWATH mode, with fragment spectra recorded in high-sensitivity mode.

3.3.1.4 Metabolomics Data processing

Metabolomics data processing was performed as described previously¹⁸⁷. Briefly, Raw data files were converted into Reifycs Abf (Analysis Base File) files and subsequent untargeted peak picking was performed using MS-DIAL software (version 3.5274)²⁰¹. Alignment was performed across all samples and the peak area of individual features were exported for further analysis using the R statistical software environment. Peak normalisation was based on QC samples, employing the method described in Wehrens *et al* 2016²⁰². All features were then combined into a single table and further normalised according to sample weight, before analysis.

3.4 Bioinformatics & Statistical Analysis

3.4.1 Functional potential prediction

To calculate predicted functional profiles based on 16S rRNA sequencing data, we utilised PICRUST2 version 2.3²⁰³. FASTA files of representative sequences and minimum sum scaled ASV tables were used as input for the command: *picrust2_pipeline.py*, which runs the full PICRUST2 pipeline, aligning and placing the sequences into a reference tree, calculation of 16S copy number, Enzyme Commission (EC) and KEGG orthologs abundances, adjustment of these by 16S abundance and finally infers MetaCyc pathways by collapsing EC numbers according to their associated metabolic pathway. All PICRUST2 data were generated in a high-performance computing environment, utilising the Linux-cluster system at the Leibniz Rechenzentrum, Garching.

3.4.2 Statistical analysis

3.4.2.1 Inferential statistics

Statistical analyses were carried out using R (R software foundation, Vienna, Austria) or GraphPad Prism (version 9.00; GraphPad Software, San Diego, CA, USA). For comparing the mean of two groups, unpaired student's *t*-test or Wilcoxon test was used where appropriate. Differences between multiple groups were tested using ANOVA or Kruskal-Wallis tests. For testing differences in frequency, Fisher's exact test was used. Effect sizes were determined using the *effsize* R package, employing Cohen's *d*, which is calculated as the mean difference divided by the standard deviation for each pairwise difference. To test for differences in beta diversity, multivariate statistical testing was conducted using Permutational multivariate analysis of variance (PERMANOVA)²⁰⁴. Trends in spatial data were detected using locally estimated scatterplot smoothing (LOESS), an extension of nonlinear regression, as implemented in the *stat_smooth* function of *ggplot2*²⁰⁵. For hierarchical clustering of metabolite intensities, Ward's method was used²⁰⁶. Multiple comparisons were controlled using the method of Benjamini & Hochberg²⁰⁷. Unless otherwise specified all data are presented as the mean \pm standard deviation, *p*-values < 0.05 were considered statistically significant (*p*<0.05=*, *p*<0.01=**, *p*<0.01=***, *p*<0.001=****).

3.4.2.2 Machine Learning

Supervised classification of mucosal and luminal microbial profiles was performed using the SIAMCAT R package, utilising ridge-regression, L1-regularised lasso regression and random forest models, which were chosen based on the ease of interpretation of model output^{140,208}. To account for differences in sample size between mucosal and luminal data, mucosal data was randomly subsampled to match luminal. Prior to model training, low prevalent features - defined as those present in less than one-third of samples - were removed, and the data transformed using a centred log-ratio transform. Feature selection was performed using an area under the curve (AUC) threshold of 0.75 and nested within the cross-validation (CV) procedure to avoid artificial inflation of model AUC. All models were trained using five-fold CV, or in case of smaller sample size (<50 total samples), leave-one-out CV (LOOCV), each with five rounds of resampling.

3.4.3 Multi-omic data integration

Integration of metabolomic and microbiota data was performed on log-transformed, scaled, and centred log-ratio (clr) transformed metabolite and ASV matrices, respectively. General associations between the microbiota and metabolites were identified using a multi-block sparse partial least squares discriminant analysis (sPLS-DA) model with seven components and 10-fold CV, as implemented in the DIABLO framework of the mixOmics R package^{209,210}. Associations between ASVs and metabolites were determined using Spearman's rank correlation, using the corrr R package²¹¹.

4. Results

4.1 Mucosa-associated microbiota dysbiosis in *nAtf6*^{IEC} mice

4.1.1 IEC-specific activation of nATF6 alters mucosa-associated microbiota in the absence of tumour formation

We previously showed transgenic nATF6 activation in the intestinal epithelium alone is enough to alter caecal microbiota, however mucosa-associated communities have been shown to be more resilient to environmental insult, possibly due to the harsh conditions of mucosal microenvironments and/or protection mediated by the mucus layer^{155,212}. To determine if nATF6 activation, in the absence of tumour presence, could alter microbial profiles we compared *nAtf6*^{tg/wt} mice to *nAtf6*^{fl/fl} controls. To assess within sample diversity, we utilised multiple alpha diversity metrics, allowing us to individually determine different components of sample diversity, namely the total number of observed ASVs (richness), evenness (Shannon Effective), and phylogenetic diversity (Faith's PD). Alpha diversity, either taxonomic (richness and Shannon Effective) or phylogenetic (Faith's PD) did not differ at 5 or 12 weeks but was significantly increased in 20-week-old *nAtf6*^{tg/wt} mice (**Figure 6A**). Next, to determine between sample differences, we utilised GUniFrac distance to measure beta diversity. This metric takes both species abundances and genetic relatedness into account while giving less weight to rare and highly abundant lineages, allowing the detection of a wider range of differences in composition between samples compared to other metrics¹⁹⁸. Beta diversity showed the opposite trend, with obvious separation in GUniFrac distance at 5 weeks, which gradually decreased with age. Despite this, PERMANOVA results indicated mucosa-associated microbiota was significantly different at all timepoints, independent of cage effects (**Figure 6B and D**). To identify ASVs affected by nATF6 activation, we performed differential abundance analysis using LEfSe¹⁹⁹. Members of *Lactobacillus*, *Odoribacter*, Lachnospiraceae, Porphyromonadaceae and Clostridiales were enriched in *nAtf6*^{tg/wt} mice, while control enriched ASVs were almost entirely classified as Lachnospiraceae, along with a single member of *Alistipes* (**Figure 6C**). Taken together this suggests nATF6 activation also modulates the mucosa-associated microbiota.

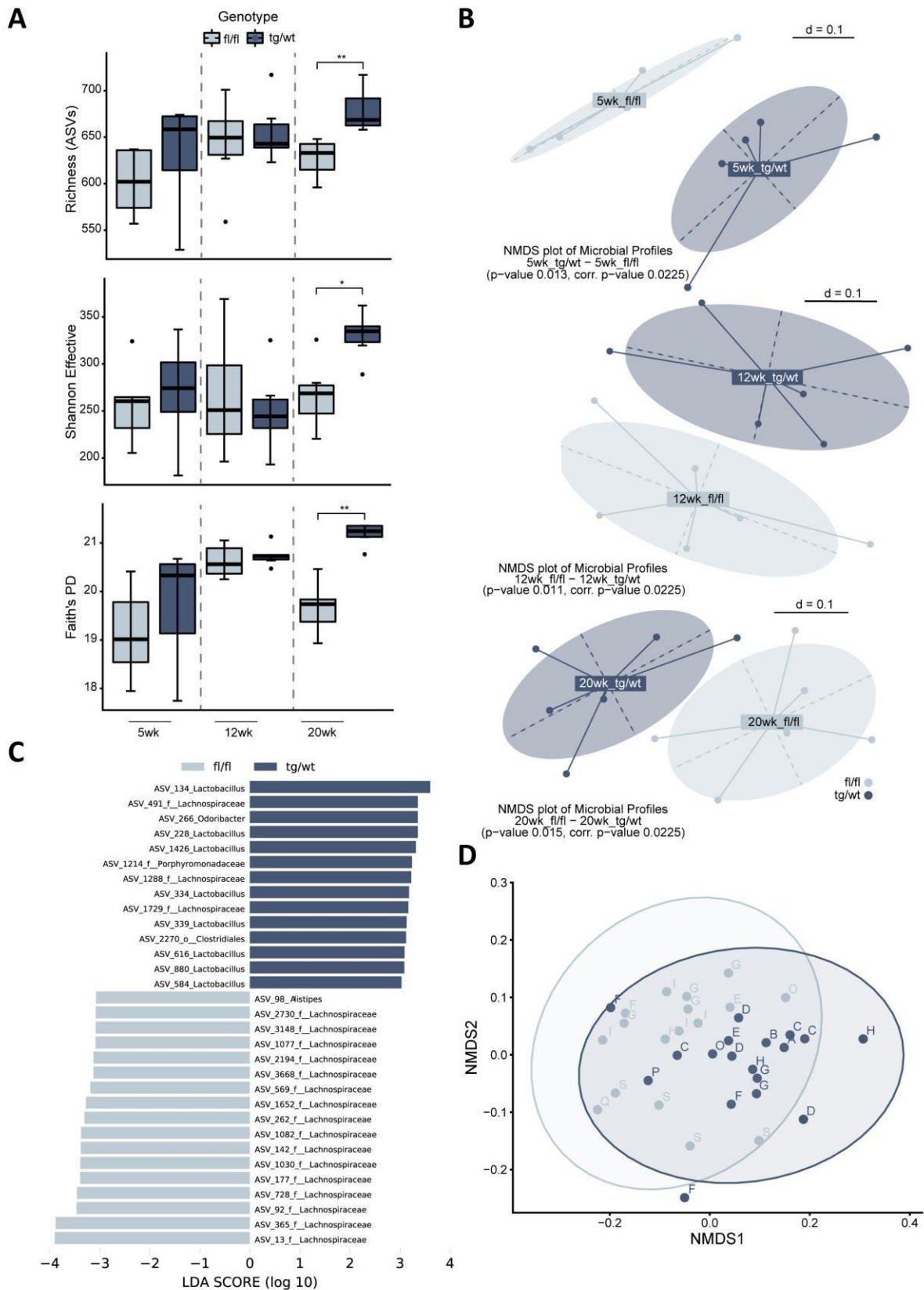


Figure 6. Monoallelic activation of *nATF6* alters the colonic mucosal microbiota. (A) Species Richness, Shannon Effective diversity, Faith's phylogenetic diversity and **(B)** beta diversity, based on generalized UniFrac distance at each timepoint between *nAtf6*^{tg/wt} mice and controls. **(C)** Differentially abundant taxa between *nAtf6*^{fl/fl} and *nAtf6*^{tg/wt} mice at 20 weeks of age. **(D)** NMDS plot of generalized UniFrac distance between *nAtf6*^{fl/fl} and *nAtf6*^{tg/wt} mice including all timepoints, labelled by age.

Since monoallelic activation of nATF6 clearly modulated the mucosa-associated microbiota, we next examined the impact of biallelic nATF6 activation alone on mucosa-associated communities, comparing microbial profiles of 5-week-old *nAtf6^{fl/fl}* to *nAtf6^{tg/tg}* mice. Firstly, to confirm if biallelic nATF6 activation did indeed alter mucosa-associated microbiota we examined alpha and beta diversity. No significant differences in any alpha diversity metric could be detected, however *nAtf6^{tg/tg}* mice did show a reduced trend across all tested metrics (**Figure 7A**). Beta diversity analysis based on GUniFrac distance however, showed a marked difference between genotypes, indicating biallelic nATF6 activation modulates community composition but not diversity (**Figure 7B**). Next, to determine whether increasing gene dose (i.e., monoallelic or biallelic) is associated with the degree of perturbation in mucosa-associated microbiota, we calculated the mean GUniFrac dissimilarity per group relative to the control mean, for all genotypes (*nAtf6^{fl/fl}*, *nAtf6^{tg/wt}*, *nAtf6^{tg/tg}*) at 5-weeks. Accordingly, both *nAtf6* genotypes were significantly different to *nAtf6^{fl/fl}* control, and an increasing gradient could be observed from *nAtf6^{tg/wt}* to *nAtf6^{tg/tg}* mice, although the difference was not statistically significant (**Figure 7C**). Finally, to examine this association in more detail, we correlated GUniFrac dissimilarity, as well as each alpha diversity metric with the expression of the *Atf6* transgene and *Grp78*, as measured by qPCR. *Grp78*, but not *Atf6* expression, negatively correlated with all alpha diversity metrics except Faith's PD, and correlated positively with mean GUniFrac dissimilarity, suggesting modulation of microbial communities occurs downstream of nATF6 activation (**Figure 7D**). Together, these data support augmented modulation of mucosa-associated microbial communities, with increasing *Atf6* gene dose.

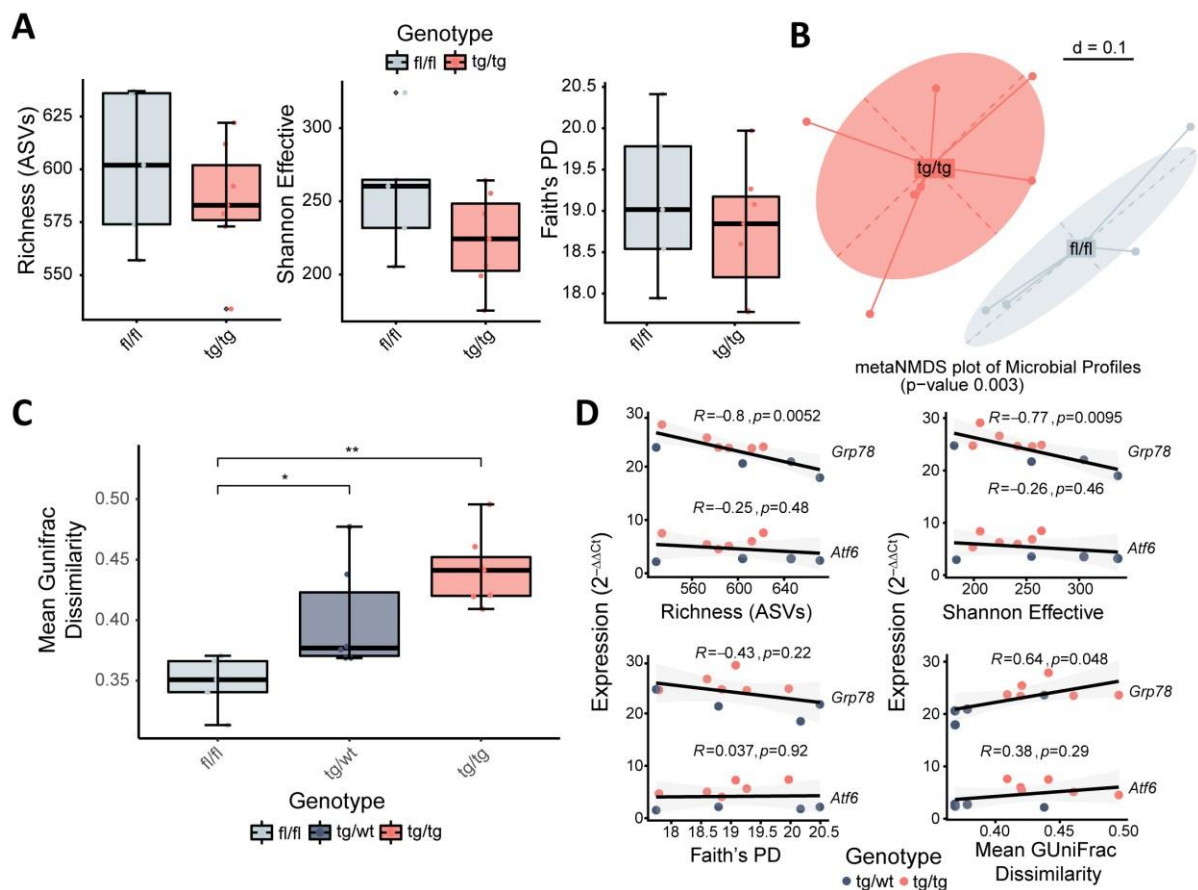


Figure 7. Increasing gene dose augments modulation of colonic mucosa-associated microbiota at the 5-week timepoint. (A) Mucosal alpha diversity in *nAtf6*^{IEC} mice comparing 5-week-old *nAtf6*^{fl/fl} and *nAtf6*^{tg/tg} mice. Left to right: Richness, Shannon Effective and Faith's Phylogenetic diversity. (B) Beta diversity profiles based on generalized UniFrac distance in between 5-week-old *nAtf6*^{fl/fl} and *nAtf6*^{tg/tg} mice. (C) Mean generalized UniFrac distance relative to control across *nAtf6*^{fl/fl}, *nAtf6*^{tg/wt} and *nAtf6*^{tg/tg} sample. (D) Correlation between relative expression, as measured by qPCR ($2^{-\Delta\Delta Ct}$), of *Atf6* and *Grp78* and community diversity metrics (Clockwise from top left: Mean Generalized UniFrac Distance, Richness, Shannon Effective and Faith's Phylogenetic diversity).

4.1.2 Mucosa-associated microbiota improves classification of phenotype

Caecal microbiota profiles provide a snapshot of one of the densest and most diverse environments in the murine host, yet may not accurately represent microbes underlying disease, particularly in conditions where the pathological site lies at the intestinal epithelium such as CRC¹⁴². Considering this, we set out to characterise tumour-associated microbiota to determine if this can improve identification of disease-associated taxa. To confirm that tumour-presence altered the mucosa-associated microbiota, we focused on 12- and 20-week-old *nAtf6*^{fl/fl} and *nAtf6*^{tg/tg} mice. We calculated various alpha diversity metrics, observing no differences between tumour-bearing *nAtf6*^{tg/tg} and *nAtf6*^{fl/fl} at either timepoint (Figure 8A). Nonparametric Spearman's rank correlation however, identified a significant correlation between tumour number and Shannon Effective diversity, indicating tumour burden has a more significant impact on diversity than tumour presence alone (Figure 8B). Analysis of beta diversity based on GUniFrac distance was used to compare microbial profiles between tumour (*nAtf6*^{tg/tg}) mice and controls (*nAtf6*^{fl/fl}). In contrast with alpha diversity, tumour presence significantly impacted mucosa-associated microbiota, with both tumour timepoints clearly

distinct from non-tumour controls (**Figure 8C**). To identify mucosa-associated bacteria associated with tumour presence, we next determined differentially abundant taxa. Tumour-enriched taxa consisted mostly of ASVs classified as *Lactobacillus* and *Turicibacter* and additionally several other single ASVs classified as Porphyromonadaceae, *Odoribacter* and *Romboutsia*. On the other hand, depleted ASVs almost entirely comprised Lachnospiraceae, as well as a single *Anaerotruncus* species, suggesting tumour presence markedly disrupts mucosa-associated microbiota (**Figure 8D**).

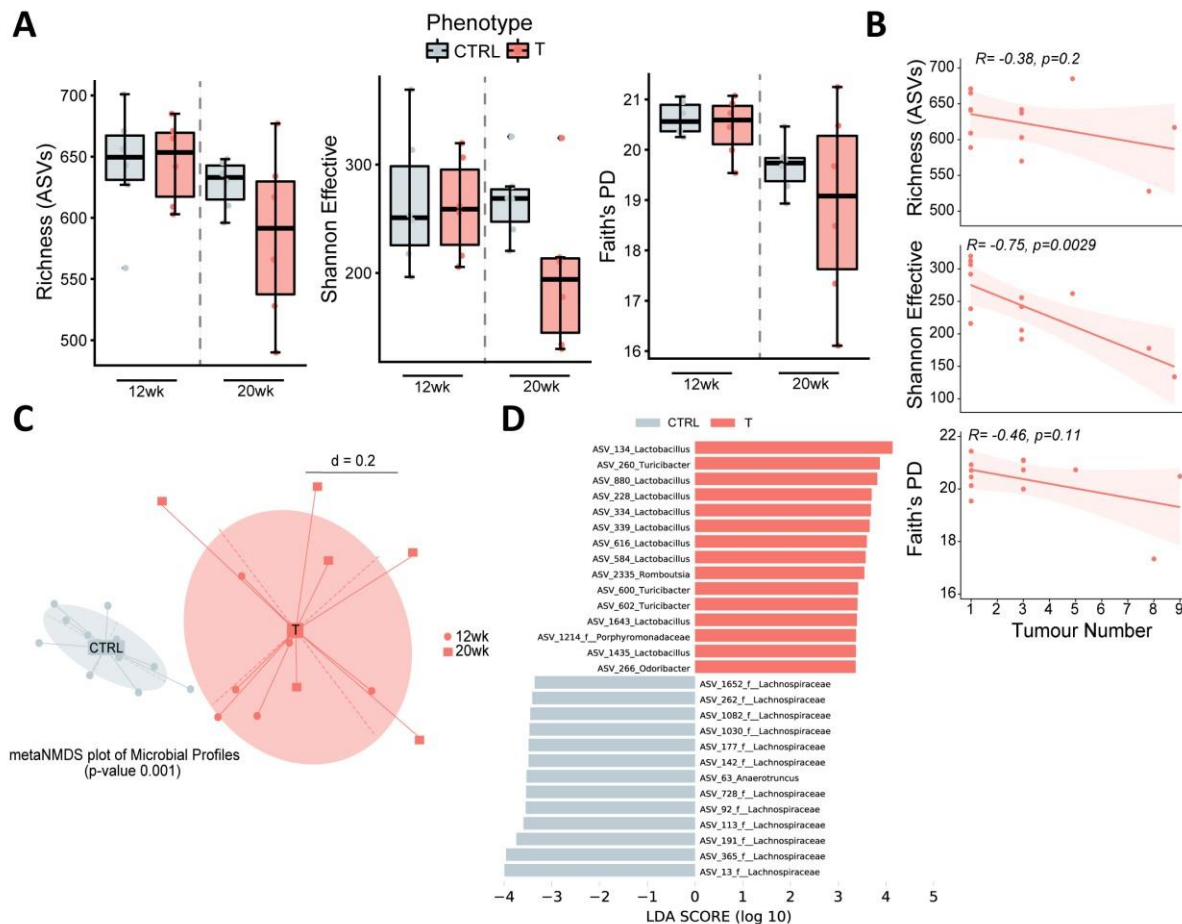


Figure 8. Mucosa-associated microbiota profiles are markedly altered in tumour-bearing *nAtf6*^{IEC} mice. (A) Mucosal alpha diversity in *nAtf6*^{IEC} mice comparing tumour and non-tumour in between 12 and 20-week samples. Left to right: Richness, Shannon Effective and Faith's Phylogenetic diversity. (B) Spearman's rank correlation between each measure of alpha diversity and tumour number. (C) Beta diversity profiles based on generalized UniFrac distance in 12 and 20-week samples. (D) Differentially abundant taxa between *nAtf6*^{fl/fl} and *nAtf6*^{tg/tg} mice at tumour timepoints, identified by LEfSe analysis (LDA threshold = 3.0).

Tumour mucosal samples were clearly altered compared to controls, however, the question remained whether mucosa-associated microbiota provides any additional utility over luminal samples in identifying tumour-associated taxa. To this end, we utilised machine learning (ML) models which facilitate recognition of complex patterns in high-dimensional datasets. Random forest (RF), L1-regularised lasso (L1L) and Ridge regression (RR) models were trained on mucosal and luminal data separately from the same mice, using repeated LOOCV to estimate model performance. Models trained on luminal data achieved AUC values of 0.97, 0.96 and 0.96, respectively. With the exception of the L1L model (AUC = 0.92), all models trained on

mucosal data achieved higher accuracy (RF = 0.99, RR = 0.97) (**Figure 9A**). Since the distribution of samples skewed towards non-tumour over tumour, Precision-Recall curves were also generated to ensure models could accurately discriminate classes. Besides the mucosal L1L model, precision-recall was high across all models, suggesting tumour samples could generally be accurately distinguished from non-tumour (**Figure 9B**). Supporting the increased utility over luminal, two out of three models trained on mucosal data showed significantly higher AUC values across all LOOCV repetitions (**Figure 9C**).

To determine features underlying the enhanced prediction capability of mucosal data, we extracted features included in at least 50% of repeats across all three model types (RF, L1L and RR). Mucosal features in tumour samples were dominated by taxa known to be associated with the murine colonic-mucosa such as *Lactobacillus* and *Alistipes*, while luminal features mostly comprised *Parabacteroides*, indicating mucosa-associated taxa could play an important role in tumorigenesis^{59,213}. (**Figure 10A and B**).

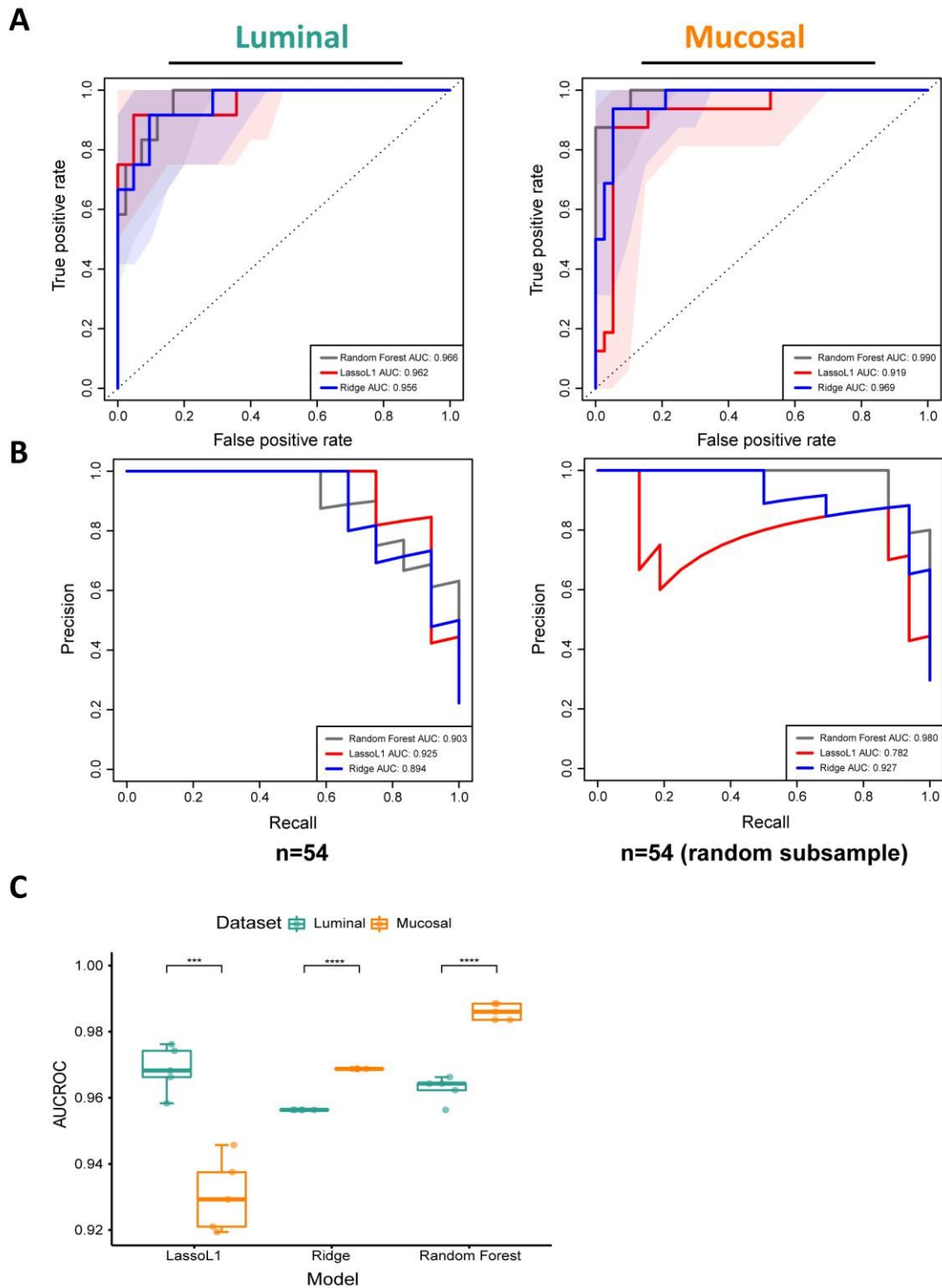


Figure 9. Machine learning models trained on mucosa-associated microbiota profiles better discriminate phenotype compared to luminal. (A) ROC curves comparing classification accuracy of Random Forest, L1-penalised Lasso and Ridge regression models built on luminal and mucosal data, respectively. Mucosal data was randomly subsampled to match luminal sample size (total $n=63$, random subsample= 54). Models were trained using repeated 5-fold cross-validation (CV). **(B)** Precision-Recall curves from each model trained. **(C)** Comparison of AUCROC values for each model trained, comparing luminal and mucosal datasets.

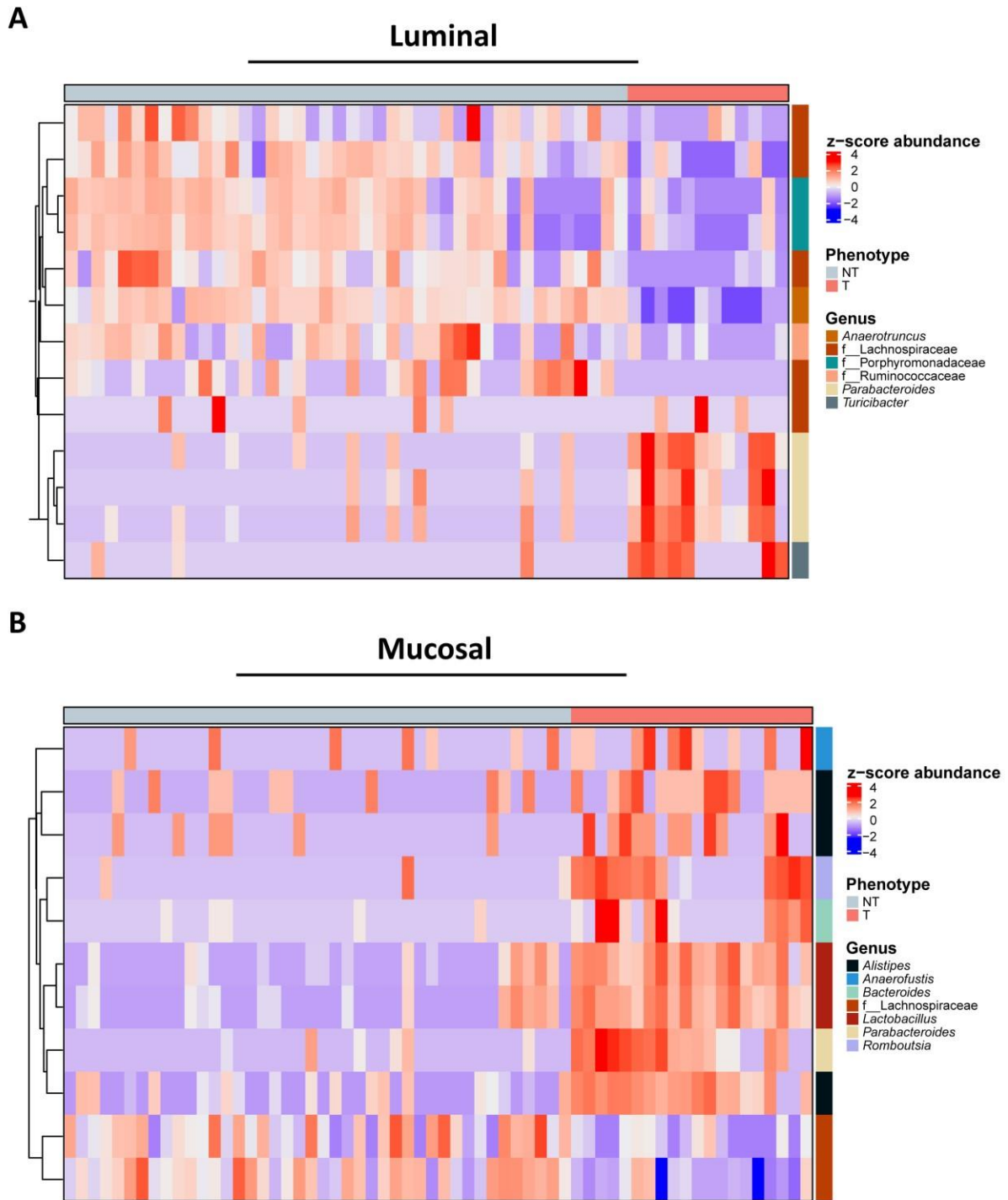


Figure 10. Consensus of discriminating features between non-tumour and tumour phenotypes across all trained models. (A) Luminal content and (B) Mucosa-associated microbiota. Heatmap depicts z-score of Log_{10} transformed abundances.

4.1.3 Analysis of fold change within and between phenotypes differentiates taxa associated with tumour initiation and onset in *nAtf6^{tg/tg}* mice

Having established the utility of mucosal over luminal data in identification of tumour-associated taxa, we next sought to differentiate those merely associated with tumour presence and those likely driving tumour-formation. Towards this end, we utilised the driver-passenger model, as a framework for microbial changes between pre-tumour and tumour onset. The driver-passenger model is a theory of microbial involvement in CRC, whereby drivers initiate tumour formation, altering the local milieu and allowing other taxa (passengers) to establish themselves in the newly created niche¹⁴³. To investigate whether microbial communities changed significantly from pre-tumour to tumour onset, we examined diversity between the 5- and 12-week timepoints in *nAtf6^{tg/tg}* mice. Both richness and Faith's PD were significantly increased at 12 weeks compared to the 5-week timepoint, while a trend of increasing Shannon Effective diversity was also observed (**Figure 11A**). Analysis of beta diversity based on GUniFrac distance however, revealed both timepoints to be distinct from one another, suggesting that microbial shifts do indeed occur from pre-tumour to tumour-onset (**Figure 11B**).

We next examined microbial changes between the pre-tumour timepoint (5-weeks) and tumour onset (12-week), reasoning those taxa that decrease in abundance from pre-tumour to tumour would putatively represent drivers and those that increase, passengers. Hence, we calculated the fold change (FC) and defined a threshold of 1.5 Log₂ FC and FDR-adjusted *p*-value of 0.05 as indicative of a significant shift. Using this method, we identified ASVs classified as *Mucispirillum*, *Anaerotruncus* and Lachnospiraceae as putative initial driver (ID) taxa in *nAtf6^{tg/tg}* mice and *Lactobacillus*, *Bacteroides*, *Parabacteroides*, *Alistipes*, *Alloprevotella* as passengers (**Figure 11C**). This suggests in *nAtf6^{tg/tg}* mice that microbial shifts occur in concordance with the driver-passenger model.

The driver-passenger model in its current form, however, only covers microbes which demonstrate clear shifts between pre-tumour and tumour onset, yet many gut microbes are highly adaptable to changing environments and would be discounted under this model, as they may not demonstrate significant shifts in abundance²¹⁴. We sought to identify a third group of taxa, which we term sustained drivers (SD), defined as those that do not change between pre-tumour and tumour-onset but are enriched in susceptible mice compared to controls. Using this approach, we additionally identified several taxa matching these criteria, mostly composed of members of *Lactobacillus*, *Odoribacter*, *Bacteroides* and *Alistipes* (**Figure 11D**). Heatmap visualisation of the identified ID, SD, and passenger taxa confirmed that shifts in their abundance were restricted to *nAtf6^{tg/tg}* mice, and not related to age or genotype factors (**Figure 11E**).

A number of existing studies, aiming to examine changes between pre-tumour and tumour onset in the context of the driver-passenger hypothesis have generally done so by comparing tumour mucosa, to non-tumour adjacent mucosa, under the assumption that driver taxa may still be present where neoplastic transformation has not yet occurred^{215–217}. However, there are conflicting reports on whether tumours and adjacent mucosa differ microbially. We therefore aimed to assess whether tumour and adjacent mucosa from the same mouse differed in terms of microbial composition and whether adjacent mucosa could be utilised to validate driver taxa. At the level of alpha diversity, no significant differences could be detected in species richness, Shannon Effective or Faith's PD (**Figure 12A**). Similarly, on the level of beta diversity microbial profiles from each sample type clustered together (**Figure 12B**). These

data suggest despite a difference in phenotype, the microbiota between tumour and adjacent mucosa did not differ. Next, to determine if the distribution of ID and SD taxa was altered between tumour and adjacent mucosa, we compared their relative abundance. Overall, abundance of ID and SD did not show any major shifts between tumour and adjacent mucosa, indicating adjacent non-tumour mucosa cannot be used to identify ID or SD (Figure 12C).

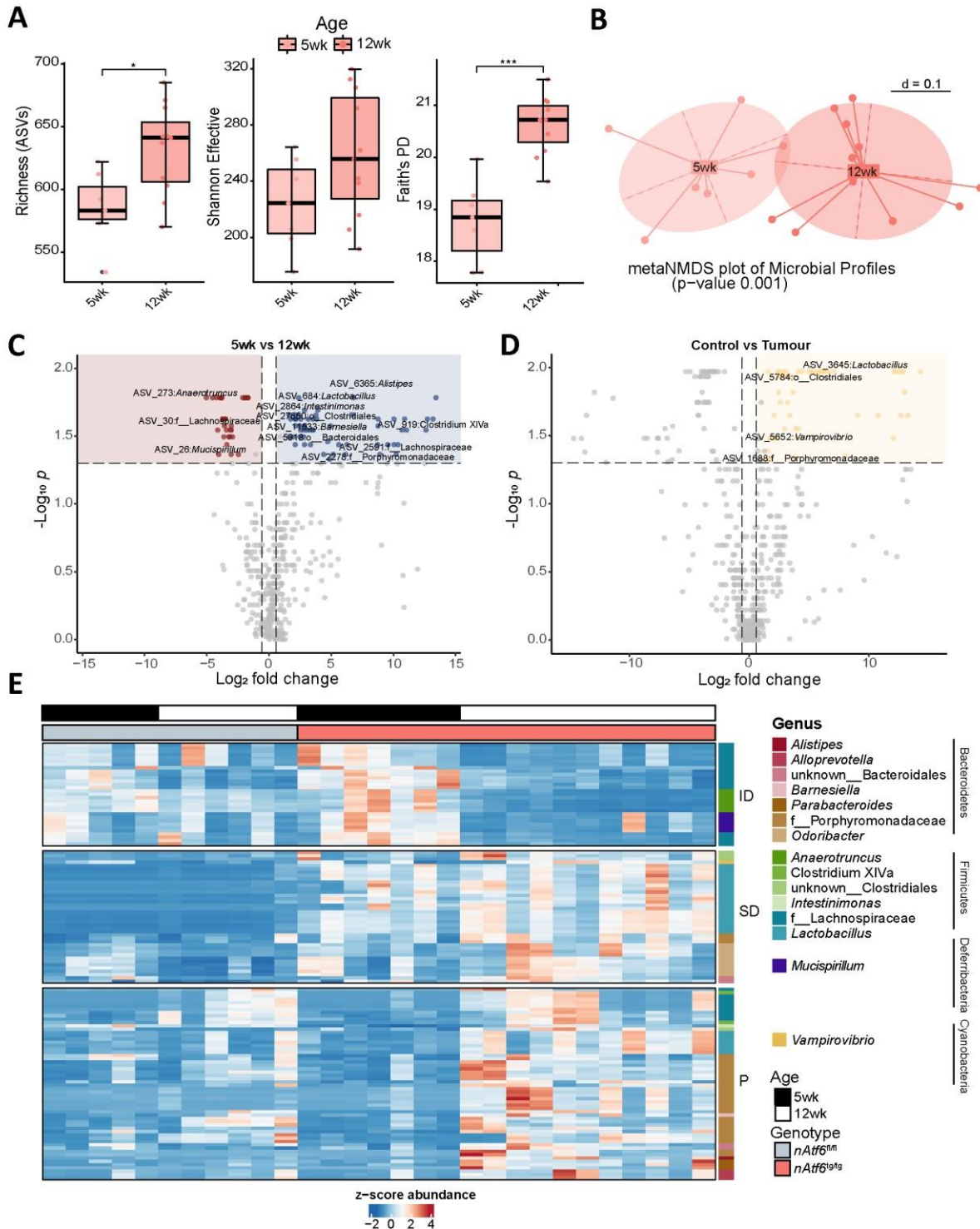


Figure 11. Identification of putative passenger and driver taxa in *nAtf6*^{tg/tg} mice. (A) Alpha diversity (left to right: Species Richness, Shannon Effective and Faith's Phylogenetic) and (B) NMDS plot of beta diversity between 5 - and 12-week timepoints in *nATF6*^{tg/tg} mice. (C) Volcano plot of differential ASVs

between 5- and 12-week *nAtf6^{tg/tg}* mice. Coloured squares highlight significant taxa with an FDR-corrected p -value <0.05 . Red represents putative initial drivers and blue putative passengers. **(D)** Volcano plot of differential ASVs between non-tumour *nAtf6^{fl/fl}* controls and *nAtf6^{tg/tg}* tumour mice. The coloured square highlights significant taxa with an FDR-corrected p -value <0.05 . Yellow represents putative sustained drivers. **(E)** Heatmap of initial driver (ID) and passenger (P) abundances at 5 and 12-weeks in *nAtf6^{fl/fl}* controls and *nAtf6^{tg/tg}* mice.

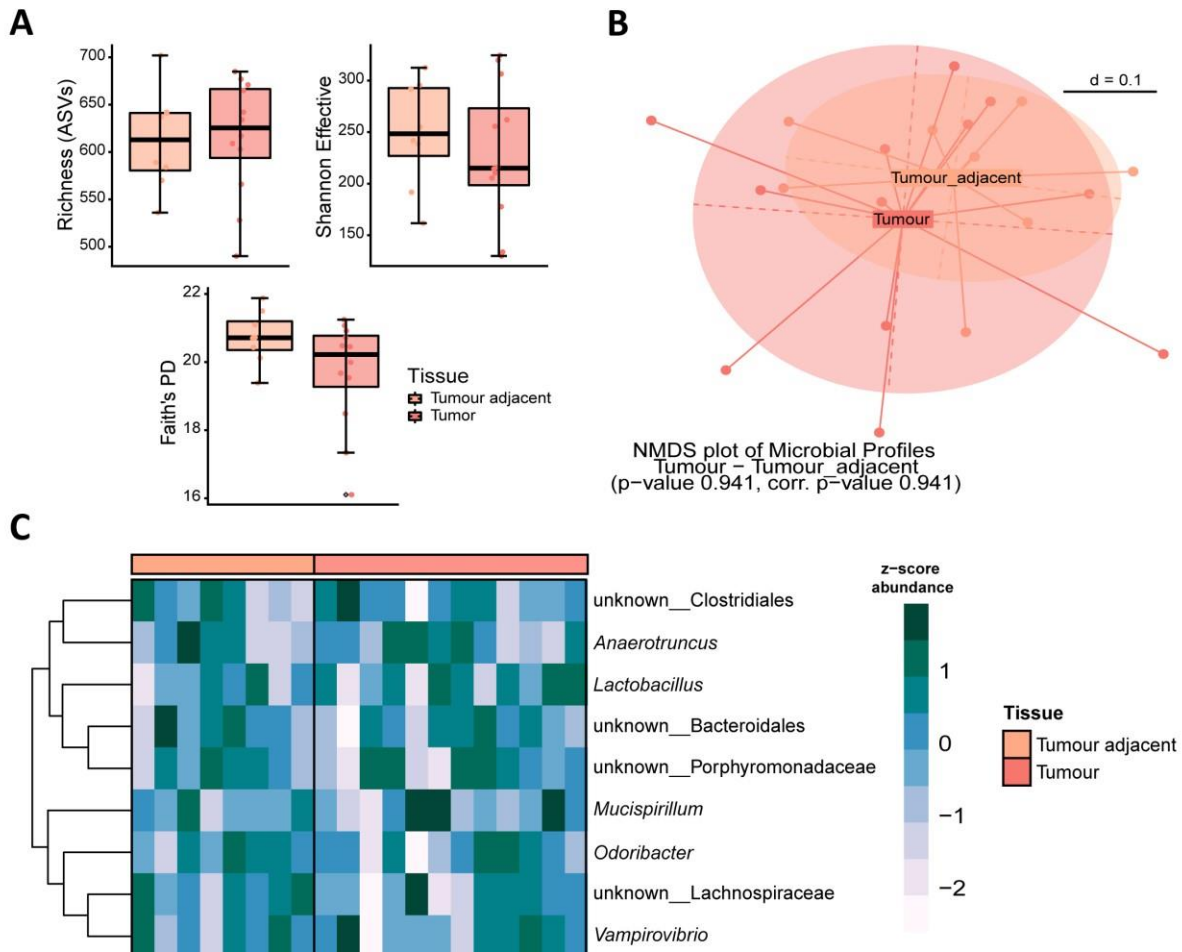


Figure 12. Driver taxa in *nAtf6^{tg/tg}* mice are not tumour specific. **(A)** Mucosal alpha diversity in tumour and adjacent samples from *nAtf6^{tg/tg}* mice. Left to right: Richness, Shannon Effective and Faith's Phylogenetic diversity. **(B)** NMDS plot of beta diversity of tumour and tumour-adjacent microbial profiles based on generalized UniFrac distance. **(C)** Heatmap of initial and sustained driver genera abundance between tumour and adjacent mucosa. Data are presented as z-scores of Log_{10} transformed abundances.

4.2 Spatial organisation of mucosa-associated microbiota in *nAtf6^{IEC}* mice

4.2.1 Microbiota profiles in *nAtf6^{IEC}* mice vary along the colon and are altered by tumour presence

Having confirmed the utility of mucosa-associated over luminal communities and identified putative driver and passenger taxa we next sought to understand how tumour presence alters spatial structure of the microbiota and where taxa of interest localise to. The spatial compartmentalisation and organisation of microbiota is thought to play an important role in

microbe-host interactions. The local mucosal environment varies along the length of the GI tract in terms of oxygen gradient, pH, and available nutrients, governing the bacteria present, which in turn may impact the host¹³. Several other studies have examined spatial microbial differences along the length of the murine colon, including in mucosa-associated communities, finding broad community level differences along the longitudinal axis and a key role for diet in modulating spatial structure^{34,218}. Despite this, spatial structuring of the murine colonic microbiota has not yet been examined in the context of tumorigenesis. Upon tumour formation, the local environment is heavily altered, which likely disrupts mucosa-associated microbiota and their spatial structuring^{219,220}. To identify spatial alterations in mucosa-associated microbiota driven by tumour formation, we sought to generate spatially resolved microbial profiles along the longitudinal axis of the colon. We defined and sampled 0.5 x 0.5cm plots from the colonic mucosa of 12-week-old *nAtf6^{tg/tg}* and *nAtf6^{fl/fl}* mice and performed 16S rRNA sequencing on each plot individually, allowing us to determine spatial structuring of microbiota at tumour onset and to map ASVs to particular sites.

We have previously shown that tumours in *nAtf6^{tg/tg}* are mainly restricted to the proximal colon, however it is not clear if certain sites are more susceptible than others¹⁵⁵. We therefore recorded the site where each tumour occurred and calculated site incidence. Tumours did not occur beyond site ten, approximating the boundary between the mid and distal colon. Incidence was relatively evenly distributed within this boundary however, peaking at 50% in sites four and eight, suggesting each site within the proximal to mid-colon may be equally susceptible to tumour development (**Figure 13A**). To investigate the impact of tumour formation on microbial diversity, we calculated the mean Shannon Effective diversity at each site in *nAtf6^{fl/fl}* and *nAtf6^{tg/tg}* genotypes. A clear reduction was observed at all sites in *nAtf6^{tg/tg}* mice, though the difference was more obvious in sites where tumours occurred (**Figure 13B**). We next examined taxonomic differences between and within genotypes at the family level. Across all sites, communities were broadly similar, implying microbial shifts occur at lower taxonomic ranks. Nevertheless, a clear expansion of taxa from the families Deferribacteraceae and Lactobacillaceae was evident in *nAtf6^{tg/tg}* mice, indicating tumour presence and/or nATF6 activation can modulate spatial structure of colonic mucosa-associated microbiota (**Figure 13C**).

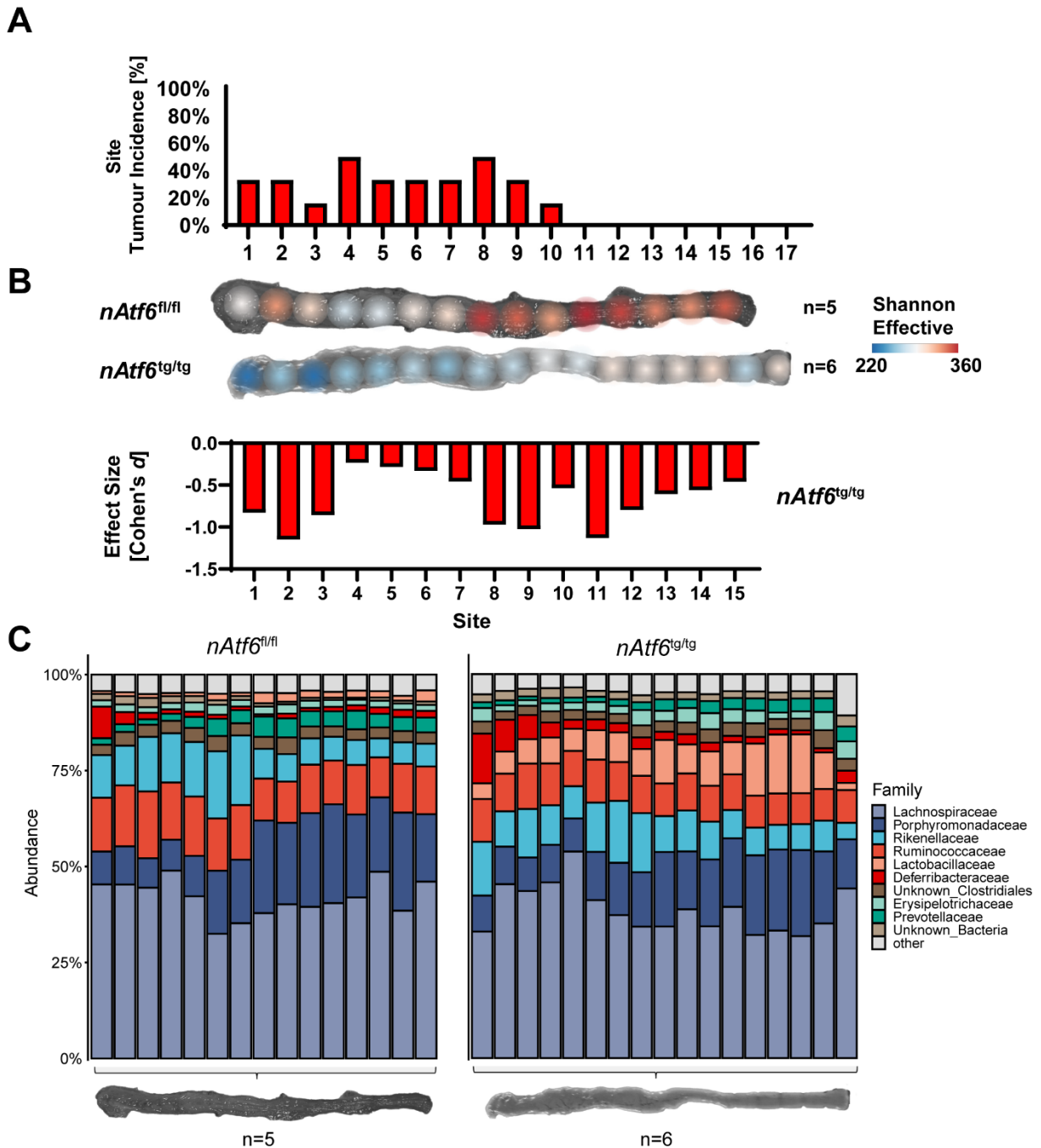


Figure 13. Mucosa-associated microbial profiles differ along the length of the colon irrespective of phenotype. (A) Tumour incidence by site from proximal to distal. (B) Mean Shannon Effective diversity at each site along the colon in *nAtf6^{fl/fl}* mice (top) and *nAtf6^{tg/tg}* mice (bottom). The magnitude of change between genotypes (Cohen's *d* effect size) is displayed beneath. (C) Family-level bacterial composition at each site in *nAtf6^{fl/fl}* (left) and *nAtf6^{tg/tg}* mice (right).

Since limited family-level shifts were observed, we investigated differentially abundant ASVs between genotypes, across all sites. Colonic length in *nAtf6^{tg/tg}* was in some cases longer than in *nAtf6^{fl/fl}* mice thus we limited our analysis to shared sites (one to 15). Site-by-site differential abundance was conducted using LEfSe analysis, using an LDA threshold of 3.0. To enable visualisation, we limited the number of taxa shown to the top five where possible, sorted by

LDA score. Striking differences were observed between genotypes, with known mucosal dwelling members from the families Lachnospiraceae and Ruminococcaceae consistently enriched in *nAtf6^{fl/fl}* mice, especially in proximal regions, while members of Porphyromonadaceae dominated distal regions. In *nAtf6^{tg/tg}* mice however, enriched taxa in all sites were dominated by members of *Lactobacillus*, *Mucispirillum* and *Odoribacter*, indicative of environment-wide microbial shifts and spatial homogenisation in tumour mice (**Figure 14**).

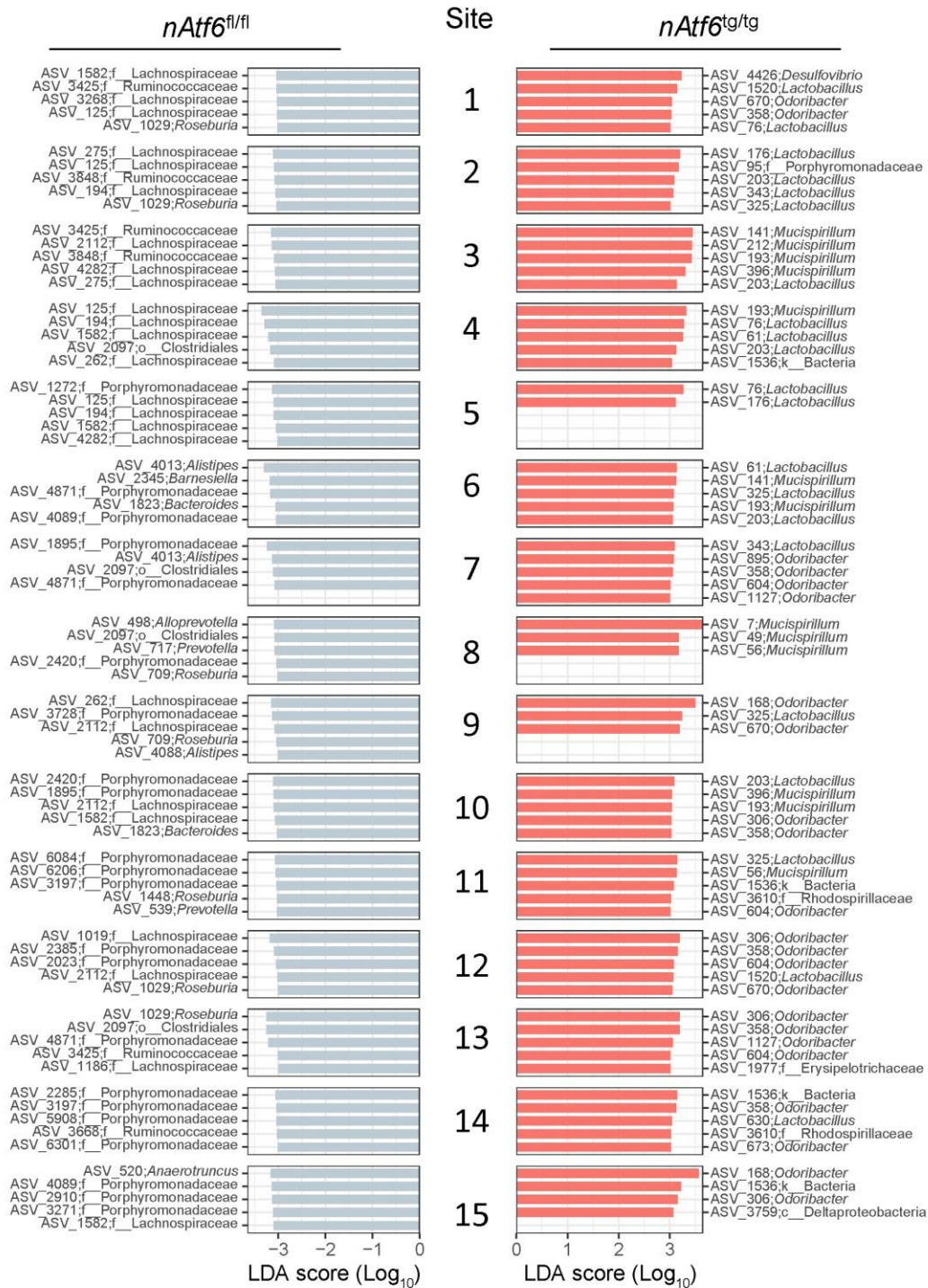


Figure 14. Differentially abundant taxa between genotypes vary by site. Site-by-site comparison of the top five differentially abundant taxa between genotypes as determined by LEfSe analysis, ranked by LDA score (LDA threshold = 3.0).

4.2.1 ATF6 activation drives spatial microbial alterations

Microbial changes were clearly observed at each site along the length of the colon in *nAtf6*^{tg/tg} mice, however not all of these sites are the same in terms of phenotype and whether these

changes were due to tumour formation, nATF6 activation or a combination remained unclear. To assess whether tumour formation or nATF6 activation was the main driver of community change, we compared alpha and beta diversity between and within genotypes. We grouped samples by mucosal phenotype into control, non-tumour (NT), tumour-adjacent (TA) and tumour (T). No significant differences were observed in Richness, although a trend of reduction compared to control was observed in all *nAtf6*^{tg/tg} sample types. Independent of mucosal phenotype, Shannon Effective diversity was significantly reduced in *nAtf6*^{tg/tg} mice compared to control, indicating a reduction in species evenness rather than total species number in nATF6 overexpressing samples (**Figure 15A**). We additionally measured total microbial load using a known concentration of artificial spike-in sequences, allowing us to calculate the total number of 16S copies¹⁸⁶. Interestingly, microbial load was increased in T mucosa compared to *nAtf6*^{fl/fl} control and NT *nAtf6*^{tg/tg} mucosa (**Figure 15B**). At the level of beta diversity, microbial profiles separated by phenotype, with shifts most apparent between all *nAtf6*^{tg/tg} mucosa and control (**Figure 15C**). These limited changes observed within susceptible *nAtf6*^{tg/tg} mucosa compared to between genotypes support a central role for nATF6 activation in the modulation of microbial community structure while suggesting a more limited role of tumour presence.

We next aimed to assess the impact of nATF6 activation and tumour formation on individual taxa. We employed LEfSe analysis comparing tumour, tumour-adjacent and non-tumour mucosa each with mucosa from *nAtf6*^{fl/fl} controls. Taxa differentially abundant in *nAtf6*^{tg/tg} varied between the three-different phenotypes, with non-tumour mostly enriched in members of *Lactobacillus*, tumour-adjacent in *Mucispirillum* and tumour tissue in Lachnospiraceae (**Figure 16A, B and C**). An enriched core set of the same ASVs was shared between all *nAtf6*^{tg/tg} mucosal samples however, comprising ASVs classified as *Mucispirillum*, *Odoribacter* and *Lactobacillus*, taxonomically resembling those previously identified as drivers. Together these results indicate that nATF6 activation supports widespread colonisation of putative driver microbes, regardless of mucosal phenotype.

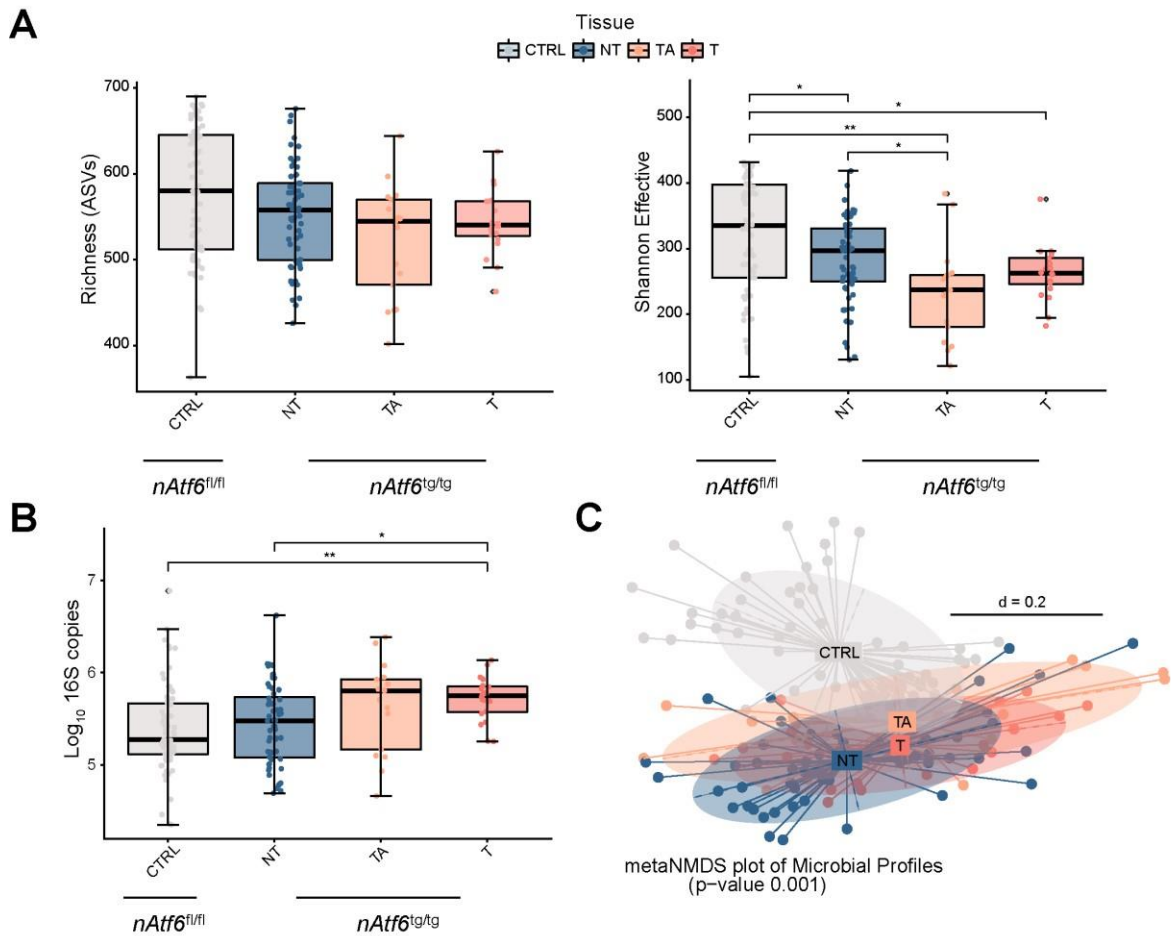


Figure 15. ATF6 activation is the major driver behind microbial community alterations. (A) Mucosal alpha diversity in *nAtf6^{IEC}* mice across various phenotypes: Control (CTRL), Non-tumour (NT), Tumour-adjacent (TA) and Tumour (T). Left to right: Richness and Shannon Effective. **(B)** Total 16S rRNA copy number as estimated by artificial spike-in DNA. Presented as Log₁₀ copy number. **(C)** NMDS plot of beta diversity, based on generalized UniFrac distance comparing control mucosa from *nAtf6^{fl/fl}* mice to Non-tumour (NT), Tumour-adjacent (TA) and tumour (T) mucosa from *nAtf6^{tg/tg}* mice.

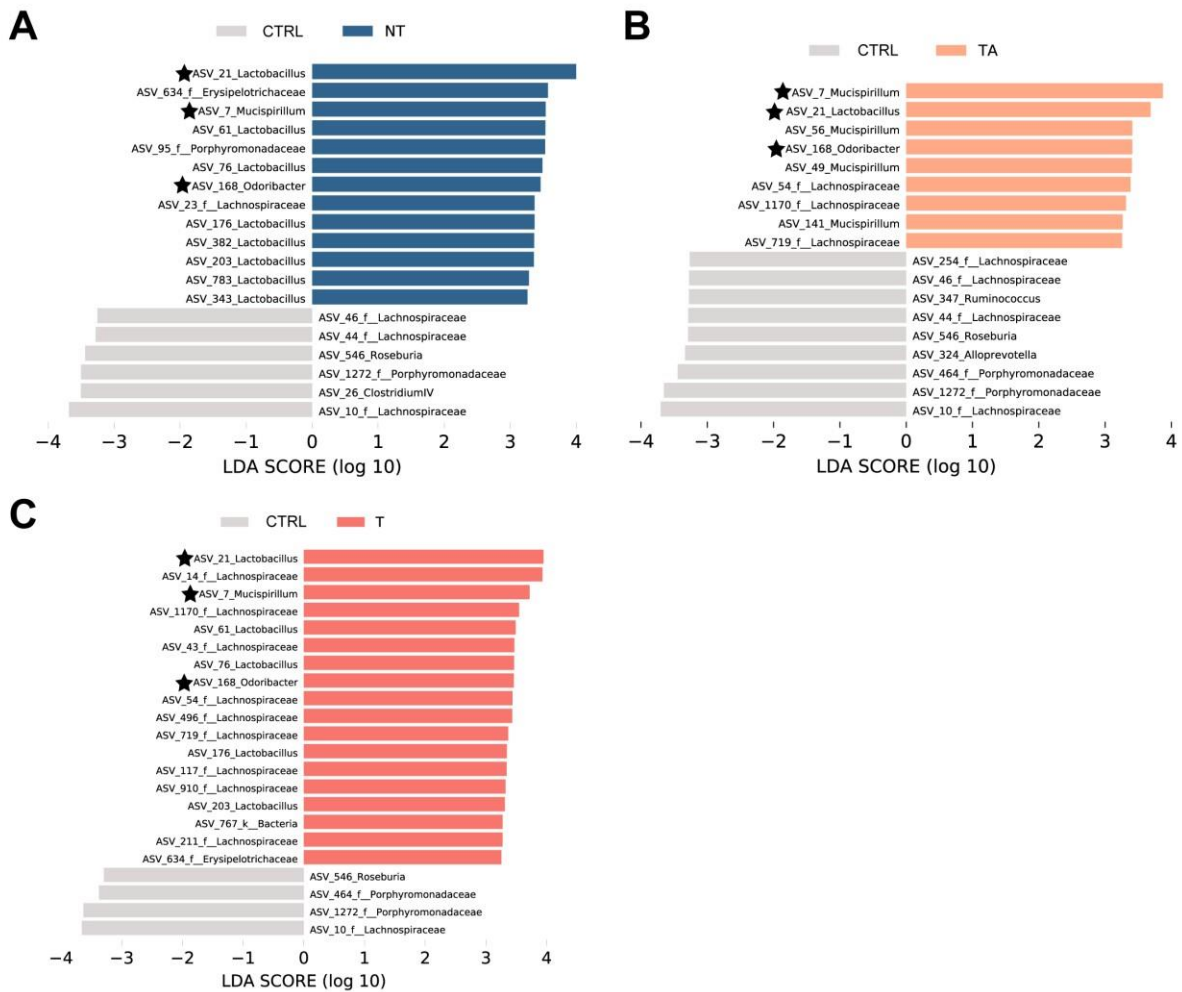


Figure 16. *nAtf6^{tg/tg}* mucosal samples share a core set of enriched taxa. LefSe analysis of differentially abundant taxa comparing (A) Non-tumour (NT) *nAtf6^{tg/tg}* mucosa, (B) Tumour-adjacent (TA) mucosa and (C) Tumour (T) mucosa, with control (CTRL) as a reference. An LDA score threshold of 3.0 was applied across all analyses. Shared *nAtf6^{tg/tg}* enriched taxa are highlighted with a star symbol.

4.2.2 Putative microbial drivers are reproducible across cohorts

As all three of the taxa in the core set of shared ASVs were taxonomically related to the previously identified ID/SD taxa, we aimed to test whether the ID, SD, as well as passenger we had identified previously, could be detected in this independent cohort of *nAtf6^{tg/tg}* mice. ASV sequences from the mucosal dataset were aligned against sequences from our spatially resolved dataset using BLAST¹⁹². To account for minor differences in error profiles between different sequencing runs, we allowed for a single mismatch between sequences. Across all *nAtf6^{tg/tg}* samples, we identified ASVs matching 100% of the ID, 80% of SD and 97% of passengers, suggesting consistency across cohorts (Figure 17A). We next examined the absolute abundance of each group along the length of the colon in *nAtf6^{tg/tg}* mice. ID load was focused in the proximal third of the colon, while both SD and passenger demonstrated similar abundance across all sites (Figure 17B). Furthermore, each of the three, core enriched ASVs, was confirmed as either an initial driver (ASV_7; *Mucispirillum*) or sustained driver (ASV_21; *Lactobacillus*, ASV_168; *Odoribacter*) taxa (Figure 17C). Together, these findings validate a robust set of ASVs in line with the driver-passenger model in *nAtf6^{tg/tg}* mice.

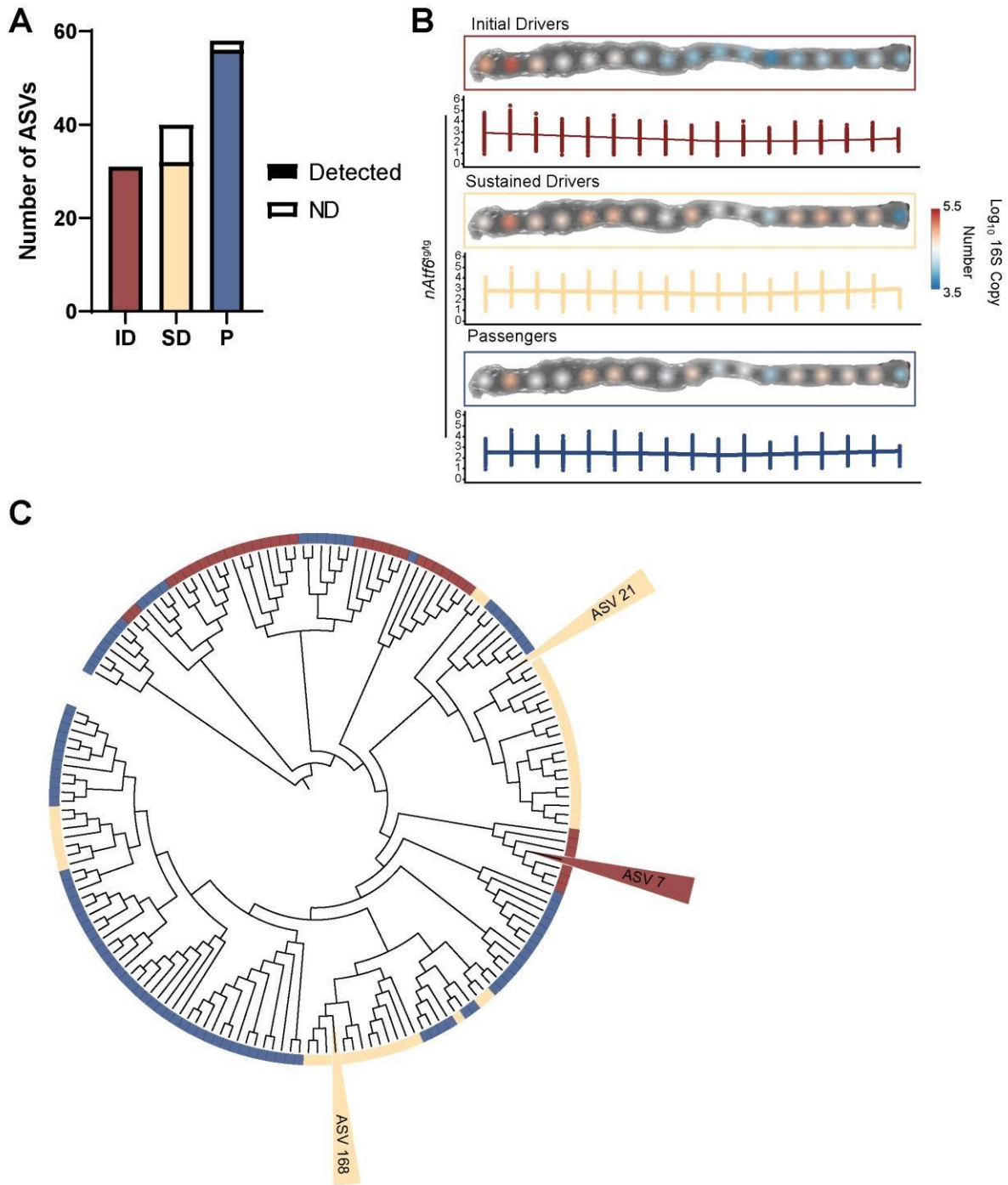


Figure 17. Driver and passenger taxa are reproducible across cohorts. (A) Number of detected/not detected (ND) driver-passenger ASVs based on BLAST alignment. Not detected refers to ASV sequences which did not generate significant alignments with any sequences in this cohort. **(B)** Log₁₀ absolute abundance of initial, sustained driver and passenger taxa along the length of the colon in *nAtf6^{tg/tg}* mice, showing distribution among *nAtf6^{tg/tg}* mice underneath. Trends are visualised using LOESS regression. **(C)** Phylogenetic tree of driver and passenger taxa, highlighting enriched core ASVs.

Next, to examine where these three ASVs localised to, we performed spatial mapping of their absolute abundance along the length of the colon. Since nATF6 activation in *nAtf6^{tg/tg}* mice appeared to support the presence of a number of ID and SD taxa we next explored the extent of this enrichment, and where these taxa localised to. Under the driver-passenger model, ID

are proposed to be indigenous, and it is possible that SD are as well. This implies these taxa colonise a healthy mucosa without a genetic or environmental trigger but may be restricted to certain locations. We therefore hypothesised that nATF6 drives expansion of ID and SD taxa. Supporting this, spatial mapping of Log_{10} transformed absolute abundance revealed a clear enrichment of each of the three driver ASVs in *nAtf6*^{tg/tg} mice, indicating nATF6 activation licences expansion of ID and SD taxa range, including into regions where tumours never occur. Notably, this expansion was more apparent in SD taxa than ID. Driver abundance was also significantly increased at multiple sites compared to *nAtf6*^{fl/fl} control; however, this did not occur at all sites, suggesting nATF6 activation may enable these taxa to expand their habitat range but not necessarily abundance. (**Figure 18A, B and C**). Together, this indicates nATF6 activation enables widespread colonic colonisation of driver microbes.

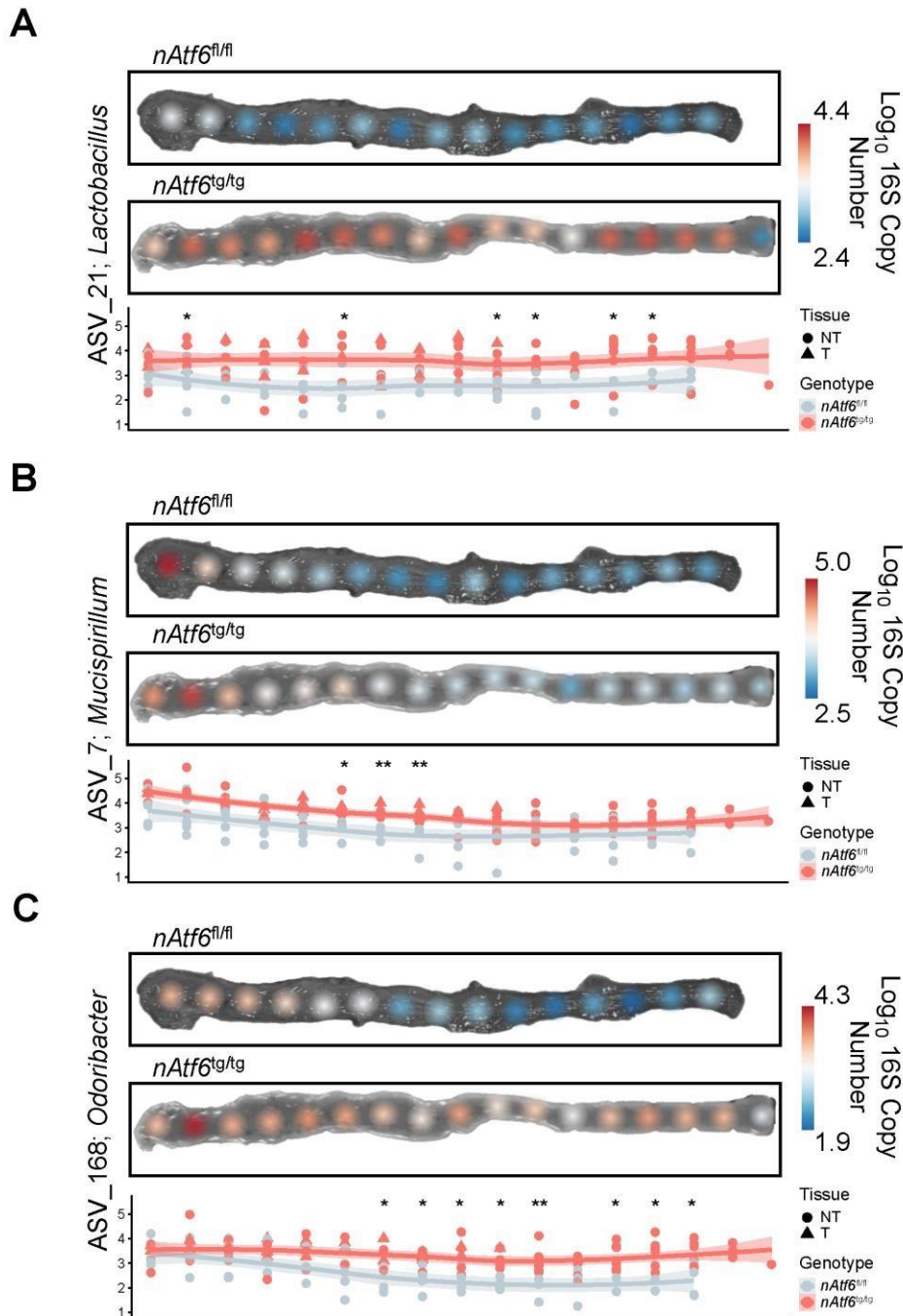


Figure 18. nATF6 activation licences widespread expansion of driver taxa. Spatial maps comparing Log₁₀ transformed mean absolute abundance of driver taxa between *nAtf6^{fl/fl}* and *nAtf6^{tg/tg}* mice, showing (A) ASV_21; *Lactobacillus*, (B) ASV_7; *Mucispirillum* and (C) ASV_168; *Odoribacter*. The distribution amongst all mice is shown underneath, using LOESS regression to visualise trends. Individual *nAtf6^{fl/fl}* samples are shown in grey, while *nAtf6^{tg/tg}* samples are shown in pink.

4.3 Characterisation of the *nAtf6*^{IEC};*Il10*^{-/-} Phenotype

4.3.1 Basic characterisation of a novel nATF6-driven model of colitis-associated cancer

We previously demonstrated that DSS-induced acute inflammation induced tumour formation in *nAtf6*^{tg/wt} mice, however, the contribution of long-term inflammatory activation remains unclear¹⁵⁵. To investigate the effect of chronic inflammation on tumour development in non-susceptible *nAtf6*^{tg/wt} mice we generated a new model crossing *nAtf6*^{IEC} mice with a global *Il10* KO (**Figure 19A and B**). Absence of *Il10* was confirmed by genotyping PCR, while HA-tag staining, and qPCR showed clear activation of the *Atf6* transgene (**Figure 19C, D and E**). Upon activation, *Atf6* induces the transcription of target genes, including *Grp78*. Thus, to confirm downstream *Atf6* signalling, we performed IHC staining for GRP78 and additionally measured its expression via qPCR. In *nAtf6*^{tg/wt;-/-} mice GRP78 staining was markedly enhanced compared to *nAtf6*^{fl/fl;-/-} control, while expression of *Grp78* mRNA was enhanced ~10 fold, clearly indicating nATF6-mediated activation (**Figure 19F and G**).

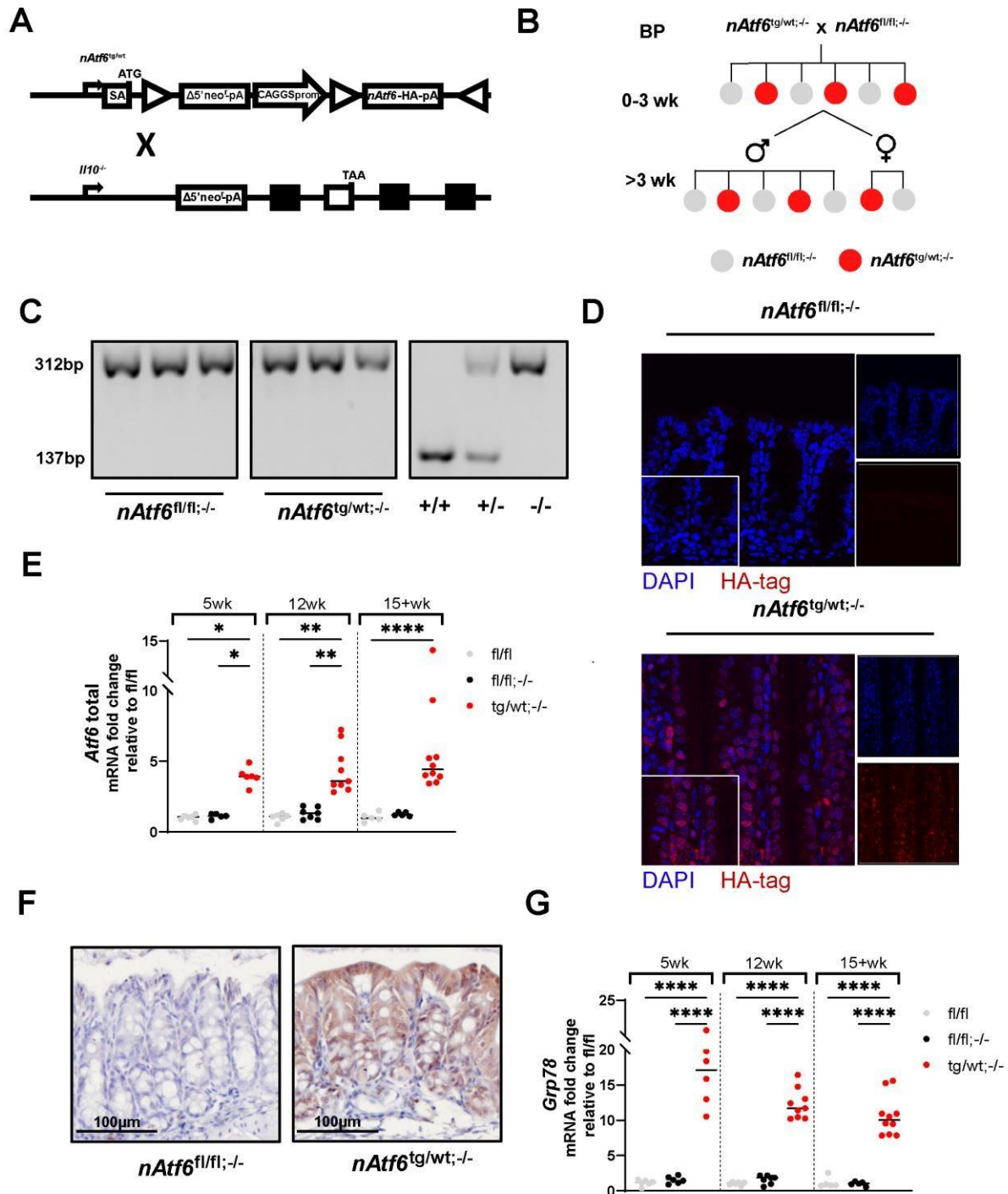


Figure 19. The *nAtf6^{IEC};Il10^{-/-}* mouse model. (A) Genomic organisation of the *nAtf6^{IEC};Il10^{-/-}* mouse model. *nAtf6* mice were generated using the highlighted targeting vector (upper) and crossed with *Il10^{-/-}* mice. **(B)** Example breeding scheme of the *nAtf6^{IEC};Il10^{-/-}* mouse line. **(C)** Confirmation of global *Il10* KO in both genotypes. **(D)** nATF6 transgene localisation in the colonic epithelium, detected by immunofluorescent staining of HA-tag. **(E)** *Atf6* (endogenous + total) mRNA expression in whole colonic tissue. **(F)** Immunohistochemical staining of GRP78 in *nAtf6^{tg/wt;-/-}* mice. **(G)** *Grp78* mRNA expression in whole colonic tissue.

We next performed phenotypic characterisation of *nAtf6^{IEC};Il10^{-/-}* mice, to assess whether *Il10* KO could lead to tumour formation. *nAtf6^{tg/wt;-/-}* mice showed very poor survival, with the majority of animals reaching abortion criteria before the age of 7 weeks, thus these mice were excluded from further experiments. Substantially reduced survival was also observed in

nAtf6^{tg/wt;-/-} mice, comparable to that of *nAtf6*^{tg/tg} mice. No reduced survival was observed in the *nAtf6*^{fl/fl;-/-} controls however, despite their predisposition to intestinal inflammation (**Figure 20A**)²²¹. With respect to tumorigenesis, the combination of monoallelic nATF6 activation and *Il10* KO clearly restored susceptibility to tumour formation in the proximal and mid-colon (**Figure 20B**). Susceptibility, however, was not restored to the level of the *nAtf6*^{tg/tg} mouse, with only approximately 70% of all mice developing tumours – termed Responders (R) - while the other animals – termed Non-responders (NR) - remained tumour-free (**Figure 20D and E**). In spite of the lack of tumour formation in NR mice, HA-tag staining of nATF6 and GRP78 staining intensity was comparable to tumour-bearing responder mice (**Figure 20C**). Tumour number did not differ between timepoints but was notably lower than the *nAtf6*^{tg/tg} reference (**Figure 20G and D**). To assess whether tumour formation was associated with enhanced proliferation, histological staining for Ki-67 was conducted. Regardless of tumour status, an increased presence of Ki-67+ IECs was observed in tumour-susceptible mice (**Figure 20F**). Taken together, these results suggest that *Il10* KO modulates the susceptibility of monoallelic *nAtf6*-overexpressing mice to colorectal tumorigenesis.

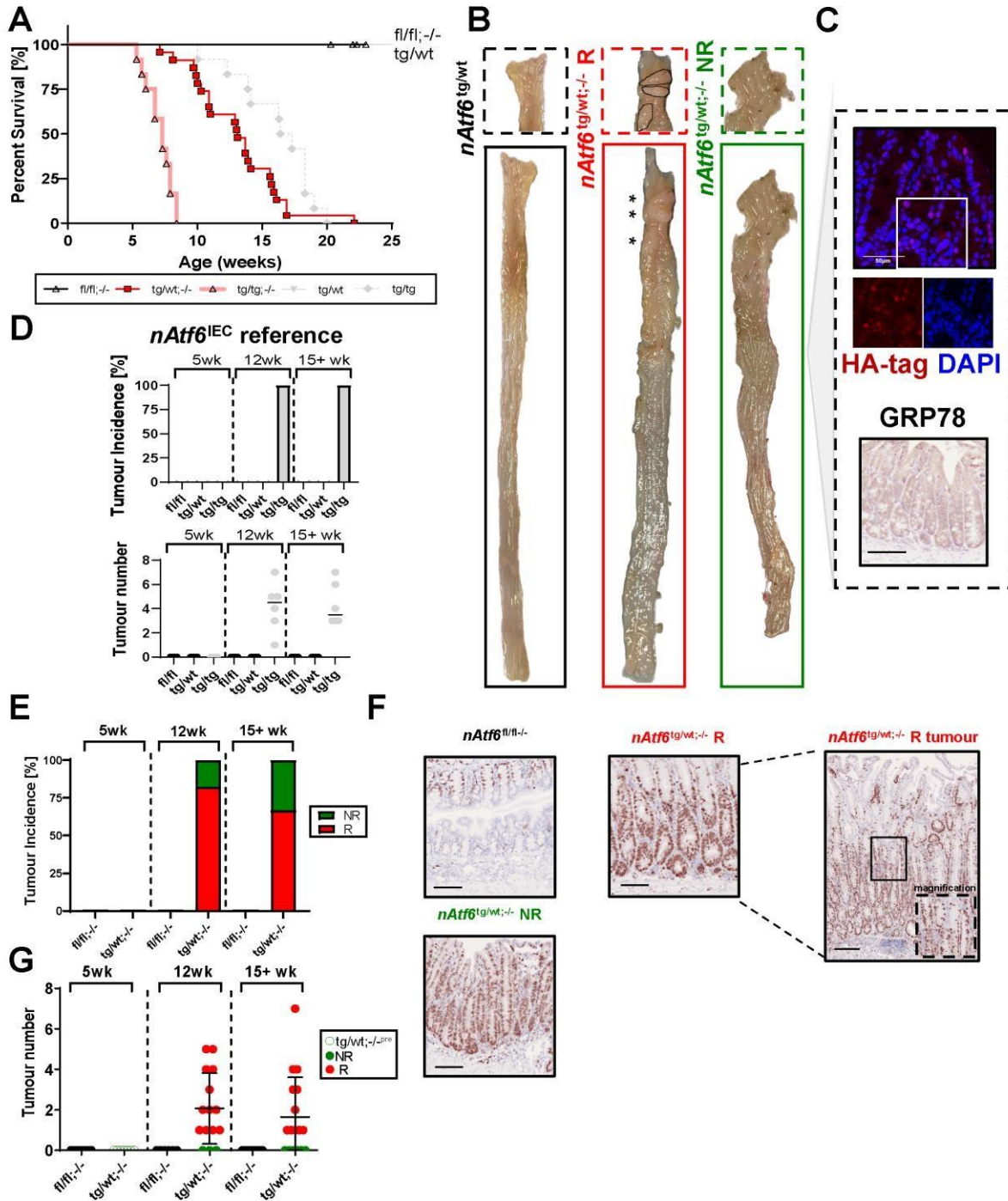


Figure 20. *I10* knockout enhances susceptibility to tumorigenesis but does not restore homozygous risk. (A) Kaplan-Meier survival curve of *nAtf6*^{fl/fl;-/-}, *nAtf6*^{tg/wt;-/-} and *nAtf6*^{tg/tg;-/-} genotypes in comparison to *nAtf6*^{IEC} mice. **(B)** Representative macroscopic images of the colon of tumour-free *nAtf6*^{tg/wt} mice, an *nAtf6*^{tg/wt;-/-} non-responder and an *nAtf6*^{tg/wt;-/-} responder (left to right) **(C)** Immunofluorescent nuclear HA-tag staining showing activation of nATF6 (top), and Immunohistochemical staining showing GRP78 activation (bottom) in NR *nAtf6*^{tg/wt;-/-} mice. **(D)** Percentage tumour incidence and number in *nAtf6*^{IEC} mice shown as a reference. **(E)** Percentage tumour incidence in *nAtf6*^{IEC}; *I10*^{-/-} mice. The percentage of non-responder (NR) mice is shown in green. **(F)** Representative IHC staining of the proliferation marker Ki67. **(G)** Tumour number in *nAtf6*^{IEC}; *I10*^{-/-} mice.

4.3.2 nATF6 activation enhances colitis in *nAtf6*^{tg/wt;-/-} mice.

Il10 KO disrupts the ability to suppress inflammation, pre-disposing mice to colitis and tumorigenesis, however mice on different backgrounds show varying degrees of severity, with C57BL/6 mice – the genetic background of *nAtf6*^{IEC};*Il10*^{-/-} mice - tending to develop a less severe phenotype²²². ATF6 on the other hand, has been shown to potentiate intestinal inflammation¹⁷². Here, we aimed to determine whether the combination of nATF6 activation and *Il10* KO led to increased intestinal inflammation. Gross measures of inflammatory status showed increased spleen weight and enhanced susceptibility to rectal prolapse in *nAtf6*^{tgwt;-/-} mice (**Figure 21A and B**). H&E staining and subsequent histological scoring revealed disruptions in tissue architecture and enhanced leukocyte infiltration compared to control (**Figure 21C and E**). NR also tended to show reduced inflammation compared to R and supporting this, tumour number correlated positively with inflammatory score (**Figure 21D**). On the molecular level, gene expression of the pro-inflammatory cytokine TNF α was upregulated at 12 and 15+ weeks. Gene expression of the chemokine IP-10 was also upregulated in *nAtf6*^{tgwt;-/-} mice but only reached significance at 12 weeks (**Figure 21F**). Other markers, such as IFN γ , IL-6, IL-12p35, IL-22 and CXCL1 were also measured, however expression of these cytokines could not be detected (data not shown). To confirm the presence of inflammation on the protein level in these mice, we measured Lipocalin-2 (LCN2) in faecal samples using ELISA. LCN2 is produced by IECs in response to inflammation, acting to limit bacterial iron acquisition and often used as a marker of inflammation in *Il10*^{-/-} mice^{223–225}. LCN2 was clearly enriched in *nAtf6*^{tgwt;-/-} mice with increased concentrations observed in both R and NR (**Figure 21G**). Taken together, these results suggest combined nATF6 activation and *Il10* KO leads to an increased tendency towards intestinal inflammation.

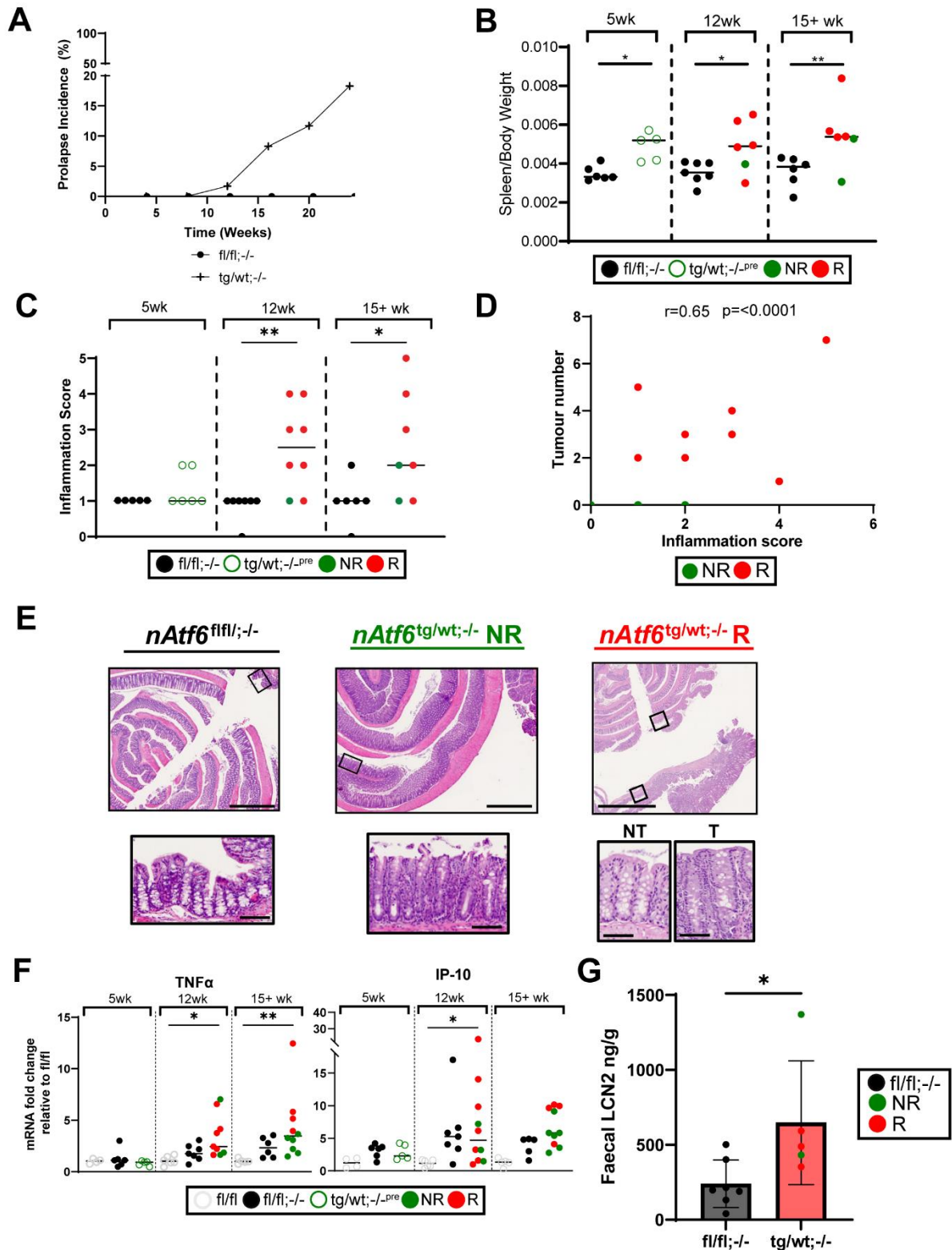


Figure 21. *nAtf6^{tg/wt;-/-}* mice display enhanced colitis compared to *nAtf6^{fl/fl;-/-}* mice. (A) Cumulative percentage prolapse incidence in *nAtf6^{IEC};Il10^{-/-}* mice. **(B)** Body-weight normalised spleen weight in *nAtf6^{IEC};Il10^{-/-}* mice. **(C)** Inflammation score in colonic Swiss roll sections. **(D)** Non-parametric Spearman's rank correlation of tumour number and inflammation score. **(E)** Representative H&E-stained sections of intestinal Swiss rolls. Left to right: *nAtf6^{fl/fl;-/-}*, *nAtf6^{tg/wt;-/-} NR* and *nAtf6^{tg/wt;-/-} R* at 2x magnification (Scale bar: 1mm). Zoomed-in sections at 20x magnification are shown below each image (Scale bar: 100 μ m). **(F)** Cytokine expression in whole colonic tissue sections. **(G)** Faecal Lipocalin2 (LCN2), as measured by ELISA.

Chronic inflammation and epithelial damage can affect secretory cell number and function in the intestinal epithelium, leading to disruption of the mucosal barrier. We previously showed biallelic activation of nATF6 in IECs led to reduced goblet cell number¹⁵⁵. To assess the combined impact of chronic inflammation and monoallelic nATF6 activation on goblet cells we performed PAS/AB staining of colonic Swiss rolls and quantified goblet cell (GC) numbers. GC numbers were noticeably reduced in the proximal colon of *nAtf6^{tg/wt;-/-}* mice, at the 12- and 15+-week timepoints but did not significantly differ at 5 weeks (**Figure 22A**). In the distal colon, the same pattern emerged, however the reduction in GC numbers was not as drastic (**Figure 22B**). To confirm impaired goblet cell function, we performed qPCR measurement of *Muc2*, which encodes the major gel-forming mucin in colonic mucus. Expression of *Muc2* was significantly reduced at all timepoints, confirming the altered GC phenotype (**Figure 22C**). GC number correlated negatively with histological score, suggesting colitis, rather than nATF6 activation was the major driver behind GC reduction in *nAtf6^{tg/wt;-/-}* mice (**Figure 22D**).

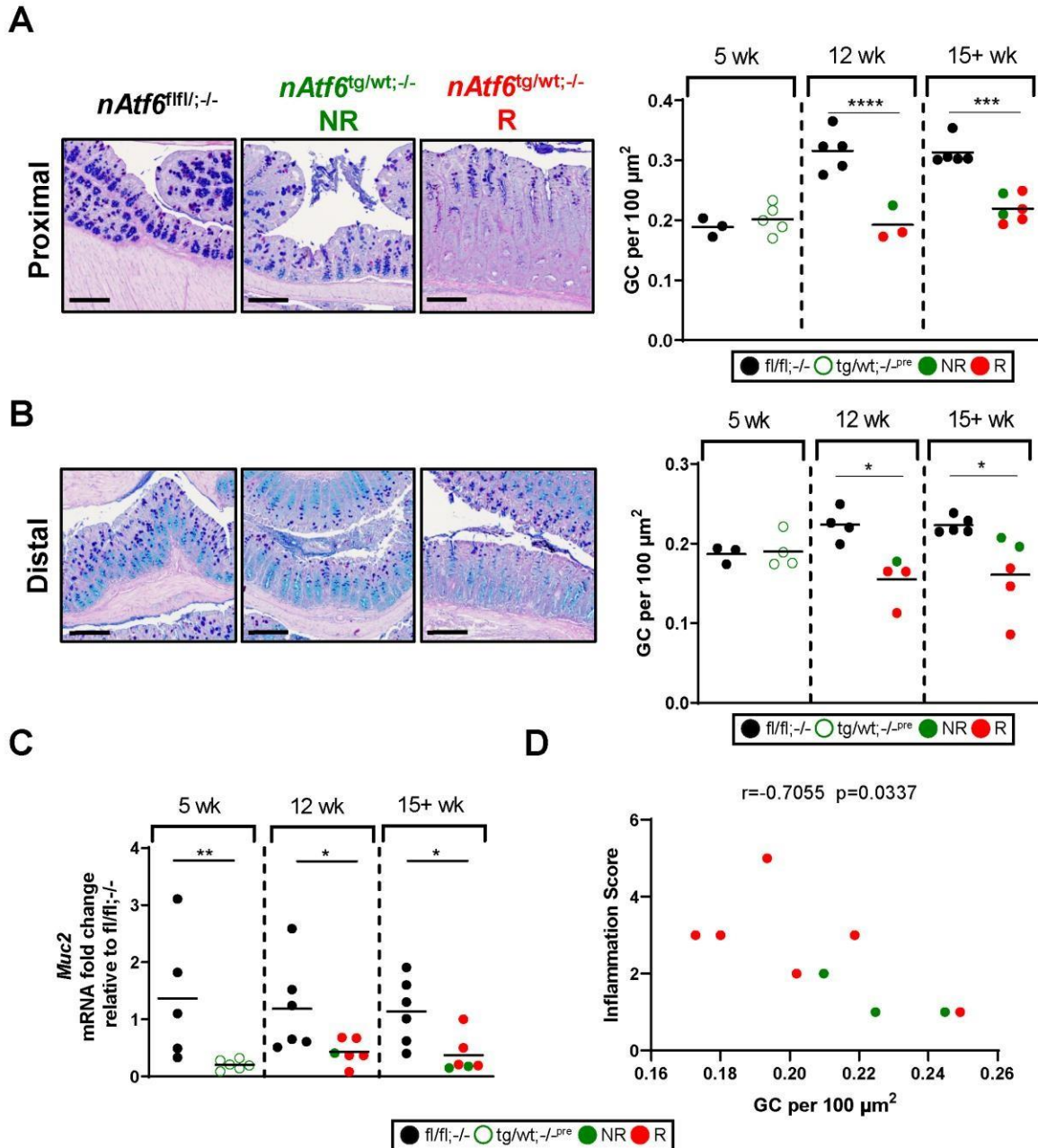


Figure 22. Goblet cell loss occurs after tumour-onset and is associated with inflammation. Representative PAS/AB staining of mucin-filled goblet cells (GC's) in the proximal (A) and (B) distal colon in $nAtf6^{fl/fl}; -/-$, $nAtf6^{tg/wt}; -/-$ NR and R mice (left to right). Scale bar: 100 μm . Dotplots (right) show the quantification of GCs per 100 μm^2 . (C) Expression of *Muc2* in whole colonic tissue extract. Data is shown relative to the $nAtf6^{fl/fl}; -/-$ control. (D) Non-parametric Spearman's rank correlation of inflammation score and GC per 100 μm^2 .

4.3.3 Tumorigenesis in $nAtf6^{tg/wt}; -/-$ mice is microbiota-driven

The intestinal mucus layer is important to limit contact between commensal bacteria and the epithelial layer. In the healthy murine colon, this is composed of two layers, a loose outer layer densely populated by microbiota and a thick inner layer mostly devoid of bacteria⁴⁷. To assess the impact of reduced goblet cell number on the thickness of the mucus layer and bacterial penetration, we performed combined Fluorescent *in situ* hybridisation (FISH with EUB338),

targeting bacterial 16S rRNA and IF staining of UEA1 lectin (Ulex Europaeus Lectin 1), to visualise mucus, on Carnoy fixed colonic tissue sections. Concordant with the reduction in goblet cell numbers, mucus thickness was reduced and bacterial penetration into the inner mucus layer was enhanced in *nAtf6^{tg/wt;-/-}* mice in the proximal colon. Additionally, the inner mucus layer appeared less structured and, in some cases, almost absent (**Figure 23A**). Interestingly, the same pattern was observed in the distal colon, however penetrating bacteria remained at a considerable distance from the epithelium and the reduction in mucus thickness was not as apparent as in proximal regions (**Figure 23B**). Taken together, these findings support microbial involvement in tumorigenesis.

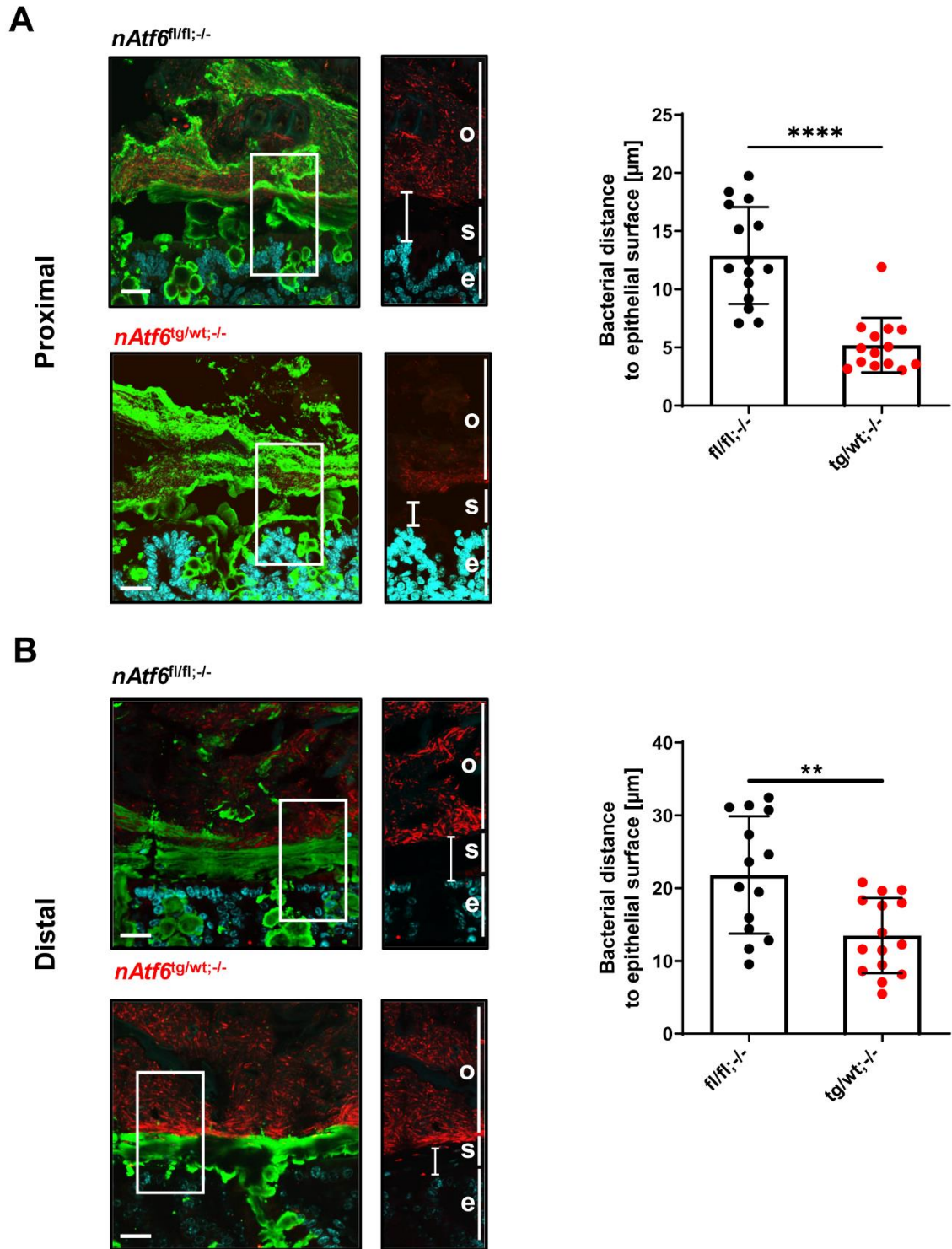


Figure 23. Enhanced mucus penetrability accompanies tumorigenesis. Fluorescent *in situ* hybridisation (FISH) targeting 16S rRNA (EUB338 probe) in combination with immunostaining of mucus (UEA1) in the proximal (A) and (B) distal colon in *nAtf6^{fl/fl}; -/-* and *nAtf6^{tg/wt}; -/-* mice (Scale bars 100 μm). Nuclei were counterstained with DAPI. Bacterial distance to the epithelium was quantified by measuring the distance to the epithelium in five different regions within the proximal colon for each mouse (right panel).

We previously showed that tumour development in the *nAtf6^{IEC}* mouse was dependent on the microbiota and that GF mice remain healthy¹⁵⁵. GF *nAtf6^{IEC}; Il10^{-/-}* mice were generated separately to this work to determine if tumorigenesis was also microbiota-dependent in this

model. GF *nAtf6*^{tg/wt;-/-} mice also did not develop tumours even after ~30 weeks of age (data not shown; Janine Kövilein, Unpublished doctoral thesis).

Since tumour formation was microbially driven, we sought to assess whether a dysbiotic microbiota could be observed in *nAtf6*^{tg/wt;-/-} mice. Caecal microbiota was collected at three endpoints, 5, 12 and 15+ weeks and profiled using 16S rRNA sequencing. To examine changes before and after tumour onset, we grouped 12 and 15-week+ samples into 12-week+ (post-tumour) and compared them to 5-week (pre-tumour). Analysis of beta diversity revealed clear separation between *nAtf6*^{tg/wt;-/-} and *nAtf6*^{fl/fl;-/-} controls, with significant separation already apparent before tumour-onset (p -value = 0.042) (**Figure 24A**). Alpha diversity analysis did not reveal any significant differences between genotypes except for Faith's PD which was decreased in 12-week+ samples (**Figure 24B**). To investigate individual taxa altered between conditions, LEfSe biomarker analysis was used at both 5-week and 12-week+ timepoints. Interestingly, abundance shifts were much more pronounced between genotypes at 5 weeks, suggesting caecal microbial dysbiosis precedes tumorigenesis in *nAtf6*^{tg/wt;-/-} mice and further implicating the microbiota in disease (**Figure 24C**).

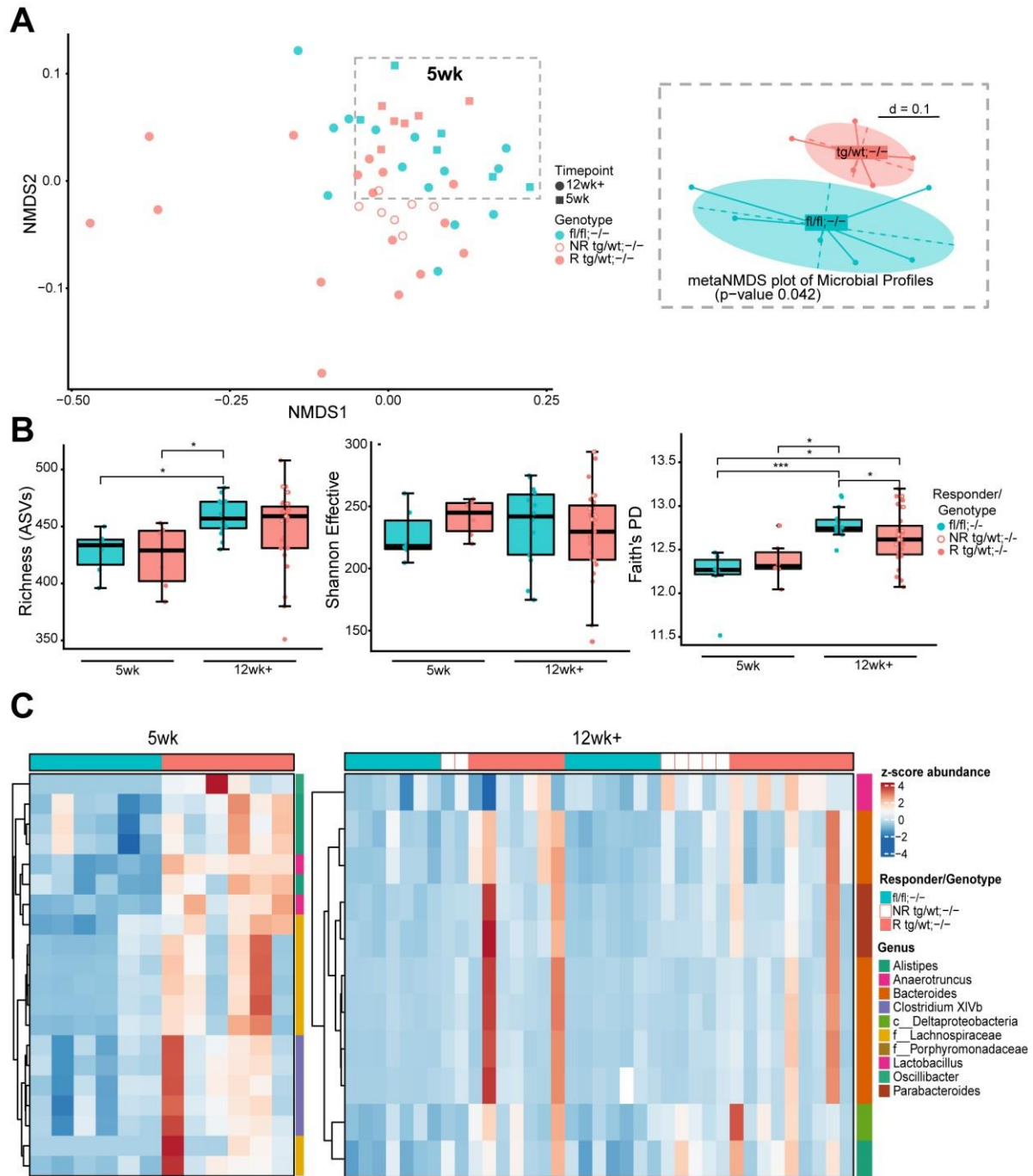


Figure 24. Caecal dysbiosis precedes tumorigenesis in *nAtf6*^{tg/wt;-/-} mice. (A) Beta diversity profiles of the microbiota in *nAtf6*^{IEC};*Il10*^{-/-} SPF mice, at 5-week (left) and 12-week+ (right) timepoints. (B) (Left to Right) Species Richness, Shannon Effective diversity, and Faith's phylogenetic diversity between 5-week and 12-week+ timepoints. (C) Heatmap of differentially abundant features between genotypes, showing 5-week (left) and 12-week+ (right) timepoints. Data are presented as z-scores of Log₁₀-transformed abundances.

4.4 Mucosa-associated dysbiosis in *nAtf6^{IEC};/I10^{-/-}* mice

4.4.1 Mucosa-associated microbiota profiles are altered regardless of tumour status in *nAtf6^{tg}/wt^{-/-}* mice.

Since *nAtf6^{tg}/wt^{-/-}* displayed many of the same phenotypic hallmarks as *nAtf6^{tg}/tg* mice, again clearly indicating involvement of mucosa-associated microbiota, we next characterised these communities in the inflammation-driven *nAtf6^{IEC};/I10^{-/-}* model. Taxonomic and phylogenetic measures of alpha diversity were significantly decreased at tumour-onset but not at 15 weeks+ (Figure 25A). Beta diversity analysis based on GUniFrac distance showed significant separation between tumour and non-tumour with microbial profiles from NR mice lying in between the two groups (Figure 25B). Differentially abundant ASVs were identified using LEfSe. Those belonging to the genera *Mucispirillum*, *Desulfovibrio* and *Bacteroides* were enriched in tumour-bearing mice, while Lachnospiraceae and *Roseburia* species were depleted, indicating tumour presence also alters microbial communities in the *nAtf6^{IEC};/I10^{-/-}* model (Figure 25C).

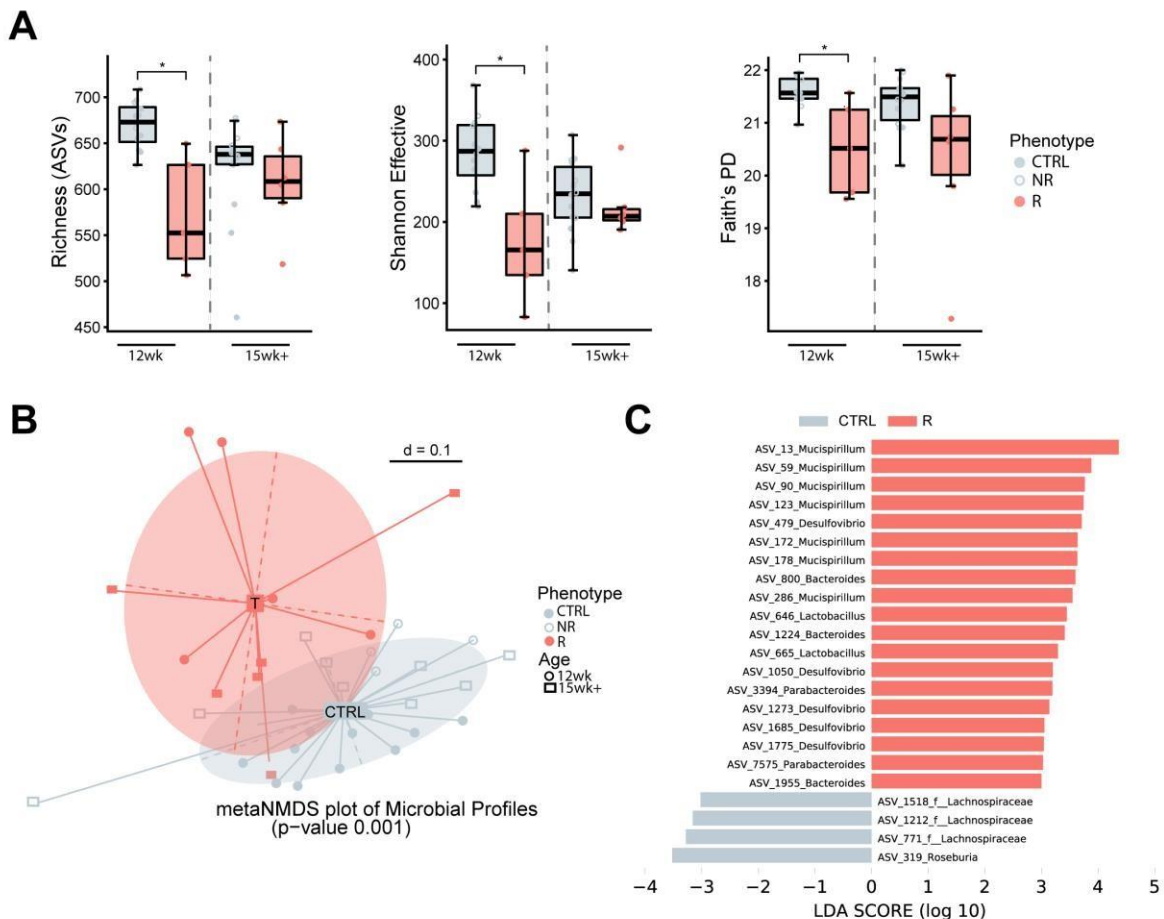


Figure 25. Mucosa-associated microbiota profiles are altered regardless of tumour status in *nAtf6^{tg}/wt^{-/-}* mice. (A) Mucosal alpha diversity in *nAtf6^{IEC};/I10^{-/-}* mice comparing tumour and non-tumour in 12-week and 15-week+ samples. Left to right: Richness, Shannon Effective and Faith's Phylogenetic diversity. **(B)** NMDS plot of beta diversity profiles based on generalized UniFrac distance in 12-week and 15-week+ plus samples. **(C)** Differentially abundant taxa between *nAtf6^{fl/fl};/-/-* and *nAtf6^{tg}/wt^{-/-}* mice, identified by LEfSe analysis. (LDA threshold = 3.0).

4.4.2 Mucosa-associated microbiota enables discrimination of microbial profiles between *nAtf6^{IEC};I10^{-/-}* mice

Since characterisation of the host phenotype did not reveal any clear differences between R and NR mice, we sought to assess microbial differences between these groups. Secondary to this, we also compared differences in caecal microbiota of R and NR to those in mucosal, to determine if this was indeed a better readout for identifying tumour-associated taxa. Significant differences in alpha diversity were observed across all metrics (Richness, Shannon Effective and Faith's PD) in mucosal but not caecal data, with the latter displaying considerable variability, particularly in responder mice (**Figure 26A and B**). Beta diversity analysis supported this, showing that caecal microbiota profiles of R from NR mice did not significantly differ, however mucosal did, even after random subsampling to an equal sample size. To ensure the robustness of this effect, we generated ten further random sub-samples of the mucosal dataset and calculated differences in GUniFrac distance using PERMANOVA. In each subset, a significant difference was observed confirming mucosal microbiota is altered across all samples (**Figure 26C and D**). Intriguingly, the same effect was also observed between *nAtf6^{fl/fl};I10^{-/-}* control mice and NR, indicating mucosal microbiota is altered even in the absence of tumours (**Figure 26E and F**). Together, these findings are suggestive of a tumour-relevant mucosa-associated microbial signature which is not present in luminal communities.

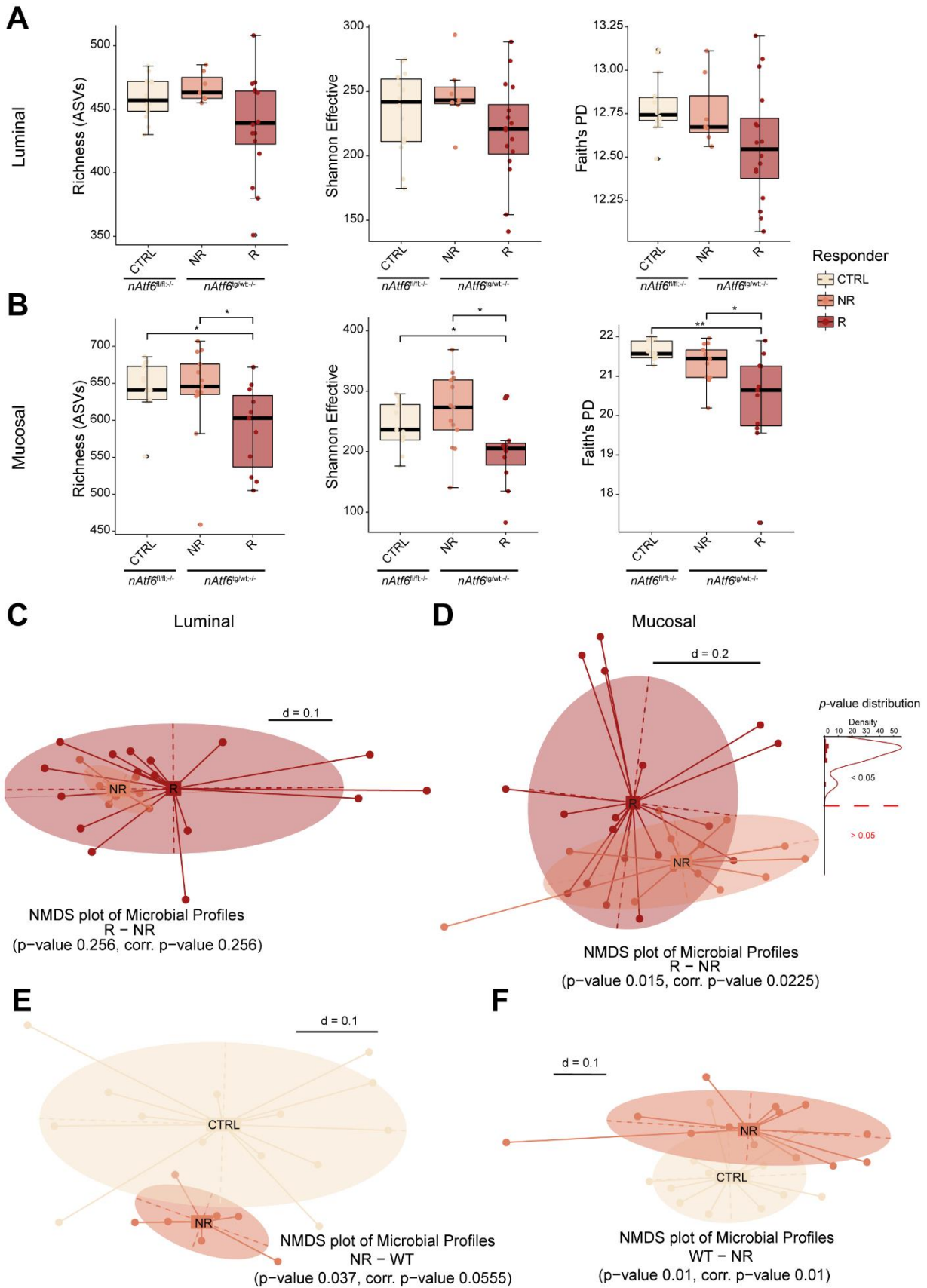


Figure 26. Mucosa-associated microbiota profiles separate responders from non-responders. (A) Luminal and (B) Mucosal alpha diversity in *nAtf6^{fl/-}* control mice, Non-responder (NR) and Responder (R) *nAtf6^{tg/wt}* mice. Left to right: Richness, Shannon Effective and Faith's phylogenetic

diversity. **(C)** NMDS plots of Luminal and **(D)** Mucosal beta diversity, between NR and R *nAtf6*^{tg/wt;-/-} mice based on generalized UniFrac distance. The distribution of PERMANOVA *p*-values from ten random subsamples of mucosal data is shown on the right. Luminal **(E)** and Mucosal **(F)** beta diversity between CTRL and NR mice.

To confirm that mucosal microbiota better represents taxa associated with disease than luminal, we trained various ML models to classify phenotype on both datasets separately, employing RF, L1L and RR models. Mucosal consistently outperformed luminal in each model tested, with mean AUC values of 0.83, 0.94 and 0.88 for the RF, L1L, and RR models respectively (**Figure 27A**). Precision and recall were also considerably higher in mucosal, indicating models trained on this data could accurately predict both classes (**Figure 27B**). Since we repeatedly performed cross-validation, we could extract AUC values for each model built and test whether the differences were statistically significant. Across all models, mucosal data was significantly better at predicting phenotype than luminal (**Figure 27C**). Surprisingly however, discriminating taxa shared between models were taxonomically similar between environments, with *Parabacteroides* species identified in both environments (**Figure 28A and B**). Taken together, these data point to an inherent tumour-associated microbial signature in mucosa-associated microbial profiles from *nAtf6*^{tg/wt;-/-} mice, which is less apparent, though still detectable in luminal communities.

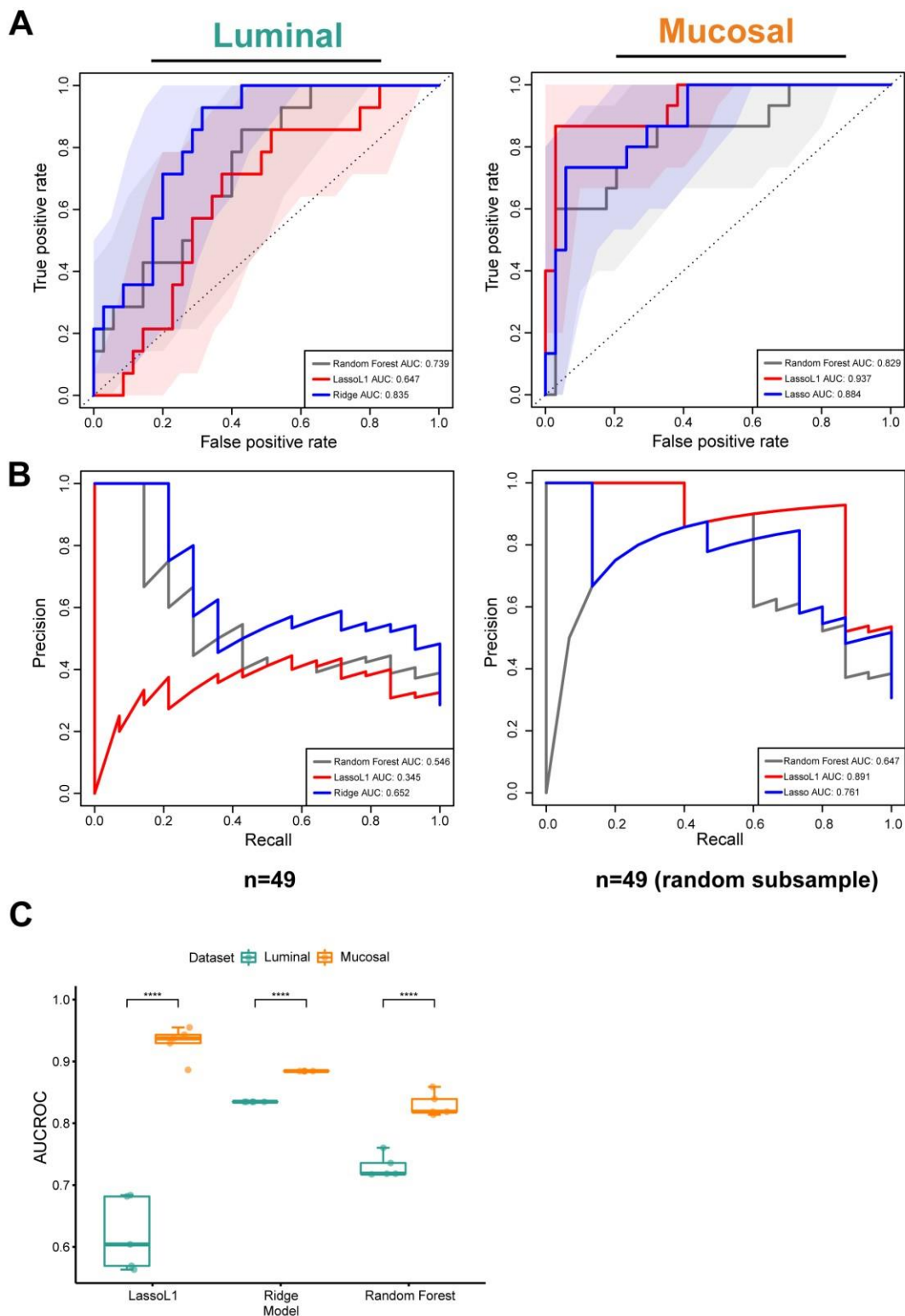


Figure 27. Machine learning models trained on mucosal microbiota profiles better discriminate phenotype compared to luminal. (A) ROC curves comparing true vs false positive rates of Random Forest, L1-penalised Lasso and Ridge regression models built on luminal and mucosal data, respectively. Mucosal data was randomly subsampled to match luminal sample size (original $n=57$, subsampled=49). Models were trained using repeated 5-fold cross-validation. **(B)** Precision-Recall curves for each of the models utilised. **(C)** Comparison of AUCROC values for each model trained, comparing luminal and mucosal datasets.

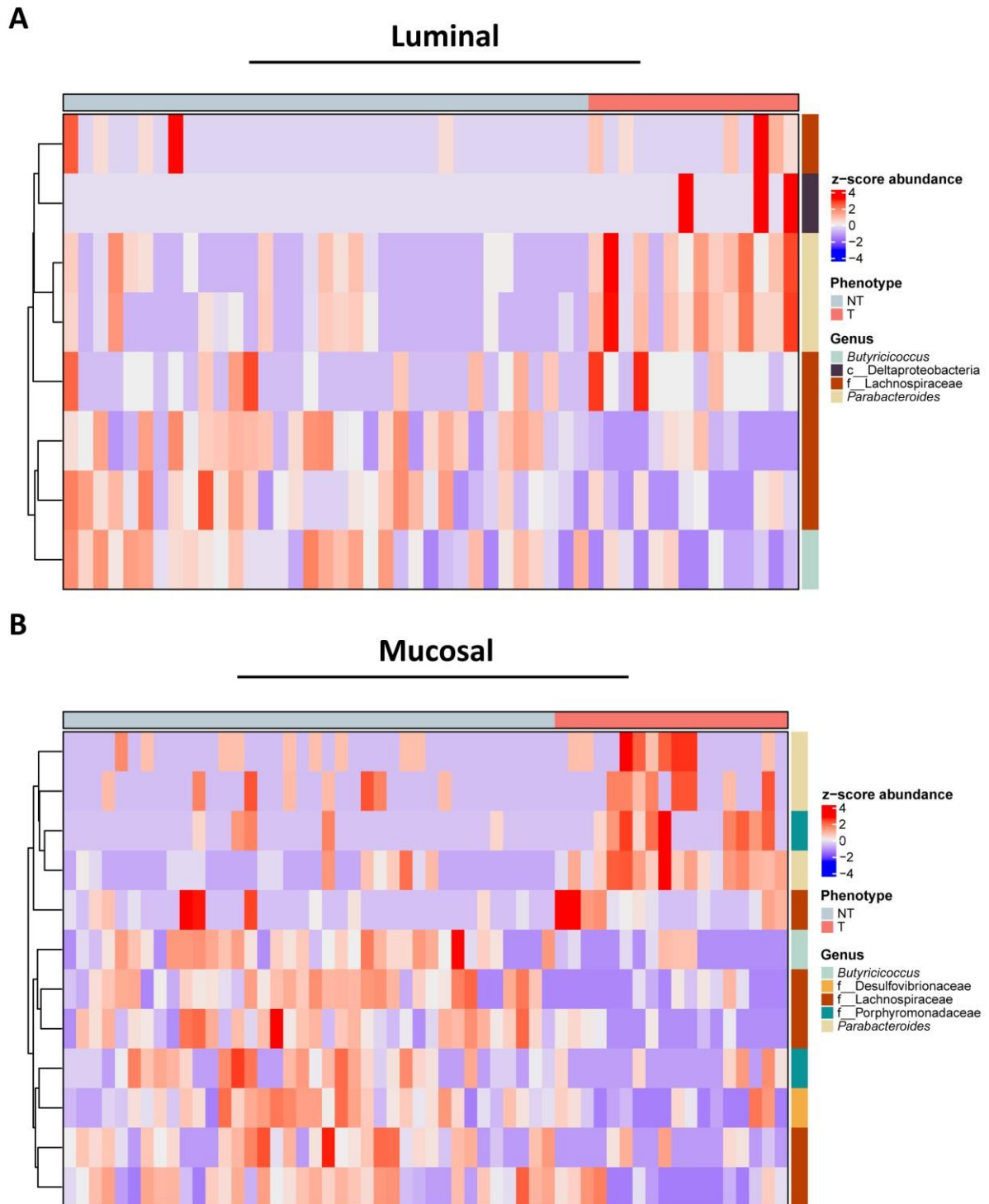


Figure 28. Consensus of discriminating features between non-tumour and tumour phenotypes across all trained models. (A) Luminal content and (B) Mucosal microbiota. Heatmap depicts z-score of Log_{10} transformed abundances.

4.4.4 Mucosa-associated populations do not shift significantly from pre- to post-tumour in *nAtf6^{tg/wt;-/-}* mice.

To date, little is known about the microbial changes which occur along the inflammation-dysplasia-carcinoma sequence. Additionally, considering the strong disruption mediated by chronic inflammation or *Il10* KO on the microbiota, it remains unclear whether the assumptions of theoretical models such as the driver-passenger model would still hold^{226–228}. Beta diversity analysis of microbial profiles between 5- and 12-week timepoints indicated microbial profiles clustered together, however were still significantly different (p -value = 0.034), indicative of minor changes between timepoints (**Figure 29A**). No passenger or driver taxa could be detected using a Log_2 FC threshold of 1.5 and FDR-corrected p -value of 0.05 between these timepoints, suggesting that in this model of CAC, major taxonomic shifts do not occur with tumour onset (**Figure 29B**). We therefore tested whether the additional group we identified in *nAtf6^{tg/tg}* mice, SD, were present. Applying this concept, we found several SD classified as Lachnospiraceae, *Lactobacillus*, *Mucispirillum* and *Bacteroides* in *nAtf6^{tg/wt;-/-}* mice (**Figure 29C**). No differences could be observed between tumour and tumour-adjacent samples, again confirming this approach could not be used to identify passengers and drivers (**Figure 29D**).

To test whether differences in the abundance of sustained-driver taxa might explain why NR mice do not develop tumours, we examined their abundance across phenotypes, comparing control, NR, tumour-adjacent and tumour mucosal samples. The majority of SD taxa, with the exception of Lachnospiraceae members, were clearly enriched in tumour and tumour-adjacent samples compared to control and NR mucosa, suggesting they may be involved in tumour formation (**Figure 29E**).

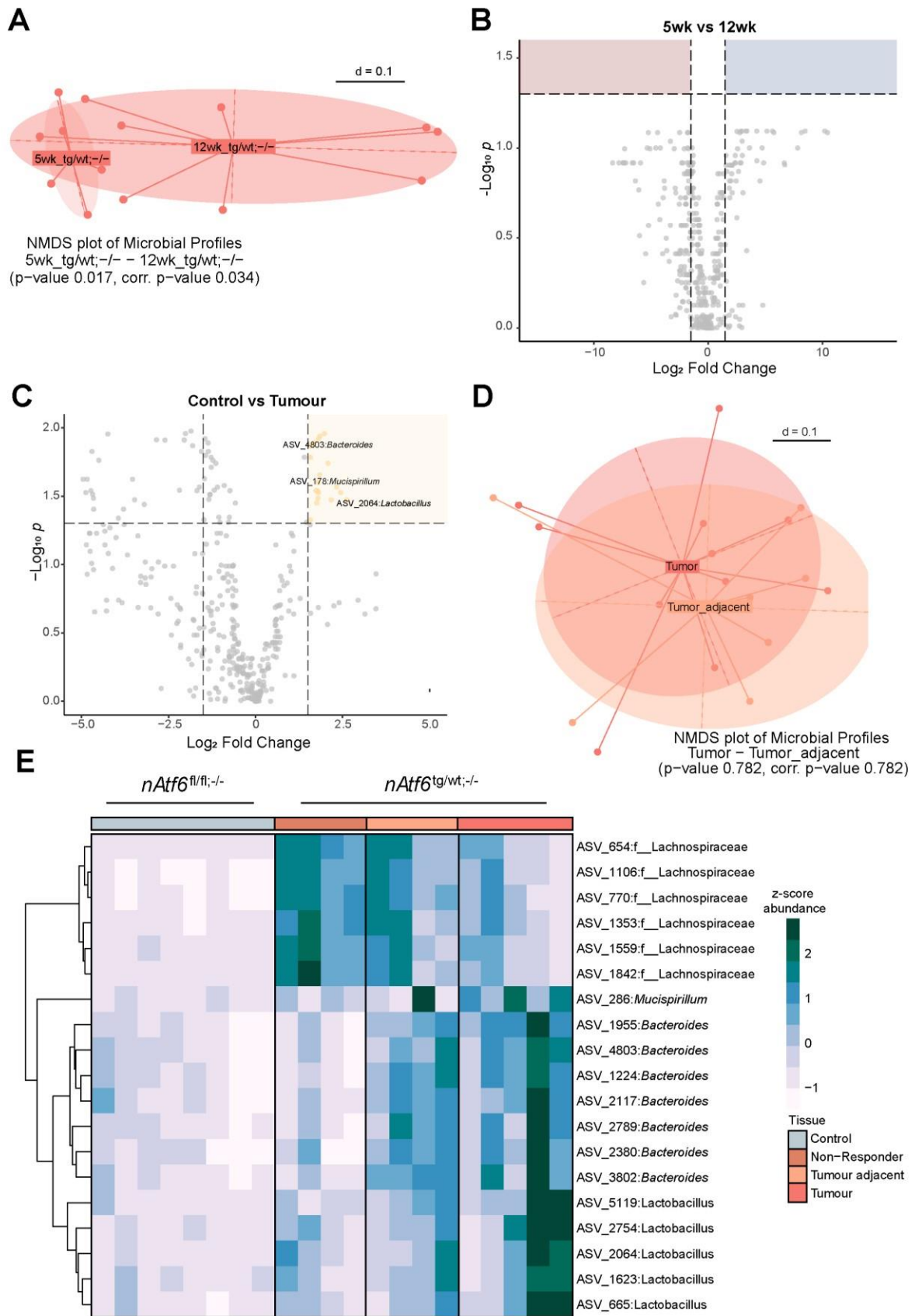


Figure 29. Mucosal microbiota shifts from pre- to post-tumour are not consistent with the driver-passenger model in *nAtf6*^{tg/wt;-/-} mice. (A) NMDS plot of beta diversity between 5 and 12-week timepoints. (B) Volcano plot of differential ASVs between pre-tumour and tumour-onset. Coloured squares highlight significant taxa with an FDR-corrected p -value <0.05. Red represents putative drivers and blue

putative passengers. **(C)** Volcano plot of differential ASVs between tumour and non-tumour which did not differ between 5-week and 12-week timepoints but were enriched in *nAtf6*^{tg/wt;-/-} mice compared to *nAtf6*^{fl/fl;-/-} controls. The yellow square in the upper right quadrant highlights significant taxa with an FDR-corrected p -value <0.05, termed sustained drivers. **(D)** Beta diversity plot of tumour and tumour-adjacent microbial profiles based on generalized UniFrac distance. **(E)** Heatmap of z-score transformed Log₁₀ abundances of sustained drivers, comparing genotypes and mucosal phenotypes.

4.5 Spatial organisation of mucosa-associated microbiota in *nAtf6*^{IEC};*Il10*^{-/-} mice

4.5.1 Spatial community structure of microbial communities does not significantly differ between *nAtf6*^{fl/fl;-/-} and *nAtf6*^{tg/wt;-/-} mice

Having characterised the mucosa and tumour-associated microbiota in *nAtf6*^{IEC};*Il10*^{-/-} mice, we next sought to better understand spatial patterns in tumour occurrence and microbial composition. We utilised the same spatial sampling design as used in the *nAtf6*^{IEC} model and again recorded tumour incidence by site. In *nAtf6*^{tg/wt;-/-} mice, tumours were concentrated in extreme proximal regions, with all responder mice developing tumours in sites 1 and 2. Tumours did not occur in the distal portion of the colon (**Figure 30A**). To gain insight into how spatial differences in tumour incidence might impact the microbiota, we calculated the mean Shannon Effective diversity at each site in each genotype and phenotype. Limited differences were observed between the *nAtf6*^{fl/fl;-/-}, *nAtf6*^{tg/wt;-/-} NR and *nAtf6*^{tg/wt;-/-} R mice, however a decrease in diversity in the distal portion of the colon was apparent in *nAtf6*^{fl/fl;-/-} as well as *nAtf6*^{tg/wt;-/-} R mice, with NR displaying the opposite trend (**Figure 30B**). Accordingly, family-level microbial compositions were generally very similar, with the exception of a notable expansion of Lactobacillaceae and Deferribacteraceae in *nAtf6*^{tg/wt;-/-} mice (**Figure 30C**).

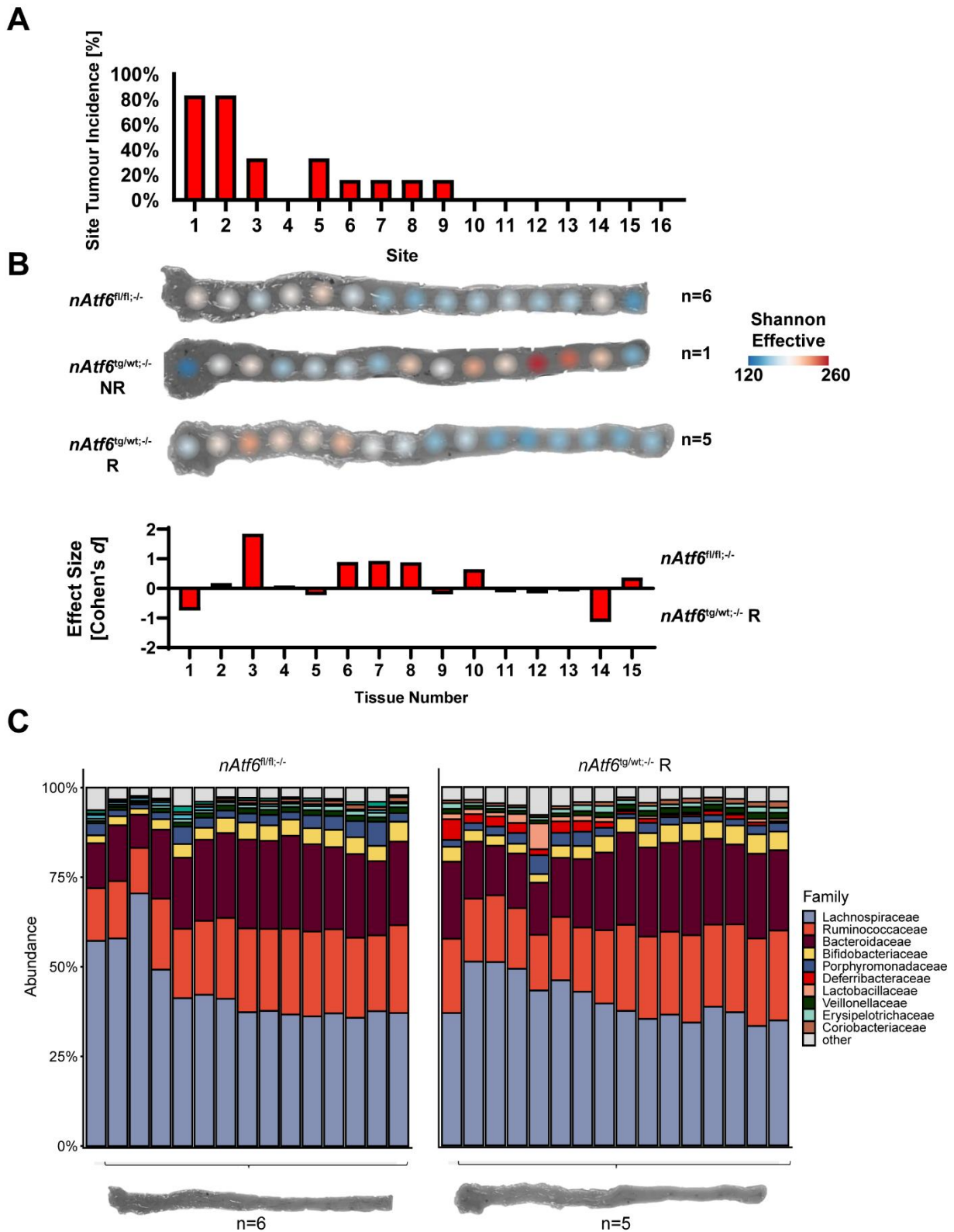


Figure 30. Mucosal microbial profiles show limited community level shifts between different sites in $nAtf6^{IEC}; Il10^{-/-}$ mice. (A) Tumour incidence by site from proximal to distal. **(B)** Mean Shannon Effective diversity at each site along the colon in $nAtf6^{fl/fl}; -/-$ (top), $nAtf6^{tg/wt}; -/-$ NR (middle) and $nAtf6^{tg/wt}; -/-$ R mice (bottom). The magnitude of change between $nAtf6^{tg/wt}; -/-$ R and $nAtf6^{fl/fl}; -/-$ controls (Cohen's d effect size) is displayed beneath. **(C)** Family level bacterial composition at each site in $nAtf6^{fl/fl}; -/-$ (left) and $nAtf6^{tg/wt}; -/-$ R mice (right).

As site-by-site differences at the community level were minimal, we explored whether differences may exist at the ASV level. To this end, we explored ASV differences between genotypes at each site using LEfSe. Since there was only a single NR mouse, these samples were excluded from further analysis. Changes in individual taxa were patchy in control sites with most changes occurring in the proximal colon. These sites were consistently enriched in Lachnospiraceae and further down the colon, *Bacteroides* ASVs. In *nAtf6*^{tg/wt;-/-} R mice a similar pattern was observed with most changes occurring in sites one to nine, where tumours occur. At these sites, enriched ASVs were mostly classified as *Mucispirillum*, *Lactobacillus* and Lachnospiraceae. Minimal or no differences were identified at many sites however, particularly in the distal portions of the colon, indicating changes in taxonomic composition occur at specific sites along the length of the colon (**Figure 31**).

A

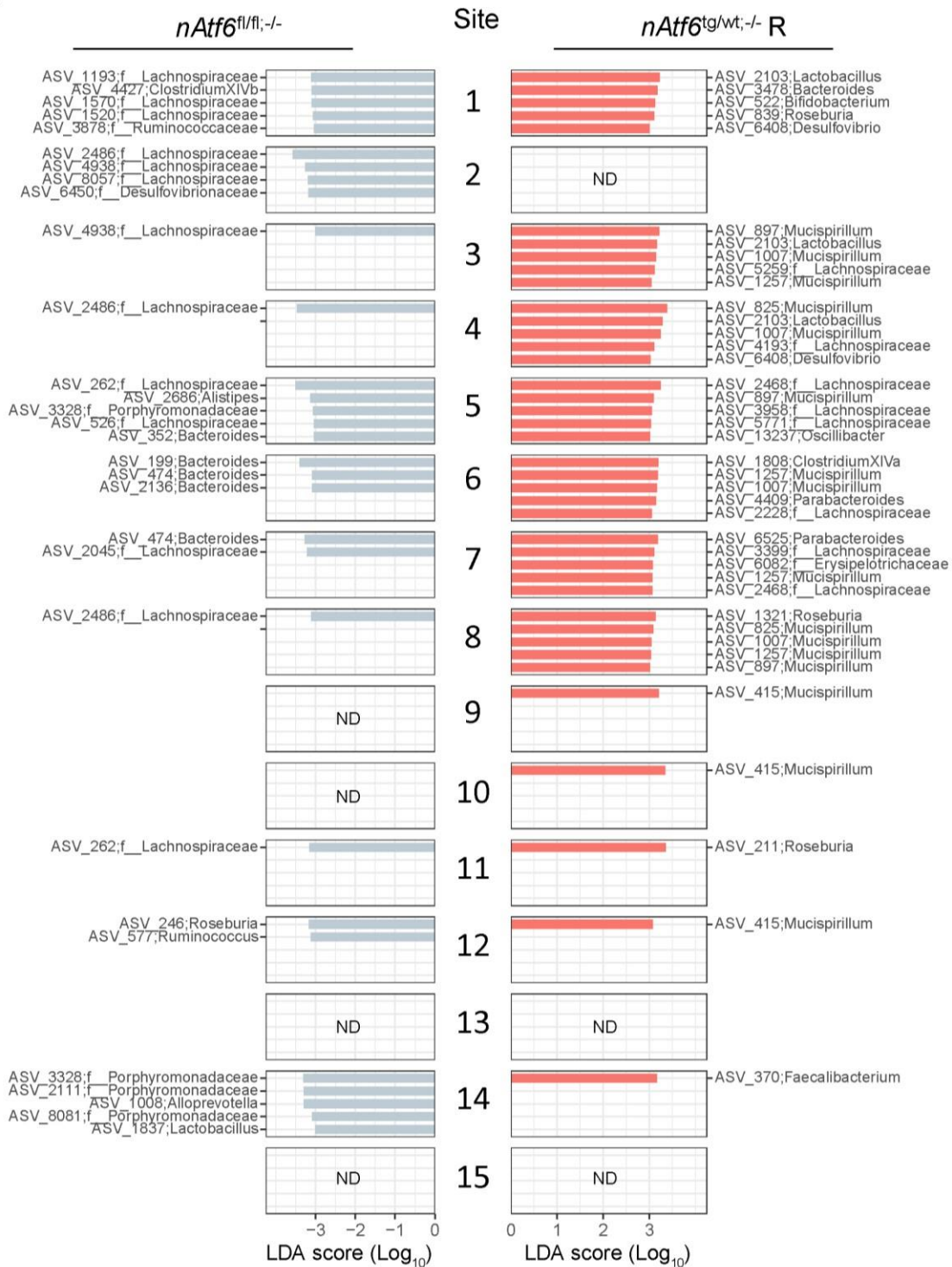


Figure 31. Differentially abundant taxa are mainly restricted to tumour-susceptible sites. Site-by-site comparison of the top five differentially abundant taxa between *nAtf6^{fl/fl};-/-* and *nAtf6^{tg/wt};-/-* responder mice as determined by LefSe analysis, ranked by LDA score (LDA threshold = 3.0). ND = not detected, indicating sites where no differentially abundant ASVs in a given phenotype were identified.

4.5.2 Mucosa-associated microbiota alterations are restricted to tumour-susceptible sites

Since tumour localisation along the colon differs between animals, we next explored whether stratifying samples accordingly would reveal further microbial shifts. We grouped mucosa samples into control, comprising samples from *nAtf6^{fl/fl};-/-* mice and *nAtf6^{tg/wt};-/-* samples into NT, TA and T and calculated alpha and beta diversity as well as microbial load. Surprisingly, Shannon Effective diversity was significantly increased in TA samples, compared to control and NT mucosa, however no other differences or trends among alpha diversity metrics were observed (**Figure 32A**). Microbial load was also similar between groups, with the exception of NT mucosa which showed significantly reduced 16S copy number compared to control mucosa (**Figure 32B**). In line with the limited differences along the length of the colon between *nAtf6^{fl/fl};-/-* and *nAtf6^{tg/wt};-/-* R, we observed no clear difference in beta diversity between genotypes, suggesting monoallelic nATF6 activation may not significantly alter microbial community structure in *Il10* KO mice. Within *nAtf6^{tg/wt};-/-* R mice, however, considerable variability was observed between tissues, particularly between T/TA and NT sites, indicating that although spatial microbial changes are limited between genotypes, tumour-relevant alterations may occur within the colon of *nAtf6^{tg/wt};-/-* mice (**Figure 32C**). Since many of the limited microbial changes that were observed, tended to be more apparent in the proximal colon, we reasoned that shifts in mucosa-associated microbiota may preferentially occur in the tumour-susceptible proximal region as opposed to the tumour-free distal (**Figure 32D**). Supporting this, tumour-susceptible regions were significantly different to non-tumour regions in terms of beta diversity (**Figure 32E**). Furthermore, tumour-susceptible regions also demonstrated increased GUniFrac dissimilarity to control, indicating changes at these sites are more pronounced (**Figure 32F**). To assess which ASVs were enriched in tumour-susceptible sites, we compared samples from these locations to *nAtf6^{fl/fl};-/-* controls. Multiple ASVs were identified as enriched in tumour-susceptible sites compared to controls. Intriguingly, the majority of these ASVs were classified as *Mucispirillum*, *Lactobacillus* and *Bacteroides*, taxonomically resembling SD taxa (**Figure 32G**). Taken together, this indicates microbial changes in *nAtf6^{tg/wt};-/-* R mice do not occur along the length of the colon but rather are focused in proximal tumour-susceptible regions.

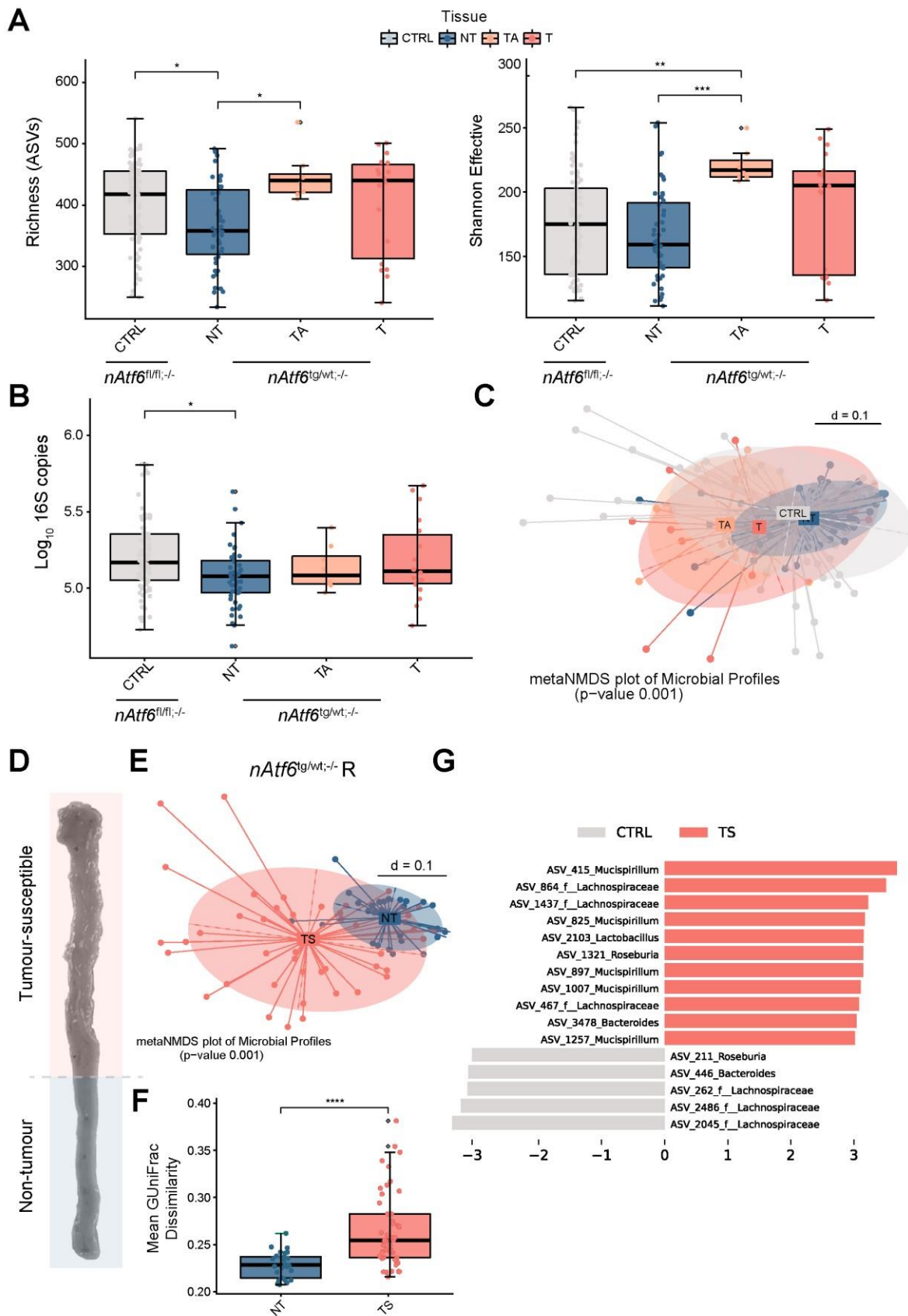


Figure 32. Alterations in mucosa-associated microbiota in *nAtf6^{tg/wt};-/-* responder mice are restricted to tumour-susceptible sites. (A) Mucosal alpha diversity in *nAtf6^{IEC};Il10^{-/-}* mice across various mucosal phenotypes. Left to right: Richness and Shannon Effective. **(B)** Total 16S rRNA copy number as estimated by artificial spike-in DNA. Presented as Log₁₀ copy number. **(C)** NMDS plot of beta

diversity, based on generalized UniFrac distance comparing mucosa between genotypes and within $nAtf6^{tg/wt;-/-}$ responder mice. **(D)** Representative macroscopic image of a longitudinally excised colon from an $nAtf6^{tg/wt;-/-}$ mouse depicting tumour-susceptible and non-tumour regions. **(E)** NMDS plot of beta diversity, based on generalized UniFrac distance comparing tumour-susceptible (TS) to non-tumour mucosa within $nAtf6^{tg/wt;-/-}$ responder mice, **(F)** Mean generalized UniFrac dissimilarity to control between tumour-susceptible (TS) and non-tumour sites in $nAtf6^{tg/wt;-/-}$ responder mice. **(G)** Differentially abundant taxa between tumour-susceptible sites and controls as determined by LEfSe analysis, ranked by LDA score (LDA threshold = 3.0).

4.5.3 Driver taxa are robust across cohorts and localise to tumour-susceptible sites

Since several of the ASVs enriched in $nAtf6^{tg/wt;-/-}$ R mice compared to $nAtf6^{fl/fl;-/-}$ taxonomically resembled those identified as SD, we aimed to test whether the taxa we previously identified were detectable in a separate cohort of mice. SD sequences from the mucosal dataset were aligned against those from the spatial cohort using BLAST. Importantly, we were able to detect all 17 sequences, while analysis of SD load (16S copy number) and distribution revealed a significant enrichment in $nAtf6^{tg/wt;-/-}$ R mice, predominantly in the proximal portion of the colon (**Figure 33A and B**). Accordingly, the total 16S copy number of SD taxa was significantly higher in the tumour-susceptible half of the colon compared to the non-tumour region (**Figure 33C**), indicating these taxa are more abundant where tumours occur. Next, to analyse the individual spatial distributions of SD ASVs, we mapped the total 16S copy number by site. Since several *Mucispirillum*, a *Bacteroides* and a *Lactobacillus* ASV were enriched in tumour-susceptible sites, we focused on these taxa. As there is only a single host-associated species in the genus *Mucispirillum*, and several ASVs may derive from the same organism we used the ASV with the highest LDA score as a representative for all ASVs classified *Mucispirillum*²²⁹. In all cases, mean abundance of SD taxa appeared increased in $nAtf6^{tg/wt;-/-}$ R mice compared to $nAtf6^{fl/fl;-/-}$ mice and mainly localised to tumour-susceptible sites. Statistical comparison indicated the *Mucispirillum* ASV was significantly increased in most of the proximal colon, while the *Bacteroides* SD ASV was only significantly enriched at a single site. Surprisingly, the *Lactobacillus* ASV did not reach significance at any site, however, an increased trend was observed in tumour samples. Taken together, these data indicate SD taxa are enriched in tumour-susceptible regions and localise to sites where tumours occur thus confirming their status as SD and further implicating them in tumorigenesis (**Figure 33D**).

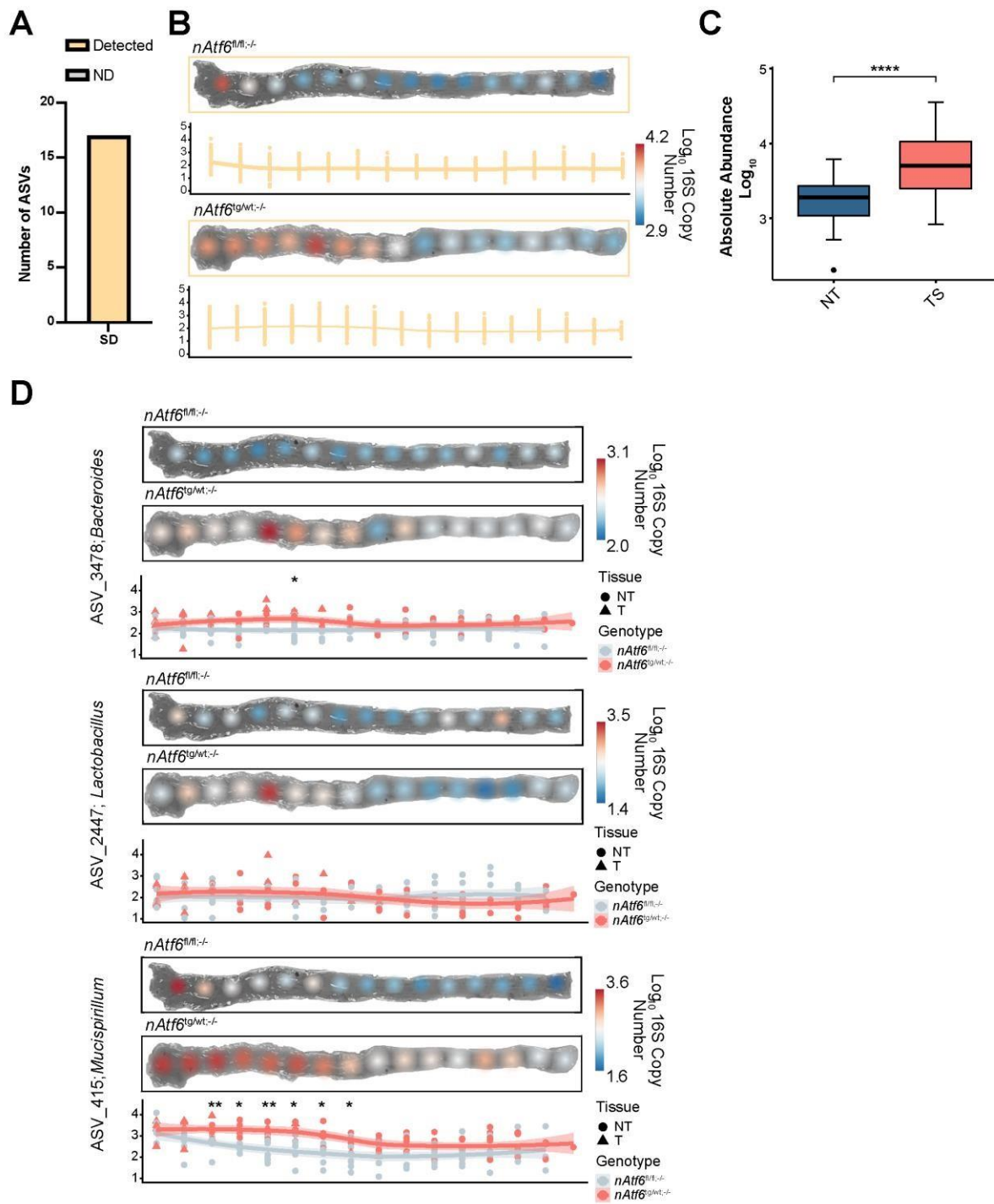


Figure 33. Sustained driver ASVs localise to tumour-susceptible sites. (A) Number of detected sustained-driver ASVs based on BLAST alignment. **(B)** Log₁₀ absolute abundance of sustained driver taxa along the length of the colon in *nAtf6^{fl/fl;-/-}* and *nAtf6^{tg/wt;-/-}* responder mice, with distribution among *nAtf6^{tg/wt;-/-}* responders depicted underneath. Trends are visualised using LOESS regression. **(C)** 16S copy number of sustained drivers comparing tumour-susceptible colonic regions to non-susceptible. **(D)** Spatial maps of Log₁₀ transformed absolute abundance of the three most abundant sustained driver ASVs in each genus between genotypes. The distribution amongst all mice is shown underneath, using LOESS regression to visualise trends. Individual *nAtf6^{fl/fl;-/-}* samples are shown in grey, while *nAtf6^{tg/wt;-/-}* responder samples are shown in pink.

4.6 Metabolomics links tumour-associated taxa to the nATF6 metabolite environment

4.6.1 nATF6 activation and associated tumour formation alters the intestinal lipid environment

nATF6 activation clearly played an important role in modulating microbial composition, however the mechanism underlying this remained unclear. Since intestinal inflammation is not observed prior to tumour formation in *nAtf6^{tg/tg}* mice we examined changes in metabolism which might explain the observed changes in microbial composition. Aberrant ATF6 signalling has been linked to metabolic alterations, particularly lipid metabolism via SREBP2¹⁶². To test the effect of nATF6 activation on the metabolome, we performed untargeted metabolomics on caecal content using LC-TOF-MS. Focusing on alterations driven by nATF6 itself, rather than that of the tumour, we compared metabolite profiles between *nAtf6^{fl/fl}* and *nAtf6^{tg/wt}* mice to assess the impact of monoallelic nATF6 activation alone on the metabolome. Principal component analysis of metabolite profiles from *nAtf6^{fl/fl}* controls and *nAtf6^{tg/wt}* mice revealed significant alteration of the caecal metabolome in *nAtf6^{tg/wt}* mice at all timepoints (**Figure 34A**). We next examined the individual metabolites underlying these differences. Differential features were determined by calculating fold change and testing for statistical significance using pairwise *t*-tests. After adjusting for multiple comparisons, differentially enriched nATF6 regulated metabolites mostly comprised lipid species such as long-chain fatty acids (LCFA), phospholipids and sphingolipids (**Figure 34B**). To examine these changes in greater detail, we focused on two of the most significantly regulated annotated features: an unknown C20 hydroxy fatty acid and Lysophosphatidylcholine O-18:1. Compared to controls, these metabolites were enriched at all timepoints supporting a role for nATF6 activation in altered lipid metabolism (**Figure 34C**).

To understand how tumour presence alters the metabolite environment in *nAtf6^{IEC}* mice, we next explored metabolite changes in the lumen, between phenotypes. Principal component analysis revealed a drastic difference in caecal metabolome between tumour-bearing mice and non-tumour mice (**Figure 35A**). Differential enrichment analysis demonstrated a depletion of peptide metabolites and sphingolipids in *nAtf6^{tg/tg}* mice at tumour timepoints, while LCFA and other complex lipid species were enriched. Both of the highly enriched metabolites (C20 hydroxy fatty acid and LPC O-18:1) in *nAtf6^{IEC}* mice were also found in tumour mice further suggesting nATF6 may drive changes in these metabolites (**Figure 35B**). Thus, to delineate the metabolites driven by nATF6 activation and those associated with tumour formation, we generated a heatmap of differential metabolites. Hierarchical clustering revealed three metabolite clusters, a non-tumour enriched group mostly comprising dipeptides and sphingolipids, an nATF6-driven group made up of lysophospholipid/phospholipid and sphingolipid metabolites. Finally, the remaining cluster appeared to be specific to *nAtf6^{tg/tg}* mice and mostly comprised lysophospholipids, and LCFA (**Figure 35C**).

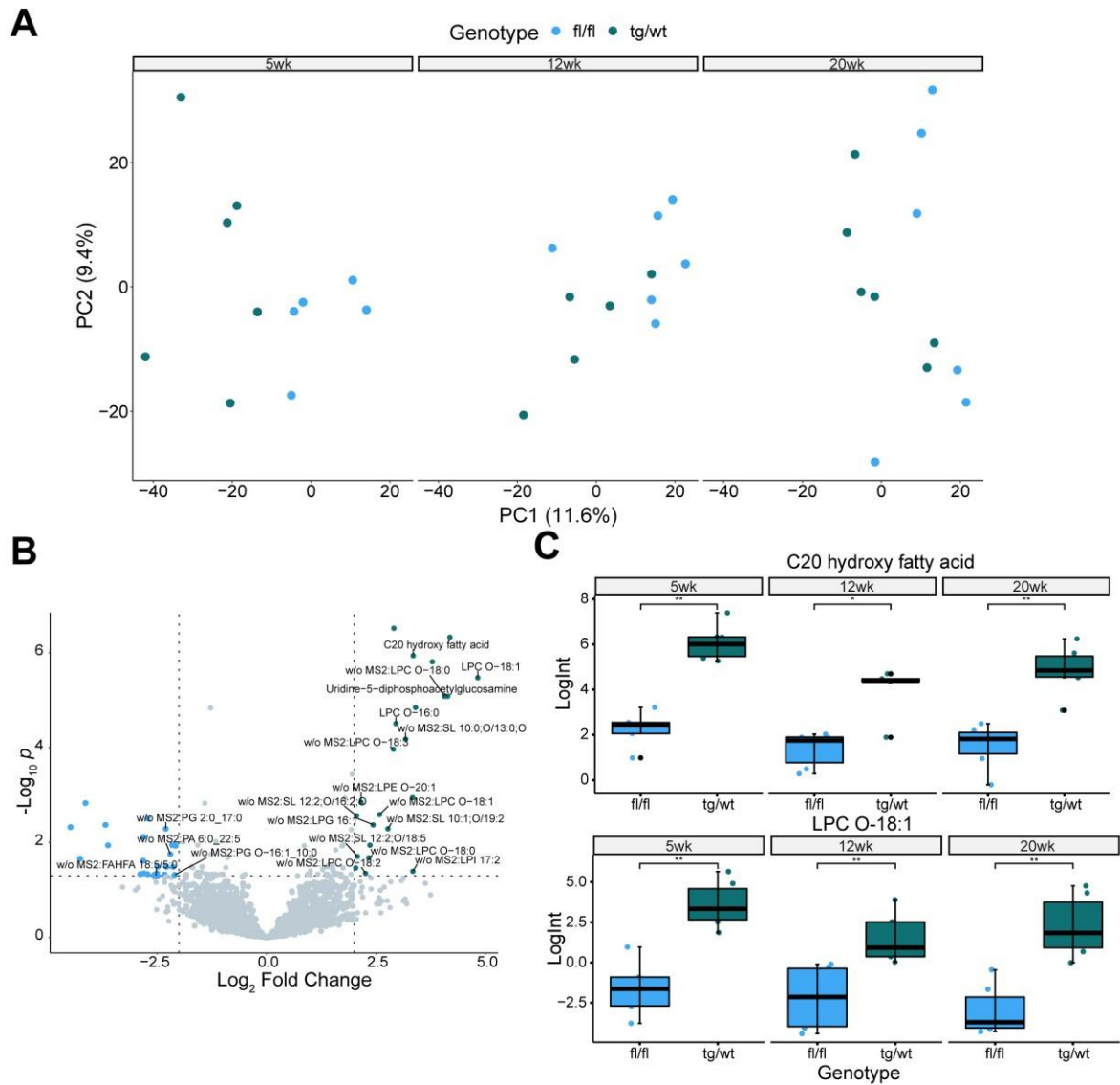


Figure 34. nATF6 activation alters the luminal metabolome, independent of tumour formation. (A) PCA of metabolite profiles comparing *nAtf6*^{fl/fl} controls (blue) to monoallelic ATF6 overexpressing *nAtf6*^{tg/wt} mice (green), at each timepoint. **(B)** Volcano plot of differentially enriched metabolites between *nAtf6*^{fl/fl} and *nAtf6*^{tg/wt} mice across all timepoints (Log_2 fold change threshold = 0.58, p -value threshold = 0.05). **(C)** Boxplots of select highly enriched metabolites in nATF6 overexpressing mice. Data are presented as normalised log-transformed feature intensity (LogInt).

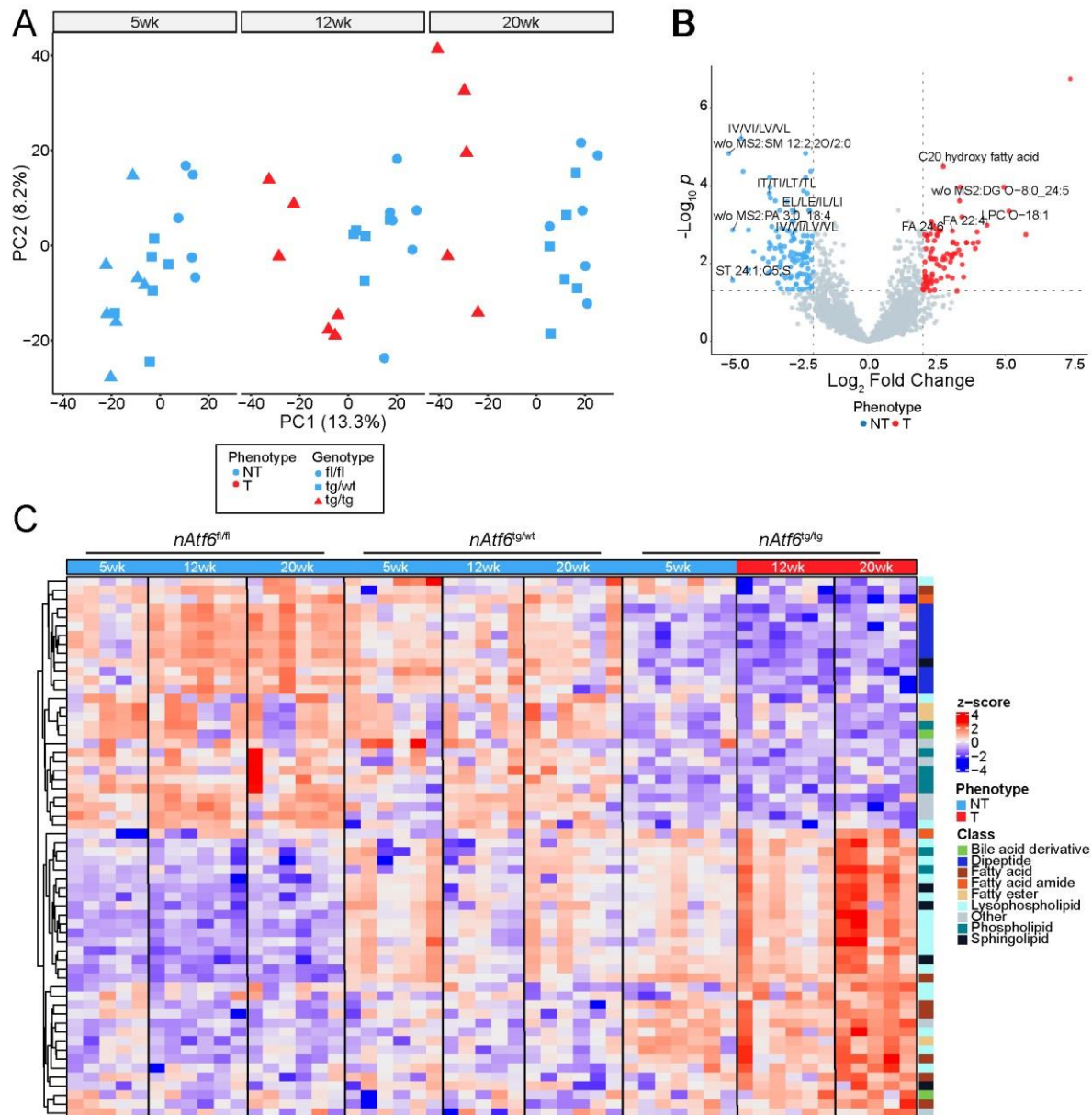


Figure 35. The *nAtf6^{tg/tg}* tumour metabolite environment is enriched in lipids and depleted in amino acids and peptides, prior to tumour onset. (A) PCA of metabolite profiles comparing tumour to non-tumour *nAtf6^{EC}* across all timepoints. (B) Volcano plot of differentially enriched metabolites between tumour and non-tumour *nAtf6^{EC}* mice (\log_2 fold change threshold = 0.58, p -value threshold = 0.05). (C) Heatmap of scaled differentially enriched metabolites, sorted by genotype and age. Rows are clustered using Ward's method.

We next sought to identify differences in caecal metabolites between phenotypes, in the *nAtf6^{EC};Il10^{-/-}* model. To assess global differences in metabolite profiles we carried out principal component analysis at each timepoint. Genotypes were similar at 5 weeks, with tumour samples deviating further from controls with increasing age. Overall, however, limited differences were observable between groups, regardless of tumour status (**Figure 36A**). We next examined metabolite differences between tumour and non-tumour mice. Caecal metabolomes of *nAtf6^{tg/wt};/-* tumour mice showed an enrichment of lipids compared to controls, with several LCFA, a fatty ester and a lysophospholipid species highly enriched at all ages. To identify nATF6-driven versus tumour-driven metabolites, we performed hierarchical clustering

of differential metabolites, utilising NR mice as a comparison. In line with the PCA analysis, no clear shifts could be observed between NR and R, nor between timepoints, implying *nATF6* activation or the combination of *I110* KO and *nATF6* activation underlies the altered metabolome (**Figure 36B**).

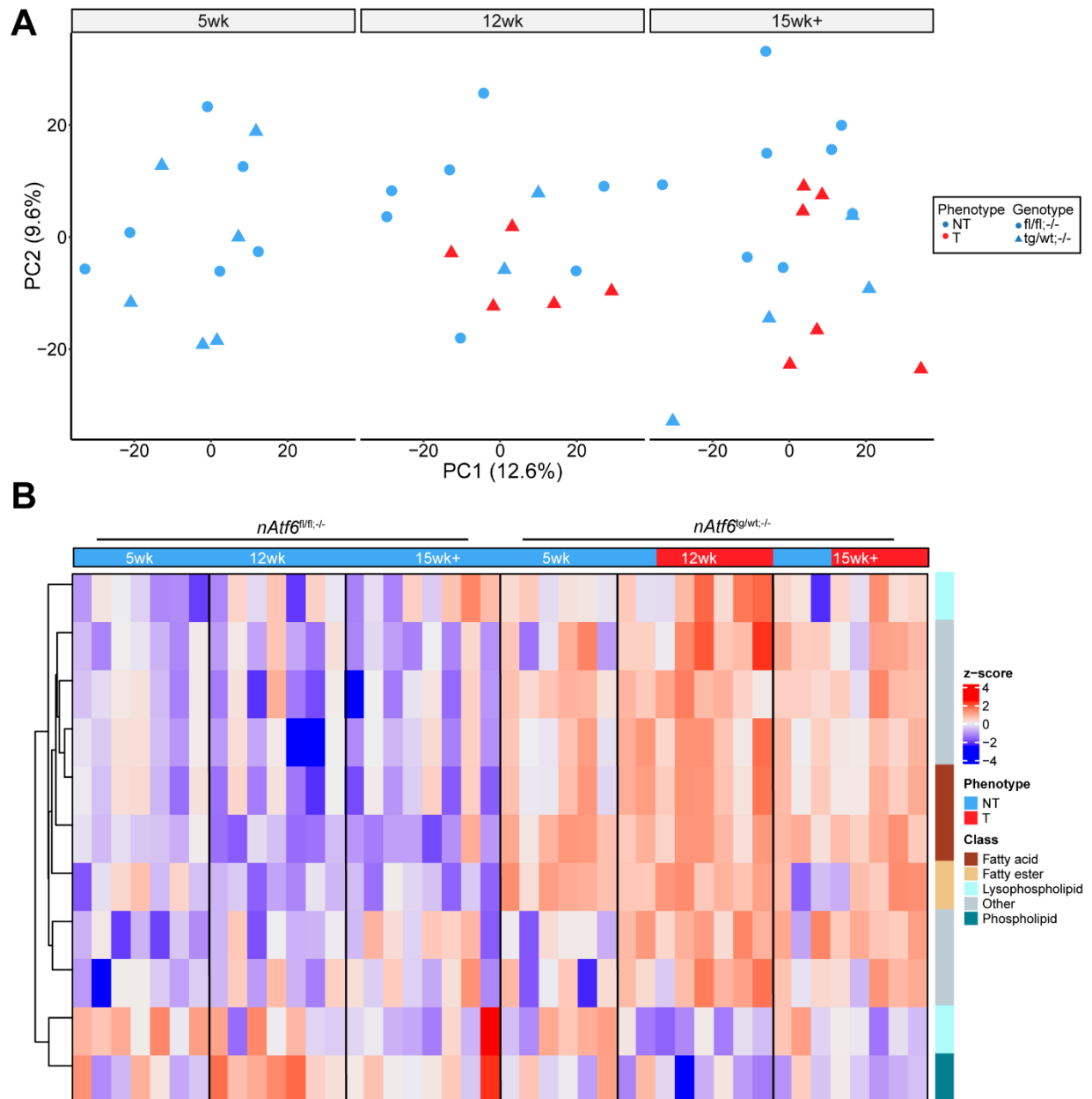
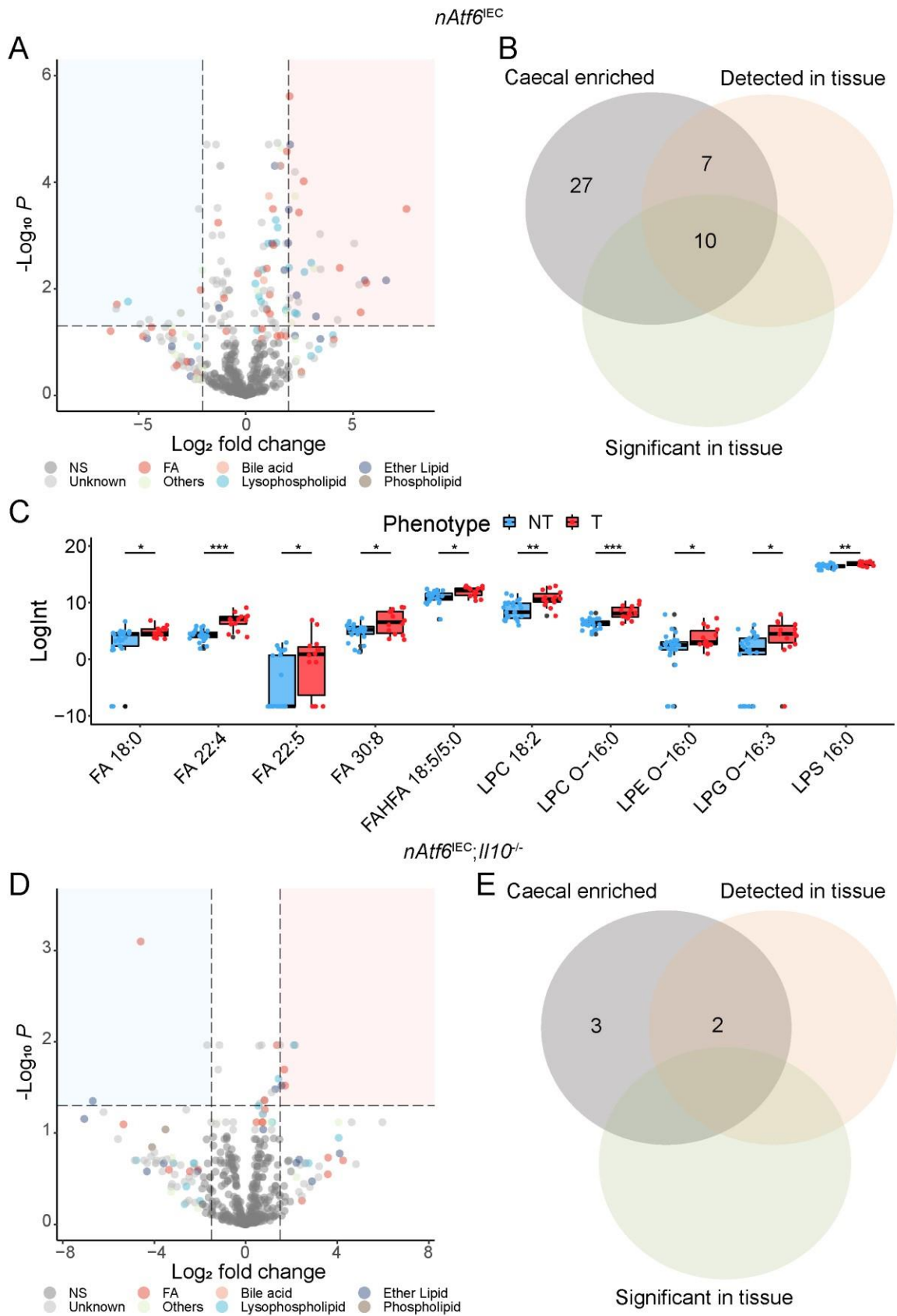


Figure 36. Fatty acid derivatives and other lipids define the *nAtf6*^{tg/wt;-/} metabolite milieu. (A) PCA of metabolite profiles comparing tumour to non-tumour *nAtf6*^{IEC};*I110*^{-/} across all timepoints. **(B)** Heatmap of scaled differentially enriched metabolites, sorted by genotype and age. Rows are clustered using Ward's method.

A clear enrichment of lipids in tumour mice could be observed in both mouse models, however, the question remained whether this lipid enrichment is host or bacteria derived as metabolite profiles from caecal content likely reflect a mixture of host and microbial metabolites. Metabolite profiles of intestinal tissue, on the other hand, are dominated by host metabolites. Thus, to exclude microbially derived metabolites, we performed untargeted metabolomics of

colonic tissue from *nAtf6*^{IEC} and *nAtf6*^{IEC};*Il10*^{-/-} mice to determine whether the caecal lipid enrichment could be recapitulated in tissue. Differential enrichment analysis between phenotypes in *nAtf6*^{IEC} revealed a clear enrichment of fatty acids, particularly LCFA, as well as lysophospholipids and their derivatives, in tumour mice implicating the host as the source of the observed lipid enrichment (**Figure 37A**). Next, to explore whether these metabolites were the same as those enriched in caecal content of *nAtf6*^{IEC} tumour-bearing mice, we matched them by annotation. Of the 27 LCFA/lysophospholipid metabolites which were detected as enriched in caecal content, 17 of them could be detected in the tissue dataset. Furthermore, 10 of these were identified as significantly different between phenotypes in *nAtf6*^{IEC} mice. Importantly, the majority also displayed the same direction of enrichment (8/10), with the exception of FA 22:4 and LPC 18:2 which were enriched in non-tumour mice in the caecal dataset (**Figure 37B and C**). We repeated this analysis for *nAtf6*^{IEC};*Il10*^{-/-} mice and similarly identified an enrichment of fatty acids and lysophospholipids in tumour mice, however, of the three metabolites from these classes, which were enriched in caecal content on *nAtf6*^{IEC};*Il10*^{-/-} tumour mice, only two were detected in tissue, and none differed between groups (**Figure 37D and E**). Taken together, this data supports the host as the source of the observed LCFA and lysophospholipid enrichment in *nAtf6*^{IEC} tumour-bearing mice but indicates in *nAtf6*^{IEC};*Il10*^{-/-} tumour mice, LCFA enrichment may be derived from elsewhere.



Venn diagram of the overlap between *nAtf6*^{EC} tumour-enriched metabolites in caecal content (grey) and those detected (orange) or significant (green) in the colonic tissue metabolome. **(C)** Boxplots of shared regulated metabolites between caecal content and tissue in *nAtf6*^{EC} tumour mice. Data are presented as normalised log-transformed feature intensity (LogInt). **(D)** Volcano plot of differentially enriched metabolites between tumour and non-tumour *nAtf6*^{EC};*Il10*^{-/-} mice (Log₂ fold change threshold = 0.58, *p*-value threshold = 0.05). **(E)** Venn-diagram of the overlap between *nAtf6*^{EC};*Il10*^{-/-} tumour enriched metabolites in caecal content (grey) and those detected (orange) or significant (green) in the colonic tissue metabolome.

4.6.2 Multi-omic integration links host-altered fatty-acid environment to tumour-associated microbiota

Next, to determine associations between the metabolome and microbiota in the context of tumorigenesis, we performed multi-omic data integration using sPLS-DA, implemented in the mixOmics R package^{209,210}. sPLS-DA is an extension of partial least-squares regression which identifies a set of features best-discriminating classes. Here we sought to investigate whether the observed metabolite changes are covariant with microbial alterations. Since 12- and 20-week-old *nAtf6*^{tg/tg} mice, showed similar changes in metabolites, especially LCFA, and lipids, we grouped them to increase sample size. In *nAtf6*^{EC} mice, sPLS-DA scores plot of the first two components of the model indicated the identified set of features separated non-tumour and tumour samples along the first component (**Figure 38A**). To examine whether the discriminating metabolites and ASVs were associated with each other, we investigated their correlations along the first component, using a threshold of 0.7, revealing a distinct cluster of tumour-enriched metabolites and ASVs (**Figure 38B**). To investigate these correlations further, we extracted fully annotated loadings from the first component of the sPLS-DA model which were predictive of tumour phenotype and then sorted by feature importance, for ASVs and metabolites separately. Microbial loadings were dominated by ASVs classified as *Lactobacillus* and additionally included several *Romboutsia* and *Turicibacter* species. Metabolite loadings on the other hand were mostly represented by LCFA and lysophospholipids, indicating that the observed enrichment of these metabolites is associated with microbial changes (**Figure 38C**).

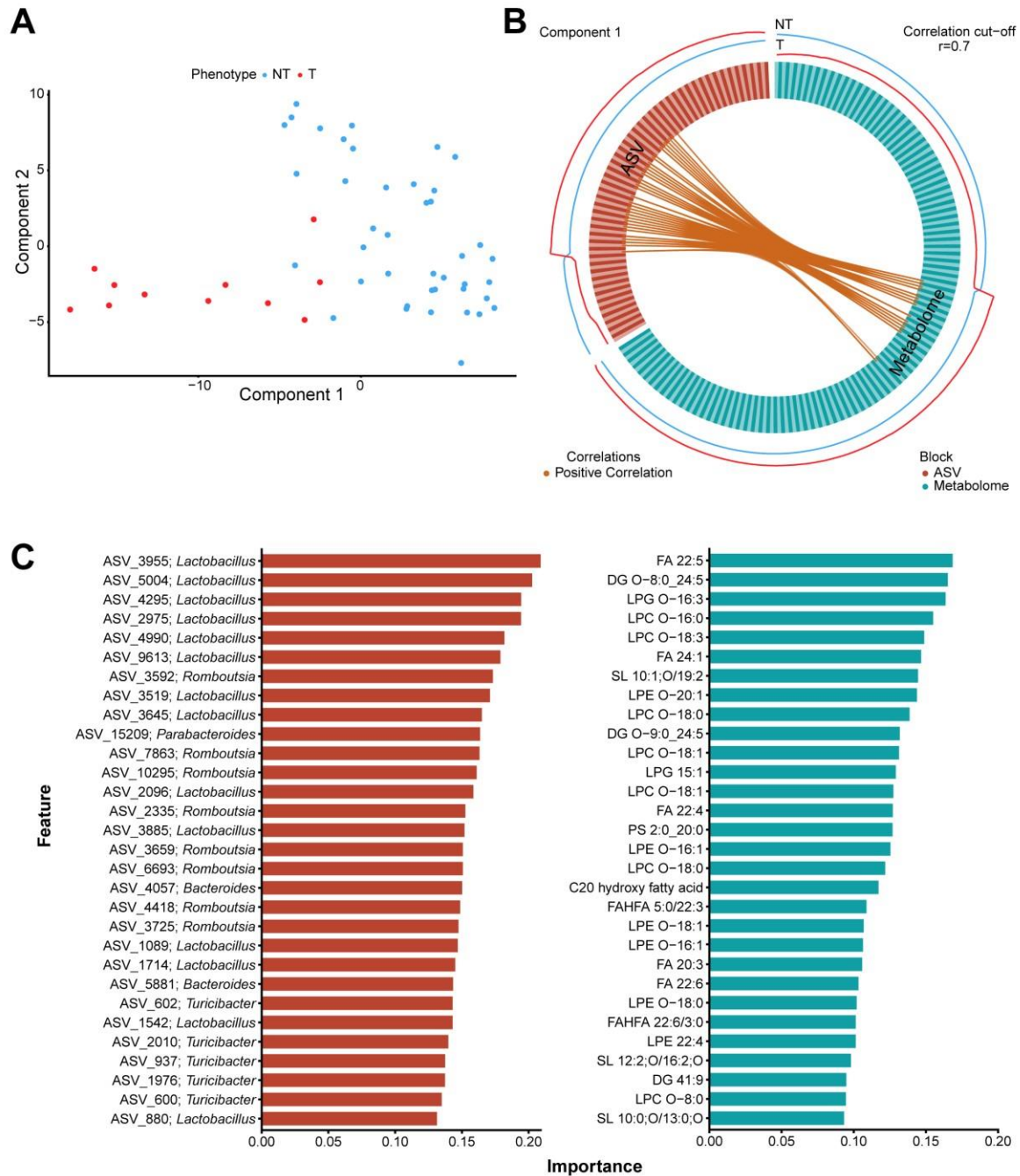


Figure 38. Multi-omic integration links tumour-associated taxa to altered lipid environment in *nAtf6*^{tg/tg} mice. (A) sPLS-DA score plot coloured by phenotype *nAtf6*^{IEC} mice. (B) Circos plot of highly correlated ($r \geq 0.7$) features between blocks. The ASV block is depicted in red, while the metabolome block is shown in light blue. Orange lines indicate a positive correlation. The height of the line surrounding the plot indicates the group (NT in blue or T in red) which a given feature is associated with. (C) Loadings plots of the top 30 annotated highly correlated features discriminating tumour from non-tumour, showing ASVs (left) and metabolites (right). Features are sorted by importance.

As metabolomic analysis revealed a similar enriched of LCFA and lysophospholipids in *nAtf6*^{tg/wt;-/-} mice as in *nAtf6*^{tg/tg}, we repeated the same integrated analysis for *nAtf6*^{IEC};*Il10*^{-/-} mice, excluding NR samples. Non-tumour and tumour samples (12 and 15-week+) clearly separated along the first component indicating the identified signature could discriminate phenotype (Figure 39A). Next, to determine associations between features, we examined intra-block correlations in the first component (ASV-Metabolome), using an absolute threshold

of 0.7. We observed a small cluster of mostly positively correlated ASVs, and metabolites enriched in tumour, with two control-enriched negative correlations (**Figure 39B**). Extracting the top 30 most important annotated features for ASVs and metabolites separately, identified *Parabacteroides*, *Desulfovibrio*, *Bacteroides* and *Mucispirillum* ASVs as discriminative, while metabolites were again dominated by LCFA and lysophospholipids (**Figure 39C**).

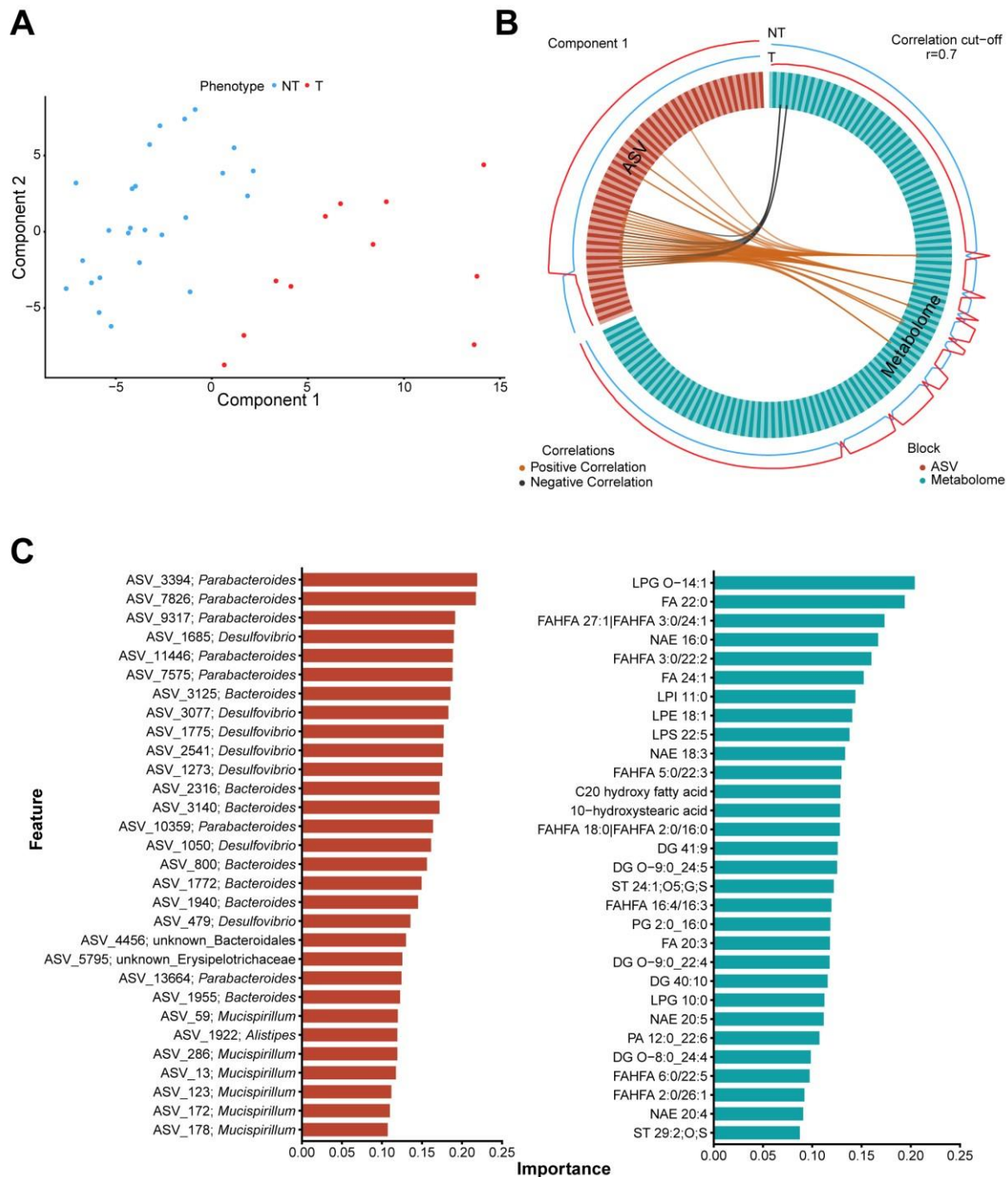


Figure 39. Multi-omic integration links tumour-associated taxa to altered lipid environment in *nAtf6*^{tg/wt;-/-} mice. (A) sPLS-DA score plot coloured by phenotype in *nAtf6*^{IEC;||10}^{-/-} mice. Non-tumour (NT), excluding NR mice is shown in blue, with Tumour (T) in red. (B) Circos plot of highly correlated ($r \geq 0.7$) features between blocks. The ASV block is depicted in red, while the metabolome block is shown in light blue. Orange lines indicate a positive correlation, while black lines indicate negative. The height of the line surrounding the plot indicates the group (NT in blue or T in red) which a given feature is

associated with. **(C)** Loadings plots of the top 30 annotated highly correlated features discriminating tumour from non-tumour, showing ASVs (left) and metabolites (right). Features are sorted by importance.

Several LCFA appeared to be linked to the microbiota in nATF6-driven tumour environments, however, it remained unclear if enrichment of these particular LCFA was driven by the tumour or by nATF6 activation. It is known that tumour cells demonstrate an increase in *de novo* fatty acid synthesis. Fatty acid synthase (FASN) – an enzyme which catalyses *de novo* fatty acid synthesis - is often upregulated in colorectal cancer and is a predictor of poor prognosis, though ATF6 can also upregulate *FASN* expression^{164,230–232}. Hence, we further examined the enrichment of microbiota correlated LCFAs at each timepoint. Several LCFA metabolites were already enriched at 5-weeks in *nAtf6*^{tg/tg} mice. Moreover, the majority of other LCFAs showed an increased trend in *nAtf6*^{tg/tg} mice but did not reach significance. The extent of enrichment also appeared to increase at tumour timepoints, with 12- and 20-week-old *nAtf6*^{tg/tg} mice displaying an enrichment of most microbiota-correlated LCFAs. An unknown C20 hydroxy fatty acid and Nervonic acid (FA 24:1) were the only LCFAs which were enriched at all three timepoints, however **(Figure 40A)**. The majority of microbiota correlated LCFAs in *nAtf6*^{tg/tg} mice, weakly correlated with tumour number, however most were not significant, with the exception of Nervonic acid and FAHFA 5:0/22:3. This indicates that rather than being the source, in this context, tumour presence increases the LCFA pool **(Figure 40B)**.

In *nAtf6*^{tg/wt;-/-} mice a similar pattern was observed and again the C20 hydroxy fatty acid was consistently enriched, along with behenic acid (FA 22:0). Interestingly, 10-hydroxystearic-acid, a known product derived from microbial fatty acid hydration of oleic acid was enriched in 12week and 15-week+ mice **(Figure 41A)**. No LCFAs significantly correlated with tumour number, suggesting an alternative source besides tumour metabolism underlies their enrichment **(Figure 41B)**. Together these data suggest LCFA enrichment occurs prior to tumour formation and is promoted by nATF6.

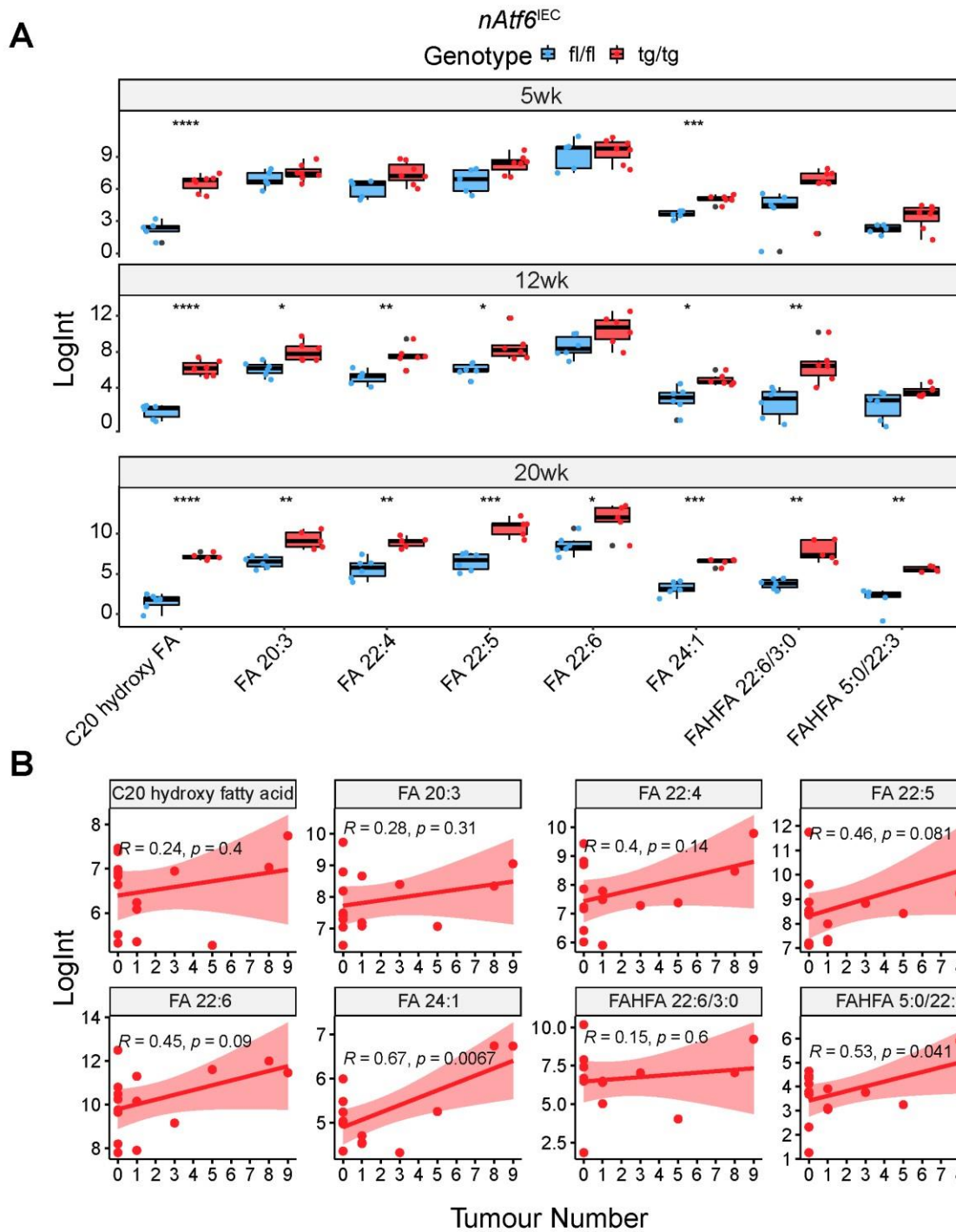


Figure 40. Microbiota correlated long-chain fatty acids are enriched in *nAtf6^{tg/tg}* mice. (A) Boxplots of microbiota-correlated Long-chain fatty acids in *nAtf6^{IEC}* mice. (B) Spearman's rank correlation between each microbiota-associated fatty acid and tumour number in *nAtf6^{tg/tg}* mice, with correlation coefficient and p -value indicated. Data are presented as normalised log-transformed feature intensity (LogInt).

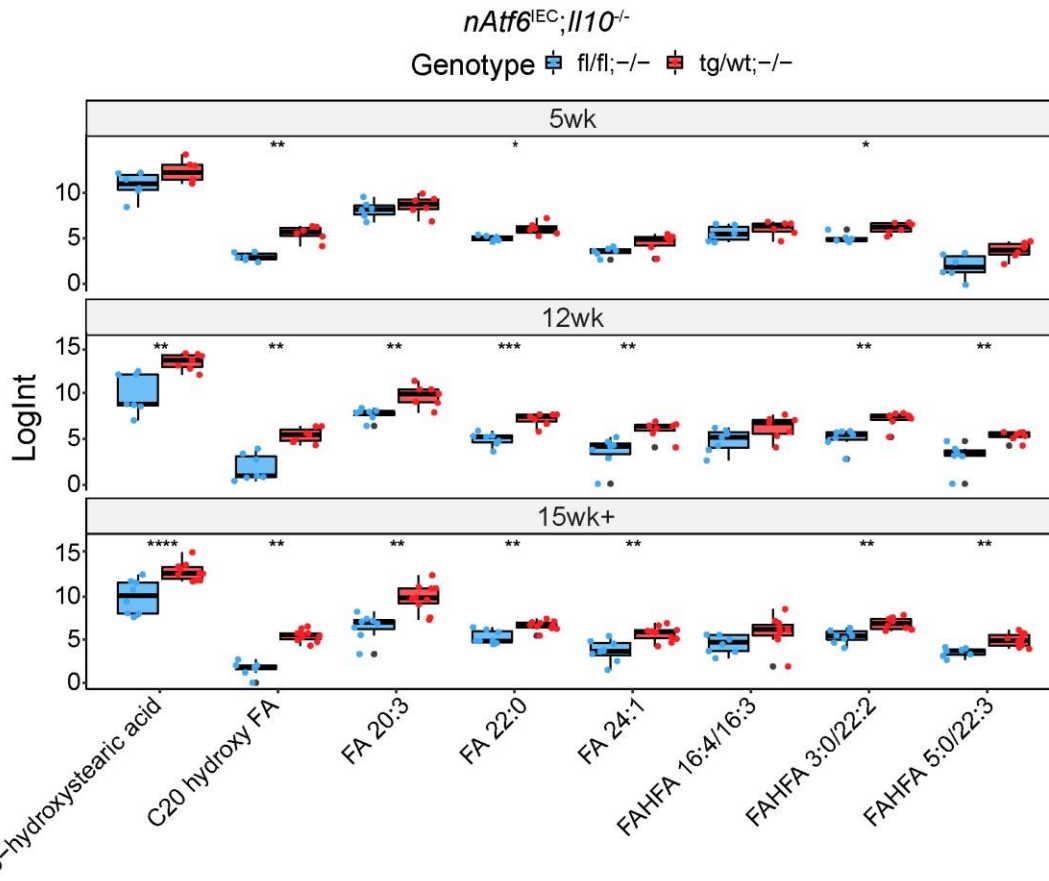
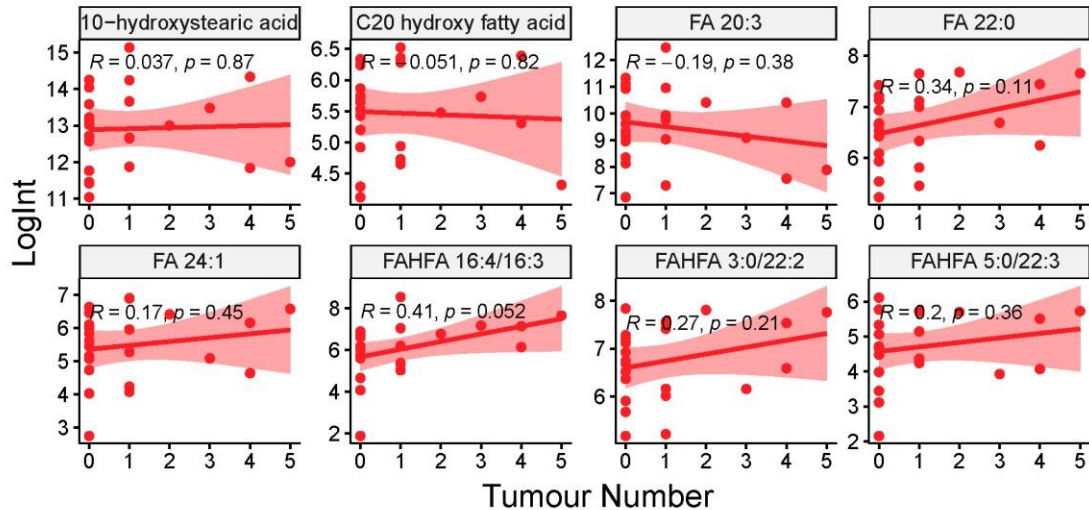
A**B**

Figure 41. Microbiota correlated long-chain fatty acids are enriched in *nAtf6^{tg/wt;-/-}* mice. (A) Boxplots of microbiota-correlated Long-chain fatty acids in *nAtf6^{EC};Il10^{-/-}* mice. **(B)** Spearman's rank correlation between each microbiota-associated fatty acid and tumour number in *nAtf6^{tg/wt;-/-}* mice, with correlation coefficient and *p*-value indicated. Data are presented as normalised log-transformed feature intensity (LogInt).

Alterations in the level of LCFA in the intestinal lumen has been shown to modulate the intestinal microbiota^{233,234}. Having established the association between these compounds and the microbiota in the tumour environment, we next explored whether they could explain the observed shifts in any of the identified driver taxa by computing correlation networks between

each of the microbiota-covariant fatty acids, clr-transformed abundance of ID/SD ASVs, as well as community metrics (alpha and beta diversity) to assess whether these compounds modified community structure, or ID/SD presence was indirectly modified by LCFA. In *nAtf6^{tg/tg}* mice, the majority of driver taxa were positively correlated with at least one LCFA at 5 weeks, however a cluster of highly positively correlated ASVs and metabolites was observed comprising an unknown 20-carbon hydroxy fatty acid, docosapentanoic acid (FA 22:5), nervonic acid (FA 24:1) and a Fatty acid ester of hydroxy fatty acid (FAHFA 22:6/3:0), along with ASVs classified as *Lactobacillus* and to a lesser extent *Mucispirillum*. Interestingly, docosahexanoic acid (FA 22:6) correlated negatively with several alpha diversity metrics (Richness and Shannon Effective) as well as positively with GUniFrac dissimilarity, suggesting this LCFA may alter community structure which could allow ID/SD taxa to colonise. Supporting this *Mucispirillum* was positively correlated with docosahexanoic acid. *Lactobacillus* species continued to show strongly positive correlations with LCFA at tumour onset supporting their status as SD, however, *Mucispirillum* ASVs only demonstrated negative correlations with LCFA. Surprisingly, *Odoribacter* taxa tended to show generally negative correlations across both timepoints (**Figure 42A**).

In *nAtf6^{tg/wt;-/-}* mice, correlations were more consistent between timepoints, likely reflecting the lack of microbial alterations, with *Mucispirillum* ASVs highly correlated with multiple FAHFA's, while *Lactobacillus* were positively correlated with almost all LCFA species, except for 10-hydroxystearic acid and an unknown C20 hydroxy fatty acid at the tumour timepoint only. *Bacteroides* however was negatively correlated with several LCFA at 5 weeks and highly correlated after tumour onset at 12 weeks (**Figure 42B**). Intriguingly, all LCFA correlated positively with Richness and Shannon Effective diversity at 5 weeks, and negatively at 12 weeks, suggesting the impact of LCFA on microbial community modulation is timepoint dependent in *nAtf6^{tg/wt;-/-}* mice. Taken together, this data suggests that the nATF6-driven increase in LCFA may modulate the colonisation of putative driver microbes, either directly or indirectly by altering community structure.

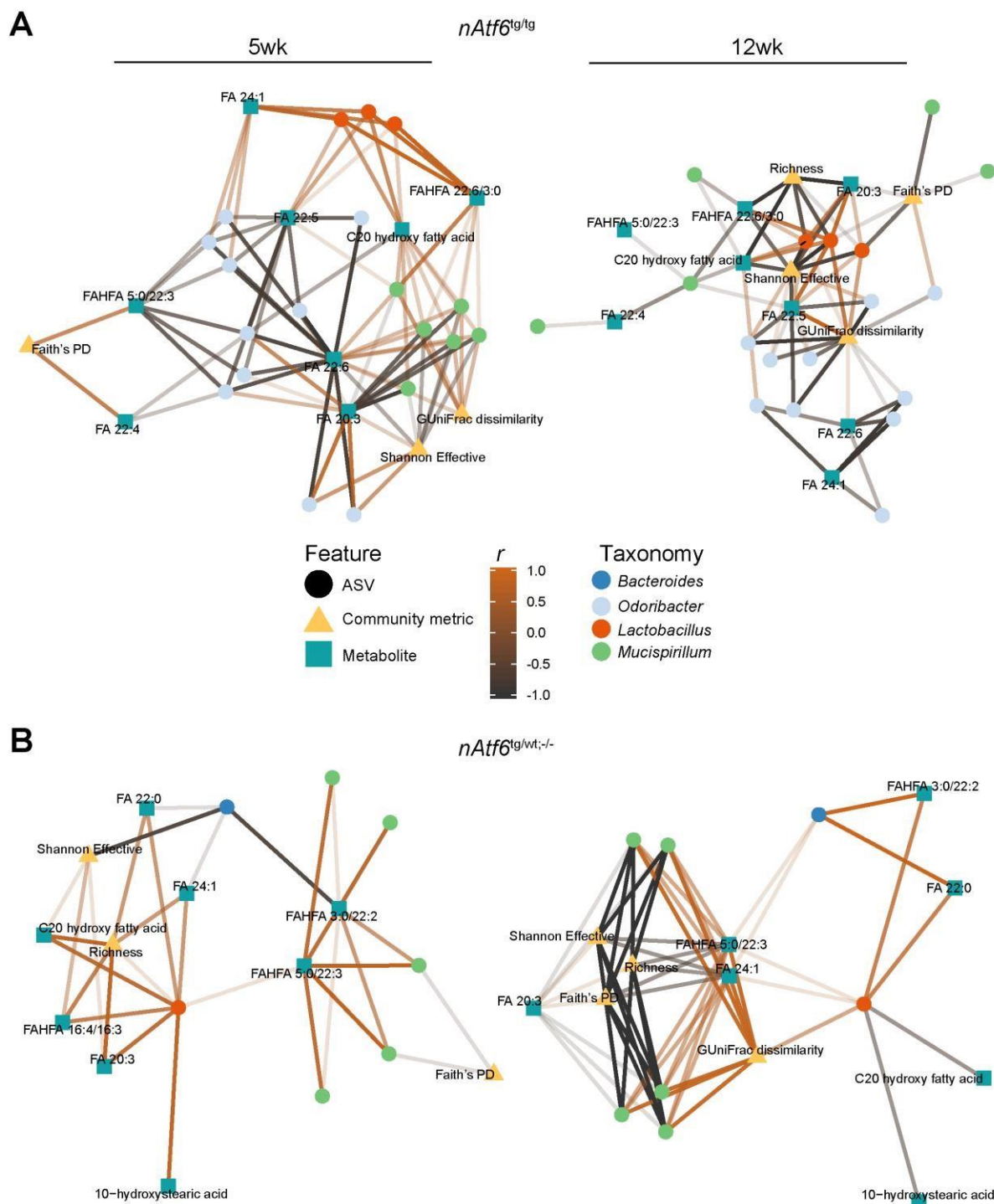


Figure 42. nATF6-enriched long-chain fatty acids modulate driver colonisation. Spearman's rank correlation networks of microbiota-correlated fatty acid species, clr-transformed driver taxa abundances and community metrics using a correlation threshold (r) of 0.5, at 5 and 12-week timepoints in **(A)** *nAtf6^{tg/tg}* and **(B)** *nAtf6^{tg/wt;-/-}* mice.

4.6.3 Long-chain fatty acids modulate microbial composition

Many LCFA compounds are toxic to bacteria, particularly gram-positive taxa which require detoxification mechanisms to grow in fatty acid-rich environments²³⁵. To understand whether this enrichment might impact microbial function, we used PICRUSt2 to generate predictions

of genomic content²⁰³. Predicted metagenomes were grouped into MetaCyc pathways and filtered to remove low-abundant pathways and assessed for differential abundance²³⁶. Focusing on changes in fatty acid metabolism, a clear enrichment of related pathways was observed in *nAtf6^{tg/tg}* mice from 5 to 12 weeks. A similar, weaker enrichment was observed between tumour and non-tumour samples (**Figure 43A**). No differentially abundant pathways were present between 5 and 12 weeks in *nAtf6^{tg/wt;-/-}* mice, reflecting the observed lack of ID and P in these mice, however fatty acid metabolism-related pathways were enriched in tumour environments compared to non-tumour, indicating a microbial response to the altered fatty acid milieu (**Figure 43B**).

The majority of identified driver taxa were gram-negative, displaying mostly positive or neutral correlations with the identified LCFA, however in both mouse models, a gram-positive *Lactobacillus* SD species was highly correlated with multiple LCFA across the 5 and 12-week timepoints. Certain Lactobacilli can survive fatty acid-rich environments via the action of hydratases such as oleate hydratase (*OhyA*; EC:4.2.1.53) which detoxify fatty acids by converting them into hydroxy fatty acids, thus we explored whether the predicted abundance of this gene differed between tumour and non-tumour mice and calculated the ratio of taxa predicted to encode *ohyA* versus those not predicted to encode *ohyA*^{237–239}. In *nAtf6^{IEC}* mice and *nAtf6^{IEC};I110^{-/-}* mice an increase in *ohyA* relative abundance was observed in tumour-bearing mice. The ratio of *ohyA*⁺ to *ohyA*⁻ was clearly increased in all *nAtf6^{tg/tg}* mucosal phenotypes compared to non-tumour controls, however in *nAtf6^{tg/wt;-/-}* mice, a significant increase was only observed in tumour-adjacent mucosa (**Figure 43C and E**). Supporting the enrichment of *ohyA*, its main product: 10-hydroxystearic acid, was increased in tumour samples in both mouse models (**Figure 43D and F**). Lastly, to associate *ohyA* presence with the capacity of Lactobacilli SD taxa to thrive in LCFA-rich environments, we examined the taxonomic contributions to predicted functional shifts in *ohyA* abundance in *nAtf6^{IEC}* and *nAtf6^{IEC};I110^{-/-}* mice. In both mouse models, *Lactobacillus* accounted for the largest share of *ohyA* relative abundance, making up 60.3% and 34.6% in *nAtf6^{tg/tg}* and *nAtf6^{tg/wt;-/-}* mice, respectively (**Figure 43G**). Taken together, these results indicate microbial adaptation to the altered fatty acid environment and support a role for nATF6-mediated LCFA enrichment in supporting colonisation of tumour-associated taxa.

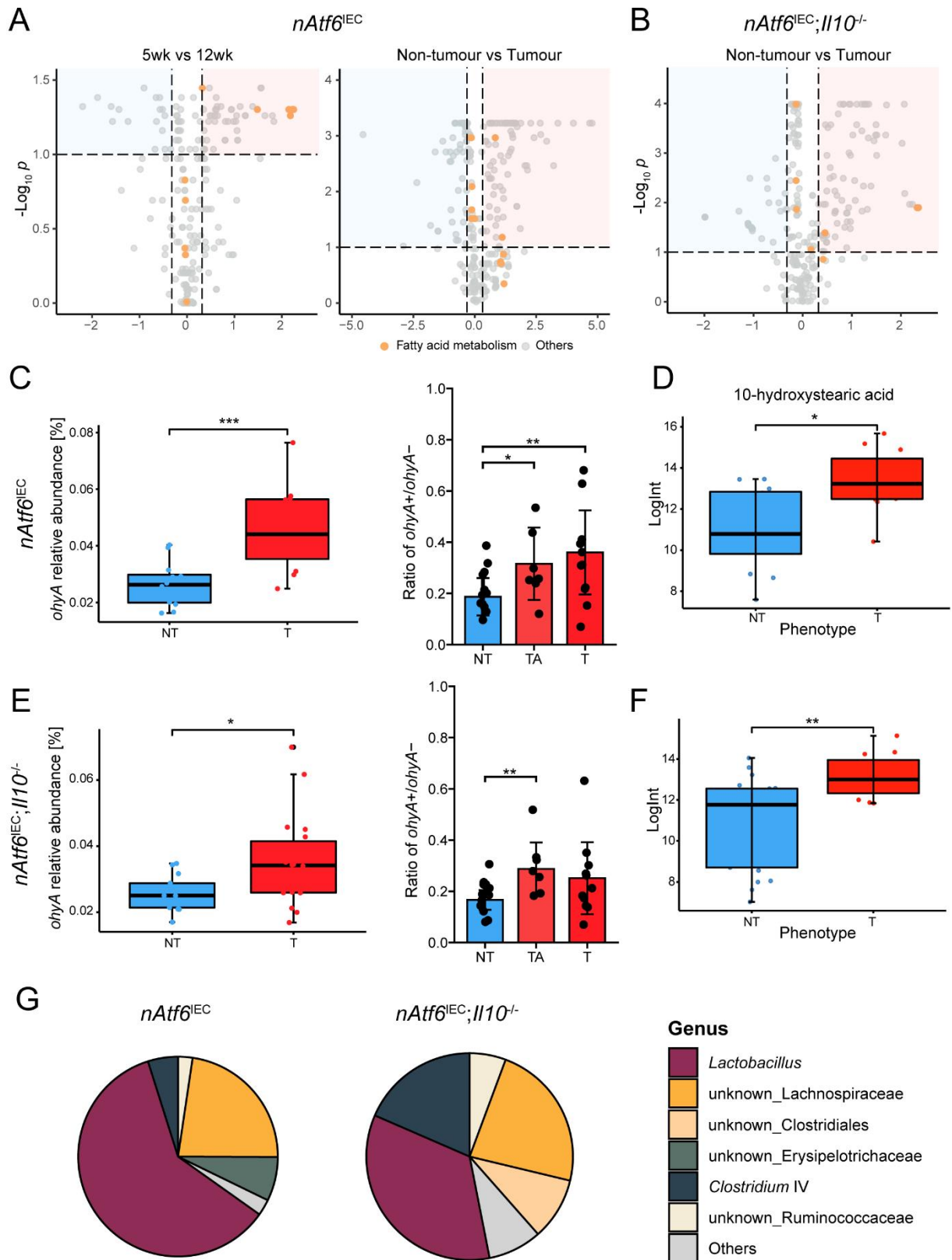


Figure 43. Evidence of microbial adaptation to the altered fatty acid environment. (A) Volcano plot of differentially abundant pathways (Log_2 fold-change 0.32, adjusted p -value < 0.05) predicted by PICRUSt2 comparing 5-week (pre-tumour) and 12-week (post-tumour) timepoints (left) in *nAtf6^{tg/tg}* mice and non-tumour controls to tumour-bearing mice (right). Pathways related to fatty acid metabolism are highlighted in yellow. **(B)** Volcano plot of differentially abundant pathways (Log_2 fold-change 0.32, adjusted p -value < 0.05) between non-tumour *nAtf6^{fl/fl}* and tumour-bearing *nAtf6^{tg/wt}* mice. Pathways related to

fatty acid metabolism are highlighted in yellow. **(C)** Relative abundance of oleate hydratase (*ohyA*) in tumour and non-tumour *nAtf6*^{EC} mice (left), and the ratio of *ohyA* positive to *ohyA* negative taxa across various mucosal phenotypes (right). **(D)** Relative intensity of the major product of *ohyA*: 10-hydroxystearic acid in caecal samples from *nAtf6*^{EC} mice. Data are presented as normalised log-transformed feature intensity (LogInt). **(E)** Relative abundance of oleate hydratase (*ohyA*) in tumour and non-tumour *nAtf6*^{EC};*Il10*^{-/-} mice (left), and the ratio of *ohyA*⁺ to *ohyA*⁻ taxa across each mucosal sample type (right). **(F)** Relative intensity of 10-hydroxystearic acid in caecal samples from *nAtf6*^{EC};*Il10*^{-/-} mice. **(G)** Top six contributors to *ohyA* relative abundance in tumour-bearing *nAtf6*^{tg/tg} (left) and *nAtf6*^{tg/wt;-/-} (right) as predicted by PICRUSt2. Minor contributors are grouped into the other category.

The increased fatty acid pool in mice with activated nATF6 appeared to instigate a microbial response as evidenced by the increased predicted abundance of *ohyA* and increased levels of its product, 10-hydroxystearic acid. The tumour microenvironment is likely also fatty acid rich however, and it remained unclear whether these LCFA could promote the same microbiota changes seen in *nAtf6*^{tg/tg} and *nAtf6*^{tg/wt} mice, in a normal caecal environment. Thus, we tested whether a cocktail of the nATF6 enriched LCFA (Nervonic; FA 24:1, Behenic; FA 22:0 and Eicosanoic acid; FA 20:0) could promote the same microbial changes in a non-tumour-susceptible environment. Accordingly, we co-cultured caecal content from three *nAtf6*^{fl/fl} and *nAtf6*^{fl/fl;-/-} controls, with this LCFA mixture to determine if we could recapitulate the observed microbial shifts (**Figure 44A**). Strikingly, LCFA co-culture led to marked, shifts in microbial composition, leading to the homogenisation of two distinct microbial environments from different mouse models (**Figure 44B and C**). Finally, to determine whether co-culture with LCFA reflected the functional changes seen *in vivo* we used PICRUSt2 to generate predicted metagenomic profiles and examined the relative abundance of *ohyA*. After 7h of coculture, *ohyA* predicted abundance was increased four-fold, regardless of genotype, indicating that LCFA likely selects for bacteria which encode this gene (**Figure 44D**). Taken together, these results recapitulate the microbial response seen *in vivo* in *nAtf6* mice and further implicate LCFA in modulating the microbiota.

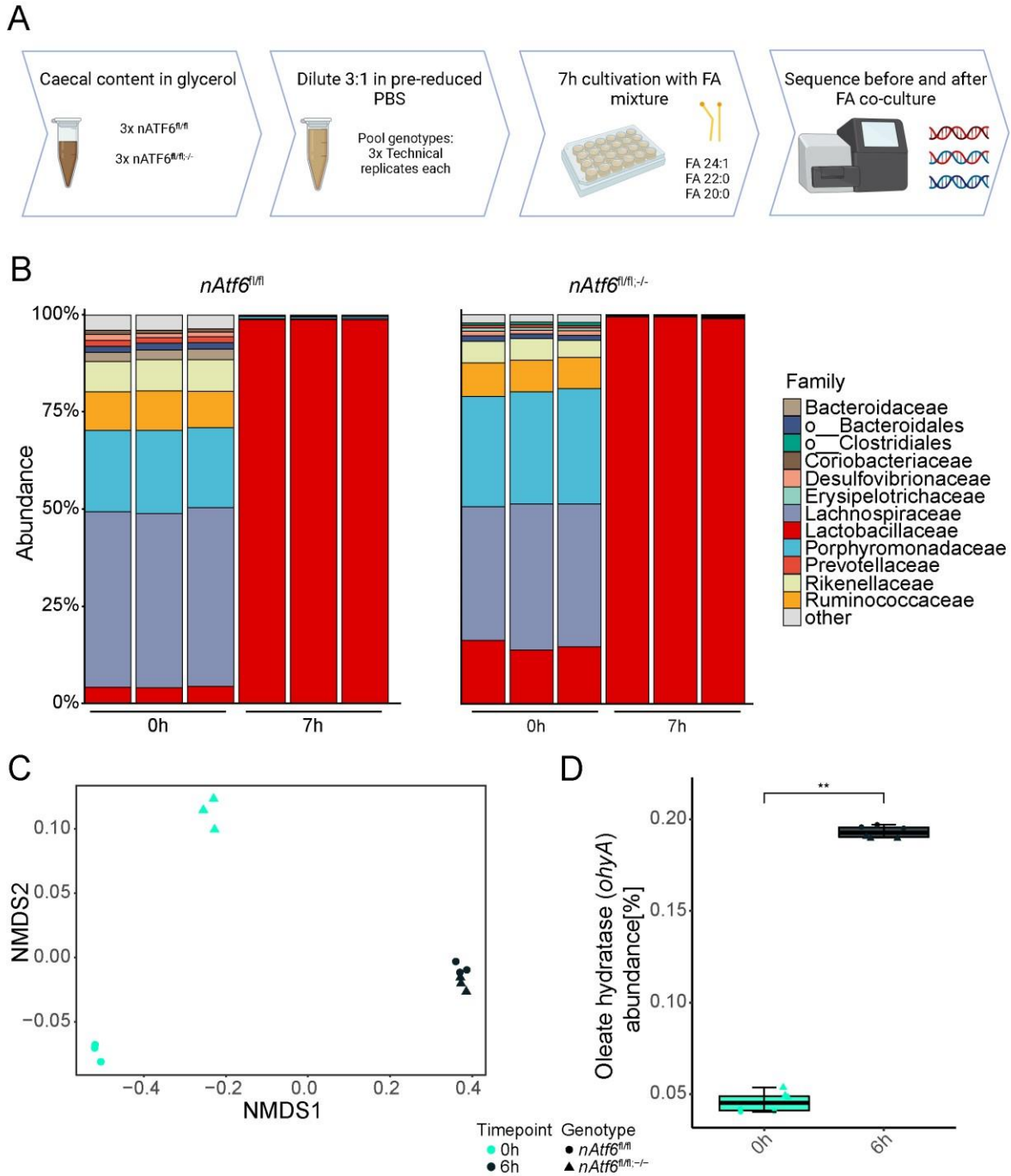


Figure 44. Ex vivo co-culture with LCFA recapitulates in vivo shifts in microbial composition and predicted function in caecal content from non-susceptible mice. (A) Experimental design schematic of caecal content and LCFA co-culture. **(B)** Taxonomic composition of ex vivo cultivated caecal content in *nAtf6^{fl/fl}* (left) and *nAtf6^{fl/-}* (right) genotypes, before and after LCFA co-culture. **(C)** NMDS plot of generalized UniFrac distance between samples at 0h after LCFA addition and 7h after in non-tumour *nAtf6^{fl/fl}* and *nAtf6^{fl/-}* samples. **(D)** Relative abundance of the predicted abundance of *ohyA* before and after LCFA addition, in both genotypes.

5. Discussion

In the present work, we comprehensively examined microbial community shifts in nATF6-driven mouse models of colorectal tumorigenesis. We additionally generated and characterised a novel microbiota-dependent model of CAC, which, alongside our *nAtf6*^{IEC} model, provides a human-relevant framework for dissecting microbe-host interactions in chronic inflammation-driven and sporadic CRC, respectively. To determine microbial factors underlying tumour formation, we characterised colonic mucosal communities, demonstrating these communities have a greater utility than luminal for identifying phenotype-relevant taxa. Additionally, we examined microbial changes occurring along tumour initiation and progression, identifying putative drivers of tumorigenesis, and spatially associating them with tumours. Finally, we utilised multi-omic data integration to link nATF6-induced changes in lipid metabolism, particularly that of LCFA, with alterations in microbiota and uncovered potential microbial adaptations to the altered lipid environment.

5.1 Identification of microbial shifts between pre-tumour and tumour onset in an nATF6-dependent model of sporadic colorectal tumorigenesis

We previously identified a microbiota-dependent tumour-promoting role of nATF6 activation, accompanied by loss of mucus structure, dysbiosis and microbial penetration close to the epithelium¹⁵⁵. However, whether this phenotype is driven by transiently colonising luminal taxa or resident mucosal microbes, and exactly which members of the microbiota are involved, was not clear. A handful of studies have identified pro-oncogenic bacteria enriched in adenoma samples; however, questions remain over whether these taxa initiate disease, as adenoma formation induces metabolic alterations which likely impact the local microenvironment, thus altering microbial composition^{124,240,241}. Furthermore, little is known about what triggers may enable putative driver taxa to expand in habitat or abundance and assume a tumour-promoting role in the first place.

Here we examined microbial changes along the sequence of tumour development in the context of nATF6-driven tumorigenesis, using the driver-passenger model as a framework. The vast majority of existing human CRC studies have profiled stool as a proxy for the gut but tumour-driver or indeed passenger bacteria may not be well represented in stool, or other luminal samples. Although murine models present the opportunity to explore additional environments, most studies have similarly examined stool or caecal content. We addressed this problem by performing 16S rRNA sequencing of both luminal and mucosal communities, extensively characterising the mucosal microbiota across tumour development and space, at tumour onset. ML classifiers trained on mucosal data achieved higher performance in discriminating phenotype in 2/3 models compared to luminal, suggestive of an inherent tumour-associated microbiota signal present in mucosal communities. Our results demonstrate that microbial profiling of luminal contents may therefore miss disease-relevant taxa, indicated by lower accuracy. In line with this finding, several discrepancies have been observed between the faecal and mucosal microbiota in CRC, though no studies have compared this directly. Nevertheless, indirect evidence indicates that certain taxa linked to CRC are detected at higher frequencies in mucosal biopsies than stool, including putative driver taxa such as *pks+* *E. coli* strains^{129,242}. Moreover, Saffarian *et al* identified a core

microbiota specific to colonic crypts in cancer patients, largely comprising aerobic non-fermentative Proteobacteria which show low abundance in faecal samples (<1%)²⁴³.

Considering the often-stringent abundance and prevalence filtering applied before applying ML on 16S rRNA sequencing data, it is likely these taxa would be removed from faecal data or may even not be detected at all^{140,244}. It is important to note however, that the ML models we trained on luminal data could still predict phenotype with high accuracy, thus in this context, disease relevant ASVs may be partially diluted out in the luminal environment, explaining the decreased AUC compared to mucosal.

To delineate taxa putatively driving tumour formation and those merely associated with tumour onset, we analysed shifts between pre-tumour and tumour-onset mucosal samples, identifying several groups of taxa which displayed abundance changes in accordance with the driver-passenger model. We additionally identified a further group, SD, which maintain their abundance between pre-tumour and tumour onset. Our study is unique in examining mucosal microbial alterations from pre-tumour to tumour onset. Existing studies examining microbial succession in the driver-passenger hypothesis have generally compared tumour mucosa to adjacent healthy mucosa, however findings from our study and others suggest that microbial communities do not significantly differ between these tissues^{142,215,216}. Furthermore, we also observed limited differences compared to non-adjacent healthy mucosa, potentially indicative of spatial homogenisation. Evidence from human patients corroborates this homogenisation; Murphy *et al* analysed the microbial communities of multiple tumours from each participant, finding profiles only differed between samples and not within, regardless of the distance between tumour sites²¹⁷. Nevertheless, others have reported differences between on and off-tumour profiles, including those without a considerable distance between sampling sites, which may suggest spatial homogenisation depends on the underlying genetic or environmental trigger^{245,246}. Indeed, although our results indicated nATF6 activation as the major driver of microbial change, tumour presence also exerted a minor additional effect, making it difficult to untangle changes driven by nATF6 activation and those driven by the tumour. Regardless, examination of microbial changes pre-tumour excludes any impact the tumour may have on surrounding microbiota and is therefore preferable for the identification of driver microbes.

By analysing putative driver taxa across multiple distinct cohorts, we narrowed down ID taxa to *Mucispirillum* and Lachnospiraceae and SD to taxa mostly comprising *Odoribacter*, *Lactobacillus*, Porphyromonadaceae. Much of these genera have previously been associated with tumorigenesis, with multiple taxa from Porphyromonadaceae enriched in CRC, while recently a Lachnospiraceae species, *Lachnoclostridium*, has been identified as a biomarker of adenoma^{105,122,247}. Similarly, in mice, an increase in *Mucispirillum* LPS transcription was associated with tumour formation in a *Smad3*^{-/-} model of CRC²⁴⁸. Interestingly and in contrast with our data, *Lactobacillus* species have generally been found to be negatively associated with and even impede or restrict tumorigenesis. For example, *L. gallinarum* produces indole metabolites which inhibit tumour formation in *Apc*^{Min/+} and AOM-DSS mouse models of tumourigenesis²⁴⁹. On the other hand, other *Lactobacillus*-derived indole metabolites have been shown to impair anti-tumour immunity in pancreatic cancer²⁵⁰. Moreover, the normally commensal *L. reuteri* can translocate across the epithelial barrier in lupus-prone hosts, worsening disease²⁵¹. This protective-harmful dichotomy in *Lactobacillus* species may therefore be context or species/strain dependent. Importantly, the driver-passenger model does not currently account for adaptable bacteria, which can adapt to changing conditions after tumour onset. Our finding of the presence of SD extends the hypothesis to incorporate these taxa. Several taxa have already been identified as both passengers and drivers, for

example, *Streptococcus gallolyticus*, which has been shown to alternatively promote tumorigenesis by increasing epithelial cell proliferation and to outcompete other resident species upon tumour formation^{252,253}. Rather than existing as both a driver and passenger, we posit that the taxa we identified and those like *S. gallolyticus* are SD, likely comprising adaptable species which adjust their metabolism to changing conditions.

We additionally found that in the absence of tumour formation, nATF6 activation alone is enough to significantly alter mucosal microbiota, in line with previous results in luminal communities. Little is known regarding how ATF6 and indeed other arms of the UPR may impact microbial composition, but given its position and function within the cell, it is likely mediated indirectly by downstream targets. Alterations in the function of other intracellular organelles, known to modulate gut microbiota composition, for example, mitochondria, influence composition via production of reactive oxygen species²⁵⁴. Similar to our previous study where nATF6 conditioned microbial communities promoted enhanced tumour formation in GF *nAtf6*^{tg/tg} mice, nATF6 activation appeared to support colonisation by ID/SD microbes¹⁵⁵. This was clearly evidenced by the similarities in differentially abundant taxa on and off tumour sites and the observed habitat expansion of ID/SD taxa, particularly *Mucispirillum*, *Lactobacillus* and *Odoribacter* ASVs. Notably, all of these taxa are normal residents of the colonic mucosa, however in *nAtf6*^{fl/fl} control mice we found them to be either lowly abundant or restricted to certain sites^{59,255}. Recently, microbial migration from distant environments to disease sites has been observed in various pathologies. For example, many oral taxa are found in tumours from CRC patients, and alterations in the oral microbiota can predict disease^{10,62}. Migration likely occurs on smaller scales as well, for example between intestinal compartments. In line with this, caecal *Lactobacillus* populations in mice have been shown to mostly be derived from the stomach^{256,257}. Such migration could also be potentially deleterious for the host under disease conditions, however, particularly when the epithelium is exposed. Expression of several receptors interfacing with microbial ligands in the colonic epithelium are highly spatially segregated, including those which have been shown to play a key role in tumorigenesis such as Wnt and Toll-like receptors^{68,258}. Similarly, the thickness of the mucus layer is not consistent throughout the colon, with the distal colon possessing a thicker and more structured mucus layer, compared to the proximal²⁵⁹. Expansion of certain bacteria outwith their normal locale may therefore lead to unwanted stimulation, potentially promoting proliferation, and instigating tumorigenesis. Collectively, these results further understanding of alterations in microbial community structure and composition occurring with tumour development and support a role for ATF6 in promoting mucosal colonisation of putatively oncogenic microbes.

5.2 Generation and microbial characterisation of a novel model for colitis-associated cancer

While the use of IL-10 deficient mice – with or without additional triggers - as a model for colitis and subsequent CAC is well established, the combined effect of disrupted immunoregulation and UPR activation are unknown. Similarly, to our knowledge only a handful of microbiota-dependent models of CAC, currently exist^{153,154}. Thus, there is a significant need for suitable animal models to study microbe-host interactions in the context of CAC. Here we introduce a novel microbiota-driven model of CAC and extensively characterise the mucosal microbiota to identify tumour-associated bacteria.

Work from our laboratory previously identified a role for nATF6 in tumour formation in homozygous mice, with heterozygous mice never developing tumours without an additional

trigger, such as acute inflammation¹⁵⁵. Our results demonstrate that chronic low-grade inflammation, induced by IL-10 loss can also trigger tumorigenesis. Comparison of gross measures of inflammation, cytokine expression and protein-based markers of intestinal inflammation revealed increased inflammation in *nAtf6* overexpressing mice, suggesting that the combination of *Il10* KO and nATF6 activation in IECs can exacerbate colitis. In line with these results, ATF6 has been previously identified as a marker of pre-cancerous lesions in CAC and has been demonstrated to increase expression of proinflammatory cytokines *in vitro*^{172,177}. In concert with an existing genetic susceptibility, nATF6 activation in IECs may therefore exacerbate inflammation or disrupt intestinal homeostasis, nudging IECs towards an oncogenic phenotype. Interestingly, despite the considerable inflammation observed at later timepoints, tumour formation was not preceded by a marked colitic phenotype in susceptible *nAtf6*^{tg/wt;-/-} mice but only became apparent after tumour onset. Histological score did, however, correlate with an increase in tumour burden, which may indicate that in this context inflammation may be more important in tumour progression, rather than initiation. Although inflammation is generally involved in both, here, other functions disrupted by the loss of IL-10 signalling may be involved. IL-10 not only suppresses pro-inflammatory signalling but also plays a key role in homeostasis of the intestinal epithelium, governing epithelial barrier integrity, IEC differentiation and mucus production^{260–263}. To assess alterations in the mucosal barrier in *nAtf6*^{tg/wt;-/-} mice we measured GC numbers and mucus penetrability, finding reduced GC numbers and decreased distance between bacteria and the epithelium post-tumour. Importantly, IL-10 maintains folding capacity under ER stress, thus the absence of IL-10 and unresolved ER stress in this model likely leads to a severely reduced capacity to produce mucus²⁶¹. As mucus plays a key role in segregating potentially immunostimulatory microbes from host cells, disruption can lead to disease. Enhanced bacterial penetration of the mucus layer, increases the possibility of bacterial interaction with the epithelium and has been found in various other mouse models of CRC as well as in patients^{153,154,264}.

To assess the role of the mucosal microbiota in tumour formation in *nAtf6*^{tg/wt;-/-} mice we again carried out microbial profiling across tumour progression, as well as spatial profiling, observing limited differences in microbial community structure overall. Certain findings from bulk mucosal data contrasted with the spatial analysis, particularly with regard to alpha diversity. Species richness, evenness and phylogenetic diversity were significantly reduced at 12 weeks in tumour compared to non-tumour in the bulk mucosal dataset, however spatial data did not support this at every site, indicating that such reductions are likely site-specific and do not occur consistently along the length of the colon. Supporting this, significant variation was observed in alpha diversity along the colon, with distal portions displaying reduced diversity independent of genotype, while stratification by region – comparing tumour-susceptible locations to controls – revealed an increase in the magnitude of microbial change. Spatially resolved data may therefore be particularly important in the context of CAC, as this method may identify alterations that analysis of individual sites cannot. Indeed, inflammation in the *Il10*^{-/-} mouse model is often patchy, which could be a plausible explanation for the observed site-specific differences²²¹. It is possible that microbial alterations as well as tumour formation, require both inflammation and nATF6 activation, and that this combination may not be present along the entire length of the colon. The majority of tumours in *nAtf6*^{tg/wt;-/-} mice occurred at sites one and two, in the extreme proximal colon, which may indicate that this combination most often occurs here. Considering the colonic mucus layer exhibits an increasing thickness and density gradient along the length of the colon, these two sites, microbial contact with the epithelium and resultant inflammation is more likely here than elsewhere in the *nAtf6*^{tg/wt;-/-} colon²⁵⁹. Subsequent work could address this by examining the spatial distribution of colitis in *nAtf6*^{tg/wt;-/-} mice. The limited microbial community changes observed between genotypes without stratification however are in contrast with other models of CAC, where generally

marked changes in both alpha and beta diversity and microbial composition are present^{153,154}. This inconsistency and the overall limited differences between genotypes may alternatively imply monoallelic nATF6 activation has a restricted impact on the microbiota in *Il10* KO mice. Given IL-10's role as a key mediator of intestinal homeostasis, particularly in tolerance to commensal antigens, its loss may therefore supplant community level effects driven by nATF6 or tumour presence^{260,265}. Our study is limited in this regard however, as we did not include *nAtf6*^{tg/wt;+/+} controls to specifically address the microbial impact of *Il10* KO.

We additionally tested the driver-passenger model in *nAtf6*^{tg/wt;-/-} mice as a framework to explain microbial changes occurring with tumour development and were unable to identify any alterations between pre-tumour and post-tumour timepoints. This would suggest this model may not explain microbial succession in CAC development. Several CRC-associated microbes have been identified as enriched in IBD patients, such as *pks+* *E. coli* and *F. nucleatum*, which may imply that the conditions present prior to tumour formation in CAC may already select for tumour-promoting microbes^{129,228,266}. Despite this, it is important to note that the microbiota under chronic inflammation shows significant temporal variability which even with our moderate sample size may impede identification of passenger and driver taxa³. Notwithstanding, under our proposed expansion of the driver-passenger model, we could identify several SD ASVs, classified as *Bacteroides*, *Lactobacillus* and *Mucispirillum*, all of which were decreased in NR mice and preferentially localised to tumour-susceptible proximal sites over distal regions. Of note, we also identified *Mucispirillum* as an ID in the *nAtf6*^{IEC} model, implicating this microbe in tumour formation in two disparate models. Yet, why *Mucispirillum* was identified as an ID in one model and SD in the other is unclear. *Mucispirillum* can utilise nitrate as a terminal electron acceptor, which is particularly abundant in the inflamed gut, promoting the expansion of pro-inflammatory bacteria, thus it may be able to sustain its presence in the *nAtf6*^{IEC};*Il10*^{-/-} model by exploiting nitrate, whereas in the *nAtf6*^{IEC} model, the comparative lack of inflammation could select against its presence^{55,155}. Nevertheless, the identification of *Mucispirillum* as a driver taxon in two different models, warrants further research, to better understand its potential role in tumorigenesis.

Microbiota-dependent models such as the *nAtf6*^{IEC};*Il10*^{-/-} are essential if the involvement of microbes in CRC and CAC is to be clinically realised as they facilitate the disentanglement of host-related and microbiota-related mechanisms. Here we have introduced and characterised a microbiota-driven mouse model for CAC, uncovered a key role for the combined action of nATF6 activation and *Il10* KO in disrupting intestinal homeostasis, and identified putative microbes driving tumour formation.

5.3 Linking shifts in metabolite environment to tumour-associated microbiota

UPR signalling has been demonstrated to play a crucial role in intestinal homeostasis^{155,172,175,267,268}. Our previous results indicated nATF6 activation can modulate the microbiota towards a pro-tumorigenic state, however the mechanism by which this occurs remained elusive. During ER stress ATF6 has been shown to promote inflammatory signalling, which could explain the altered intestinal microbiota, yet tumorigenesis occurs independent of inflammation in *nAtf6*^{IEC} mice, and additionally did not precede tumour formation in *nAtf6*^{IEC};*Il10*^{-/-} mice, indicating nATF6 alters the intestinal environment by other means^{155,228}. Since tumour formation and indeed nATF6 activation brings marked changes in metabolism, we focused on the metabolome as a potential actor in mediating these alterations^{269,270}. Here we performed untargeted metabolomics to determine nATF6-driven alterations in the intestinal

metabolite milieu and additionally integrated mucosal microbiota profiles with metabolomic profiles of luminal content to understand how nATF6 activation and associated tumour formation modulate microbial communities, as well as identifying metabolite alterations which may modulate driver colonisation.

Untargeted LC-TOF-MS measurement of caecal samples revealed a significant impact of nATF6 activation on the luminal metabolite environment, leading to a clear enrichment of fatty acid and phospholipid derivatives. A similar lipid enrichment was observed in tumour samples from both *nAtf6*^{tg/tg} mice and *nAtf6*^{tg/wt;-/-} mice, however many of these lipid species were already enriched before tumour onset, implicating nATF6 as the main driver behind this observation. What remains less clear however, is the exact source of this lipid enrichment and whether this arises from host or microbial changes. Untargeted measurement of tissue samples from *nAtf6*^{tg/tg} mice revealed an increased presence of similar LCFA and lysophospholipids as in caecal content suggesting these metabolites are host-derived in this model, however this was not observed in *nAtf6*^{tg/wt;-/-} mice. Notably, fewer differential metabolites were identified in *nAtf6*^{tg/wt;-/-} compared to control than in *nAtf6*^{tg/tg} mice, in both caecal and tissue. Since *I110* KO has been demonstrated to induce alteration in fatty acids (including LCFA), this may mask differences between genotypes, while still increasing the overall lipid pool compared to *I110* positive and floxed *nAtf6* WT mice^{271,272}. Future work could address this by comparing *nAtf6*^{fl/fl;-/-} and *nAtf6*^{tg/wt;-/-} mice to *nAtf6*^{fl/fl;+/+} mice to comprehensively characterise metabolite changes induced by both *I110* KO and monoallelic nATF6 activation. Nevertheless, although bacteria produce both LCFA and lysophospholipids, the observed enrichment in *nAtf6*^{tg/wt;-/-} is likely also host-derived²⁷³. Despite the fact that the same metabolites enriched in caecal content of tumour-bearing *nAtf6*^{tg/wt;-/-} mice, were not observed in colonic tissue, several other lysophospholipids and fatty acid derivatives were increased, suggesting altered host metabolism is at least partially involved.

It is well established that the UPR plays an important role in lipid metabolism, particularly XBP1 which upregulates genes involved in lipid synthesis upon activation^{160,161,274}. ATF6 thus also indirectly modulates lipid metabolism via activation of XBP1²⁷⁵. Less is known however, regarding the specific contribution of ATF6 to lipid metabolism. Recently, a non-canonical lipid-mediated mechanism of ATF6 activation was discovered which triggers an alternative transcriptional program, upregulating genes involved in lipid metabolism¹⁶⁵. Similarly, *nAtf6* OE in the zebrafish liver has been demonstrated to upregulate fatty acid synthase (*FASN*) and other targets involved in fatty acid elongation¹⁶⁴. In line with our observation, this would support a role for nATF6 activation in modulating lipid and in particular fatty acid metabolism. How exactly an nATF6-driven increase in fatty acid synthesis or elongation would translate into increased fatty-acid pool in the intestinal lumen, however, remains uncertain. Under normal physiological circumstances, colonic epithelial cells renew approximately every 3-days, where old cells undergo apoptosis and are shed into the lumen²⁷⁶. Increased epithelial shedding is evident in multiple pathologies, particularly those involving disruption to the gut barrier, thus an increase in nATF6-driven fatty acid synthesis and/or elongation and concomitant shedding could underlie the observed altered luminal LCFA pool^{277,278}. Indeed, the observed increase in LCFA host tissue in *nAtf6*^{tg/tg} mice would support this. Considering the observation that the majority of regulated metabolites in *nAtf6* mice, comprised lysophospholipids and LCFAs, an alternative explanation is that nATF6 activates phospholipases, which could cleave epithelial membrane phospholipids, into lysophospholipids and fatty acids, releasing them into the intestinal lumen²⁷⁹. Activation of the ATF6 α isoform of ATF6 has been shown *in vitro* to increase the release of arachidonic acid via the phospholipase *Pla2g4a* in prostate cancer-derived cells, which may indicate that ATF6 can regulate phospholipase activity in other cell types as well. Furthermore, we previously demonstrated Signal transducer and activator of transcription

3 (STAT3) activation in *nAtf6^{tg/tg}* mice, of which the phospholipases *Pla2g2a* and *Pla2g5* are known targets^{155,280}. Regardless of the mechanism, nATF6 activation clearly increases the luminal LCFA pool, altering the intestinal microenvironment.

To examine how changes in the metabolite environment relate to microbial alterations we performed multi-omic integration of mucosal microbiota data with luminal untargeted metabolomic data. In both mouse models, the altered fatty acid environment and associated increase in LCFA appeared to be linked to the microbiota and correlated with driver abundance. Many free fatty acids, in particular, LCFA are bactericidal or bacteriostatic, thus changes in the concentrations of these compounds in the gut are likely to have a knock-on effect on the microbiota^{281,282}. Indeed, shifts in intestinal LCFA composition have previously been linked to changes in microbiota. In a mouse model of alcoholic liver disease, dysbiosis was accompanied by a reduction in saturated LCFA levels, while LCFA supplementation restored a normal microbiota²³⁴. Moreover, microbially derived LCFA can themselves, alter microbiota composition. Here we observed a marked increase in LCFA levels. Considering the antibacterial effects of these compounds it is possible that the nATF6-driven increase in LCFA leads to a population-level decline of normal microbiota, allowing more FA-resistant taxa to gain a foothold. This is further supported by the fact that many taxa which are normally resident in the colonic mucosa such as Lachnospiraceae and Ruminococcaceae are Gram-positive taxa, which are particularly sensitive to LCFA-induced toxicity, and accordingly, both were decreased in tumour environments²⁸³. In parallel with this, the majority of ID and SD taxa identified were Gram-negative, which are intrinsically more resistant to killing by LCFA²⁸⁴. Intriguingly, *Lactobacillus* was identified as a SD in both models and appeared to be highly correlated with the observed fatty acid enrichment in *nAtf6* overexpressing environments. Lactobacilli are Gram-positive but can survive in fatty acid-rich environments via the action of *ohyA*^{235,285}. Predicted genomic content suggested an enrichment of *ohyA* in tumour samples, with its major product 10-hydroxystearic acid also tumour enriched. Further supporting these observations, *ex vivo* co-culture of caecal content from non-tumour *nAtf6^{fl/fl}* and *nAtf6^{fl/fl;-/-}* mice, led to *Lactobacillus*-dominated communities, reflecting the compositional changes seen in *nAtf6^{tg/tg}* and *nAtf6^{tg/wt;-/-}* genotypes. Subsequent work could confirm this *in vivo* via gavage of LCFA to control genotypes and measuring microbial composition to determine if these changes recapitulate those seen in *nAtf6^{tg/tg}* and *nAtf6^{tg/wt;-/-}* mice. Together, this may indicate a novel mechanism by which nATF6 activation can indirectly alter the composition of the gut microbiota and promote colonisation by driver microbes. It is important to note however, that our study did not directly address the impact of these LCFA on individual microbes, thus the observed associations could also reflect indirect effects.

Concluding Remarks

A growing list of bacteria are increasingly linked to CRC, yet distinguishing between those directly instigating tumour formation, those promoting progression and bystanders remains challenging. Moreover, mechanisms which might licence driver bacteria to initiate tumorigenesis are also poorly understood. This work is unique in employing the driver-passenger model as a framework to separate putative tumour-initiating taxa from bystanders and also represents one of the first efforts to investigate microbial changes occurring in parallel with tumour development experimentally in CRC and CAC mouse models. We find evidence of marked shifts in microbial composition between pre-tumour and tumour onset in inflammation-independent models of CRC but not in CAC and propose that understanding of microbial succession in CRC should be extended to include adaptable microbes which can survive or potentially thrive in changing environmental conditions. Furthermore, we provide a

method by which individual ASVs can be directly mapped to tumour sites within the colon, further facilitating identification of tumour-associated microbiota. Additionally, we link nATF6-activation-mediated changes in the intestinal fatty acid pool to changes in the microbiota. Given the potent antimicrobial activity of many fatty acids, particularly longer chain varieties, this may represent a previously unknown method by which the host can alter the microbiota, and by which dysbiosis is induced in disease states.

Future work could expand on this by examining the identified putative driver taxa's ability to promote disease. Colonisation with single driver strains or consortia in susceptible GF mice, for example, would conclusively prove whether or not these taxa are true drivers, either alone or in combination. Furthermore, feeding experiments with microbiota-correlated LCFAs in non-susceptible mice would allow validation of their microbial composition-altering effects, as well as addressing whether exogenous administration of these metabolites can increase the abundance of driver microbes. Building upon this as a framework for identification of tumour-promoting microbes and environmental mechanisms underlying their colonisation will facilitate the development of targeted interventions to prevent or impede CRC.

List of Figures

Figure 1. Mucosa-associated bacteria and mucosal defence.	4
Figure 2. Pathogenesis of Colorectal and colitis-associated cancer.	8
Figure 3. The driver-passenger model of colorectal tumorigenesis.....	11
Figure 4. Activation of the endoplasmic reticulum unfolded protein response.....	13
Figure 5. Schematic of spatial organisation experimental design.....	22
Figure 6. Monoallelic activation of nATF6 alters the colonic mucosal microbiota.	29
Figure 7. Increasing gene dose augments modulation of colonic mucosa-associated microbiota at the 5-week timepoint.	31
Figure 8. Mucosa-associated microbiota profiles are markedly altered in tumour-bearing <i>nAtf6</i> ^{IEC} mice.....	32
Figure 9. Machine learning models trained on mucosa-associated microbiota profiles better discriminate phenotype compared to luminal.....	34
Figure 10. Consensus of discriminating features between non-tumour and tumour phenotypes across all trained models.	35
Figure 11. Identification of putative passenger and driver taxa in <i>nAtf6</i> ^{tg/tg} mice.	37
Figure 12. Driver taxa in <i>nAtf6</i> ^{tg/tg} mice are not tumour specific.	38
Figure 13. Mucosa-associated microbial profiles differ along the length of the colon irrespective of phenotype.....	40
Figure 14. Differentially abundant taxa between genotypes vary by site.....	42
Figure 15. ATF6 activation is the major driver behind microbial community alterations.	44
Figure 16. <i>nAtf6</i> ^{tg/tg} mucosal samples share a core set of enriched taxa.	45
Figure 17. Driver and passenger taxa are reproducible across cohorts.	46
Figure 18. nATF6 activation licences widespread expansion of driver taxa.	48
Figure 19. The <i>nAtf6</i> ^{IEC} ; <i>Il10</i> ^{-/-} mouse model.	50
Figure 20. <i>Il10</i> knockout enhances susceptibility to tumorigenesis but does not restore homozygous risk.	52
Figure 21. <i>nAtf6</i> ^{tg/wt;-/-} mice display enhanced colitis compared to <i>nAtf6</i> ^{fl/fl;-/-} mice.	54
Figure 22. Goblet cell loss occurs after tumour-onset and is associated with inflammation.	56
Figure 23. Enhanced mucus penetrability accompanies tumorigenesis.	58
Figure 24. Caecal dysbiosis precedes tumorigenesis in <i>nAtf6</i> ^{tg/wt;-/-} mice.....	60
Figure 25. Mucosa-associated microbiota profiles are altered regardless of tumour status in <i>nAtf6</i> ^{tg/wt;-/-} mice.....	61
Figure 26. Mucosa-associated microbiota profiles separate responders from non-responders.	63
Figure 27. Machine learning models trained on mucosal microbiota profiles better discriminate phenotype compared to luminal.	65
Figure 28. Consensus of discriminating features between non-tumour and tumour phenotypes across all trained models.	66
Figure 29. Mucosal microbiota shifts from pre- to post-tumour are not consistent with the driver-passenger model in <i>nAtf6</i> ^{tg/wt;-/-} mice.....	68
Figure 30. Mucosal microbial profiles show limited community level shifts between different sites in <i>nAtf6</i> ^{IEC} ; <i>Il10</i> ^{-/-} mice.	70
Figure 31. Differentially abundant taxa are mainly restricted to tumour-susceptible sites.	72

Figure 32. Alterations in mucosa-associated microbiota in <i>nAtf6</i> ^{tg/wt;-/-} responder mice are restricted to tumour-susceptible sites.	74
Figure 33. Sustained driver ASVs localise to tumour-susceptible sites.	76
Figure 34. nATF6 activation alters the luminal metabolome, independent of tumour formation.	78
Figure 35. The <i>nAtf6</i> ^{tg/tg} tumour metabolite environment is enriched in lipids and depleted in amino acids and peptides, prior to tumour onset.	79
Figure 36. Fatty acid derivatives and other lipids define the <i>nAtf6</i> ^{tg/wt;-/-} metabolite milieu.	80
Figure 37. Fatty acid and lysophospholipids are enriched in colonic tissue of <i>nAtf6</i> ^{tg/tg} and <i>nAtf6</i> ^{tg/wt;-/-} mice.	82
Figure 38. Multi-omic integration links tumour-associated taxa to altered lipid environment in <i>nAtf6</i> ^{tg/tg} mice.	84
Figure 39. Multi-omic integration links tumour-associated taxa to altered lipid environment in <i>nAtf6</i> ^{tg/wt;-/-} mice.	85
Figure 40. Microbiota correlated long-chain fatty acids are enriched in <i>nAtf6</i> ^{tg/tg} mice. ...	87
Figure 41. Microbiota correlated long-chain fatty acids are enriched in <i>nAtf6</i> ^{tg/wt;-/-} mice.	88
Figure 42. nATF6-enriched long-chain fatty acids modulate driver colonisation.	90
Figure 43. Evidence of microbial adaptation to the altered fatty acid environment.	92
Figure 44. <i>Ex vivo</i> co-culture with LCFA recapitulates <i>in vivo</i> shifts in microbial composition and predicted function in caecal content from non-susceptible mice.	94

List of Tables

Table 1. Mouse lines and genotype descriptions.	16
Table 2. Genotyping primers used in this study.	17
Table 3. Antibodies used for immunofluorescence and immunohistochemistry and their respective concentrations.	19
Table 4. Protocol for cDNA synthesis.	20
Table 5. qPCR primers used in this study.	20
Table 6. Spike-in DNA sequences.	24

Glossary

16S	16S rRNA small ribosomal subunit. Composed of conserved and hypervariable regions which can be used as a molecular barcode to profile bacteria and archaea.
Alpha diversity	Species diversity within samples. Measures include Richness, Shannon Effective, and Faith's phylogenetic diversity.
ASV	Amplicon sequence variant. Error-corrected reads representing true biological sequences. Also known as Exact sequence variants (ESV), sub-operational-taxonomic units (sOTU) or zero-radius operational-taxonomic units (zOTU)
AUC (AUCROC)	Area under the receiver operator curve. A measure used to assess the predictive accuracy of a machine learning model to discriminate between classes.
Beta diversity	Between sample diversity. A measure of how dissimilar samples are.
Clr	Centred-log ratio transform. A data transformation computed by calculating the log-ratio of each taxon relative to the geometric mean.
Cross-validation (CV)	A method of estimating the generalizability of a machine learning model by resampling a subset of data.
Dysbiosis	An altered microbiota state usually associated with disease and often characterised by low diversity and an increased abundance of potentially harmful microbes.
Evenness	A diversity measure reflecting the proportion of species of equal abundance within a sample.
Faith's phylogenetic diversity	A phylogenetic measure of alpha diversity, calculated as the sum of branch lengths in a phylogenetic tree.
False-discovery rate	Proportion of findings falsely identified as significant.
Germfree (GF)	Animal models lacking a natural microbiome.
Gnotobiotic	Animal models with a known, defined microbiome composition. Includes GF and consortia or mono-colonized animals.
GUniFrac	Generalized UniFrac. An extension of the UniFrac algorithm which gives less weight to rare and highly abundant lineages, allowing more broad changes in composition to be detected.
Microbiota	A microbial population in a given, well-defined environment.
Microbiome	Refers to a group of microorganisms, their genomes, metabolic activity, habitat, and environmental conditions.
Metabolome	The small-molecule component of the microbiota.
Metagenome	The entire gene content present in a given sample, including functional potential.
NMDS	Non-metric multidimensional scaling. An ordination method which attempts to represent distances between samples, as closely as possible in a low-dimensional space.
PERMANOVA	Non-parametric multivariate ANOVA. Used to assess differences in beta diversity between groups.
Machine learning	Use of data and algorithms to learn from data, identify patterns and generate predictions without explicit instruction.
Shannon Effective diversity	Alpha diversity measure. Accounts for the number of observed species as well as evenness. Expressed as the exponent of the Shannon index.
Richness	Alpha diversity metric measuring the total number of observed species/ASVs.

ROC

Receiver-operator curve. Visual aid to assess machine learning classifier performance, plotting true positive rate against false positive rate.

Abbreviations

<i>A. muciniphila</i>	<i>Akkermansia muciniphila</i>
ASV	Amplicon sequence variant
ATF6	Activating transcription factor 6
APC	Adenomatous polyposis coli
AUC	Area Under the Curve
BA	Bile acid
<i>B. fragilis</i>	<i>Bacteroides fragilis</i>
BLAST	Basic Local Alignment Search Tool
BMI	Body Mass Index
BRAF	B-Raf proto-oncogene/v-Raf murine sarcoma viral oncogene homolog B
CAC	Colitis-associated cancer
CD	Crohn's Disease
CFU	Colony Forming Unit
CIP2A	Cellular Inhibitor of PP2A
CLR	Centred-log ratio transformation
CRC	Colorectal cancer
CXCL1	C-X-C motif chemokine ligand 1
DAB	3,3'-Diaminobenzidine
dNTP	Deoxy-nucleotide triphosphate
<i>E. coli</i>	<i>Escherichia coli</i>
EC	Enzyme Commission
EDTA	Ethylenediaminetetraacetic acid
eIF2 α	Eukaryotic Initiation Factor 2
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated protein degradation
ETBF	Enterotoxigenic <i>Bacteroides fragilis</i>
<i>F. nucleatum</i>	<i>Fusobacterium nucleatum</i>
FASN	Fatty acid synthase
FC	Fold-change
FDR	False-discovery rate
FFPE	Formalin-fixed paraffin-embedded
FISH	Fluorescent <i>in situ</i> hybridisation
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC	Goblet cell
GF	Germ-free
GI	Gastrointestinal
GRP78	Glucose-regulated protein 78
GUniFrac	Generalized UniFrac
H&E	Haematoxylin & Eosin
HA	Hemagglutinin
HDI	Human development index
HMP	Human Microbiome Project
IBD	Inflammatory Bowel Disease
ID	Initial driver

IEC	Intestinal epithelial cell
IF	Immunofluorescence
IFN γ	Interferon-gamma
IgA	Immunoglobulin A
IHC	Immunohistochemistry
IL	interleukin
IP-10/CXCL10	Interferon gamma-induced protein 10/C-X-C motif chemokine ligand 10
IRE1 α	Inositol requiring enzyme 1-alpha
ISC	Intestinal stem cell
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	Knockout
KRAS	KRAS proto-oncogene/Kirsten Rat Sarcoma Viral Oncogene Homologue
L1L	L1-regularised lasso regression
<i>L. reuteri</i>	<i>Lactobacillus reuteri</i>
LCFA	Long-chain fatty acid
LC-TOF-MS	Liquid Chromatography-Time of Flight-Mass Spectrometry
LDA	Linear discriminant analysis
LEfSe	Linear discriminant analysis Effect Size
LI	Large intestine
LOOCV	Leave-one-out cross-validation
LPCN2	Lipocalin-2
LPS	Lipopolysaccharide
ML	Machine learning
MMLV	Moloney Murine Leukaemia Virus Reverse Transcriptase
MS	Mass Spectrometry
MyD88	Myeloid differentiation primary response 88
NF- κ B	Nuclear factor-kappa B
NLR	Nod-like receptor
NMDS	Non-Metric Multidimensional Scaling
NR	Non-responder
OE	Overexpression
<i>ohyA</i>	oleate hydratase
PAS/AB	Periodic acid Schiff-alcian blue
PBS	Phosphate Buffered Saline
PCA	Principle Component Analysis
PCR	Polymerase Chain Reaction
PERK	PKR-like ER kinase
pH	potential of hydrogen
PICRUSt2	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2
<i>Pla2g</i>	Phospholipase A2
PRR	Pattern recognition receptor
QC	Quality control
qPCR	Quantitative Polymerase Chain Reaction
R	Responder
RDP	The Ribosomal Database Project
RF	Random Forest
RIDD	Regulated IRE1 α -dependent decay

ROC	Receiver operating characteristic Curve
RR	Ridge regression
rRNA	Ribosomal ribonucleic acid
RT	Room Temperature
S1P/S2P	Site 1/2 protease
SCFA	Short Chain Fatty Acids
SD	Sustained driver
SI	Small intestine
SPF	Specific Pathogen Free
sPLS-DA	Sparse partial least-squares discriminant analysis
SREBP2	Sterol regulatory element-binding protein 2
STAT3	Signal transducer and activator of transcription 3
<i>S. gallolyticus</i>	<i>Streptococcus gallolyticus</i>
TCGA	The Cancer Genome Atlas
TLR	Toll-like receptor
TNF α	Tumour necrosis factor alpha
TP53	Tumor protein P53/Tumour suppressor P53
TRIF	TIR-domain-containing adapter-inducing interferon- β
UC	Ulcerative Colitis
UEA1	Ulex Europaeus Lectin 1
UHPLC	Ultra-High Performance Liquid Chromatography
UPL	Universal Probe Library
UPR	Unfolded protein response
V3/V4	16S hypervariable regions 3 and 4
WCA	Wilkins-Chalgren-Anaerobe Agar
WT	Wildtype
XBP1	X-box binding protein 1

References

1. Berg, G. *et al.* Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**, 1–22 (2020).
2. Metwaly, A., Reitmeier, S. & Haller, D. Microbiome risk profiles as biomarkers for inflammatory and metabolic disorders. *Nat. Rev. Gastroenterol. Hepatol.* **19**, 383–397 (2022).
3. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).
4. Zhou, W. *et al.* Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature* **569**, 663–671 (2019).
5. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
6. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
7. Gacesa, R. *et al.* Environmental factors shaping the gut microbiome in a Dutch population. *Nature* **604**, 732–739 (2022).
8. Zhernakova, A. *et al.* Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
9. Turnbaugh, P. J. *et al.* The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Sci. Transl. Med.* **1**, 1–19 (2009).
10. Schmidt, T. Extensive Transmission of Microbes along the Gastrointestinal Tract. *eLIFE* (2019).
11. Martinez-Guryn, K., Leone, V. & Chang, E. B. Regional Diversity of the Gastrointestinal Microbiome. *Cell Host Microbe* **26**, 314–324 (2019).
12. Tian, L. *et al.* Deciphering functional redundancy in the human microbiome. *Nat. Commun.* **11**, 6217 (2020).

13. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2015).
14. Olsson, L. M. *et al.* Dynamics of the normal gut microbiota: A longitudinal one-year population study in Sweden. *Cell Host Microbe* 1–14 (2022).
15. Groussin, M., Mazel, F. & Alm, E. J. Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host Microbe* **28**, 12–22 (2020).
16. Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E. & Borisy, G. G. Biogeography of a human oral microbiome at the micron scale. *Proc. Natl. Acad. Sci.* **113**, E791–E800 (2016).
17. Wilbert, S. A., Mark Welch, J. L. & Borisy, G. G. Spatial Ecology of the Human Tongue Dorsum Microbiome. *Cell Rep.* **30**, 4003-4015.e3 (2020).
18. Tuganbaev, T., Yoshida, K. & Honda, K. The effects of oral microbiota on health. *Science* **376**, 934–936 (2022).
19. Bik, E. M. *et al.* Molecular analysis of the bacterial microbiota in the human stomach. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 732–737 (2006).
20. Engstrand, L. & Lindberg, M. Helicobacter pylori and the gastric microbiota. *Best Pract. Res. Clin. Gastroenterol.* **27**, 39–45 (2013).
21. Delgado, S., Cabrera-Rubio, R., Mira, A., Suárez, A. & Mayo, B. Microbiological Survey of the Human Gastric Ecosystem Using Culturing and Pyrosequencing Methods. *Microb. Ecol.* **65**, 763–772 (2013).
22. Gu, S. *et al.* Bacterial Community Mapping of the Mouse Gastrointestinal Tract. *PLoS ONE* **8**, (2013).
23. Jung, C., Hugot, J.-P. & Barreau, F. Peyer’s Patches: The Immune Sensors of the Intestine. *Int. J. Inflamm.* **2010**, 1–12 (2010).
24. Boonjink, C. C. G. M. *et al.* High temporal and inter-individual variation detected in the human ileal microbiota. *Environ. Microbiol.* **12**, 3213–3227 (2010).

25. Szarka, L. A. & Camilleri, M. Methods for the assessment of small-bowel and colonic transit. *Semin. Nucl. Med.* **42**, 113–123 (2012).
26. Hayashi, H., Takahashi, R., Nishi, T., Sakamoto, M. & Benno, Y. Molecular analysis of jejunal, ileal, caecal and rectosigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J. Med. Microbiol.* **54**, 1093–1101 (2005).
27. Zoetendal, E. G. *et al.* The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J.* **6**, 1415–1426 (2012).
28. van den Bogert, B. *et al.* Diversity of human small intestinal Streptococcus and Veillonella populations. *FEMS Microbiol. Ecol.* **85**, 376–388 (2013).
29. I., I. *et al.* Specific microbiota direct the differentiation of Th17 cells in the mucosa of the small intestine. *Cell Host Microbe* **4**, 337–349 (2008).
30. Suzuki, T. A. & Nachman, M. W. Spatial heterogeneity of gut microbial composition along the gastrointestinal tract in natural populations of house mice. *PLoS ONE* **11**, 1– 15 (2016).
31. Lagkouvardos, I. *et al.* Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome* **7**, 28 (2019).
32. Marteau, P. *et al.* Comparative Study of Bacterial Groups within the Human Cecal and Fecal Microbiota. *Appl. Environ. Microbiol.* **67**, 4939–4942 (2001).
33. Nava, G. M., Friedrichsen, H. J. & Stappenbeck, T. S. Spatial organization of intestinal microbiota in the mouse ascending colon. *ISME J.* **5**, 627–38 (2011).
34. Riva, A. *et al.* A fiber-deprived diet disturbs the fine-scale spatial architecture of the murine colon microbiome. *Nat. Commun.* **10**, 4366 (2019).
35. Pedron, T. *et al.* A Crypt-Specific Core Microbiota Resides in the Mouse Colon. *mBio* **3**, e00116-12-e00116-12 (2012).

36. Daniel, N., Lécuyer, E. & Chassaing, B. Host/microbiota interactions in health and diseases—Time for mucosal microbiology! *Mucosal Immunol.* **14**, 1006–1016 (2021).
37. Paone, P. & Cani, P. D. Mucus barrier, mucins and gut microbiota: the expected slimy partners? *Gut* [gutjnl-2020-322260](https://doi.org/10.1136/gutjnl-2020-322260) (2020).
38. Pabst, O., Cerovic, V. & Hornef, M. Secretory IgA in the Coordination of Establishment and Maintenance of the Microbiota. *Trends Immunol.* **37**, 287–296 (2016).
39. Uchimura, Y. *et al.* Antibodies Set Boundaries Limiting Microbial Metabolite Penetration and the Resultant Mammalian Host Response. *Immunity* **49**, 545-559.e5 (2018).
40. Okumura, R. *et al.* Lypd8 promotes the segregation of flagellated microbiota and colonic epithelia. *Nature* **532**, 117–121 (2016).
41. Bergström, J. H. *et al.* Gram-positive bacteria are held at a distance in the colon mucus by the lectin-like protein ZG16. *Proc. Natl. Acad. Sci.* **113**, 13833–13838 (2016).
42. Propheter, D. C., Chara, A. L., Harris, T. A., Ruhn, K. A. & Hooper, L. V. Resistin-like molecule β is a bactericidal protein that promotes spatial segregation of the microbiota and the colonic epithelium. *Proc. Natl. Acad. Sci.* **114**, 11027–11033 (2017).
43. Lewis, J. D. *et al.* Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn’s Disease. *Cell Host Microbe* **18**, 489–500 (2015).
44. Tran, H. Q., Ley, R. E., Gewirtz, A. T. & Chassaing, B. Flagellin-elicited adaptive immunity suppresses flagellated microbiota and vaccinates against chronic inflammatory diseases. *Nat. Commun.* **10**, 5650 (2019).
45. Vaishnava, S. *et al.* The Antibacterial Lectin RegIII Promotes the Spatial Segregation of Microbiota and Host in the Intestine. *Science* **334**, 255–258 (2011).
46. Barr, J. J. *et al.* Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10771–6 (2013).
47. Hansson, G. C. & Johansson, M. E. V. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Gut Microbes* **1**, 51–54 (2010).

48. Johansson, M. E. V. *et al.* Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut* **63**, 281–291 (2014).
49. van der Post, S. *et al.* Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis. *Gut* **68**, 2142–2151 (2019).
50. Velcich, A. *et al.* Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science* **295**, 1726–9 (2002).
51. Yuan, C., Graham, M., Staley, C. & Subramanian, S. Mucosal Microbiota and Metabolome along the Intestinal Tract Reveal a Location-Specific Relationship. *mSystems* **5**, 1–11 (2020).
52. Savage, D. C., Dubos, R. & Schaedler, R. W. The gastrointestinal epithelium and its autochthonous bacterial flora. *J. Exp. Med.* **127**, 67–76 (1968).
53. Davis, C. P., Mulcahy, D., Takeuchi, A. & Savage, D. C. Location and description of spiral-shaped microorganisms in the normal rat cecum. *Infect. Immun.* **6**, 184–192 (1972).
54. Daniel, N., Lécuyer, E. & Chassaing, B. Host/microbiota interactions in health and diseases—Time for mucosal microbiology! *Mucosal Immunol.* **14**, 1006–1016 (2021).
55. Loy, A. *et al.* Lifestyle and Horizontal Gene Transfer-Mediated Evolution of *Mucispirillum schaedleri*, a Core Member of the Murine Gut Microbiota. *mSystems* **2**, e00171-16 (2017).
56. Li, H. *et al.* The outer mucus layer hosts a distinct intestinal microbial niche. *Nat. Commun.* **6**, 8292 (2015).
57. Ivanov, I. I. *et al.* Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria. *Cell* **139**, 485–498 (2009).
58. Wang, G. *et al.* A Surface Protein From *Lactobacillus plantarum* Increases the Adhesion of *Lactobacillus* Strains to Human Epithelial Cells. *Front. Microbiol.* **9**, 1–9 (2018).

59. Galley, J. D. *et al.* The structures of the colonic mucosa-associated and luminal microbial communities are distinct and differentially affected by a prolonged murine stressor. *Gut Microbes* **5**, 748–760 (2015).
60. Goto, Y. *et al.* Segmented filamentous bacteria antigens presented by intestinal dendritic cells drive mucosal Th17 cell differentiation. *Immunity* **40**, 594–607 (2014).
61. Zoetendal, E. G. *et al.* Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl. Environ. Microbiol.* **68**, 3401–3407 (2002).
62. Flemer, B. *et al.* The oral microbiota in colorectal cancer is distinctive and predictive. *Gut* gutjnl-2017-314814 (2017).
63. Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019).
64. Darwich, A. S., Aslam, U., Ashcroft, D. M. & Rostami-Hodjegan, A. Meta-Analysis of the Turnover of Intestinal Epithelia in Preclinical Animal Species and Humans. *Drug Metab. Dispos.* **42**, 2016–2022 (2014).
65. Rath, E. & Haller, D. Intestinal epithelial cell metabolism at the interface of microbial dysbiosis and tissue injury. *Mucosal Immunol.* **15**, 595–604 (2022).
66. Martens, E. C., Neumann, M. & Desai, M. S. Interactions of commensal and pathogenic microorganisms with the intestinal mucosal barrier. *Nat. Rev. Microbiol.* **16**, 457–470 (2018).
67. Eshleman, E. M. & Alenghat, T. Epithelial sensing of microbiota-derived signals. *Genes Immun.* **22**, 237–246 (2021).
68. Price, A. E. *et al.* A Map of Toll-like Receptor Expression in the Intestinal Epithelium Reveals Distinct Spatial, Cell Type-Specific, and Temporal Patterns. *Immunity* 1–16 (2018).
69. Abraham, C., Abreu, M. T. & Turner, J. R. Pattern Recognition Receptor Signaling and

- Cytokine Networks in Microbial Defenses and Regulation of Intestinal Barriers: Implications for Inflammatory Bowel Disease. *Gastroenterology* **162**, 1602-1616.e6 (2022).
70. Schlee, M. *et al.* Induction of human β -defensin 2 by the probiotic *Escherichia coli* Nissle 1917 is mediated through flagellin. *Infect. Immun.* **75**, 2399–2407 (2007).
 71. Lotz, M. *et al.* Cytokine-mediated control of lipopolysaccharide-induced activation of small intestinal epithelial cells. *Immunology* **122**, 306–315 (2007).
 72. Krautkramer, K. A., Fan, J. & Bäckhed, F. Gut microbial metabolites as multi-kingdom intermediates. *Nat. Rev. Microbiol.* **19**, 77–94 (2021).
 73. Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P. & Macfarlane, G. T. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **28**, 1221–1227 (1987).
 74. Thangaraju, M. *et al.* GPR109A Is a G-protein–Coupled Receptor for the Bacterial Fermentation Product Butyrate and Functions as a Tumor Suppressor in Colon. *Cancer Res.* **69**, 2826–2832 (2009).
 75. Guo, C. *et al.* Bile Acids Control Inflammation and Metabolic Disorder through Inhibition of NLRP3 Inflammasome. *Immunity* **45**, 802–816 (2016).
 76. Schaubert, J. *et al.* Expression of the cathelicidin LL-37 is modulated by short chain fatty acids in colonocytes: relevance of signalling pathways. *Gut* **52**, 735–741 (2003).
 77. Fellows, R. *et al.* Microbiota derived short chain fatty acids promote histone crotonylation in the colon through histone deacetylases. *Nat. Commun.* **9**, 1–15 (2018).
 78. Singh, B., Halestrap, A. P. & Paraskeva, C. Butyrate can act as a stimulator of growth or inducer of apoptosis in human colonic epithelial cell lines depending on the presence of alternative energy sources. *Carcinogenesis* **18**, 1265–1270 (1997).
 79. Byndloss, M. X. *et al.* Microbiota-activated PPAR- γ signaling inhibits dysbiotic *Enterobacteriaceae* expansion. *Science* **357**, 570–575 (2017).
 80. Nepelska, M. *et al.* Commensal gut bacteria modulate phosphorylation-dependent

- PPAR γ transcriptional activity in human intestinal epithelial cells. *Sci. Rep.* **7**, 43199 (2017).
81. Pathak, P. Intestine farnesoid X receptor agonist and the gut microbiota activate Gprotein bile acid receptor-1 signaling to improve metabolism. *Hepatol. Baltim. Md* **68**, 1574–1588 (2018).
 82. Li, W. *et al.* A bacterial bile acid metabolite modulates Treg activity through the nuclear hormone receptor NR4A1. *Cell Host Microbe* **29**, 1366-1377.e9 (2021).
 83. Kawamata, Y. *et al.* A G Protein-coupled Receptor Responsive to Bile Acids *. *J. Biol. Chem.* **278**, 9435–9440 (2003).
 84. Waclawiková, B. *et al.* Gut bacteria-derived 5-hydroxyindole is a potent stimulant of intestinal motility via its action on L-type calcium channels. *PLoS Biol.* **19**, e3001070 (2021).
 85. Babidge, W., Millard, S. & Roediger, W. Sulfides impair short chain fatty acid beta-oxidation at acyl-CoA dehydrogenase level in colonocytes: implications for ulcerative colitis. *Mol. Cell. Biochem.* **181**, 117–124 (1998).
 86. Attene-Ramos, M. S. *et al.* DNA damage and toxicogenomic analyses of hydrogen sulfide in human intestinal epithelial FHs 74 Int cells. *Environ. Mol. Mutagen.* **51**, 304– 314 (2010).
 87. Cao, Y. *et al.* Commensal microbiota from patients with inflammatory bowel disease produce genotoxic metabolites. *Science* **378**, eabm3233 (2022).
 88. Couturier-Maillard, A. *et al.* NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer. *J. Clin. Invest.* **13**, 155–156 (2013).
 89. Rakoff-Nahoum, S. & Medzhitov, R. Regulation of Spontaneous Intestinal Tumorigenesis Through the Adaptor Protein MyD88. *Science* **317**, 124–127 (2007).
 90. Vijay-Kumar, M. *et al.* Deletion of TLR5 results in spontaneous colitis in mice. *J. Clin. Invest.* **117**, 3909–3921 (2007).

91. Jenkins, B. J. Multifaceted Role of IRAK-M in the Promotion of Colon Carcinogenesis via Barrier Dysfunction and STAT3 Oncoprotein Stabilization in Tumors. *Cancer Cell* **29**, 615–617 (2016).
92. Hall, A. B. *et al.* A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Med.* **9**, 1–13 (2017).
93. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA. Cancer J. Clin.* **72**, 7–33 (2022).
94. Venugopal, A. & Stoffel, E. M. Colorectal Cancer in Young Adults. *Curr. Treat. Options Gastroenterol.* **17**, 89–98 (2019).
95. Weinberg, B. A. & Marshall, J. L. Colon Cancer in Young Adults: Trends and Their Implications. *Curr. Oncol. Rep.* **21**, 3 (2019).
96. Kitisin, K. & Mishra, L. Molecular Biology of Colorectal Cancer: New Targets. *Semin. Oncol.* **33**, 14–23 (2006).
97. O'Keefe, S. J. D. *et al.* Fat, fibre and cancer risk in African Americans and rural Africans. *Nat. Commun.* **6**, (2015).
98. Mehta, R. S. *et al.* Association of Dietary Patterns With Risk of Colorectal Cancer Subtypes Classified by Fusobacterium Nucleatum in Tumor Tissue. *JAMA Oncol.* **02215**, 1–7 (2017).
99. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
100. Deng, Y. Rectal Cancer in Asian vs. Western Countries: Why the Variation in Incidence? *Curr. Treat. Options Oncol.* **18**, (2017).
101. Siegel, R. L. *et al.* Colorectal cancer statistics, 2020. *CA. Cancer J. Clin.* **70**, 145–164 (2020).
102. Song, M., Chan, A. T. & Sun, J. Influence of the Gut Microbiome, Diet, and Environment on Risk of Colorectal Cancer. *Gastroenterology* (2019).
103. Kim, N. H. *et al.* Prevalence of and Risk Factors for Colorectal Neoplasia in

- Asymptomatic Young Adults (20–39 Years Old). *Clin. Gastroenterol. Hepatol.* **17**, 115–122 (2019).
104. Ijssennagger, N. *et al.* Gut microbiota facilitates dietary heme-induced epithelial hyperproliferation by opening the mucus barrier in colon. *Proc. Natl. Acad. Sci.* **112**, 201507645 (2015).
105. Thomas, A. M. *et al.* Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat. Med.* **25**, 667–678 (2019).
106. Helmus, D. S., Thompson, C. L., Zelenskiy, S., Tucker, T. C. & Li, L. Red Meat-derived heterocyclic amines increase risk of colon cancer: A population-based case-control study. *Nutr. Cancer* **65**, 1141–1150 (2013).
107. Shivappa, N. *et al.* Dietary inflammatory index and colorectal cancer risk—a meta-analysis. *Nutrients* **9**, (2017).
108. Ortiz, A. P., Thompson, C. L., Chak, A., Berger, N. A. & Li, L. Insulin resistance, central obesity, and risk of colorectal adenomas. *Cancer* **118**, 1774–1781 (2012).
109. Tsong, W. H. *et al.* Cigarettes and alcohol in relation to colorectal cancer: the Singapore Chinese Health Study. *Br. J. Cancer* **96**, 821–827 (2007).
110. Duricova, D., Burisch, J., Jess, T., Gower-Rousseau, C. & Lakatos, P. L. Age-related differences in presentation and course of inflammatory bowel disease: An update on the population-based literature. *J. Crohns Colitis* **8**, 1351–1361 (2014).
111. Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M. & Wallace, M. B. Colorectal cancer. *The Lancet* **394**, 1467–1480 (2019).
112. Li, J., Ma, X., Chakravarti, D., Shalpour, S. & DePinho, R. A. Genetic and biological hallmarks of colorectal cancer. *Genes Dev.* **35**, 787–820 (2021).
113. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
114. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).

115. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
116. Boutin, A. T. *et al.* Oncogenic Kras drives invasion and maintains metastases in colorectal cancer. *Genes Dev.* **31**, 370–382 (2017).
117. Porter, R. J., Arends, M. J., Churchhouse, A. M. D. & Din, S. Inflammatory Bowel Disease-Associated Colorectal Cancer: Translational Risks from Mechanisms to Medicines. *J. Crohns Colitis* **15**, 2131–2141 (2021).
118. Brentnall, T. A. *et al.* Mutations in the p53 gene: an early marker of neoplastic progression in ulcerative colitis. *Gastroenterology* **107**, 369–378 (1994).
119. Moparthi, L. & Koch, S. Wnt signaling in intestinal inflammation. *Differentiation* **108**, 24–32 (2019).
120. Robles, A. I. *et al.* Whole-Exome Sequencing Analyses of Inflammatory Bowel Disease-Associated Colorectal Cancers. *Gastroenterology* **150**, 931–943 (2016).
121. Richard, M. L. *et al.* Mucosa-associated microbiota dysbiosis in colitis associated cancer. *Gut Microbes* **9**, 131–142 (2018).
122. Wirbel, J. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **25**, 679–689 (2019).
123. Wirbel, J. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **25**, 679–689 (2019).
124. Yachida, S. *et al.* Metagenomic and metabolomic analyses reveal distinct stage specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* **25**, 968–976 (2019).
125. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* **10**, 766 (2014).
126. Kostic, A. D. *et al.* Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res.* **22**, 292–298 (2012).
127. Kostic, A. D. *et al.* *Fusobacterium nucleatum* Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host Microbe* **14**, 207–215

- (2013).
128. Wu, S. *et al.* A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat. Med.* **15**, 1016–1022 (2009).
 129. Prorok-Hamon, M. *et al.* Colonic mucosa-associated diffusely adherent afaC+ *Escherichia coli* expressing IpfA and pks are increased in inflammatory bowel disease and colon cancer. *Gut* **63**, 761–70 (2014).
 130. Rubinstein, M. R. *et al.* *Fusobacterium nucleatum* Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/ β -Catenin Signaling via its FadA Adhesin. *Cell Host Microbe* **14**, 195–206 (2013).
 131. Brewer, M. L. *et al.* *Fusobacterium* spp. target human CEACAM1 via the trimeric autotransporter adhesin CbpF. *J. Oral Microbiol.* **11**, 1–16 (2019).
 132. Boleij, A. *et al.* The *Bacteroides fragilis* toxin gene is prevalent in the colon mucosa of colorectal cancer patients. *Clin. Infect. Dis.* **60**, 208–215 (2015).
 133. Chung, L. *et al.* *Bacteroides fragilis* Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host Microbe* **23**, 421 (2018).
 134. Coughnoux, A. *et al.* Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut* 1–11 (2014).
 135. Lavelle, A. *et al.* Fecal microbiota and bile acids in IBD patients undergoing screening for colorectal cancer. *Gut Microbes.* **14**, 2078620 (2022).
 136. Wirbel, J. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **25**, 679–689 (2019).
 137. Powolny, A., Xu, J. & Loo, G. Deoxycholate induces DNA damage and apoptosis in human colon epithelial cells expressing either mutant or wild-type p53. *Int. J. Biochem. Cell Biol.* **33**, 193–203 (2001).
 138. Yazici, C. *et al.* Race-dependent association of sulfidogenic bacteria with colorectal cancer. *Gut* **66**, 1983–1994 (2017).

139. Nguyen, L. H. *et al.* Association Between Sulfur-Metabolizing Bacterial Communities in Stool and Risk of Distal Colorectal Cancer in Men. *Gastroenterology* **158**, 1313–1325 (2020).
140. Wirbel, J. *et al.* Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. *Genome Biol.* **22**, 93 (2021).
141. Zeller, G. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol* **10**, 766 (2014).
142. Flemer, B. *et al.* Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut* gutjnl-2015-309595 (2016).
143. Tjalsma, H., Boleij, A., Marchesi, J. R. & Dutilh, B. E. A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects. *Nat. Rev. Microbiol.* **10**, 575– 582 (2012).
144. Iftekhar, A. *et al.* Genomic aberrations after short-term exposure to colibactin-producing *E. coli* transform primary colon epithelial cells. *Nat. Commun.* **12**, 1003 (2021).
145. Sarshar, M. *et al.* Genetic diversity, phylogroup distribution and virulence gene profile of pks positive *Escherichia coli* colonizing human intestinal polyps. *Microb. Pathog.* **112**, 274–278 (2017).
146. Garza, D. R. *et al.* Metabolic models predict bacterial passengers in colorectal cancer. *Cancer Metab.* **8**, 1–13 (2020).
147. Tomkovich, S. *et al.* Human colon mucosal biofilms from healthy or colon cancer hosts are carcinogenic. *J. Clin. Invest.* **129**, (2019).
148. Bürtin, F., Mullins, C. S. & Linnebacher, M. Mouse models of colorectal cancer: Past, present and future perspectives. *World J. Gastroenterol.* **26**, 1394–1426 (2020).
149. Su, L. K. *et al.* Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* **256**, 668–670 (1992).
150. Evans, J. P. *et al.* From mice to men: Murine models of colorectal cancer for use in translational research. *Crit. Rev. Oncol. Hematol.* **98**, 94–105 (2016).

151. Dove, W. F. *et al.* Intestinal neoplasia in the Apc(Min) mouse: Independence from the microbial and natural killer (beige locus) status. *Cancer Res.* **57**, 812–814 (1997).
152. Washington, M. K. *et al.* Pathology of rodent models of intestinal cancer: Progress report and recommendations. *Gastroenterology* **144**, 705–717 (2013).
153. Slowicka, K. *et al.* Zeb2 drives invasive and microbiota-dependent colon carcinoma. *Nat. Cancer in press*, (2020).
154. Han, J. *et al.* ZFP90 drives the initiation of colitis-associated colorectal cancer via a microbiota-dependent strategy. *Gut Microbes* **13**, 1–20 (2021).
155. Coleman, O. I. *et al.* Activated ATF6 Induces Intestinal Dysbiosis and Innate Immune Response to Promote Colorectal Tumorigenesis. *Gastroenterology* **155**, 15391552.e12 (2018).
156. Grootjans, J., Kaser, A., Kaufman, R. J. & Blumberg, R. S. The unfolded protein response in immunity and inflammation. *Nat. Rev. Immunol.* **16**, 469–484 (2016).
157. Coleman, O. I. & Haller, D. ER Stress and the UPR in Shaping Intestinal Tissue Homeostasis and Immunity. *Front. Immunol.* **10**, 2825 (2019).
158. Hetz, C., Axten, J. M. & Patterson, J. B. Pharmacological targeting of the unfolded protein response for disease intervention. *Nat. Chem. Biol.* **15**, 764–775 (2019).
159. Jacquemyn, J., Cascalho, A. & Goodchild, R. E. The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. *EMBO Rep.* **18**, 1905–1921 (2017).
160. So, J. *et al.* Silencing of Lipid Metabolism Genes through IRE1 α -Mediated mRNA Decay Lowers Plasma Lipids in Mice. *Cell Metab.* **16**, 487–499 (2012).
161. Sriburi, R. *et al.* Coordinate regulation of phospholipid biosynthesis and secretory pathway gene expression in XBP-1(S)-induced endoplasmic reticulum biogenesis. *J. Biol. Chem.* **282**, 7024–7034 (2007).
162. Zeng, L. *et al.* ATF6 modulates SREBP2-mediated lipogenesis. *EMBO J.* **23**, 950–958 (2004).

163. Chen, X. *et al.* Hepatic ATF6 increases fatty acid oxidation to attenuate hepatic steatosis in mice through peroxisome proliferator-activated receptor α . *Diabetes* **65**, 1904–1915 (2016).
164. Howarth, D. L. *et al.* Activating Transcription Factor 6 Is Necessary and Sufficient for Alcoholic Fatty Liver Disease in Zebrafish. *PLoS Genet.* **10**, (2014).
165. Tam, A. B. *et al.* The UPR Activator ATF6 Responds to Proteotoxic and Lipotoxic Stress by Distinct Mechanisms. *Dev. Cell* **46**, 327–343.e7 (2018).
166. Rao, J. *et al.* Activating Transcription Factor 6 Mediates a Pro-Inflammatory Synergy Between ER Stress and TLR Activation in the Pathogenesis of Liver Ischemia Reperfusion Injury. *Transplantation* **98**, 40 (2014).
167. Yamazaki, H. *et al.* Activation of the Akt-NF- κ B Pathway by Subtilase Cytotoxin through the ATF6 Branch of the Unfolded Protein Response. *J. Immunol.* **183**, 1480–1487 (2009).
168. Gaudette, B. T., Jones, D. D., Bortnick, A., Argon, Y. & Allman, D. mTORC1 coordinates an immediate unfolded protein response-related transcriptome in activated B cells preceding antibody secretion. *Nat. Commun.* **11**, 723 (2020).
169. Gass, J. N., Jiang, H.-Y., Wek, R. C. & Brewer, J. W. The unfolded protein response of B-lymphocytes: PERK-independent development of antibody-secreting cells. *Mol. Immunol.* **45**, 1035–1043 (2008).
170. Zhao, F. *et al.* Disruption of Paneth and goblet cell homeostasis and increased endoplasmic reticulum stress in *Agr2*^{-/-} mice. *Dev. Biol.* **338**, 270–279 (2010).
171. Kaser, A. *et al.* XBP1 Links ER Stress to Intestinal Inflammation and Confers Genetic Risk for Human Inflammatory Bowel Disease. *Cell* **134**, 743–756 (2008).
172. Stengel, S. T. *et al.* Activating Transcription Factor 6 Mediates Inflammatory Signals in Intestinal Epithelial Cells Upon Endoplasmic Reticulum Stress. *Gastroenterology* (2020).
173. Lee, J. & Ozcan, U. Unfolded Protein Response Signaling and Metabolic Diseases. *J. Biol. Chem.* **289**, 1203–1211 (2014).

174. Zhang, H. S. *et al.* The endoplasmic reticulum stress sensor IRE1 α in intestinal epithelial cells is essential for protecting against colitis. *J. Biol. Chem.* **290**, 15237– 15336 (2015).
175. Kaser, A. *et al.* XBP1 Links ER Stress to Intestinal Inflammation and Confers Genetic Risk for Human Inflammatory Bowel Disease. *Cell* **134**, 743–756 (2008).
176. Ranjan, K., Hedl, M., Sinha, S., Zhang, X. & Abraham, C. Ubiquitination of ATF6 by disease-associated RNF186 promotes the innate receptor-induced unfolded protein response. *J. Clin. Invest.* **131**, (2021).
177. Hanaoka, M. *et al.* Expression of ATF6 as a marker of pre-cancerous atypical change in ulcerative colitis-associated colorectal cancer: a potential role in the management of dysplasia. *J. Gastroenterol.* (2017).
178. Liu, C.-Y. Y. *et al.* ER stress-related ATF6 upregulates CIP2A and contributes to poor prognosis of colon cancer. *Mol. Oncol.* **12**, 1706–1717 (2018).
179. Robertson, S. J. *et al.* Comparison of Co-housing and Littermate Methods for Microbiota Standardization in Mouse Models. *Cell Rep.* **27**, 1910-1919.e2 (2019).
180. Lobner, E. Epithelial Cell-Specific Induction of Activating Transcription Factor 6 Signaling Promotes Intestinal Dysbiosis and Colonic Tumorigenesis. (2018).
181. Moolenbeek, C. & Ruitenber, E. J. The ‘Swiss roll’: A simple technique for histological studies of the rodent intestine. *Lab. Anim.* **15**, 57–59 (1981).
182. 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 (2016).
183. Khaloian, S. *et al.* Mitochondrial impairment drives intestinal stem cell transition into dysfunctional Paneth cells predicting Crohn’s disease recurrence. *Gut* gutjnl-2019319514 (2020).
184. Godon, J. J., Zumstein, E., Dabert, P., Habouzit, F. & Moletta, R. Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis. *Appl. Environ. Microbiol.* **63**, 2802–13 (1997).

185. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
186. Tourlousse, D. M. *et al.* Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res.* **45**, gkw984 (2016).
187. Metwaly, A. *et al.* Integrated microbiota and metabolite profiles link Crohn's disease to sulfur metabolism. *Nat. Commun.* **11**, 1–15 (2020).
188. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
189. Lagkouvardos, I. *et al.* IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci. Rep.* **6**, 33721 (2016).
190. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–8 (2013).
191. Edgar, R. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv* 081257 (2016).
192. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
193. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
194. Davis, N. M., Proctor, D., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 221499 (2018).
195. Lagkouvardos, I., Fischer, S., Kumar, N. & Clavel, T. Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ* **5**, e2836 (2017).
196. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive

- Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **8**, (2013).
197. Faith, D.P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 1–10 (1991).
 198. Chen, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **28**, 2106–2113 (2012).
 199. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
 200. Nearing, J. T. *et al.* Microbiome differential abundance methods produce different results across 38 datasets. *Nat. Commun.* **13**, 1–16 (2022).
 201. Hiroshi Tsugawa, Tomas Cajka, Tobias Kind, Yan Ma, Brendan Higgins, Kazutaka Ikeda, Mitsuhiro Kanazawa, Jean VanderGheynst, Oliver Fiehn, and M. A. MS-DIAL: Data Independent MS/MS Deconvolution for Comprehensive. *Nat Methods* **12**, 523– 526 (2015).
 202. Wehrens, R. *et al.* Improved batch correction in untargeted MS-based metabolomics. *Metabolomics* **12**, 88 (2016).
 203. Douglas, G. M. *et al.* PICRUST2 for prediction of metagenome functions. *Nature Biotechnology* vol. 38 685–688 Preprint at (2020).
 204. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
 205. Wickham, H. *Ggplot2: Elegant graphics for data analysis.* (Springer International Publishing, 2016).
 206. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
 207. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 (1995), Publi. *J. R. Stat. Soc.* **57**, 289–300 (1995).

208. Sidak, D., Schwarzerová, J., Weckwerth, W. & Waldherr, S. Interpretable machine learning methods for predictions in systems biology from omics data. *Front. Mol. Biosci.* **9**, (2022).
209. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).
210. Singh, A. *et al.* DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**, 3055–3062 (2019).
211. Kuhn, M., Jackson, S. & Cimentada, J. corrr: Correlations in R. (2022).
212. Duncan, K., Carey-Ewend, K. & Vaishnava, S. Spatial analysis of gut microbiome reveals a distinct ecological niche associated with the mucus layer. *Gut Microbes* **13**, 1–21 (2021).
213. Donaldson, G. P. *et al.* Gut microbiota utilize immunoglobulin A for mucosal colonization. *Science* **800**, eaaq0926 (2018).
214. Watson, A. R. *et al.* Adaptive ecological processes and metabolic independence drive microbial colonization and resilience in the human gut. *bioRxiv* 2021.03.02.433653 (2021).
215. Geng, J. *et al.* Co-occurrence of driver and passenger bacteria in human colorectal cancer. *Gut Pathog.* **6**, 1–5 (2014).
216. Wang, Y. *et al.* Analyses of potential driver and passenger bacteria in human colorectal cancer. *Cancer Manag. Res.* **12**, 11553–11561 (2020).
217. Murphy, C. L. *et al.* Mapping the colorectal tumor microbiota. *Gut Microbes* **13**, 1–10 (2021).
218. Sheth, R. U. *et al.* Spatial metagenomic characterization of microbial biogeography in the gut. *Nat. Biotechnol.* (2019).
219. Brown, D. G. *et al.* Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer Metab.* **4**, 11 (2016).

220. Chiche, J., Brahim-Horn, M. C. & Pouyssegur, J. Tumour hypoxia induces a metabolic shift causing acidosis: a common feature in cancer. *J. Cell. Mol. Med.* **14**, 771–794 (2010).
221. Kühn, R., Löhler, J., Rennick, D., Rajewsky, K. & Müller, W. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**, 263–274 (1993).
222. Mähler, M. & Leiter, E. H. Genetic and environmental context determines the course of colitis developing in IL-10-deficient mice. *Inflamm. Bowel Dis.* **8**, 347–355 (2002).
223. Flo, T. H. *et al.* Lipocalin 2 mediates an innate immune response to bacterial infection by sequestering iron. *Nature* **432**, 917–921 (2004).
224. Goetz, D. H. *et al.* The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. *Mol. Cell* **10**, 1033–1043 (2002).
225. Toyonaga, T. *et al.* Lipocalin 2 prevents intestinal inflammation by enhancing phagocytic bacterial clearance in macrophages. *Sci. Rep.* **6**, 1–13 (2016).
226. Kitamoto, S. *et al.* Dietary l-serine confers a competitive fitness advantage to Enterobacteriaceae in the inflamed gut. *Nat. Microbiol.* **5**, 116–125 (2020).
227. Cevallos, S. A. *et al.* Increased Epithelial Oxygenation Links Colitis to an Expansion of Tumorigenic Bacteria. *mBio* **10**, 1–10 (2019).
228. Arthur, J. C. *et al.* Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota. *Science* **338**, 120–123 (2012).
229. Herp, S., Durai Raj, A. C., Salvado Silva, M., Woelfel, S. & Stecher, B. The human symbiont *Mucispirillum schaedleri*: causality in health and disease. *Med. Microbiol. Immunol. (Berl.)* **210**, 173–179 (2021).
230. Kuhajda, F. P. *et al.* Fatty acid synthesis: A potential selective target for antineoplastic therapy. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 6379–6383 (1994).
231. Ecker, J. *et al.* The Colorectal Cancer Lipidome: Identification of a Robust Tumor-Specific Lipid Species Signature. *Gastroenterology* **161**, 910-923.e19 (2021).

232. Ogino, S. *et al.* Cohort study of fatty acid synthase expression and patient survival in colon cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **26**, 5713–5720 (2008).
233. Zhao, L. *et al.* Saturated long-chain fatty acid-producing bacteria contribute to enhanced colonic motility in rats. *Microbiome* **6**, 1–16 (2018).
234. Chen, P. *et al.* Supplementation of saturated long-chain fatty acids maintains intestinal eubiosis and reduces ethanol-induced liver injury in mice. *Gastroenterology* **148**, 203–214.e16 (2015).
235. Subramanian, C., Frank, M. W., Batte, J. L., Whaley, S. G. & Rock, C. O. Oleate hydratase from *Staphylococcus aureus* protects against palmitoleic acid, the major antimicrobial fatty acid produced by mammalian skin. *J. Biol. Chem.* **294**, 9285–9294 (2019).
236. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **42**, 459–471 (2014).
237. Bevers, L. E., Pinkse, M. W. H., Verhaert, P. D. E. M. & Hagen, W. R. Oleate hydratase catalyzes the hydration of a nonactivated carbon-carbon bond. *J. Bacteriol.* **191**, 5010–5012 (2009).
238. Hirata, A. *et al.* A novel unsaturated fatty acid hydratase toward C16 to C22 fatty acids from *Lactobacillus acidophilus*. *J. Lipid Res.* **56**, 1340–1350 (2015).
239. Kuhl, G. C. *et al.* Oleate Hydratase in *Lactobacillus delbrueckii* subsp. *bulgaricus* LBP UFSC 2230 Catalyzes the Reversible Conversion between Linoleic Acid and Ricinoleic Acid. *Microbiol. Spectr.* **9**, 1–12 (2021).
240. Dejea, C. M. *et al.* Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592–597 (2018).
241. Kim, M. *et al.* Fecal metabolomic signatures in colorectal adenoma patients are associated with gut microbiota and early events of colorectal cancer pathogenesis. *mBio* **11**, 1–16 (2020).

242. Buc, E. *et al.* High Prevalence of Mucosa-Associated *E. coli* Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS ONE* **8**, (2013).
243. Saffarian, A. *et al.* Crypt- and Mucosa-Associated Core Microbiotas in Humans and Their Alteration in Colon Cancer Patients. *mBio* **10**, 1–20 (2019).
244. Yasuda, K. *et al.* Biogeography of the Intestinal Mucosal and Luminal Microbiome in the Rhesus Macaque. *Cell Host Microbe* **17**, 385–391 (2015).
245. Mo, Z. *et al.* Meta-analysis of 16S rRNA Microbial Data Identified Distinctive and Predictive Microbiota Dysbiosis in Colorectal Carcinoma Adjacent Tissue. *mSystems* **5**, 1–16 (2020).
246. Nakatsu, G. *et al.* Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat. Commun.* **6**, 1–9 (2015).
247. Liang, J. Q. *et al.* A novel faecal *Lachnoclostridium* marker for the non-invasive diagnosis of colorectal adenoma and cancer. *Gut* 1–10 (2019).
248. Daniel, S. G., Ball, C. L., Besselsen, D. G., Doetschman, T. & Hurwitz, B. L. Functional Changes in the Gut Microbiome Contribute to Transforming Growth Factor β -Deficient Colon Cancer. *mSystems* **2**, e00065-17 (2017).
249. Sugimura, N. *et al.* *Lactobacillus gallinarum* modulates the gut microbiota and produces anti-cancer metabolites to protect against colorectal tumourigenesis . *Gut* gutjnl-2020-323951 (2021).
250. Hezaveh, K. *et al.* Tryptophan-derived microbial metabolites activate the aryl hydrocarbon receptor in tumor-associated macrophages to suppress anti-tumor immunity. *Immunity* **55**, 324-340.e8 (2022).
251. Zegarra-Ruiz, D. F. *et al.* A Diet-Sensitive Commensal *Lactobacillus* Strain Mediates TLR7-Dependent Systemic Autoimmunity. *Cell Host Microbe* **25**, 113-127.e6 (2019).
252. Kumar, R. *et al.* *Streptococcus gallolyticus* subsp. *gallolyticus* promotes colorectal tumor development. *PLoS Pathog.* **13**, e1006440 (2017).
253. Taylor, J. C. *et al.* A type VII secretion system of *Streptococcus gallolyticus* subsp.

- gallyticus contributes to gut colonization and the development of colon tumors. *PLOS Pathog.* **17**, e1009182 (2021).
254. Yardeni, T. *et al.* Host mitochondria influence gut microbiome diversity: A role for ROS. *Sci. Signal.* **12**, eaaw3159 (2019).
255. Wu, M. *et al.* The differences between luminal microbiota and mucosal microbiota in mice. *J. Microbiol. Biotechnol.* **30**, 287–295 (2020).
256. Walter, J. Ecological role of lactobacilli in the gastrointestinal tract: Implications for fundamental and biomedical research. *Appl. Environ. Microbiol.* **74**, 4985–4996 (2008).
257. Walter, J. *et al.* A high-molecular-mass surface protein (Lsp) and methionine sulfoxide reductase B (MsrB) contribute to the ecological performance of *Lactobacillus reuteri* in the murine gut. *Appl. Environ. Microbiol.* **71**, 979–986 (2005).
258. Neumann, P. A. *et al.* Gut commensal bacteria and regional Wnt gene expression in the proximal versus distal colon. *Am. J. Pathol.* **184**, 592–599 (2014).
259. Kamphuis, J. B. J., Mercier-Bonin, M., Eutamène, H. & Theodorou, V. Mucus organisation is shaped by colonic content; a new view. *Sci. Rep.* **7**, 8527 (2017).
260. Landuyt, A. E. *et al.* ICOS ligand and IL-10 synergize to promote host–microbiota mutualism. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–11 (2021).
261. Hasnain, S. Z. *et al.* IL-10 promotes production of intestinal mucus by suppressing protein misfolding and endoplasmic reticulum stress in goblet cells. *Gastroenterology* **144**, 357–368.e9 (2013).
262. Jenkins, B. R. *et al.* Loss of interleukin-10 receptor disrupts intestinal epithelial cell proliferation and skews differentiation towards the goblet cell fate. *FASEB J.* **35**, (2021).
263. Sun, X. *et al.* Decline in intestinal mucosal IL-10 expression and decreased intestinal barrier function in a mouse model of total parenteral nutrition. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **294**, 139–147 (2007).
264. Dejea, C. M. *et al.* Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18321–6 (2014).

265. Shouval, D. S. *et al.* Interleukin-10 Receptor Signaling in Innate Immune Cells Regulates Mucosal Immune Tolerance and Anti-Inflammatory Macrophage Function. *Immunity* **40**, 706–719 (2014).
266. Strauss, J. *et al.* Invasive potential of gut mucosa-derived fusobacterium nucleatum positively correlates with IBD status of the host. *Inflamm. Bowel Dis.* **17**, 1971–1978 (2011).
267. Niederreiter, L. *et al.* ER stress transcription factor Xbp1 suppresses intestinal tumorigenesis and directs intestinal stem cells. *J. Exp. Med.* **210**, 2041–2056 (2013).
268. Heijmans, J. *et al.* ER Stress Causes Rapid Loss of Intestinal Epithelial Stemness through Activation of the Unfolded Protein Response. *Cell Rep.* **3**, 1128–1139 (2013).
269. Moncan, M. *et al.* Regulation of lipid metabolism by the unfolded protein response. *J. Cell. Mol. Med.* jcmm.16255 (2021).
270. Snaebjornsson, M. T., Janaki-Raman, S. & Schulze, A. Greasing the Wheels of the Cancer Machine: The Role of Lipid Metabolism in Cancer. *Cell Metab.* 1–15 (2019).
271. Martin, F.-P. J. *et al.* Metabolic Assessment of Gradual Development of Moderate Experimental Colitis in IL-10 Deficient Mice. *ACS Publ.* (2009).
272. Lin, H.-M. *et al.* Metabolomic Analysis Identifies Inflammatory and Noninflammatory Metabolic Effects of Genetic Modification in a Mouse Model of Crohn's Disease. *J. Proteome Res.* **9**, 1965–1975 (2010).
273. Parsons, J. B. & Rock, C. O. Bacterial Lipids: Metabolism and Membrane Homeostasis. *Prog. Lipid Res.* **52**, 249–276 (2013).
274. Lee, A.-H., Scapa, E. F., Cohen, D. E. & Glimcher, L. H. Regulation of Hepatic Lipogenesis by the Transcription Factor XBP1. *Science* **320**, 1492–1496 (2008).
275. Bommasamy, H. *et al.* ATF6 induces XBP1-independent expansion of the endoplasmic reticulum. *J. Cell Sci.* **122**, 1626–1636 (2009).
276. Heath, J. P. Epithelial cell migration in the intestine. *Cell Biol. Int.* **20**, 139–146 (1996).
277. Hagiwara, C., Tanaka, M. & Kudo, H. Increase in colorectal epithelial apoptotic cells in patients with ulcerative colitis ultimately requiring surgery. *J. Gastroenterol. Hepatol.*

- Aust.* **17**, 758–764 (2002).
278. Zeissig, S. *et al.* Downregulation of epithelial apoptosis and barrier repair in active Crohn's disease by tumour necrosis factor α antibody treatment. *Gut* **53**, 1295–1302 (2004).
279. Lambeau, G. & Gelb, M. H. Biochemistry and Physiology of Mammalian Secreted Phospholipases A2. *Annu. Rev. Biochem.* **77**, 495–520 (2008).
280. Wittkopf, N. *et al.* Activation of intestinal epithelial stat3 orchestrates tissue defense during gastrointestinal infection. *PLoS ONE* **10**, 1–14 (2015).
281. Desbois, A. P. & Smith, V. J. Antibacterial free fatty acids: Activities, mechanisms of action and biotechnological potential. *Appl. Microbiol. Biotechnol.* **85**, 1629–1642 (2010).
282. Pujo, J. *et al.* Bacteria-derived long chain fatty acid exhibits anti-inflammatory properties in colitis. *Gut* **70**, 1088–1097 (2021).
283. Yoon, B. K., Jackman, J. A., Valle-González, E. R. & Cho, N. J. *Antibacterial free fatty acids and monoglycerides: Biological activities, experimental testing, and therapeutic applications. International Journal of Molecular Sciences* vol. 19 (2018).
284. Galbraith, H., Miller, T. B., Paton, A. M. & Thompson, J. K. Antibacterial Activity of Long Chain Fatty Acids and the Reversal with Calcium, Magnesium, Ergocalciferol and Cholesterol. *J. Appl. Bacteriol.* **34**, 803–813 (1971).
285. Yang, B. *et al.* Myosin-cross-reactive antigens from four different lactic acid bacteria are fatty acid hydratases. *Biotechnol. Lett.* **35**, 75–81 (2013).

Acknowledgements

Firstly, I would like to thank Prof. Dirk Haller for providing me with the opportunity to join the Nutrition and Immunology lab and carry out my PhD under his supervision. Thank you for always challenging me and giving me the space and resources to develop and explore my scientific interests even if they may have veered away from the initial plan. It may not have always been easy but working in your lab has definitely shaped me into a better and more resilient scientist. In addition, I would like to thank Dr. Olivia Coleman for your supervision and scientific support, as well as your patience and the great deal of techniques you introduced me to. Of course, I owe a great deal to everyone else in the lab, without whom this work would not have been possible. In particular, I would like to thank Nico Gebhardt for his technical support and advice, without which I would have never been able to solve the many experimental problems I had along the way. Many of my experiments would not have been possible were it not for the fantastic support of Silvie, Simone, Sigrid, Steffie, Marlene and Sabine and the rest of the Animal team. Similarly, the completion of this work owes a great deal to everyone at the Microbiome Core facility, namely PD Dr. Klaus Neuhaus, Dr. Ilias Lagkourdos, Caroline Ziegler, and Angela Sachsenhauser. Thank you for the advice, the technical support and for processing all of the huge number of samples I brought over for sequencing during my time in the lab. Next, I would like to thank Sevana, Mohammed, Baraa, Niu and Marjolein for being there for me through the ups and downs and for providing a welcome relief from the stress of the lab more times than I can count. Thank you to everyone else in the Haller Lab as well, I really appreciate working with such a great group of people. Finally, I would like to thank my friends and family, and most of all my partner Leila. You were what got me through when times were tough and were always there to listen to my complaints and help me with problems, as well as celebrating my successes. As difficult an experience as this PhD was, I'm glad to have gone through it and came out of the other side and it certainly wouldn't have been possible without everyone mentioned here.

Curriculum Vitae

ADAM SORBIE

FEODOR-LYNEN STR. 17, MUNICH, GERMANY, 81377

TELEPHONE 089 440046203 adam.sorbie@med.uni-muenchen.de adam.sorbie.github.io

EDUCATION

2016 – current – PhD – Technical University of Munich

Activating transcription factor 6-altered fatty acid metabolism shapes the tumour-associated microbiota.

Supervisor: Prof. Dr. Dirk Haller

2015 – 2016 – MSc – The University of Glasgow

Infection Biology with a specialisation in Bacteriology – Merit

Thesis: Characterisation of 24 recent clinical *Escherichia coli* isolates from patients with Crohn's Disease.

Supervisor: Dr Donal Wall

2011 – 2015 – BSc – Edinburgh Napier University

Biological Sciences – 1st class honours

PROFESSIONAL EXPERIENCE

March 2016 – September 2016 – Writer/Editor – The GIST

During my master's I contributed to the student-led science magazine: The Glasgow insight into science and technology. I helped edit articles, also publishing one of my own.

Article: <https://the-gist.org/2016/04/can-antibiotics-actually-make-an-infection-worse/>

November 2015 – May 2016 – Tutor – Clever Tutors Ltd

Provided one-to-one tuition for high school-level Biology students, helping them to understand key concepts and prepare for exams.

June 2015 – August 2015 – Research intern – Edinburgh Napier University

Investigating the environmental extent of MRSA and VRE in coastal waters and sand around Edinburgh.

TEACHING EXPERIENCE

Teaching assistant – Experimental Immunology and Pathology Practical course – 2019-2020

Taught basic microbiology skills such as aseptic technique, bacterial culture, and enumeration to master's students.

Tutorial Leader – Introduction to the Unix shell – February 2020

Prepared and led a short 2-hour course introducing the basics of the Unix shell and bash scripting. This course is available here:

https://github.com/adamsorbie/unix_shell_course2020-02-14

Tutorial Leader - Introduction to High-performance computing on the LRZ Linux Cluster – July 2020

Prepared and led a short course introducing the basics of working and submitting jobs in a cluster environment. This course is available here: <https://github.com/adamsorbie/intro-toHPC-course-2020>

PUBLICATIONS & PRESENTATIONS

Publications

Coleman, Olivia I., Elena M. Lobner, Sandra Bierwirth, **Adam Sorbie**, Nadine Waldschmitt, Eva Rath, Emanuel Berger, Ilias Lagkourdos, Thomas Clavel, Kathleen D. McCoy, Achim Weber, Mathias Heikenwalder, Klaus-Peter Janssen, and Dirk Haller. 2018. "Activated ATF6 Induces Intestinal Dysbiosis and Innate Immune Response to Promote Colorectal Tumorigenesis." *Gastroenterology* 155(5):1539-1552.e12.

Littmann, Maria, Katharina Selig, Liel Cohen-Lavi, Yotam Frank, Peter Hönigschmid, Evans Kataka, Anja Mösch, Kun Qian, Avihai Ron, Sebastian Schmid, **Adam Sorbie**, Liran Szlak, Ayana Dagan-Wiener, Nir Ben-Tal, Masha Y. Niv, Daniel Razansky, Björn W. Schuller, Donna Ankerst, Tomer Hertz, and Burkhard Rost. 2020. "Validity of Machine Learning in Biology and Medicine Increased through Collaborations across Fields of Expertise." *Nature Machine Intelligence* 2(1):18–24.

Published abstracts

A. Sorbie, O.I Coleman, J. Kövilein, D. Haller "OP053 MUCOSA-ADHERENT MUCISPIRILLUM SPP. DRIVE ATF6-DEPENDENT COLORECTAL TUMOURIGENESIS IN MURINE MODELS." *UEG Journal* 9(S8): pp. 44.

Oral presentations

UEG Week 2021, Online

Poster presentations

11th Seeon Conference "Microbiota, Probiotics and Host", 2018 in Seeon, Germany

12th Seeon Conference "Microbiota, Probiotics and Host", 2019 in Seeon, Germany