Taxonomic and Functional Signatures of the Active Human Gut Microbiota in Homeostasis and Colorectal Cancer

Lamaudiere, Matthew

Award date:
2023

Awarding institution:
Coventry University

Taxonomic and Functional Signatures of the Active Human Gut Microbiota in Homeostasis and Colorectal Cancer

Matthew T. F. Lamaudière

Supervisor, Dr Igor Morozov

February 2023

Coventry University, Centre for Sport, Exercise and Life Sciences

&

University Hospitals Coventry and Warwickshire, NHS Trust

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy

Certificate of Ethical Approval

Applicant:

Matthew Lamaudiere

Project Title:

Taxonomic and Functional Signatures of the Active Human Gut Microbiota in Homeostasis and Colorectal Cancer.

This is to certify that the above-named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Medium Risk.

Date of approval:

14 October 2019

Project Reference Number:

P95024

Acknowledgements

Firstly, I would like to thank Igor, who to call a supervisor would be an insult. The last five and a half years of working together have seen some of the most difficult times of my life, but also the most rewarding and enjoyable by far, in no small part to his guidance and unconditional encouragement. For this I am grateful. I would also like to give a big thank you to Gareth, for always pushing through and finding a solution when needed and for always providing a fresh perspective. Thank you to Ramesh and Steven for continually supporting my development and providing feedback and advice. Thank you also to all the technical staff at Coventry, particularly Hugh and Keith for always facilitating the work.

Mum, Dad, Fiona, Gramps, Tanisha, Adam and Valerie. Thank you for always having belief that I could achieve something if I really wanted it, and I have always wanted nothing more than to make you all proud. You have always been there when I needed you to take the weight off my shoulders. I love you all so much.

List of Abbreviations

3'-reversibly tagged dNTPs (3'-RTa-dNTPs), aberrant NDRG family member 4 (NDRG4), absorbance units (AU), Acinetobacter baumannii, aconitate hydratase B (acnB), acyl-homoserine lactone (AHL), adenosine monophosphate (AMP), adjusted P-value (P_{adj}), ammonia (NH_3^+), Amplicon Sequence Variant (ASV), antibiotic (AB), antimicrobial peptides (AMPs), arginyl-tRNA synthetase/ligase (argS), auto-inducer 2 (AI-2), Benjamini-Hochberg (BH), body mass index (BMI), bone morphogenetic protein 3 (BMP3), carbon (C), carbon dioxide (CO_2), carboxyspermidine decarboxylase (CASDC), carboxyspermidine dehydrogenase (CASDH), Carcinoembryonic antigen (CEA), caspase recruitment domains (CARDs), catenin-β1 (CTNNB1), Cell-free DNA (cfDNA), chromosomal instability (CIN), cluster of differentiation (CD), colitis-associated cancer (CAC), colorectal cancer (CRC), competence and sporulation factor (CSF), complimentary DNA (cDNA), Comprehensive Antimicrobial Resistance Database (CARD), CpG island methylator phenotype (CIMP), Crohn's disease (CD), C-type lectin receptors (CLRs), cycle-inhibiting factor (CIF), cyclin-dependent kinase 1 (CDK1), cyclooxygenase-2 (COX-2), cytokine interleukin (IL), cytolethal distending toxin (CDT), Cytotoxic necrotising factor (CNF), damage-associated molecular patterns (DAMPs), dendritic cells (DCs), dimension (Dim), DL-Dithiothreitol (DTT), double stranded (ds), dihydrolipoyltranssuccinylase: E2 component of the 2-oxoglutarate dehydrogenase (sucB), E-cadherin (CDH1), effective half-maximal concentration (EC_{50}), electron transport chain (ETC), Embden-Meyerhof-Parnas (EMP), Enterococcus faecium, enterohemorrhagic Escherichia coli (EHEC), enteropathogenic Escherichia coli (EPEC), Entner–Doudoroff (ED), environmental DNA (eDNA), European Nucleotide Archive (ENA), example epidermal growth factor receptor (EGFR), Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter species (ESKAPE), false discovery rate (FDR), familial adenomatous polyposis (FAP), adenomatous polyposis coli (APC), family with sequence similarity 123B FAM123B frizzled class receptor (10FZD10), ferric iron (Fe^{3+}), ferrous iron (Fe^{2+}), flavin mononucleotide (FMN), gastrointestinal (GI), G-coupled receptor proteins (GCPRs), genomic DNA (gDNA), germ-free (GF), glucagon like peptide-1 (GLP-1), glutamate decarboxylase (GAD), glyceraldehyde-3-phosphate (GADP), Helicobacter pylori neutrophil-activating protein (HP-NAP), high fat diets (HFD), histone deacetylases (HDACs), horizontal transfer (HT), hydrogen chloride (HCl), hydrogen ion (H^+), hydrogen peroxide (H_2O_2), hydrogen sulfide (H_2S), hydroxyl radicals (·OH), hypochlorous acid (HOCl), IFN regulatory factors (IRFs), immunoglobulin A (IgA), inflammatory bowel disease (IBD), innate lymphoid cells (ILC), inorganic pyrophosphate (PPi), inorganic sulphate (SO_4^{2-}), insertion-deletion loop (IDL), interferon-γ (IFN-γ), interleukin-1 receptor-associated kinases (IRAKs), interleukin-1 receptors (TIR), intestinal epithelial cells

(IECs), intestinal gluconeogenesis (IGN), isocitrate dehydrogenase A (icdA), Kyoto Encyclopedia of Genes and Genomes (KEGG), leucine-rich repeats (LRR), linear discriminant analysis (LDA), linear discriminant analysis effect size (LEfSe), lipopolysaccharide (LPS), lithocholic acid (LCA), log2 fold change standard error (lfcSE), loss of heterozygosity (LOH), low-density lipoprotein receptor-related protein (5LRP5), Luria-Bertani (LB), macrophages (Mφ), malate dehydrogenase (mdh), messenger RNA (mRNA), methane (CH_4), methicillin-resistant Staphylococcus aureus (MRSA), microbiota-parasympathetic nervous system (PSNS), microsatellite instability (MSI), million (M), mismatch repair (MMR), mitogen-activated protein kinase (MAPK), muramyl dipeptide (MDP), National Health Service (NHS), natural killer (NK), nicotinamide adenine dinucleotide + hydrogen (NADH), nitrate (NO_3^-), nitric oxide (NO), nitrogen (N), NOD-like receptors (NLR), non-enterotoxigenic Bacteroides fragilis ($ETBF^-$), not applicable (NA), nuclear factor of activated T cells cytoplasmic-3 (NFATc3), Operational Taxonomic Units (OTUs), oxygen (O_2), paired-end raw reads (Raw PE), past medical history (PMH), pathogen-associated molecular patterns (PAMPs), pattern recognition receptors (PRRs), peptide YY (PYY), peptidoglycan (PG), peroxisome proliferator-activated receptor gamma (PPAR-γ), phenylacetyl-coenzyme A (phenylacetate-CoA), Phosphate-buffered saline (PBS), phosphatidylinositol 3 kinase (PI3KCA), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit-α (PI3KCA), phosphatase and tensin homologue (PTEN), polyamines (PAs), polyketide synthase (pks), polysaccharide A (PSA), principle component analysis (PCA), principle coordinate analysis (PCoA), protospacer adjacent motif (PAM), Pseudomonas aeruginosa and Enterobacter spp (ESKAPE), queuosine (Q), quorum sensing (QS), reactive nitrogen species (RNS), reactive oxygen species (ROS), receptor-interacting protein (RIP), Resistance Gene Identifier (RGI), retinoic acid-inducible gene I-like receptors (RIG-1), ribosomal DNA (rDNA), ribosomal RNA (rRNA), room temperature (Tr), S-adenosyl-L-methionine-dependent methyltransferase (SAM MTase), secreted frizzled-related protein (SFRP), secreted immunoglobulin A (sIgA), selenium (Se), sequencing (seq), Shine-Dalgarno (SD), short-chain fatty acids (SCFAs), single stranded (ss), SMAD family member 4 (SMAD4), sodium (Na), sodium chloride (NaCl), somatic copy number alterations (SCNAs), spermine oxidase (SMO), spermine synthase (SMS), Streptococcus gallolyticus member bacteria (SGMB), sulphate-reducing bacteria (SRB), superoxide ($O_2^{\cdot -}$), superoxide dismutase (SOD), T cell factor (TCF), T follicular helper cells (Tfh), TGFβ receptor 2 (TGFBR2), thousand (k), tight junctions (TJs), toll like receptor (TLR), transfer RNA (tRNA), transforming growth factor-β (TGFβ), tricarboxylic acid (TCA), Tris/Borate/EDTA (TBE), tumour microenvironment (TME), tumour necrosis factor α (TNF-α), ulcerative colitis (UC), University Hospital Coventry and Warwickshire NHS Trust (UHCW), whole genome sequencing (WGS), zonula occludens-1 (TJP1/ZO-1) and γ-aminobutyrate (GABA).

Contents

1.0 Abstract DNA-based studies have revealed the composition of the gut microbiota is changed in colorectal cancer (CRC) and implicated in its development and progression. However, these analyses are influenced by dead and dormant cell DNA, potentially leading to the overestimation of certain species contribution to the disease. Furthermore, gene expression is a regulated process and can be microbe-specific for the same stimuli which is overlooked by DNA-based analysis of functional potential. Therefore, to overcome these limitations, metatranscriptomes and 16S rDNA of faecal microbiota from CRC patients (n=10) and non-CRC volunteers (n=10) were subject to high throughput sequencing to characterise active microbial taxonomy and metabolic functions. These analyses revealed microbial abundance and activity are not always comparable. We describe sub-populations in both cohorts of 'hyper-active' (at least an order of magnitude higher levels of transcript relative to gene abundance) and dormant species, including CRC-associated *F. nucleatum* and *B. fragilis*, and probiotic genera *Lactobacillus* and *Bifidobacterium* respectively. This suggests that the CRC niche regulates species-specific changes in gene expression independently of genomic abundance. Surprisingly, the dominant and mostly conserved activity of the microbiome in both cohorts was metabolism of reactive oxygen species (ROS), a hallmark of inflammation, arguing the microbial population is responsible for maintaining physiological levels of ROS in the gut. Yet, the observed dysregulation of certain bacterial ROS scavenging pathways during CRC suggests a potentially transient mechanism for the gradual accumulation of genetic lesions that lead to disease development over time in an otherwise healthy individual. Taxonomic analysis also uncovered diminished activity of butyrate-producing bacteria and enhanced activity of clinically relevant ESKAPE, oral cavity and Enterobacteriaceae pathogens with no previous association to the malignancy. Functional analysis of expressed genomes uncovered a potential contribution of gut microbiota to known patient deficiencies for ferrous iron, carnitine, and folate, among others. This analysis also showed overexpression of multiple virulence factors, particularly genes related to host colonisation, biofilm formation, quorum sensing, genetic exchange, acid stress, and a highly expressed antibiotic (AB) resistome. *In vitro* investigation of bacterial acid and AB resistance determinants (particularly of ESKAPE members) in response to CRC-specific environmental pressures exposed mechanisms of possible long-term transcriptional memory. Microbes which likely faced previous oxidative, osmotic and/or acid stresses within the CRC gut increased expression of these determinants upon repeat exposure to a significant degree in comparison to the same species isolated from healthy individuals. This pioneering work uncovered the CRC-specific changes to active population composition and metabolic pathway expression. Crucially, from this we propose the mechanisms by which the human gut microbiota may protect against or drive CRC and reveal that exposure of the microbiota to CRC-specific pressures may condition them to adapt more readily to future stresses.

2.0 Introduction

Colorectal cancer (CRC) is one of the most prevalent, environment-driven cancers worldwide with a distinct lack of early diagnostic markers, killing >2,400 people daily, with nearly 2 million new cases annually. CRC has a poorly understood, multifactorial aetiology with over 95% of cases developing in people with no genetic predisposition (1). However, CRC is a genetic disease; mutations in tumour suppressor genes (2) and the DNA mismatch repair (MMR) mechanisms (3) promote uncontrolled cellular proliferation, primarily in glandular epithelial cells and drive the onset of CRC. Environmental, lifestyle factors and most recently the microbial community of the gut (the microbiota) have been associated with the disease. It has been proposed that the dysregulation of host immune functions promotes inflammation, epithelial damage, and stepwise accumulation of genetic lesions, leading to colonic crypt hyperplasia and CRC (4). However, the underpinning microbial activity *in vivo* which may lead to this disease is yet to be uncovered.

The relevance of gastrointestinal (GI) bacteria to health and disease is becoming ever more understood, and now termed the germ-organ due to its function and ultimate influence over human health (5). The GI microbiota is emerging as a primary environmental determinant of health (6), forming natural defences against pathogens and controlling host immune function, possessing anti-carcinogenic properties, and influencing host metabolism. The microbiota are also the primary source of numerous essential metabolites of the host, the balanced production of which is critical to homeostasis of the gut (7). Intestinal microorganisms and their anaerobically synthesised metabolic products (e.g. short-chain fatty acids, SCFAs) directly interact with the host to determine the metabolic activity of host colonocytes and in turn maintain anaerobic conditions and healthy mucus for commensal (beneficial) microbial growth (6). Changes to the composition of the microbiota, termed dysbiosis are related to CRC (8–11) while the prevalence of active gut microorganisms, their metabolism and the mechanisms by which their expressed genomes modulate CRC remains poorly understood. Therefore, elucidation of both the functional composition of the disease-associated microbiota and their active metabolic pathways is imperative for understanding microbial roles in host physiology and pathology.

The GI tract is the most densely colonised part of the body, the colon accommodates over 70% of all human microbes, numbering as many as 10^{14} (100 trillion) individual microorganisms (12). This community, which comprises bacteria, archaea, viruses, fungi, protozoa, and helminths, is termed the microbiota. The combined genome of these microbes has been termed the microbiome by Joshua Lederberg, the Nobel laureate, however this is now interchangeable with ‘microbiota’ (13). Understanding the impact of microorganisms on

human health guided a number of efforts to identify pathogens, leading to the development of culture-dependent methods for isolation and characterisation of individual microbes (14). The focus of microbial studies has long been on establishing compositions of bacterial communities via cultivation, isolation and characterisation of specific bacteria in laboratory settings (15). However, this approach has several limitations, including favouring the selection of certain species over others due to specific growth conditions/requirements; the majority of the GI microbes (>75%) are unculturable *in vitro* (16, 17). Through the advent of culture-independent, DNA-based techniques such as DNA hybridisation, PCR, and sequencing (18, 19), slow culture-dependent methods of identification of uncultured bacteria have become obsolete. A much more diverse population of GI microbiota has been characterised through sequencing of the bacterial 16S rRNA genes (rDNA) belonging to the 30S small ribosomal subunit (20). 16S rDNA sequences are present in all prokaryotes and archaea, are relatively short (~1.5kb) and possess both ultra-conserved and hypervariable regions (Fig. 4) (21). Similar sequences of variable regions are clustered into Operational Taxonomic Units (OTUs) and now Amplicon Sequence Variant (ASV) which are used for taxonomic profiling through 16S rRNA and other databases (e.g. Green Genes, Silva or the Ribosomal Database Project) (22, 23), allowing identification of bacteria based on sequence similarity, from family with >90% identity to genus (>95%) and species level-resolution (>97%) (24). The arrival of novel high-throughput profiling techniques, such as DNA-seq (shotgun), which allowed the quantification of gene abundance, has significantly improved our current understanding of the complexity and the functional potential of the GI microbial community, associating specific bacteria with disease. However, DNA-based taxonomic and functional analyses of the microbiota have a fundamental limitation, namely the inability to discriminate between dormant (persister/quiescent cells) and active species and genes.

2.1 Genetics, Epigenetics and Signalling Pathways of Colorectal Cancer Pathogenesis

Of the >1.9M new cases and ~935k deaths annually (as of 2020) (25), studies have shown that the pathology of those suffering from colorectal cancer has four (five including mixed subtype) distinct molecular subtypes. These four subtypes have three distinct mutational origins, ~95% occur sporadically (~25% of which have a family history of CRC but have no known genetic predisposition, suggesting a potential role of common lifestyle factors) and only 5% of cases are of a hereditary nature (26–28). Genetic and critically, numerous environmental factors (e.g. age, gut microbiota, diet, obesity and physical activity) have been associated with disease onset (29). Colon inflammation which triggers production of reactive oxygen species (ROS) and reactive nitrogen species (RNS) by innate immune cells (e.g. macrophages (Mφ) and neutrophils) or colonocytes, may lead to DNA damage and genome instability (30) as oxygen/nitrogen radicals supress DNA MMR mechanisms (31). Defects in

epithelial barrier function, the epithelium and bilayer of mucus that prevents paracellular translocation of microbes to the lamina propria and blood supply, allows bacterial colonisation and invasion and further enhances inflammation (11, 32–35). This, in turn, promotes synthesis of cytidine deaminase, CDA, which has been found to mutate e.g. *p53* and *c-MYC*, key cancer-associated genes and therefore promotes tumorigenesis (36). Mutations in these factors are associated with the acquisition of the cancer hallmarks of intestinal epithelial cells (IECs) (Fig. 1) (37). This results in the stepwise accumulation of genetic mutations and changes in epigenetic regulation (modifications of DNA structure and function without changing its sequence) (38, 39), leading to enhanced expression of oncogenes, down-regulation of tumour suppressor genes, and limiting replication of DNA.

Key Hallmarks of CRC

- **Avoiding immune destruction**, immune suppression in the tumour microenvironment by induction of local cytokines
- **Evading growth suppressors**, mutation and down-regulation of growth-inhibiting factors and their receptors
- **Genome instability and mutation**, inactivation of DNA repair mechanisms
- **Enabling replicative immortality**, inhibition of mechanisms that induce senescence and induction of telomerase activity
- **Deregulating cellular energetics**, aerobic glycolysis (Warburg phenomenon) and glutaminolysis
- **Tumour-promoting inflammation**, induction of growth-promoting and angiogenesis-promoting factors by secreted proteins made by local inflammatory cells
- **Inducing angiogenesis**, induction of the formation of new blood vessels
- **Resisting cell death**, escape from autonomous and paracrine mediators of apoptosis and other forms of cell death (necrosis or necroptosis)
- **Activating invasion and metastasis**, remodelling of the extracellular matrix to promote cell motility and induction of epithelial–mesenchymal transition

Fig. 1 | Key cancer hallmarks exhibited by intestinal epithelial cells during CRC. Genetic and environmental factors can contribute to the acquisition of cancerous traits through the stepwise accumulation of genetic lesions (mutations) and epigenetic modifications, leading to up-regulation of oncogenes and inactivation of tumour suppressor genes. Such loss of genetic/epigenetic stability is a striking feature of neoplastic alterations in intestinal epithelial cells (IECs), leading to CRC (40). Colorectal cancer cells generally originate from a single stem cell, containing cancerous mutations, which is located within the crypt (41) which undergo clonal proliferation, expanding the number of cancer stem cells, the first step in tumorigenesis. This clonal expansion leads to an epithelium with diminished immune function via loss of small intestinal Paneth cells, as well as Goblet cells, owing to a loss of mucin and antimicrobial peptide production, as the cancerous stem cells are immature, termed colonic crypt hyperplasia. Furthermore, these non-differentiated IECs have reduced oxygen consumption due to a metabolic shift from β-oxidation towards anaerobic glycolysis. Adapted from (42).

It is noteworthy that certain mutations are more common to different regions of the colorectum, for instance the proximal colon is featured more frequently with *BRAF* mutations, while the distal colon features *APC*, *KRAS* and *p53* mutations more often (43). The (epi)genetic instability can stem from microsatellite instability (MSI), chromosomal instability (CIN), CpG island (regions of DNA with multiple CpG dinucleotides within and around 40% of all promoters) methylator phenotype (CIMP), and somatic copy number alterations (SCNAs) (44), which also display anatomical biases (Fig. 2).

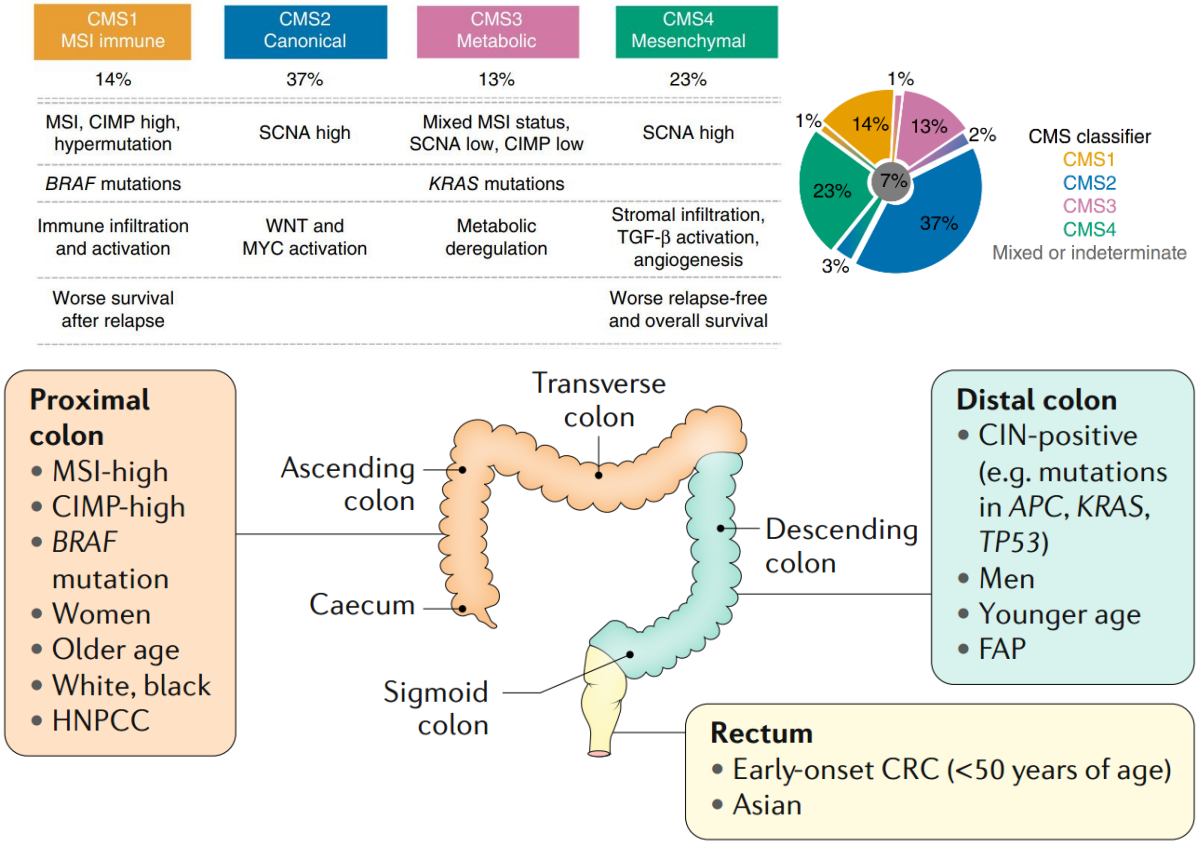


Fig. 2 | Consensus molecular subtypes of colorectal cancer. Genetic and phenotypic features of the different CMSs, their prevalence, anatomical, age and geographical frequencies. The association of rectum specific CRC with Asian populations is likely of genetic origin rather than geographical, as migrant Asian populations in e.g. the USA demonstrate increased risk of CRC onset compared to other ethnicities (45). CMS, consensus molecular subtype; CIN, chromosomal instability; MSI, microsatellite instability; CIMP, CpG island methylator phenotype; SCNA, somatic copy number alterations. Adapted from (43, 46).

Loss of DNA MMR, a mechanism that detects and corrects DNA replication errors in S-phase, wherein DNA polymerase incorporates the incorrect number of nucleotides in a long repetitive sequence, such as a microsatellite, can manifest as microsatellite instability, MSI. MSI exhibits a hyper mutable phenotype in their repetitive motifs. This is caused by deficiency of DNA mismatch repair pathways which recognise, and correct mismatches of DNA bases due to DNA replication errors, resulting in the formation of insertion-deletion loops (IDLs). These IDLs lead to the gradual accumulation of single base pair mismatches in both non-coding regions (such as promoters) and coding microsatellites. 15% of all CRC cases present with MSI, with 12% of these cases a consequence of acquired hypermethylation of the promoter of *hMLH1*, suppressing the efficiency of the MMR pathway (47). The *hMLH1* encoded protein is essential in forming heterodimers with other factors involved in MMR. Gene silencing is often a consequence of CIMP when excessive DNA methylation occurs within the oncogene promoter region (e.g. CpG islands), hence sporadic cases of MMR suppression take place almost exclusively as a consequence of CIMP-associated methylation of *hMLH1* (48).

Chromosome instability (CIN) was initially defined as abnormalities in copy numbers of chromosomes in clonal populations of CRC cells (49) and was predicted to, at first, occur at the stage of adenoma (Fig. 3), while playing a larger role in carcinomas (50). Increased frequency of errors in chromosome segregation during cell division and telomere dysfunction are the key features of CIN. Multiple studies have proposed several potential molecular bases for the CIN phenotype, with distinct segregation error characteristics including anaphase laggards, multipolar spindles, chromatin bridges and loss of heterozygosity (51). Moreover, CIN has also been associated with chromosomal rearrangements/translocations (52) including chromothripsis, distinguished by widespread chromosomal rearrangements and fluctuating copy numbers. CIN can affect expression of key genes, including *APC*, *p53, N*- and *K-RAS*, and *PI3KCA* (53), which are critical to controlling cell homeostasis. *APC*, the loss of which may also be a basis of CIN, regulates WNT/β-catenin signalling which drives expression of genes involved in tumorigenesis and invasion. Mutations in *p53*, the gene encoding for the main cell cycle check point regulator, result in uncontrolled cell cycle progression, leading to enhanced cell proliferation (54). *RAS* and *PI3KCA* mutations stimulate proliferation via permanent activation of MAP kinase (53). The CIN pathway is typically the cause of CRC, accounting for up to 85% of all cases (40). It is noteworthy that mutations in some genes (e.g. *APC* and *SMAD4*, a transcription factor which mediates TGF-β signal transduction) are common in all genome instability phenotypes, arguing that they are *the* major driving force behind CRC, whilst mutations in other genes are restricted to specific phenotypes (e.g. *BRAF* mutations are associated with CIMP-driven CRCs) (55).

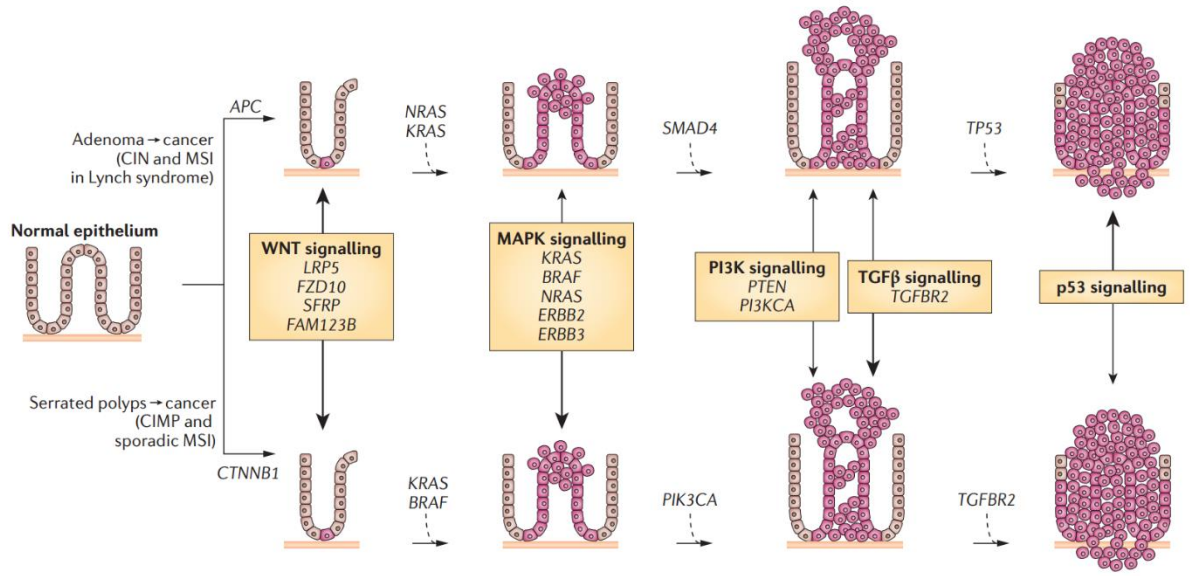


Fig. 3 | **Colorectal cancer progression pathways**. FAP, familial adenomatous polyposis (top) and HNPCC, non-polyposis colorectal cancer (bottom) pathways of CRC development include progression from healthy IECs to irregular crypt foci, proceeded by polyps followed by development from early cancer to malignant cancer. The sporadic pathway (top) includes the pathogenesis of tubular adenomas which progress to adenocarcinomas, while the alternate pathway (bottom) involves serrated polyps and their evolution to serrated CRC. Some mutated or epigenetically modified genes are common to both pathways, however others are distinct (e.g. CIMP (CpG island methylator phenotype) and *BRAF* mutations occur exclusively within the serrated pathway). The deregulated signalling pathways, overactive or inhibited, during CRC progression are indicated (arrow width signifies the significance of the contribution of said pathway to CRC pathogenesis). Lynch syndrome or hereditary nonpolyposis colorectal cancer, HNPCC; IECs, Intestinal epithelial cells; familial adenomatous polyposis, FAP; adenomatous polyposis coli, APC; chromosomal instability, CIN; catenin-β1, CTNNB1; family with sequence similarity, 123B FAM123B; frizzled class receptor, 10FZD10; low-density lipoprotein receptor-related protein, 5LRP5; mitogen-activated protein kinase, MAPK; microsatellite instability, MSI; phosphatidylinositol 3-kinase, PI3KCA; phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit-α, PI3KCA; phosphatase and tensin homologue, PTEN; secreted frizzled-related protein, SFRP; SMAD family member 4, SMAD4; transforming growth factor-β, TGFβ; TGFβ receptor 2, TGFBR2. Adapted from (42).

The WNT pathway is critical for cellular growth, tight junction (TJ) assembly (epithelial barrier) and stem-cell development via e.g. *c-MYC* (56), thus upon its alteration, cells can gain cancerous phenotypes. Loss of function of the WNT pathway can have differing effects depending upon the progression of the cancer, the pathway is often deregulated at the primary stages of CRC development, and the resultant neoplasia causes colonic crypt hyperplasia, beginning the cascade of deterioration of the intestinal epithelial barrier (57). This is critical in allowing pathogenic bacterial adhesion to the epithelium as well as exposing IECs to damaging microbial metabolites, leading to excessive cellular damage, inflammation and decreased genetic stability. The latter stages of disease progression are featured with weakened TJs due to loss of WNT signalling (58), leading to supressed cellular adhesion, migration and eventual metastasis (57). The timing and indeed the extent to which the WNT pathway is compromised, either via *APC* (adenomatous polyposis coli) inactivation (a negative regulator of the pathway with an 81% rate of mutation in a non-hyper mutated phenotype and 53% in the hyper mutable phenotype) (59) or other factors in the pathway (60) indicates whether the cancer is colitis-associated (CAC) or sporadic. In sporadic CRC WNT activation via *APC* mutation occurs during the early stages while the opposite is true of CAC regarding timing of *APC* inactivation (61). Loss of APC activity triggers the development of a cancerous phenotype as it targets β-catenin (required for adherens junction assembly) for destruction (acting as part of the scaffold for the β-catenin degradation complex). Formation of Tcf/Lef transcription factor complexes follows accumulation of β-catenin in the cytoplasm which promotes transcription following nuclear translocation (62). These complexes initiate DNA binding in a gene-specific manner while β-catenin acts to activate transcription (63). However, β-catenin is not used as a prognostic marker as it is generally overexpressed during CRC, while over-expression of *c-MYC* is seen as a good prognostic marker, denoting probability of metastasis (64).

Cell survival and proliferation are controlled by the MAPK/ERK and PI3K/AKT/mTOR pathways (65, 66). Therefore, tumour cells wherein these pathways are deregulated possess an increased capacity for proliferation and evading apoptosis (67). Common mutations within the pathways include *BRAF* and *KRAS*, aberrations causing the related proteins to become constitutively active regardless of any upstream stimuli, for example epidermal growth factor receptor (EGFR) activation (68). Additionally, EGFR homologs shown to undergo recombination(s) in the cytoplasmic, transmembrane and kinase domains (L858R and T790M), as well as being upregulated by ETF (EGFR-specific transcription factor) (69) can contribute to MAPK/ERK pathway activation in cancer (70). In CRC development, mutations in the GTPase proto-oncogene *KRAS* are often an early event affecting the MAPK/ERK signalling pathway. *KRAS* mutations appear in around 40% of cases and are associated with

poor prognosis and survival rates (71). The proto-oncogene *BRAF*, encoding a serine/threonine kinase, possesses mutations more often associated with MSI^+ CIMP-associated CRC. The predominant *RAF* mutation in CRC is BRAF-D594G, rather than BRAF-V600E, which is linked to poor prognosis with 15% of early and 6% of metastatic cases possessing the mutation (72).

Silencing of *p53* is detrimental to the cell, in around 40-50% of all CRC cases (50% in sporadic and 85% in CAC (73)) a mutation in the *p53* master tumour suppressor is found (74). Inactivation/deregulation of *p53* can be achieved either through loss-of-function mutations (e.g. R273H, R248W/Q and S241F) which can be introduced by cytosine deaminases (54), epigenetic inhibition of transcription due to CIMP (75) or gain-of-function mutations (such as R248W). Gain-of-function mutations (such as R248W) prolong NF-κB expression in a TNF-α dependent manner, increasing CAC and IEC damage (76), enhancing proliferation, survival, and invasion (77). In tumours, viral or host proteins (MDM2 and 4) can bind and inactivate TP53 causing replication arrest to fail (78). Hence, the cancer cells are less genetically stable and prone to increased chromosomal rearrangements and further accumulation of mutations, resulting in rapid production of malignant clones. This would explain why DNA-containing viruses do not induce damage responses, they can highjack TP53 and the active host replication machinery in S-phase in order to facilitate their own replication. Mutations in *p53* are also crucial in the transition from adenoma to carcinoma throughout pathogenesis (79). *p53* inactivation weakens TJs, disrupting the barrier function of colonocytes, thus enhancing NF-κB-dependent inflammation (80). Its silencing suppresses DNA repair mechanisms and enhances expression of pro-inflammatory IL-6, activation of STAT3 by phosphorylation and TNF signalling (81). Inflammation-dependent production of RNS, such as nitric oxide (NO), stabilises TP53 which in turn blocks *NOS2* expression and NO production, hence attenuating tumorigenesis (82) via cell cycle arrest, DNA repair and apoptosis as well as reducing oxidative/nitrosative stresses. Furthermore, TP53 drives oxidative phosphorylation in IECs, which is critical for maintaining the anaerobic gut environment, thus suppressing dysplasia (31). A mutation in the *p53* gene in the onset of CAC generally results in loss of heterozygosity (LOH), leading to complete loss-of-function of *p53*. *p53* LOH in turn allows progressive accumulation of DNA mutations in genes such as *APC*, resulting in carcinoma development (61). Interestingly, some bacteria, such as *Shigella flexneri* are capable of deregulating expression of *TP53* and suppress DNA repair in a *TP53*-dependent manner, hence increasing the risk of CRC development (83). The distinct status of the *p53* mutation determines the aggressiveness or metastatic features (e.g. invasive depth) of the CRC. The ~60% of CRC patients with wild-type *p53* have far better prognosis than those without, in part due to the chemo-resistance conferred by *TP53* CRC (84). However, *p53* is not used as a marker of CRC

as data regarding its association are statistically insufficient. Instead, *KRAS* and *APC* mutations (85) alongside bone morphogenetic protein 3 (*BMP3*), aberrant NDRG family member 4 (*NDRG4*) and the methylation/presence of β-actin, with more sensitive analytical tools are currently preferred (86). Molecular screening via biomarkers is an ever-evolving utility which relies upon innovative studies to analyse the metagenomics of tumour cells as well as the GI bacteria which closely influence its pathology and *vice versa*, with the microbiome directly affected by the disease.

2.2 Composition and Structure of the Gastrointestinal Microbiota

The human GI tract is home to the most diverse and dense population of microbes in/on the body. Within one individual, the ~10^{14} bacteria outnumber human cells by a ratio of 9:1, therefore bacteria have the potential to have a major influence on human health and disease and establishing their compositions in the gut has long been a challenging objective. The use of 16S rRNA genes to identify specific bacterial species dates back to 1977 (87), this principle underpinned the first system of sequence-based taxonomy and has had a remarkable impact on the field of microbiology. Infrequent mutations of certain bacterial rRNA gene regions due to their conserved function as scaffolds to build up the ribosome made 16S ribosomal DNA sequences a primary target to classify different bacterial species (21, 88). The composition and diversity of the gut microbiota was firstly studied through Sanger sequencing of full-length 16S rRNA genes which discovered a great number of uncharacterised gut species (around 75% of the currently classified microbiome) (16, 89). Further development of parallel deep-sequencing techniques, sequencing of short, from 70 to 500 nucleotides reads length led to a higher sensitivity and resolution producing more comprehensive estimates of the GI microbiota composition (20, 90). The conserved 16S rRNA gene regions are interspersed with nine variable (V) regions (V1-9) (Fig. 4) which can be used for taxonomic alignment, hence the V regions targeted have high inter-species disparity. The use of different V regions to classify taxonomy, e.g. V1-V3, V3-V4 or V4-V5 should be considered with respect to the native ecosystem. For example, the primary application of V1-V3 sequencing is for environmental taxonomic analysis of microbiomes, alongside the different sequencing platforms used to generate DNA reads (e.g. illumina, 454 and pyrosequencing) (91, 92). For identification of clinically relevant species and strains the length of the PCR fragments should be >500 bp to span the 16S rRNA gene hyper-variable regions (93). Amplification of specific regions by PCR allows identification of bacterial taxonomy of meta-samples, the data can be quantified and used for generating a broad profile of microbiomes. Yet, the use of 16S rDNA for taxonomic analysis has been found to have multiple limitations. rRNA gene-based sequencing can generate biased PCR due to the conserved nature of rRNA genes among closely related species, lack of standardisation because of the use of different variable (V1-

V9) regions as well as difficulties in OTUs assignment (94, 95). Importantly, given the high prevalence of horizontal gene transfer (HT) in bacterial communities, utilisation of the 16S rDNA as a single marker gene for phylogenetic analysis may underestimate diversity of such communities (96). Shotgun whole genome sequencing (WGS) can generate information on the genome structure of the whole microbial community, providing the genetic potential of the microbiome of the sample, through alignment of representative species identified by 16S sequencing with their genome (97). WGS builds up our knowledge on the composition of genes present in the microbiome, hence adding more detail to the phylogenetic characterisation and the microbial potential within specific settings.

The gut microbiome is extremely heterogenous, varying in composition from one individual to another (16). Microbial diversity can be influenced by myriad factors, from host ethnicity, age, geographical location, diet, exercise to even sleep pattern (98–101). Interestingly, age only appears to become a critical determinant of microbial composition in those >90 years of age (102), remaining relatively stable throughout life (103), despite previous studies reporting changes in the compositional diversity from earlier ages (104). However, more transient disruptions can also alter microbiome status, such as infection, drug treatment, stress, and acute gastrointestinal conditions (105–108). These fluctuating factors have the potential to amplify any differences in microbiome composition observed between individuals or critically, groups of different health statuses. The consequence of pathologies is often to lower estimates of microbial diversity (109–111); however, even 'healthy' individuals can often intermittently be exposed to transient factors, sometimes concurrently. Controlling for this in volunteer recruitment can be very difficult as there are only so many excluding criteria which can be applied, particularly for more subjective criteria such as stress or sleep quality. Reporting of e.g. dietary changes, alcohol consumption, or an individual's memory of previous drug treatments may also pose limitations to the design of associative studies. These fundamental limitations in the field can partially be addressed through changes in methodological approaches, such as correctional research or longitudinal studies and specific analytical tools, including principal component and co-ordinate analyses (PCA and PCoA) of microbiome diversity to account for potential co-founding affects (112). Moreover, as it is more stable relative to community composition, assessing the microbial metatranscriptome for both taxonomic and functional profiling reduces the likelihood of transient perturbations to the community influencing analyses. Although functional potential of the microbiome remains fairly consistent relative to species composition (101), the transcriptome appears more stable (113), therefore making it an important tool for understanding condition specific changes to the microbiome.

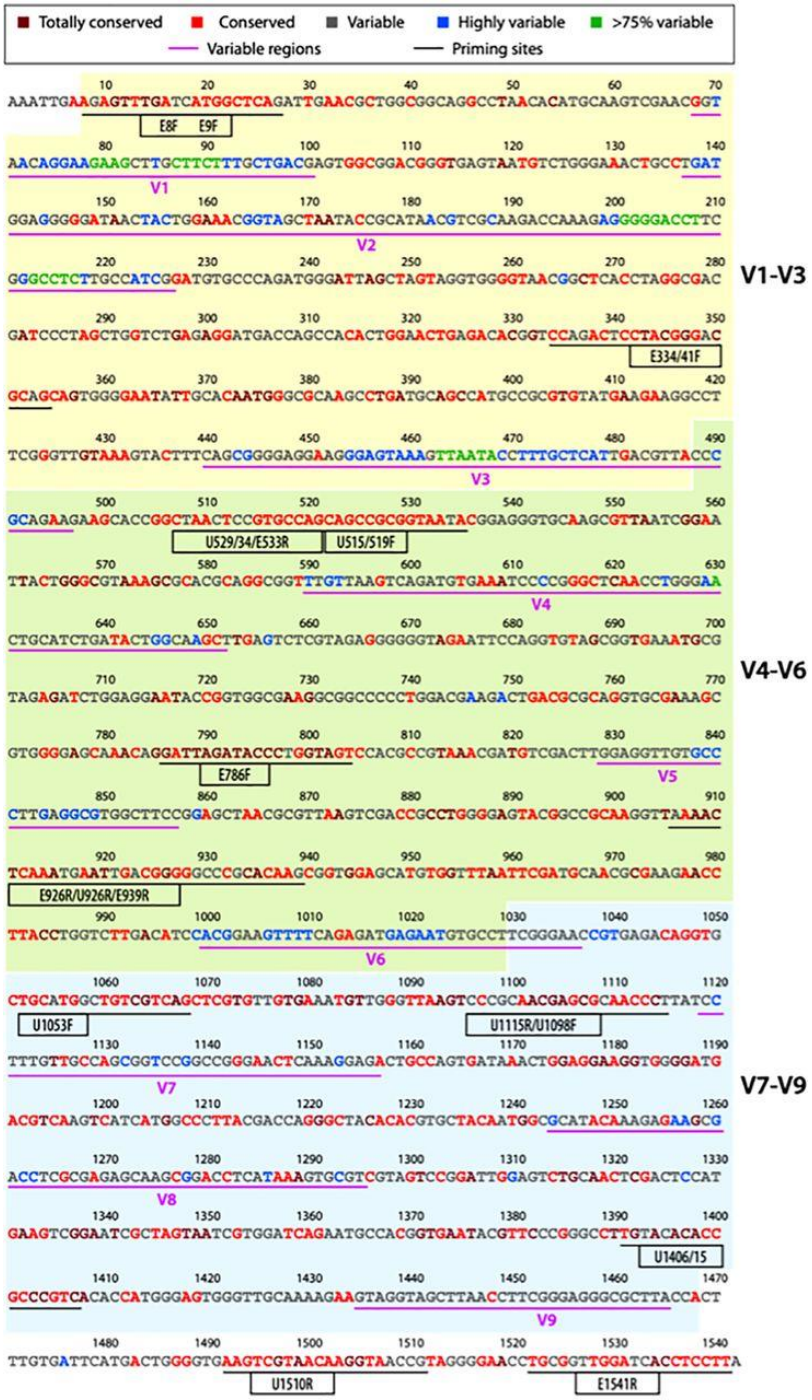


Fig. 4 | **Conserved and variable regions and features of the 16S rRNA gene.** Representative 16S rRNA gene (of *Escherichia coli*) showing positions of conserved regions, (hyper)variable (V) regions of which there are 9 and primer annealing sites for sequencing. Burgundy letters: totally conserved nucleotides; Red letters: conserved nucleotides; Black letters: variable nucleotides/sequences; Blue letters: highly variable nucleotides/sequences; Green nucleotides: ≥75% (hyper) variable nucleotides/sequences: Underlined in pink: variable regions; Underlined in black: primer annealing sites. Adapted from (92).

Illumina sequencing and pyrosequencing are sequencing platforms based on synthesis methods (sequencing through synthesis); single bases are detected upon their incorporation into the nascent DNA strand during polymerisation. In the case of Illumina sequencing single stranded (ss) fragmented DNA templates are hybridised with both 5' and 3' probes (oligos) which are the bases of primer annealing. A complementary strand is synthesised from the primed oligos using 3'-reversibly tagged dNTPs (3'-RTa-dNTPs). For each base the 3'-hydroxyl group is esterified with a distinct fluorescent residue tag, which when excited emits a distinct wavelength of light or absorbs a specific wavelength of light (114). Once detected, the pattern of emission or absorption spectra are interpreted as the sequence of four nucleotides. Pyrosequencing (pyrophosphate-based sequencing between 300 and 500 nts long) utilises the inorganic pyrophosphate (PPi) generated through the formation of the 5'-3' phosphodiester bond during polymerisation. PPi is converted to ATP in the presence of adenosine 5'-phosphosulfate, by the action of ATP sulphurylase, which subsequently drives luciferase to convert the substrate luciferin to oxyluciferin, generating light which is proportional to the amount of ATP generated (115). Each round of PCR a different base is used, thus the sequence of light emitted per round of PCR is analysed via a pyrogram and interpreted as the sequence of four nucleotides.

The continuous use of DNA-seq significantly increases the quality, number and size/depth of databases and hence requires high-throughput bioinformatics tools to infer the biological relevance of the growing amounts of data (116, 117). Analyses of high-throughput data requires the deployment of sophisticated analytical suites, e.g. Trimmomatic (118) for quality filtering and MG-RAST (119) for mapping millions of DNA reads against a non-redundant database. These can be used in combination with more specialised bioinformatics tools depending upon the purposes, whether it is BOWTIE-2 for alignment (120), DESeq2 for differential gene expression (121) or BLASTp analysis for prediction of gene function (122). The predicted functional open reading frames can be further aligned against specific databases of functional networks, such as KEGG (Kyoto Encyclopedia of Genes and Genomes), STRING (functional protein association networks) or CAZymes (123). Chief among these limitations however is the properties of the starting material itself, DNA, the identification of bacteria through the presence of its DNA via rRNA genes or WGS (116). DNA analysis does not distinguish whether DNA is extracted from viable or dead cells as well as whether the predicted genes are active (transcribed) or not, hence does not allow for the differentiation between the active and dormant bacterial subgroups, a critical determinant of the microbial community that can either promote health or disease. Metagenomic analysis is limited in scope and only provides the potential held the microbiome. For example, analysis of the antibiotic meta-resistome, the array of resistance genes within the population, will generate

comprehensive information on the potential of resistance determinants of the whole community, however not the active proportion contributing to any resistant phenotype (108). Only through sequencing the expressed genome (e.g. mRNA) can the metabolically active bacterial proportion be established (124).

Metatranscriptome sequencing can allow identification of the dormant (when used in combination with metagenome analysis) and active microbial communities, thus facilitating the study of shifting active populations at different stages of disease and treatment. Metagenome abundance has been previously shown to have little, if any impact on differences in gene/pathway transcription (125), suggesting that a transcription-based approach is more condition-sensitive and reliable (116). Transcription is a highly sensitive and regulated process and may significantly fluctuate between cells within a community (126). This implies that the same signal within the community may regulate the expression of the same gene/operon differently in otherwise genetically similar organisms. This further highlights a fundamental limitation of metagenome analysis wherein a species' functional potential will be assessed through gene abundance. The gene sequence of distinct strains of bacteria may be the same or mostly the same, however due to other differences, e.g. the presence of unique genetic elements, each strain may respond at the level of transcription to the same environmental stimuli differently (127). Differences in gene expression may also be a result of spatial heterogenicity of the gut environment, e.g. a mixture of healthy colon and tumour microenvironment (TME), involving localised nutrients and stimuli, as well as specific microbial cell-cell communication, including cross-feeding interactions (128). The tumour microenvironment provides greater access to host cell surface glycans, eDNA, and O_2, limited simple sugar availability due to tumour cell uptake, the environment is also known to be much more acidic with higher rates off cell-cell communication and potential immune infiltration (129–131). Therefore, despite this diverse gut environment, metatranscriptome-based taxonomic analyses provides functional insights into microbial responses to alterations to the environmental niche in disease.

Based mainly on metagenomic analyses the diverse bacterial population within the gut is dominated by the two major phyla, Firmicutes and Bacteroidetes. Genomes of certain *Clostridia* ($Gram^+$ Firmicutes) and *Bacteroidia* ($Gram^-$ Bacteroidetes) classes, dominant obligate anaerobes of the GI microbiota, encode for an array of hydrolytic enzymes (glycoside hydrolases) that metabolise a broad spectrum of fibre, complex dietary polysaccharides, hence benefitting the host (132).

Firmicutes are the largest bacterial division, containing over 250 genera (e.g. *Lactobacillus, Mycoplasma, Bacillus* and *Clostridium*) some being obligate anaerobes, such

as *Clostridium* spp. while others are aerobic, including *Bacillus* spp. or facultative anaerobes *Lactobacillus* spp.. This phylum is responsible for *n*-butyrate (a common SCFA) production (especially *Faecalibacterium* and *Clostridium* spp.), a source for acetyl-CoA and the primary energy currency of the colon (133) which is required to maintain the hypoxic environment of the intestine (134). Normal epithelial cells use *n*-butyrate as an oxidative C-source for mitochondrial β-oxidation (Fig. 5, C2) followed by the tricarboxylic acid (TCA) cycle, displaying a growth promoting phenotype. Tumour cells do not metabolise *n*-butyrate for their growth due to the Warburg effect, namely enhanced anaerobic glycolysis as a result of a reliance on glucose as a carbon and energy source, resulting in an increased production of lactate (Fig. 5, C1) (135, 136). This, in turn, accumulates *n*-butyrate in the nucleus and inhibits histone deacetylation, thus repressing the growth of tumour colonocytes (137). Therefore, *n*-butyrate producing bacteria play a key role in maintaining homeostasis of the gut (138). The second phylum, Bacteroidetes, consists of 20 different genera (e.g. *Prevotella, Bacteroides* and *Porphyromonas*) of both obligate anaerobes (*Bacteroides*) and aerobes (*Arenibacter*). Members of the Bacteroidetes phylum are also SCFA-producing bacteria (mostly propionate) and play a key role in e.g. fermentation of indigestible complex carbohydrates (e.g. xylan, mannan, xyloglucan or starch) due to encoding a multitude of carbohydrate-active enzymes (polysaccharide lyases and glycoside hydrolases, CAZYmes) (132).

Most members of the GI microbiota demonstrate a complimentary symbiotic relationship with the host (139) and are vital in protection against pathogen colonisation through maintaining the host epithelial barrier (140), while providing pathogens with their by-products to promote their growth (141), and aiding in the development of the immune system (139). Access to attachment sites on intestinal epithelial cells involves competition between different bacteria that can colonise the mucus using it as a carbon source (142). A diet in which fibre is depleted can lead to the colonisation of mucin-eroding bacteria *Citrobacter rodentium* on IECs. Bacteria lacking mucolytic capability, including *C. difficile*, *E. coli* and *S. Typhimurium*, can expand their presence in the gut due to their ability to metabolise sialic acid and fucose, which are cleaved from mucins by sialidases encoded by e.g. *Ruminococcus gnavus* (143). Some microorganisms utilise metabolites produced by other microbes to support their growth, termed cross-feeding. $Gram^+$ Clostridia *Anaerostipes caccae* and *Eubacterium hallii* convert lactate and acetate synthesised by *Bifidobacteria* and *Lactobacilli* or *R. bromii*, correspondently, into *n*-butyrate. Strains of commensal *E. coli* can inhibit growth of pathogenic enterohemorrhagic *E. coli* (EHEC) by either consuming proline, hence depleting this amino acid and preventing EHEC growth (144) or by secreting the toxin bacteriocin (145). By-products of commensal bacteria, such as secondary (deconjugated) bile acids, can be used by pathogens, e.g. *C. difficile* to activate spore germination (146). Through activation of

the host immune system, commensal bacteria can also protect the gut against pathogens. *B. thetaiotaomicron* can trigger production of α-defensins by small intestinal Paneth (HD5 and 6) cells and neutrophils (HNP1-4) through promoting expression of matrilysin, a metalloproteinase that cleaves prodefensins, a cysteine-rich AMP (antimicrobial peptide) precursor (147), which kill bacteria via membrane disruption. Other human AMPs, such as β-defensins (hBD) are continually expressed in the gut and other organs exposed to the environment. Epithelial cells can constitutively express some hBDs, e.g. hBD-1 while expression of others, including hBD-2-4 is regulated by inflammation (148). Bacteria can also suppress pathogens by locally modulating the level of O_2, hence reducing the virulence of aerobes e.g. *Shigella flexneri* (144). Members of *Bacteroides* and some *Clostridia* species induce production of mucosal secreted immunoglobulins, namely sIgA (149), which protects the gut against pathogen colonisation through binding and coating bacteria, limiting their exposure to the surface of IECs (150) or via interaction with bacterial sIgA receptors, precluding formation of biofilms (151). Segmented filamentous bacteria, SFB adhere to the epithelium and protect the gut against pathogen attachment through their ability to interact and activate the host immune system, including activation of sIgA and the development of anti-inflammatory T_{reg} cells (152). Commensal bacteria can also modulate colonisation of pathogens through changing mucosal properties, for example *B. thetaiotaomicron, Lactobacillus casei* and *R. gnavus* E1, can modulate expression of glycosyltransferase (153–155), hence altering mucin glycosylation.

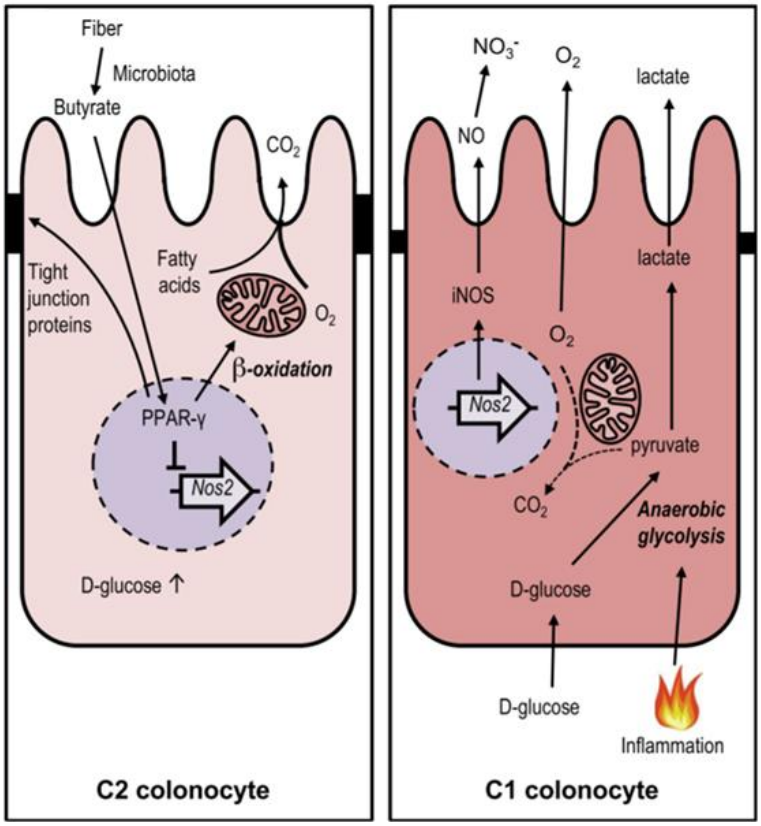


Fig. 5 | **Metabolism of colonocytes (C2 and C1) is regulated in response to infections, SCFAs availability, IEC damage and chronic inflammation**. Metabolism of IECs in response to bacterial metabolites, such as SCFAs or inflammation may resemble M2 and M1 metabolism of Mφs, respectively. C1 metabolism is featured with low levels of AMP and enhanced anaerobic glycolysis, a high level of glucose consumption and lactate production in contrast to high levels of AMP and β-oxidation in M2 and C2. Adapted from (134).

Members of the gut microbiota can also enhance epithelial integrity, control IEC proliferation and cell renewal, including mucin-degrading *A. muciniphila* (156), *L. rhamnosus* (157) and *L. plantarum* (158). Specific members of the mammalian microbiota possess anti-inflammatory properties, hence potentially reduce the risk of CRC development. *E. coli*, *Bacteroides* spp. and *Lactobacillus* spp. can suppress activation of NF-κB signalling through preventing degradation of its inhibitor IκB-α (by inhibiting polyubiquitylation), resulting in a hypo-responsive phenotype of IECs (159). The anti-inflammatory properties of *B. thetaiotaomicron* stem from its ability to induce PPAR-γ, peroxisome proliferator-activated receptor γ (160), antagonising pro-inflammatory NF-κB-dependent transcription through relocating its subunit, p65/RelA to the cytoplasm, hence precluding synthesis of pro-inflammatory proteins, including TNF-α (tumour necrosis factor-α), IL-1β and some adhesion proteins (161). PPAR-γ, a nuclear receptor, which is expressed in terminally differentiated IECs as well as in dendritic cells (DCs) and macrophages (Mφ), is a primary regulator of lipid metabolism, T_{h17} (RORγt/C2^{+}IL17A^{+}) and T_{reg} (Foxp3^{+}) cells differentiation (162) which has long been shown to suppress inflammatory responses (163). PPAR-γ can be activated by multiple mechanisms, responding to oxidised *n*-butyrate (164), activation of TLR (toll like receptor)-4 by lipopolysaccharides, LPS (165) or production of IL-4/IL-13 by T_{h2} cells (166). Beneficial effects of *n*-butyrate are mediated via its receptor, GPR109A which induces PPAR-γ-mediated anti-inflammatory responses via promoting T_{reg} differentiation (167), FOXP3- and IL-10-secreting T cell differentiation by triggering acetylation of the H3 histone in the promoter regions of *FOXP3*, therefore suppressing inflammation (168). This, as well as activating IL-18 expression in colonocytes, enhances the epithelial barrier function (169). The potential loss of these anti-inflammatory mechanisms of bacterial species cannot be understated in the context of inflammatory CRC development.

2.3 Global mechanism of bacterial killing post-stress exposure through ROS accumulation

It was found that ciprofloxacin (fluoroquinolone antibiotic) triggers production of superoxide by *S. aureus*, which in turn contributes to the accumulation of ROS that very often kill bacteria (170). In the subsequent 20 years after this finding numerous studies have identified similar responses to specific antibiotics in e.g. *E. coli*, *E. faecalis, P. aeruginosa*, *B. subtilis*, *A. baumannii*, *S. pneumonia, M. tuberculosis* and *S. enterica*. This common post-stress response to e.g. aminoglycosides, β-lactams and fluoroquinolones (171) has led to a global model of bacterial killing by antibiotics in a ROS accumulation mediated manner (172). However, while the outcome is the same, ROS accumulation triggered by antibiotic damage leading to cell death, the exact response and mechanisms of post-antibiotic ROS production are complex and vary between species and antibiotic and are not fully understood. It is

noteworthy that this model does not discount the lethal effects of the target-specific actions of antibiotics, rather provides an additional causative mechanism to their lethality.

Some antibiotics such as nitrofurantoin and metronidazole act directly through their own reduction by e.g. NADH-dependent reductases, generating reactive species $O_2^{\cdot -}$ and subsequent H_2O_2 and $OH^{\cdot}$ which cause lethality by directly damaging DNA/RNA (173). Other antibiotics alter microbial central metabolism, cellular respiration, and iron metabolism as a consequence of their activity, inducing microbial ROS production. It was shown that removal of TCA-cycle genes (specifically in mutant strains Δ*acnB* (aconitate hydratase B), Δ*icdA* (isocitrate dehydrogenase A), Δ*sucB* (dihydrolipoyltranssuccinylase, E2 component of the 2-oxoglutarate dehydrogenase) and Δ*mdh* (malate dehydrogenase) responsible for the formation of NADH, increased the NAD^+:NADH ratio and decreased $O_2^{\cdot -}$ production, hence allowing bacterial survival post treatment with bactericidal drugs, such as the quinolone norfloxacin. The protective effects were solely observed upon knockout of genes, the products of which act immediately upstream of production of the first reduced dinucleotide, and the first step of NADH synthesis (172). These findings demonstrate the regulatory effects of antibiotics over microbial metabolism and subsequent lethal levels of ROS accumulation. Interestingly, bacterial overexpression of reducing enzymes such as catalase can attenuate antibiotic ROS-mediated killing and hence activation of reducing factors provides bacteria a degree of tolerance. This was shown when *E. coli* deficient for *soxS*, *katG* and *ahpC* ROS-reducing enzyme encoding genes were found to be more susceptible to antibiotics, in this case elevating the lethality of ampicillin and kanamycin (171).

2.4 Interplay between the intestinal mucosal epithelium, immunity and microbes

CRC is a disease of accumulated genetic lesions (49, 50, 174), of which in >90% cases are underpinned by environmental factors (43, 175, 176) that drive chronic inflammation (34, 177–179) over a number of decades (27, 180). The healthy human colon is not generally inflamed despite being constantly exposed to the largest and most dense microbial population (181). These microbes are capable of modulating both homeostasis and the inflammatory state of the colon and have a profound effect on the metabolism of colonocytes of the mucosa (138, 182, 183) (Fig. 6). IECs are protected from direct contact with luminal microbes by a mucus bilayer barrier, defects in which very often lead to colonic inflammation by allowing microbial access (184). The colonic mucus consists of a thick, impermeable (by bacteria) inner layer, covered by a loose outer layer, which unlike the inner layer, is accessible to bacteria. The mucosal structure is based around MUC2, a glycosylated gel-forming protein, mucin, produced by Goblet cells (142). The mucus mucin-type O-glycans, such as N-acetylglucosamine and N-acetylgalactosamine are further modified with sialic acid and fucose, representing ~80% of MUC2 mass (185). These heavily altered O-glycans on MUC2 serve as

sources of carbohydrates (energy) and sites for attachment by bacteria, hence potentially contributing to the composition of the luminal microbiota (142) which in turn modulates the integrity and metabolism of the colonocytes. Germ-free (GF) mice have been shown to possess a very thin mucosal layer the thickness of which is restored upon exposure to bacterial LPS or peptidoglycan (PG) (35). Furthermore, impaired IEC turnover was also restored upon transplantation of GF rodents with healthy microbiota, e.g. *Faecalibacterium prausnitzii* and *B. thetaiotaomicron* (186). This strongly argues that the GI microbiota are a key modulator of IEC metabolism. Under homeostasis bacteria promote colonocyte renewal (e.g. by *Lactobacilli rhamnosus*) (187), integrity (e.g. by *A. muciniphila* and *Lactobacillus plantarum*) (188), co-ordinate mucus production (e.g by *B. thetaiotaomicron* and *F. prausnitzii*) (189) or remodel the pattern of mucin glycosylation, hence altering structures of attachment sites (by e.g. *B. thetaiotaomicron*, *L. casei* and *R. gnavus*). Colonocytes in turn, maintain their 72hr life cycle and high oxygen consuming metabolism (via mitochondrial β-oxidation), hence promoting epithelial hypoxia enabling obligate anaerobes to dominate within the GI microbial community. This symbiosis suppresses enhanced metabolic activities of immune cells and thus, prevents undue inflammation (190). In response to the emergence of pathogens, epithelial injury or depletion of dietary fibre, which may lead to decreased production of beneficial fermentation products (e.g. fatty acids or polyamines), the host immune system becomes increasingly activated. This would generally lead to differentiation of myeloid monocytes of the innate immune system (DCs (dendritic cells) and Mφs) and production of inflammatory cytokines, including interleukin (IL)-23 by DCs (Fig. 6A) or IL-18 and IL-1β by inflammasomes in Mφ and IECs (Fig. 6B) (191, 192). Inflammasomes are protein complexes within the cytoplasm that sense endo- and exogenous pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patterns (DAMPs) (endogenous danger molecules). This signals the adaptive immune system to develop $CD138^+$ (*c*luster of *d*ifferentiation, a specific subset of cell-surface receptors that are recognised by antibodies and used as phenotypic markers, defining types and stages of differentiation of immune cells) B cells and a specific subset of $CD4^+$ T cells (Fig. 6A), T helper T_{h1}, T_{h2}, T_{h17} cells and regulatory T_{reg} cells, which in turn can regulate the activity of the innate immune system (147). $CD138^+$ B cells produce sIgA antibodies (Fig. 6A), killing pathogens and bacteria that have infiltrated the lamina propria by enhancing cytotoxic responses (61) while IL-23, IL-1β and potentially IL-21 produced by T follicular helper cells (T_{fh}) induces differentiation of $CD3^+CD4^+$ T_{h17} and innate lymphoid cells (e.g. ILC3) (193). In response to stimuli, DCs activate naïve $CD3^+CD4^+$ and $CD3^+CD8^+$ T cells by presenting antigens to T cells to initiate the immune response, leading to polarisation of T cells into $CD4^+$ T_{h1}, T_{h2} or T_{regs} or $CD8^+$ cytotoxic T cells (194). $CD3^+CD8^+$ cytotoxic T cells harbour cytotoxic granules and promote apoptosis through secretion of granulysin, granzymes and perforin cytotoxic factors, upon interaction with

damaged cells. T_{h1} is involved at the initial stage of epithelial injury by activating a type 1 response via interferon-γ (IFN-γ) primarily, sending signals to $CD8^+$ T cells, Mφ and $CD57^+$ NK (natural killer) cells. T_{h2} is usually active at later stages of damage repair to lessen inflammation with a type 2 response through secreting anti-inflammatory cytokines, IL-4, IL-5, IL-10 and IL-13, to control Mφ and B cells function (195). Expression of IL-12p70, the dominant DC-produced cytokine, upon activation of specific pattern recognition receptors (PRRs) such as toll-like receptor (TLR)-9 by CpG DNA or TLR-4 by bacterial LPS signals polarisation of T_{h1} (inducting IFN-γ production) while the absence of which will drive T_{h2} cell development (196). T_{h17} and IECs secrete cytokines IL-17A and IL-22, the adequate expression of which is critical for IEC homeostasis, including tissue repair and protecting the colon against pathogen colonisation (Fig. 6B).

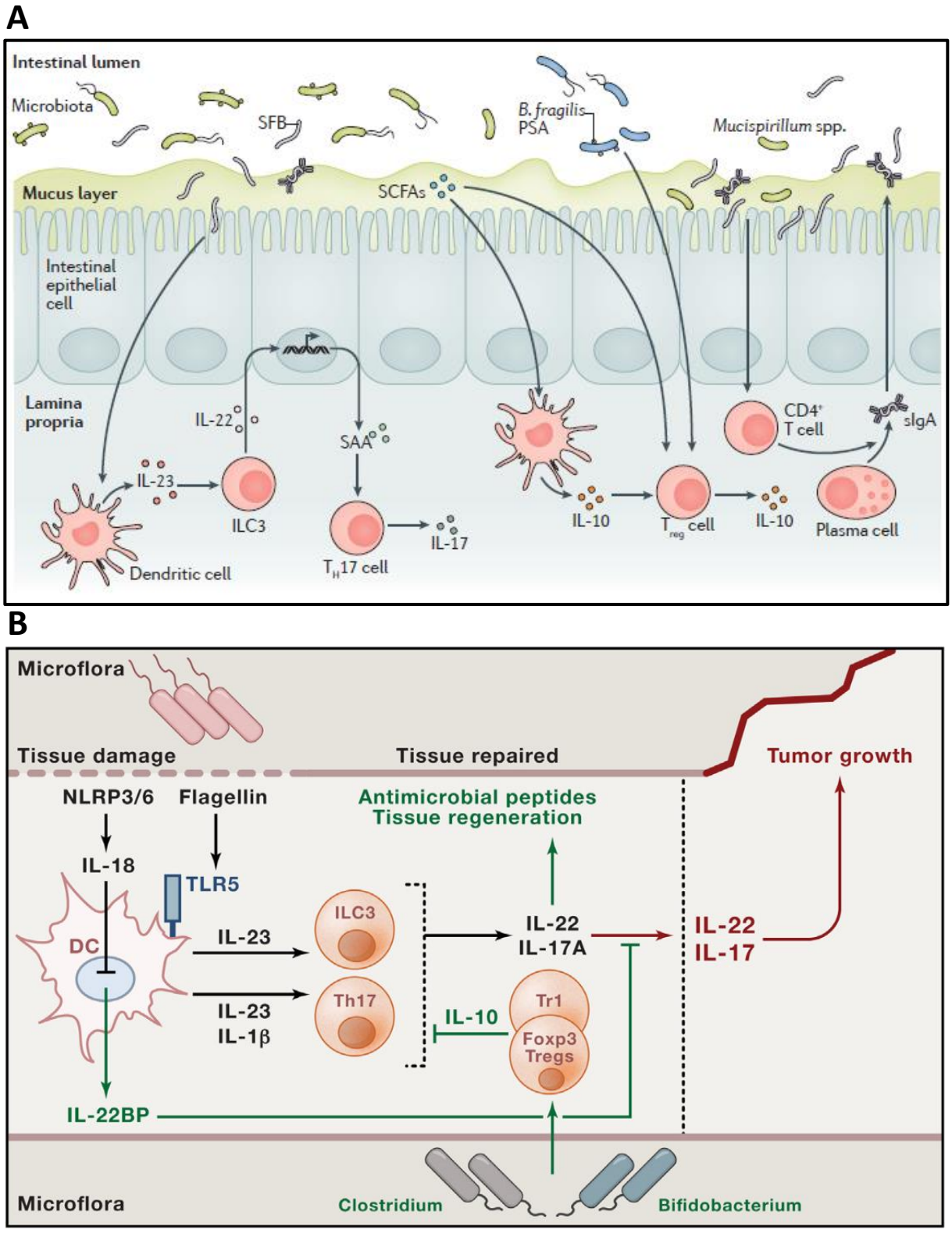


Fig. 6 | Microbiota, its metabolites and damaged IECs trigger protective inflammatory responses. **A**. The mucus layer protects colonocytes from the luminal microbiota the composition of which is modulated by the host immune system. During homeostasis, commensal anaerobes produce fermentation products (such as SCFAs) which are crucial for barrier function and protection against pathogen colonisation via induction of T_{reg} development and expression of anti-inflammatory cytokines, such as IL-10. Changes in interactions between commensal bacteria and the epithelial cells prompts

inflammation via activation of a variety of immune responses (197). DCs and epithelial cells in response to lipoproteins and flagellin that activate the TLR-5 produces IL-23 (198). In response to bacterial metabolites (e.g. histamine or spermine) or epithelial injury by e.g. conjugated bile acids or pathogens, NLRP6 inflammasomes secrete IL-18, IL-1β, promoting production of epithelial AMPs. IL-23 and IL-1β activate ILC3 (199) and $CD3^+CD4^+$ ($RORC2^+$) T (T_{h17}) cells which secrete IL-17A and IL-22 (a member of the IL-10 family) cytokines (200), protecting the gut against pathogen colonisation. Bacterial (*Lactobacilli*)-derived Trp metabolites (indole derivatives, L-kynurenine) can also activate ILC3 and IL-22 synthesis, protecting the colon against fungal colonisation and mucus from damage (201). Activity of T_{h17} cells is controlled by compensatory mechanisms (e.g. via PPAR-γ, a critical mediator of T_{h17} and T_{reg} cells differentiation, inhibiting expression of adhesion proteins, TNF-α or IL-1β) (162) the failure of which leads to chronic inflammation and tumour growth (202). SFB, segmented filamentous bacteria (152) trigger IECs to release SAA, serum amyloid A, which induces the development of T_{h17} cells and IgA production (203). $CD4^+$ T follicular helper (T_{FH}) cells regulate the microbial population in the intestinal lumen by activating plasma cells ($CD138^+$ B cells) to secrete sIgA (204). Mucosal-proximally localised SFB and *Mucispirillum* spp. are coated by sIgA, eliciting a T cell-mediated adaptive response (205). $Foxp3^+$ T_{reg}, $CD3^+CD4^+Foxp3^+$ IL-10 producing T_{reg} cells; Tr1, $CD3^+CD4^+Foxp3^-$ IL-10-producing type 1 T regulatory cells; sIgA, secreted immunoglobulin A; PD1, the inhibitory receptor programmed cell death protein 1. **B**. NLRP3/6 inflammasomes produce IL-18 in response to epithelial damage and block the expression of IL-22BP, which de-represses IL-22 activity generated by RORγt/$C2^+$ T_{h17} and ILC3 upon activation of TLR-5 in a lipoprotein- and flagellin-dependent manner. TLR-5 signalling responds to *Enterobacteriaceae* which are in close proximity to IECs. At the early stage of bacterial infection (e.g. EHEC and EPEC *E. coli*, A/E bacterial pathogens) IL-22 directly induces the host antimicrobial responses, promoting expression of lectins, e.g. RegIIIγ and RegIIIβ, members of the antimicrobial Reg family, and S100A8/9 calprotectins, by Paneth cells (206). Uncontrolled expression of IL-22 and IL-17A leads to chronic inflammation and CRC. Members of the commensal microbiota, such as *Clostridium* and *Bifidobacterium* can manipulate the development of T cells, inducing polarisation of $Foxp3^+$ T_{regs} to avoid the antimicrobial activity of T_h immunity. AMPs, antimicrobial peptides; ILC3, innate lymphoid cells; DCs, dendritic cells; SCFAs, short chain fatty acids; PSA, polysaccharide A; IL, cytokine interleukin; EHEC, enterohemorrhagic; EPEC, enteropathogenic. Adapted from (193).

Initial host responses to epithelial barrier function disruptions involve activation of immune and epithelial cell metabolism, producing an array of cytokines which induce specific signalling pathways, such as NF-κB (by TNF-α, IL-1β and IL-17A) (61) and STAT3 (by IL-6 and IL-22) (207). These signalling pathways promote IEC proliferation and block apoptosis, leading to restoration of the integrity of the epithelial barrier, host defence and management of the microbial composition. Differentiated IECs, in turn, continue to utilise bacterial fermentation products to drive their metabolism towards oxidative phosphorylation and mitochondrial β-oxidation, ensuring a strictly anaerobic environment, thus controlling anaerobiosis and the growth of obligate anaerobic bacteria in the gut.

Uncontrolled activation of T_{h17} cells can lead to inflammatory pathologies. T_{h17}-produced cytokines, IL-6, TGF-β, IL-1β or IL-21 promote polarisation of naïve $CD3^+CD4^+$ T cells into RORγt/$RORC2^+$ T_{h17}, further activating STAT3 and NF-κB leading to colitis associated cancer (202). Development of T_{h17} and production of its cytokines is regulated by another type of $CD3^+CD4^+$ T cell, namely T_{regs} via the secretion of IL-10 (Fig. 6A) and DCs via production of IL-22 binding protein, IL-22BP (Fig. 6B) (208). During infection or epithelial damage through recognition of microbial- or self-antigens (produced by damaged tissues), T_{regs} modulate a balance between the pathogen and the host, often favouring survival and expansion of the pathogen by attenuating excessive immune responses (209). T_{reg} cells are characterised by expression of *Foxp3* encoding for forkhead box P3 (208), producing IL-10, a key anti-inflammatory cytokine which controls the differentiation of T_{h17} and the subsequent production of IL-17. Proliferated $CD3^+CD4^+$ T cells, which produce IL-17A, express IL-10Rα, the IL-10 receptor-α via which IL-10 modulates activity of T cells (191). Interestingly, some GI bacteria (e.g. *Clostridium*, *Bifidobacteria* or *Lactobacillus* genera) can induce proliferation of $Foxp3^+$ T_{regs} and reduce differentiation of T_{h17} cells, limiting the production of IL-17 (Fig. 6B) (210), suggesting their role in anti-inflammatory processes. These SCFAs-producing bacteria induce PPAR-γ, the epithelial receptor (211), which represses expression of RORγt/RORC2 in T cells (212), hence inhibiting T_{h17} proliferation while inducing transcription of *Foxp3* that regulates the function of T_{regs} through enhancing acetylation of the H3 histone in the promoter locus of the gene, hence promoting survival of $Foxp3^+T_{reg}$ cells (213).

Bacteroides fragilis can protect colonocytes from sustained inflammation in response to *Helicobacter hepaticus* colonisation through activation of the TLR-2 receptor by its polysaccharide A and stimulate differentiation of an inducible subset of regulatory T cells, namely $Foxp3^-$Tr1 (Fig. 6B), resulting in secretion of high amounts of IL-10 and suppression of IL-17A production (181). *Bifidobacterium,* such as *B*. *longum* and *B*. *breve* can also induce proliferation of IL-10^+Tr1 cells in the colon of mice by inducing TLR-2 signalling that activates IL-23-producing DCs, thus restricting the immune response during infection (214). Hence,

based on *in vivo* studies in mice it is apparent that during homeostasis, colonocytes drive their metabolism towards high oxygen consumption and regulate the microbial composition of the gut through controlled expression of key immune cytokines, which induce AMP (adenosine monophosphate) secretion and IEC regeneration through crosstalk with the GI microbiota. AMP is an indicator of the energy status of the cell, an increase of which reflects a reduction in energy production, promoting oxidative phosphorylation and β-oxidation by the mitochondria. The microbiota, in turn, influence the development of the host immune system by different mechanisms (via metabolites or conserved bacterial signatures) to harmonise their co-existence with the host and limit pro-tumorigenic inflammation as failure to control IL-22 and IL-17A expression can drive the onset of tumorigenesis. Elucidation of the taxonomy of metabolically active microbiota and their pattern of gene expression in healthy individuals and CRC patients will help to establish the molecular mechanisms by which specific microorganisms can protect gut homeostasis alongside bacterial determinants that trigger inflammation.

2.5 Microbial Dysbiosis Associated with Colorectal Cancer

The microbiota of the gastrointestinal tract contribute to CRC, individually and as a community, directly and indirectly through numerous potential mechanisms (215). This includes amending the rate of epithelial cell turnover and proliferation, affecting the development of the host immune system, metabolising host- and dietary-derived products which may lead to DNA damage and hence genomic instability. Additionally, the ability of some bacteria to communicate with each other through secretion of chemical signals, so-called quorum sensing, can lead to formation of biofilms, promoting inflammation and hence activating a number of the host signalling pathways (e.g. Wnt/β-catenin, STAT3, SMAD4 and MAPK). Biofilm positive tissues of cancer patients have been observed with increases to both IL-6 (including its signal transducer) and STAT3 signalling when compared to biofilm negative normal tissues of the same host (216). This is known to lead to a stepwise accumulation of mutations, a primary feature of CRCs (217–220). As sequencing technologies advance, as does the abundance of research regarding microbial meta'omics advancing our knowledge of how and which microbe-host interactions, down to the species level, modulate disease (221). As this next wave of research comes into focus, it is evident that the gut microbiota are intrinsically involved in CRC inception, progression, and treatment efficacy. The germ-organ undergoes substantial global compositional changes in prevalence of particular clades of microbiota (commonly termed dysbiosis) in patients with CRC (222). However, the current understanding of the dysbiotic microbiome is based mainly on metagenome analyses, showing only the altered genome potential of the microbiota and not changes in its metabolic activity.

Constituents of the oral cavity microbiome (e.g. *Fusobacterium nucleatum*, *Prevotella intermedia*, *Porphyromonas asaccharolytica* and *Parvimonas micra*) can be translocated to the gut, are more prevalent in CRC and act as accurate predictors of disease progression (9, 223), suggesting a potential causative role in disease development. Balanced GI microbiota promote homeostasis, while overgrowth or depletion of some microorganisms can be detrimental to human health (147). There are a plethora of studies linking specific bacterial species to CRC (8, 9, 11, 224–228). Specific bacteria linked to CRC through metagenomic analyses of CRC patients and mouse model systems revealed CRC-related enrichment of specific bacteria including pro-carcinogenic strains of *Enterococcus faecalis*, *Escherichia coli, Fusobacterium nucleatum, Streptococcus gallolyticus* (formerly *S. bovis*) member bacteria (SGMB) and *Bacteroides fragilis* and potential mechanisms by which they may promote inflammation and the development of adenocarcinomas (Fig. 7 and Fig. 8) (229–232).

Meta-analysis of CRC faecal metagenomic data established a set of 29 core bacteria substantially enriched among CRC patient microbiomes, namely *Anaerococcus obesiensis*/*vaginalis*, *Clostridium bolteae*/*clostridioforme*, *Clostridium symbiosum*, *F. nucleatum* subspecies *animalis*, *Fusobacterium* species oral taxon 370, *Gemella morbillorum*, *Hungatella hathewayi*, *Parvimonas micra*, *Parvimonasspecies*, *Peptostreptococcus stomatis*, *Porphyromonas asaccharolytica*, *Porphyromonas somerae*, *Porphyromonas uenonis*, *Prevotella intermedia*, *Prevotella nigrescens*, *Ruminococcus torques*, *Solobacterium moorei*, *Subdoligranulum* species, uknown *Dialister*, unknown *Anaerotruncus*, unknown *Clostridiales*, unknown *Clostridiales*, unknown *Peptostreptococcaceae*, unknown *Porphyromonas*, unknown *Porphyromonas*, unknown *Porphyromonas* and *vincentii* and *nucleatum* (222). Alteration in the composition of specific bacteria in the gut at different stages of CRC was reported, showing a selection for Fusobacteria and Proteobacteria (e.g. *E. coli* and *Campylobacter*) at earlier stages of cancer (transformation of polyps to adenoma) alongside Actinobacteria *Atopobium parvulum* (the family *Coriobacteriaceae*) and *Actinomyces odontolyticus* (the family *Actinomycetaceae*) (233, 234). The elevated abundance of carcinogenic cyclomodulin^{+} *E. coli* and *B. fragilis* was found at late stages CRC while *F. nucleatum* and *Solobacterium moorei* are enriched at all stages of the disease (235). Therefore, changes in abundance of specific bacteria in the gut at different stages of CRC could be used as potential pathogenesis biomarkers. Yet, the activity of these species *in vivo* has not been investigated, only their DNA-based abundance.

While many bacterial species are seen enriched in CRC it was found that some specific bacteria are depleted, including butyrate-producing *Clostridia*, the family Lachnospiraceae *(*genera *Roseburia, Blautia* and *Anaerostipes), Eubacterium* spp.*,* the genus *Clostridium* and *Feacalibacterium* (*F. prausnitzii*). *F. prausnitzii* (Ruminococcaceae family) exert an anti-

inflammatory influence by suppressing the NF-κB pathway (236), mediating epithelial barrier function via coordinating mucus production (189) and preventing overgrowth of commensal and problematic bacteria, including Gram$^-$ pathogens, a distinct feature of dysbiosis (237). For example, lactic acid-producing bacteria *Bifidobacterium* and *Lactobacillus*, obligate and facultative anaerobes correspondently are also depleted, hence paving the way for expansion of *E. coli* and *Listeria* due to decreased acetate and lactic acid production (238). It has also been shown that CRC-associated microbiota possess altered phylogenetic α-diversity (diversity within the community) when compared to that of non-cancerous individuals, meaning the richness (variability) and distribution (evenness) of microbes present within the diseased gut varies. In some cases, it has been shown that the phenotypic α-diversity in the CRC gut is greater than during homeostasis (10), while in other cases - less (16, 239). This disparity can be attributed to the heterogenous nature of the human gut microbiome, and the disparity in approaches taken within the field, such as sample collection and storage, DNA isolation and sequencing (16S rRNA gene or shotgun sequencing) bioinformatic pipelines and statistical analysis used (240–243).

One of the most thoroughly investigated members of the microbiota, *Helicobacter pylori,* is present in more than 50% of the population and may play a major role in CRC. *H. pylori* is a prominent inducer of chronic inflammation and is a risk factor for the development of colonic adenocarcinomas, gastric ulcers and mucosa-associated lymphoid tissue lymphomas (244). Infection of *H. pylori* is eminent in the adenoma stage upon comparison to polyp-free one and the correlation of *H. pylori* infection with CRC is more pronounced in advanced stages of tumorigenesis, including multiple adenomas (245).

Dysbiosis may be a consequence of a disruption to gastric acid (HCl) production caused by *H. pylori.* Colonisation of *H. pylori* activates production of gastrin by G cells, a short hormonal peptide of 14, 17 or 34 amino acids in length that stimulates proliferation of mucosal cells (224) and damages IECs through promoting hyperplasia of Goblet cells, increasing the risk of CRC development (246). It has also been found that enhanced HCl secretion correlates with the colonisation of other pro-oncogenic bacteria *E. faecalis* and *B. fragilis*, hence potentially enhancing the risk of carcinogenesis (247). The host immune system protects the colon against bacterial localisation via the rapid production and secretion of reactive oxygen species (ROS), known as an ‘oxidative burst’, by phagocytes, neutrophils and Mφs to kill invading bacteria (248). DNA damage induced by ROS (e.g. hypochlorous acid (HOCl) and ·OH, hydroxyl radicals) and RNS, attempting to kill the invading *H. pylori* by Nox catalysis in the phagosome. Nox donates an electron to molecular O_2 within the phagosome or outside the phagocyte to yield $O_2^{\cdot -}$, superoxide, resulting in the production of microbicidal oxidants H_2O_2 via superoxide dismutase (SOD)-mediated reduction. H_2O_2 passively permeates the cell

membrane due to its non-polar nature, and is subsequently converted to HOCl, 100-fold more toxic than H_2O_2. In the presence of Fe^{2+} or Cu^{+} ions H_2O_2 also reacts non-enzymatically with $O_2\cdot^{-}$ to form ·OH (249). ROS and RNS are known contributors to CRC development as they damage DNA and are produced as a direct response of the neutrophils protecting the host, prompted by the *H. pylori* neutrophil-activating protein (HP-NAP) virulence factor (250), from the invading pathogen, however unsuccessfully (249). *Helicobacter* spp. can survive the acidic host stomach environment due to the production of urease, the enzyme that locally increases the pH from 2 to around 6-7, minimising the detrimental effect of acidic pH on bacteria (251). Other virulence factors, e.g. CagA either phosphorylates STAT3 or enhances SHP2-ERK1/2 binding, in phosphorylation-dependent manners (252). VacA (253), BabA (254) and SabA (255) encoded by the genome of *H. pylori* are considered more harmful than those without due to their potential to facilitate cellular proliferation and colonisation, corpus atrophy and intestinal metaplasia and are negatively associated with neutrophil infiltration respectively. $Toxin^{+}$ strains of *H. pylori* can inhibit the activity of the tumour suppressor p53 and disrupt the switch between the JAK/STAT and SHP2/ERK pathways via gp130 receptor, promoting CRC (256). Finally, despite the historically controversial link between this pathogen and CRC, *H. pylori* is known to promote the secretion of numerous pro-inflammatory cytokines IL-1β, IL-8, IL-6, interferon gamma (IFN-γ) and TNF-α, direct and potent causes of inflammation-driven disease (257). It is worth noting, that the relationship of *H. pylori* with CRC is causal, the length of infection correlates with the risk of cancer development, consistent with chronic inflammation of the colon over years (between 15-30 years) playing a part in stepwise accumulation of genetic lesions of epithelial cells (258). In light of comprehensive data from a multitude of sources, *H. pylori* has now been considered as a critical determinant of gastric health and can affect GI homeostasis through a number of mechanisms, actively promoting neoplasia within the colon, the significance of which was recognised by scientific communities leading to the 2005 Nobel Prize for Marshall and Warren (259).

E. coli can be divided into four major phylogenetic groups, A, B1, B2 and D, the latter two of which often carry virulence factors and are pathogenic, associated with increased intestinal inflammation and inflammation-independent carcinogenic pathways (Fig. 7) (260, 261). In conjunction with other bacteria known to expand during CRC, mucosa-associated pks^{+} pathogenic *E.* coli NC101 triggers invasive carcinoma in $IL\text{-}10^{-/-}$ mice. These mice with reduced inflammatory mediators show a progression to a cancerous phenotype associated with *E. coli*-produced colibactin (hybrid non-ribosomal polyketide-peptide encoded by the 54kb polyketide synthase (*pks*) pathogenicity island) (215, 262). Colibactin causes DNA adducts, covalent modifications of DNA (e.g. an N3-substituted adenine adduct) via alkylation of DNA, furthermore, it induces phosphorylation of H2AX histone (at Ser139) in response to double

stranded (ds) DNA breaks, a marker of CIN (263). These DNA adducts may directly amend regulation of transcription of tumour suppressors and/or oncogenes. However, CRC-related B2 and D *E. coli* can also induce colonic inflammation through their ability to survive in macrophages by activating expression of COX-2, a pro-carcinogenic mediator in a colibactin-independent manner (264), a chronic inflammatory marker and promotes angiogenesis, inhibition of apoptosis and elevated proliferation (239). Importantly, it appears that pro-oncogenic properties of *pks*$^{+}$ *E. coli* strains relate to the level of expression of the *pks* island and not to the abundance of the bacterium, highlighting the importance of studying the expressed microbiome. Other B2 pathogenic strains of *E. coli* produce toxins such as CIF (cycle-inhibiting factor) which mediate the transition of the G2/M checkpoint, resulting in cell cycle arrest via inhibiting the phosphorylation of cyclin-dependent kinase 1 (CDK1) (265). Cytotoxic necrotising factor (CNF) toxin, which is also produced by *E. coli* B2 induces COX-2 activation transiently and expression of Rho GTPases, Cdc42, Rac and RhoA causing DNA damage via glutamine deamidation (266). Additionally, *E. coli*-produced cytolethal distending toxin (CDT), which can be synthesised by various Gram^{-} bacteria, has DNase activity, causing dsDNA breaks and subsequent cell cycle arrest and apoptosis if the damage exceeds the mismatch repair capacity of the MMR pathway (267). *pks* islands individually have the capability of inducing DNA damage while not inhibiting the DNA damage response (230). However, genotoxic cyclomodulins and genotoxins produced by *E. coli* B2 strains cause downregulation of attachment-mediated DNA mismatch repair MSH2 and MLH1 proteins, hence affecting a major DNA damage response mechanism closely associated with the inception of CRC (268).

Non-enterotoxigenic (ETBF^{-}) strains of *Bacteroides fragilis* are important symbiotic constituents of the GI microbiota and metabolise a wide range of polysaccharides, promoting colonic homeostasis (269). However, ETBF^{+} *B. fragilis* initiate rapid colitis, colonic mucosal inflammation, and its presence induces colonic tumorigenesis in multiple intestinal neoplasia mice. The *B. fragilis* enterotoxin-encoding gene *bft* which encodes for fragilysin its major pathogenic factor. The *bft* gene is localised within the *BfPAI* locus (the *B. fragilis* pathogenicity island) and encodes a 21kDa Zn-dependent metalloprotease (270). Epithelial and immune STAT3 selective activation is prompted by ETBF along with colitis, leading to IEC transformation and inflammation. ETBF binds IECs and causes the cleavage of E-cadherin, disrupting the paracellular epithelial cell barrier, possibly through the degradation of tight/adherens junction proteins e.g. zonula occludens-1 (TJP1/ZO-1) (271). Degradation of E-cadherin, a tumour suppressor, stimulates β-catenin localisation to the nucleus and activates Wnt/β-catenin and NF-κB signalling and increasing pro-inflammatory signalling through STAT3 signalling and Th_{17} and γδ T lymphocytes, IL-8 and IL-17 production, as well

as triggering IEC proliferation (229). *B. fragilis* can also modulate host immune responses through polysaccharide A (PSA), the zwitterionic structure facilitates growth and colonisation as well as interactions with the host and other bacterial species. PSA interacts with TLR-2, a receptor occurring on $CD11^+$ DCs and is subsequently recognised by naïve T cells. $CD4^+$ T cells in turn express TGF-β, which induces $FOXP3^+$ in $CD4^+$ T_{regs} to produce anti-inflammatory IL-10 while blocking Th_{17} and Th_1 from producing pro-inflammatory IL-17 and IFN-γ respectively, therefore facilitating colonisation (272). ETBF can also induce DNA damage through activating expression of spermine oxidase, SMOX in IECs, inducing ROS production (226). Biofilm-positive tumours often contain ETBF, *pks*$^+$ *E. coli* and *F. nucleatum* which can facilitate the recruitment other bacteria (273). This biofilm formation can further increase risk of CRC through elevated inflammation, pro-oncogenic epithelial permeability, facilitating bacterial antigen translocation and promoting pro-inflammatory cytokine production e.g. increased epithelial IL-6 expression (274). Expression of adhesins and other virulence factors can also aid these bacteria in effectively colonising the epithelium. Degradation of mucus due to inflammation, hyperplasia or infection/colonisation by other pathogens during dysbiosis and an enhanced oxygen level increases adhesion of *E. coli* and *B. fragilis*, hence increasing the risk of the cancer development (273).

Sulphate-reducing (SRB) $Gram^-$ Proteobacteria primarily consist of *Deltaproteobacteria*, of the orders Syntrophobacterales, Desulfobacterales and Desulfovibrionales (275). The second largest group including Firmicutes of the genera *Desulfobacter, Desulfotomaculum*, *Desulfosporomusa*, *Desulfovibrio, Desulfosporosinus, Desulfomonas*, *Desulfococcus* and *Bilophila* (*B. wadsworthia*, catalase$^+$ and bile acid-resistant obligate anaerobe) all utilise sulphate in anaerobic respiration. Lactate, acetate and pyruvate are used for the reduction of SO_4^- that is tumour promoting, as well as the following product of reduction, hydrogen sulphide, H_2S, which is highly genotoxic, hence causing damage to host DNA. In addition to damaging host DNA, H_2S regulates IEC proliferation, apoptosis and differentiation through modulation of RAS/MEK/ERK signalling (276). High fat and protein diets increase the rates of taurine conjugation which in turn selects for *B. wadsworthia* causing its expansion within the colon, due to its resistance to conjugated secondary bile acids (277). Consequently, further dysbiosis towards bile acid conjugating bacteria promotes ROS production and induction of NF-κB, cell damage and elevated rates of IEC proliferation leading CRC.

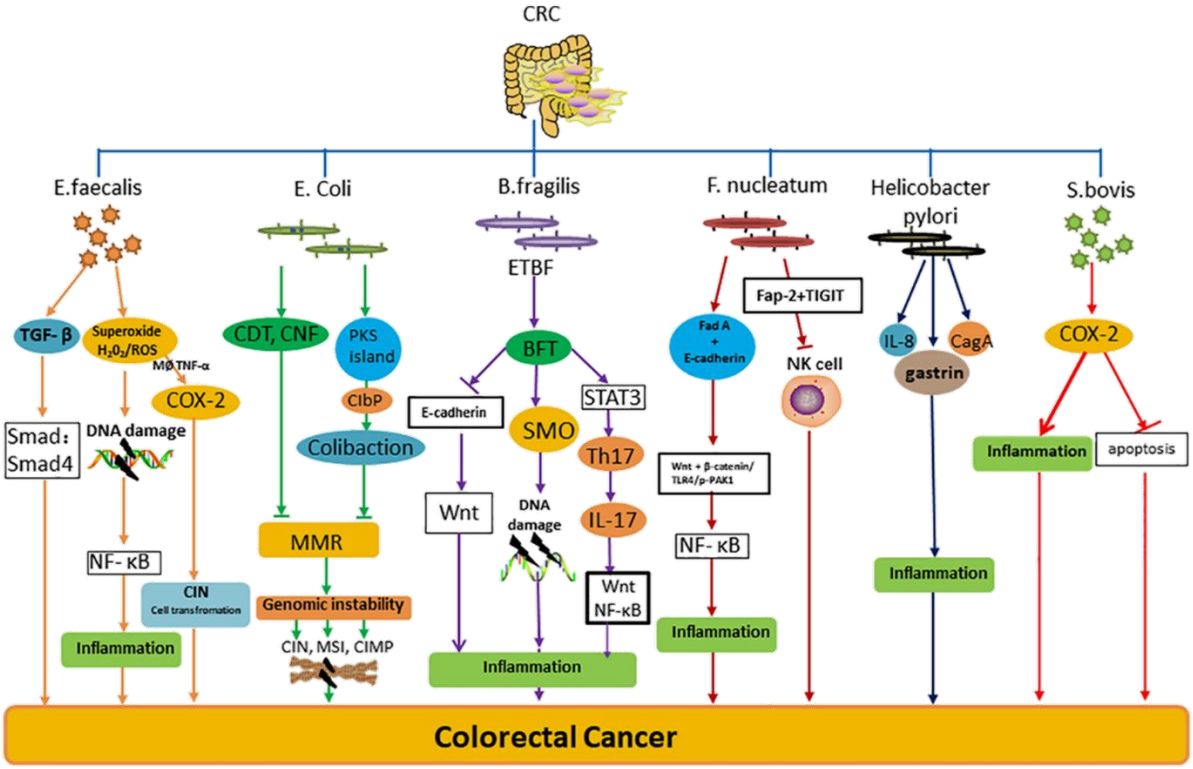


Fig. 7 | Role of individual gut microbes implicated in CRC development. *Enterococcus faecalis* causes inflammation leading to expression of TGF-β in IECs, thus triggering SMAD4 signalling. *E. faecalis* strains produce excessive amounts of free radicals, extracellular superoxide and H_2O_2, causing damage to host DNA. *E. faecalis* also appear to activate the COX-2 pathway and the release of the macrophage-derived TNF-α through ROS and 4-hydroxynonenal production, causing cell transformation and endogenous mutagen, trans-4-hydroxy-2-nonenal production, resulting in chromosomal instability and carcinogenesis (278). The genome of certain *E. coli* B2 EPEC strains encodes for CDT and CNF, virulence factors that can cause DNA damage leading to genome instability. Additionally, the attached *E. coli* represses MMR, leading to elevated short-term mutations and long-term tumorigenesis. *pks*$^+$ *E. coli* can also induce chromosomal aberrations via the expression of genomic toxin-containing polyketide-synthase island genes, *clbDEFG*, *ClbP* and *clbH-A*$_1$-synthesising colibactin, a prominent carcinogen. Ultimately, this brings about CIMP, MSI and CIN, which results in cancer Enterotoxigenic (ET)BF *Bacteroides fragilis* expressed the toxin BFT which binds to IECs and promotes the E-cadherin cleavage, thereby augmenting the Wnt/NF-κB pathway and the release of pro-inflammatory agents, destroying DNA. Concurrently, ETBF, through activation of STAT3 signalling activates the development of Th_{17} cells that produce IL-17 and promotes colon tumorigenesis. EBFT causes the rapid expression of spermine oxidase, SMO and facilitates ROS production damaging the IEC DNA, leading to tumour formation. Through forming a complex with E-cadherin on IECs, *Fusobacterium nucleatum* prompts FadA, an activator of WNT/β-catenin/TLR-4 signalling, upregulating expression of oncogenes, including activation of expression of non-coding (nc)RNAs, e.g. mir-21. *F. nucleatum* can stimulate inflammation through releasing RNA into IECs which in turn activates NF-κB. *F. nucleatum* via another adherin Fap2, binds to and inhibit NK, causing colorectal tumorigenesis. *Helicobacter pylori* promotes the chronic gastritis-induced metastasis of CRC. Furthermore, at early

stages of CRC infection of *H. pylori*, bacterial colonisation can cause colonic IEC damage through e.g. inflammation mediated by IL-8. Colon inflammatory responses can also be induced by *cagA*$^{+}$ virulent strains of *H. pylori* which encode for CagA, the cytotoxin-associated gene A gene, hence leading to CRC. The *Streptococcus gallolyticus* (*S. bovis*) antigen, such as pilus protein (the *pil1* gene) promotes COX-2 expression that, along with prostaglandins, induces angiogenesis and cell proliferation, while also inhibiting apoptosis, hence stimulating carcinogenesis. TGF-β, transforming growth factor β; IECs, intestinal epithelial cells; COX-2, cyclooxygenase-2; TNF-α, tumour necrosis factor α; EPEC, enterohemorrhagic *E. coli*; CDT, cytolethal distending toxin; CNF, cytotoxic necrotizing factors; NK, natural killer cells; CIMP, CpG island methylator phenotype; MSI, microsatellite instability; CIN, chromosomal instability; MMR, mismatch repair genes; ETBF, enterotoxigenic *Bacteroides fragilis*; BTF, fragilysin; TLR-4, toll-like receptor 4. Adapted from (4).

Enterococcus faecalis, a commensal Gram$^+$ facultative anaerobe belongs to the Firmicutes phylum and has been linked to CRC due to a high prevalence in patient faecal samples (279). *E. faecalis* produces excessive amounts of free $O_2{\cdot}^-$ via the autoxidation of membrane associated demethylmenaquinone, an anaerobic electron transporter, also used by *E. coli* (280). Additionally, both colonic and ileal-derived samples from *E. faecalis* colonised rats detected ·OH and thiyl radicals (228). These oxidants may be a central source of CIN through the involvement of the pro-inflammatory COX-2 pathway, concomitant with sporadic adenomatous polyps and CRC. Furthermore, it was shown that *E. faecalis*-colonised mice displayed an increased TGF-β level, hence activating SMAD4/TGF-β signalling, essential in regulation of growth and the cell cycle in IECs (281). Increases in immunosuppressive TGF-β relates to a loss of TLR-2 production and the inhibition of NF-κB-dependent pro-inflammatory signalling in an IL-10-dependant manner. Colonisation of *E. faecalis* leads to inflammation (282), promoting aneuploidy and CIN in primary IECs, a common feature of tumour cells.

A feature of the TME of CRC patients is bacteria invasion, SGMB, a member of the Firmicutes phylum, colonises adenomatous and cancerous tissue in about 20% of CRC patients and it is argued that its induction of tumour progression occurs through the COX-2 pathway (232). Pro-inflammatory mediators IL-1β, COX-2 and IL-8 are elevated in SGMB-colonised patients relative to control or SGMB$^-$ patients. Colonisation and invasion of colon tumours by SGMB argues a probable active role in inflammation-based tumour development or dissemination through, not exclusively IL-1β, COX-2, and IL-8 (283). This coincides with an elevated occurrence of aberrant colonic crypts and the development of polyps in rats. Carcinogenic properties of SGMB are also associated with suppressed IEC barrier function. The *S. gallolyticus pil1* locus encodes for Pil1, a pilus protein which, through its collagen-binding domain, enhances translocation through paracellular epithelial cells (284). Interestingly, this *S. gallolyticus* subsp *gallolyticus* translocation does not trigger epithelial IL-1β and IL-8 responses, hence providing the bacterium with a growth advantage in the TME by evading host innate immune responses.

Metagenomic analyses have revealed associations between *Fusobacterium nucleatum* and CRC development. *F. nucleatum*, a Gram$^-$ anaerobic periodontal pathogen (285) of the Fusobacteria phylum (286, 287) is present at elevated concentrations in faecal (222, 288–290) and tumour tissue (238, 291–293) samples of CRC patients. *F. nucleatum* leads to the activation of immune responses and an increased level of inflammatory mediators (225). *F. nucleatum* is a prominent biofilm organiser with the ability to adhere to abiotic surfaces as well as host and bacterial cells via expression of a set of adhesins, such as FadA and Fap2 (294). Biofilm-containing CRC tumours feature enhanced IECs permeability, allowing bacterial signals (e.g. lipid A, lipopolysaccharides, flagella, microbial RNA and DNA)

to translocate, be recognised by PRRs and activate production of pro-inflammatory agents, such as IL-6 (295). The bacterium was found to be enriched at the early, pre-malignant stage of carcinogenesis, arguing for a potential role in initiation of tumorigenesis (295), consistent with its ability to adhere to epithelial cells and facilitate colonisation of other bacteria, including *E. coli* and *B. fragilis*. Furthermore, *F. nucleatum* may serve as a physical scaffold in supporting microbial shifts in the colon, modulating dysbiosis, thus promoting neoplasia over time.

Microarray analysis of gene expression in mice has shown induction of TLR-4/MyD88/NF-κB signalling in CRC tumour cells infected with *F. nucleatum*, as well as *in vitro*, where it was established that *F. nucleatum* enhances expression of *miR-21* via the same TLR-4-mediated pathways (225). *miR-21* is considered an onco-miR, non-coding microRNA, due to its contribution to many cancerous phenotypes across a number of cancers, for example promoting cell invasion and growth through the downregulation of PTEN, a tumour suppressor (296). Cell invasion, the direct entry of bacteria to the epithelium, can be facilitated by FadA through the modulation of the E-cadherin (CDH1) signalling pathway and subsequent activation of an array of transcription factors, e.g. NF-κB, T cell factor (TCF), c-MYC, E-cadherin and cyclin D1 lead to increased colonocyte proliferation (297). Ultimately, FadA is responsible for enhanced proliferation of tumour cells and the release of pro-inflammatory cytokines mediated by its binding to E-cadherin. *F. nucleatum* invasion of host cells can be achieved through the release of RNA into the host cells which is recognised by a cytoplasmic PRR, RIG-1, retinoic acid–inducible gene I, which up-regulates NF-κB signalling and activation of inflammation (298). Fap2, another adhesin produced by *F. nucleatum* binds surface Gal-GalNAc, a tumour-call disaccharide sugar motif, and helps in localising to tumour cells and prevents the killing of its own and these cells by NK cells. Interaction of Fap2 with TIGIT, the immune receptor, of NK and T cells, therefore, plays a major immunosuppressive role, contributing to cancer cell immune evasion. *F. nucleatum* promotes myeloid cell differentiation and infiltration, of e.g. $CD11b^+$ DCs, macrophages, myeloid-derived suppressor cells and neutrophils, increasing the expression of pro-inflammatory genes (287). Other immunomodulatory *F. nucleatum* functions that drive CRC, *n*-butyrate metabolism via unknown mechanisms involving its butyryl-CoA dehydrogenase, can suppress $CD4^+$ T-cell activity, hence reducing immunity to pathogen colonisation (288). This argues that *F. nucleatum* infection not only influences the TME, but also has an effect on tumour and hence disease progression as a whole.

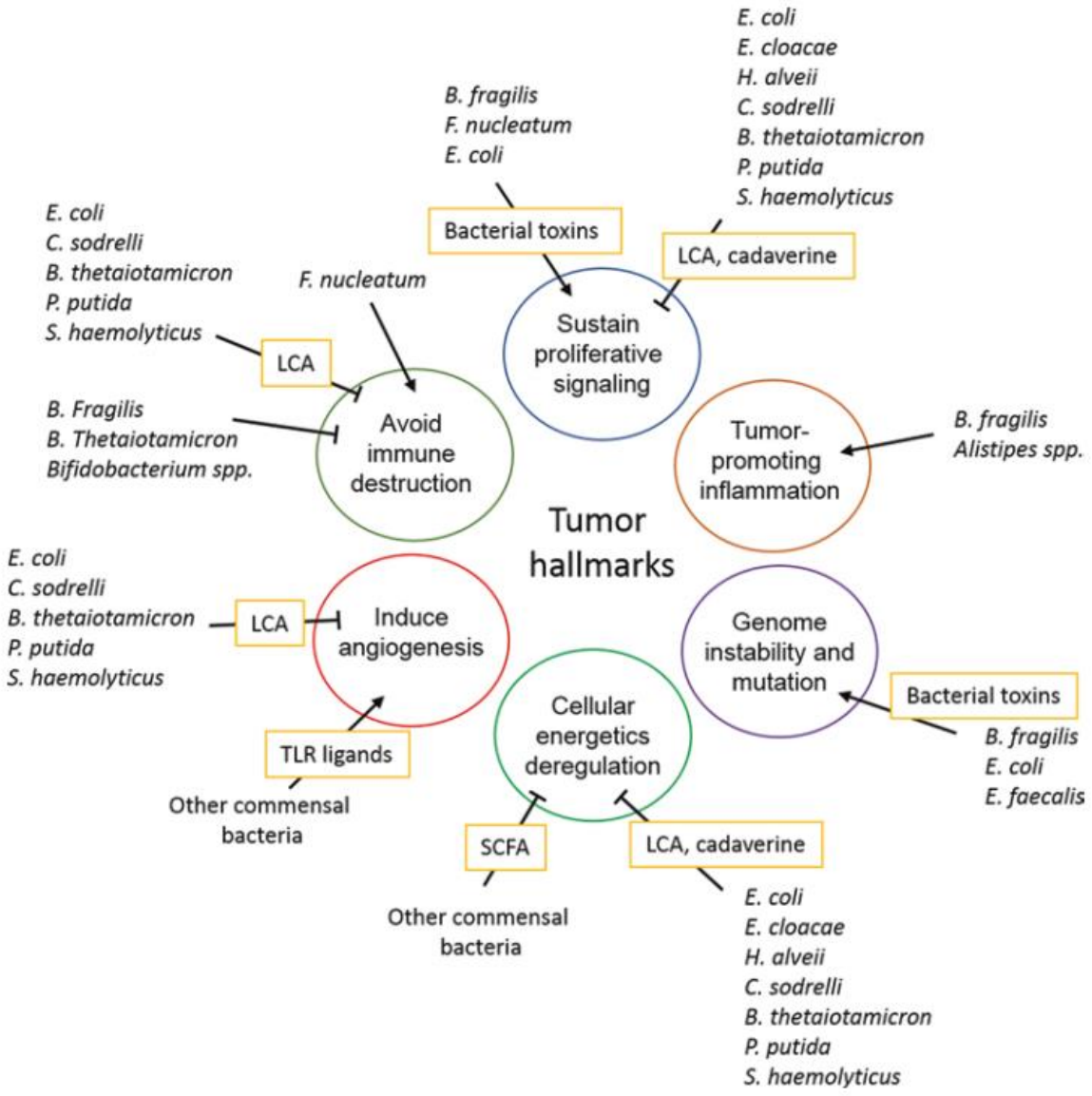


Fig. 8 | Potential roles of specific microbes on influencing cancer hallmarks. Microbial metabolites and microbe-host interactions driving tumour hallmarks. LCA, Lithocholic acid; TLR, toll like receptor; SCFA, short chain fatty acid; Adapted from (129).

2.6 Biofilms and Colorectal Cancer

Biofilms, a 3D bacterial plaque formed of multiple bacterial layers interlinked with a nucleic acid matrix, result from microbial surface adherence and organisation, involving communication by way of quorum sensing. Over 50% of CRC patients are positive for biofilms of the colonic mucosa, with different prevalence depending upon the site of cancer, right-sided (cecum, ascending colon up to the hepatic flexure) (Fig. 9A) being positive significantly more frequently (299). It is noteworthy that ~13% of healthy individuals are also colonic biofilm positive. The formation of biofilms represents bacterial cooperation for the survival of the community rather than the typical survival of the planktonic individual bacterium. This is achieved through multiple avenues, for example the physical barrier the biofilm affords microbes within the structure against any changing external environment, including drugs, limiting their permeability and hence delivery to the bacteria. This platform also gives the bacteria greater means for host invasion. Additionally, consistent cell-cell interaction provides microbes the opportunity to exchange genetic material, crucial for the dissemination of possible virulence and antibiotic resistance determinants, conferring any recipients' selective advantages.

Bacteria compete within the biofilm due to the cell-cell interactions, via production of ROS, toxins and antibiotics, such as novobiocin (299) (Fig. 9B), a non-ribosomally synthesised peptide antibiotic targeting DNA gyrase (topoisomerase II), GyrB. This in turn provides stress and selective pressures for more resistant species of bacteria and the spread of resistance determinants. Moreover, antibiotic penetration is limited due to numerous biofilm matrix components being antibiotic targets themselves. For example, antibiotics will target matrix polysaccharides, environmental (e)DNA, proteins (including DNA-binding proteins), proteoglycans, glycoproteins, lipids and LPS. Early/primary biofilm colonisers (see below) will often become persister cells, bacteria which transiently become antibiotic tolerant through arresting their growth (300), becoming dormant or quiescent, effectively rendering antibiotic targets inactive and hence the antibiotic ineffective. These cells can later regain their activity (germinate if sporulation has occurred) once the antibiotic pressure has passed.

Common early/primary biofilm colonisers include *Actinomyces* spp. and oral *Streptococcus* spp., however, unlike dental biofilms/plaques, mucosal and luminal biofilms of the large intestine harbour a diverse bacterial population. Bacteroidetes predominate, followed by Clostridia, Fusobacteria, Beta- and Gammaproteobacteria populations (Fig. 9C) (301). As previously stated, common biofilm forming species such as *F. nucleatum* and others have a strong association with CRC and its development, namely through their harbouring of pro-oncogenic genes (FadA and Fap2 in the case of *F. nucleatum*). Unsettlingly, introducing biofilms with these same members, as identified in humans, induces a cancerous phenotype

in healthy mice (302). ETBF as well as oral pathogens, *Parvimonas micra*, *Peptostreptococcus stomatis* and *F. nucleatum* all form a cohort of biofilm forming species however, *B. fragilis* and human oral microbes cohabitate tumoral biofilms in only 8% of cases. Spatial separation of microbial taxa throughout the colon may also play a role in biofilm formation, e.g. *Clostridium ramosum* is predominantly enriched in right-sided tissue biopsy (216), suggesting potentially different functional characteristics of biofilms depending upon taxonomic composition. There are however shared functional characteristics predicted though the extrapolation of 16S rRNA gene abundance to whole genome content of biofilm enriched species. These include peptidoglycan, cytoskeleton protein and peptidase biosynthesis and sporulation, while coinciding with a decline in flagellar motility (by e.g. Enterobacteriaceae) (Fig. 9B). Whether these and other genomically enriched biofilm features represent regulated processes at the level of transcription under CRC conditions is yet to be established.

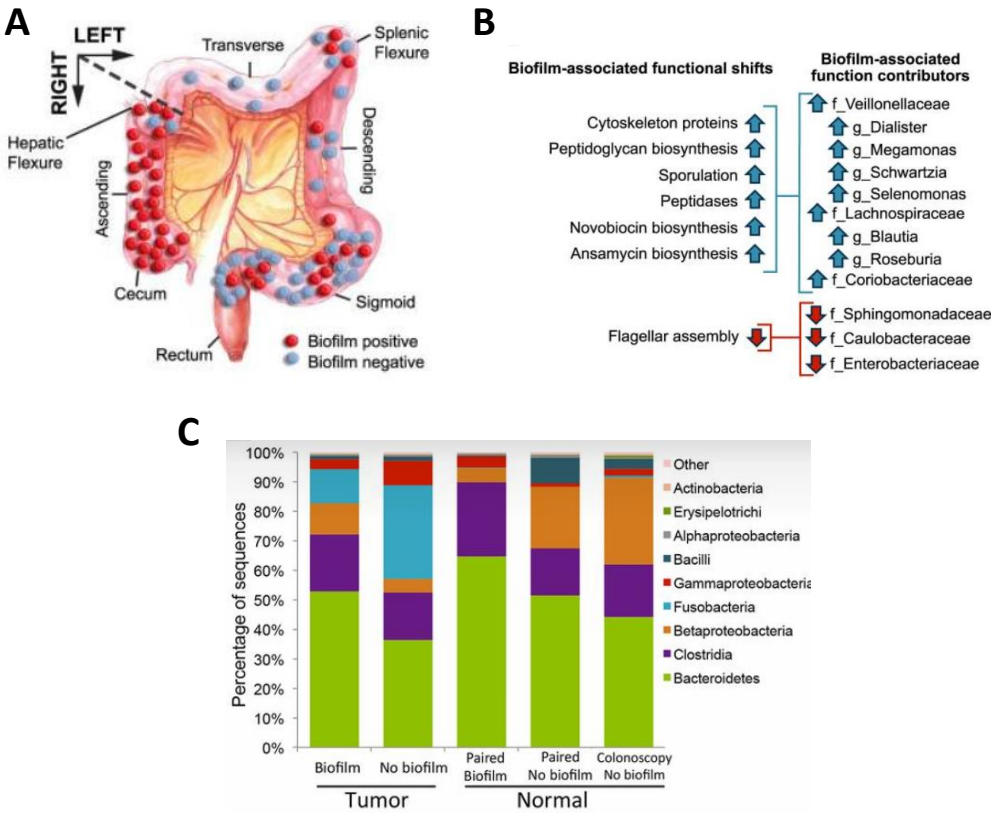


Fig. 9 | Sites and functional characteristics of colorectal cancer biofilms. A. Common sites of biofilm formation within the colon of CRC patients, predominantly right sided. **B.** Genome enrichment of functional features in biofilm associated taxa and the contributing taxonomic family or genus. **C.** DNA-based proportional composition of tumoral, non-tumoral biofilms and general colonic microbial population. Adapted from (299, 301).

2.7 Microbial Signals and the Host Epithelial Barrier

While many enteric bacteria display health-promoting characteristics (140, 303), however, some members of the microbiota may be detrimental to host health under altered conditions, such as diet, drug treatments and epithelial access. Therefore, it is of critical importance for the host to recognise and protect beneficial bacteria and eliminate infected/damaged cells and pathogens. All microorganisms are recognised by the immune system through sensing conserved microbial structures, such as LPS, lipoproteins, flagella, lipid-A alongside DNA and RNA fragments, comprising MAMPs, microbe-associated molecular patterns (304). These microbial signatures and endogenous compounds of damaged cells (DAMPs) are detected by PRRs (305), which signal an innate immune response to tolerate or trigger anti-inflammatory/anti-viral responses (304) by activating maturation of antigen presenting cells, such as DCs followed by the corresponding adaptive immune response (306). PRRs can be expressed on the membranes of IECs (non-immune cells) as well as innate immune cells, e.g. Mφ and DCs (307). Receptors for these ligands include TLR, NLR (NOD-like receptors), RLR, retinoic acid-inducible gene I-like receptors (RIG-1) and CLRs (C-type lectin receptors) (Fig. 10 and Fig. 11). Under physiological anaerobiosis, PRR-sensed beneficial bacteria or self-antigens signal the host immune system to release an appropriate level of specific cytokines to maintain tolerance.

Germline-encoded PRRs require adaptive proteins to prompt a specific immune response through activating transcription factors, including AP-1, NF-κB and IFN-γ (308). Localisation of PRRs, whether it is the epithelial membrane (TLR-1, 2, 4, 5, 6, and CLRs), the endosome (TLR-3, 7/8, 9 and 11), the cytoplasm (NLRs, RLRs or MDA-5, melanoma-differentiation-associated gene 5) or the mitochondria (the inflammasome NLRPs), reflects the molecular patterns of recognition (307). The epithelial cells of the intestine form a protective monolayer alongside a mucus layer comprised of mostly modified glycoproteins (309), physically separating the gut microbiota from the intestinal tissue.

Toll-like receptors (TLRs) are members of a superfamily of interleukin-1 receptors (TIR), and comprise a transmembrane and extracellular domain, 16-28 hydrophobic LRR modules, leucine-rich “LxxLxLxxN” repeats (310) and are involved in binding of PAMPs (311). TLRs are situated on the cell surface or intracellular space and are sensitive to fragments of microbial unmethylated CpG DNA, viral RNA/DNA and pathogen ligands from bacteria, fungi, helminths and protozoa (312, 313). PAMPs activate specific TLRs, which involves dimerisation and recruitment of adaptors, instigating downstream signalling and coordinated immune responses (Fig. 10). TLR-adaptor interactions through interleukin-1 receptor-associated kinases (IRAK1, 2 and 4) or IFN regulatory factor 3 (IRF3) trigger NF-κB signalling or IFN-γ-mediated anti-viral responses correspondently (314) (Fig. 11). During homeostasis

expression of IEC-localised TLRs (TLR-1, 2, and 6) is not generally induced, instead they are active basolaterally, TLR-5 is active on the base/side of polarised IECs or in the endosome, while TLR-3, 7 and 9 preclude interactions and activation of TLR-5 by PAMPs (307, 315). It was shown that such low-level recognition of microbial signatures by IEC-localised TLRs enhances protection against damage to the gut epithelium and requires induction of IL-6 and the chemokine KC-1 (316).

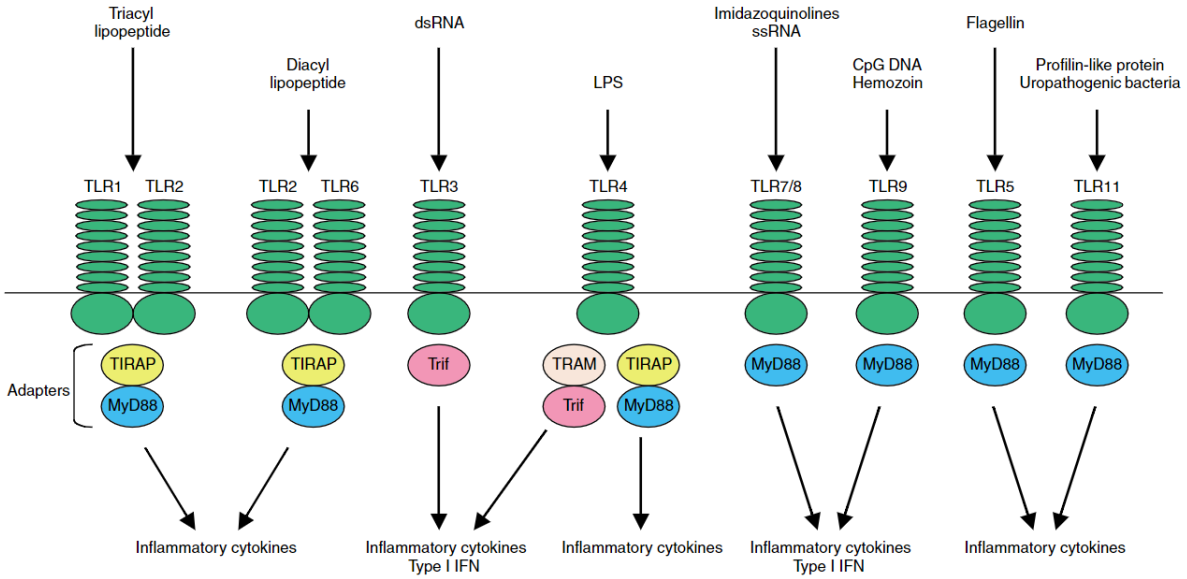


Fig. 10 | 10 human TLRs recognise different ligands and mediate tolerant or inflammatory responses via specific adaptors. TLR-2 alongside TLR-6 or TLR-1 distinguishes between the diacyl and triacyl lipopeptide molecular patterns, respectively. Viral dsRNAs are recognised by TLR-3, triggering type I IFN response. LPS, a signature of pathogens, are recognised by TLR-4 and depending upon the ligand can trigger either an anti-viral (IFN-γ) or inflammatory (cytokine-mediated) responses using different adaptors. Recognition of ssRNA and imidazoquinolines as well as bacterial and viral unmethylated CpG DNA recognised by TLR-7/8 and TLR-9, respectively, inducing IFN-γ responses via MyD88 signalling. Bacterial flagellin is recognised by TLR-5. Specific adaptor proteins recognise specific TLR to transduce the signal, except TLR-4, which recruits the adapter depending upon the origin of the ligand. TLR-1/2 and TLR-2/6 utilise TIRAP/MAL and MyD88 as adapters. TLR-3 employs Trif. TLR-4 uses TIRAP/MAL, TRAM, MyD88 and Trif while TLR-11, TLR-9, TLR-7/8 and TLR-5 use solely MyD88. Inflammatory responses are MyD88-dependent, while type I IFN responses are mainly mediated by Trif. Type I IFN signalling in a MyD88-dependent manner in plasmacytoid (p)DCs are prompted by TLR-7/8 and TLR-9 upon recognition of viral nucleic acids. Adapted from (308).

Activation and expression of TLRs strictly depends upon the nature of the ligand of specific microbes (Fig. 10). Liporpoteins/lipopetides which are always produced by pathogens, bacteria, including Mycoplasma (a bacterial genus of the phylum Mycoplasmatota (formerly Tenericutes) that lacks a cell wall), viruses as well as zymosan, a yeast glucan are recognised by a heterodimeric complex of TLR-2 with TLR-1 or TLR-6 (317, 318). Liporpotein/lipopetide-activated TLR-2/1 and TLR-2/6 induces NF-κB signalling, inflammasome assembly and promotes apoptosis of monocytes while activating phagocytosis of Mφ in response to peptidoglycans of yeast and $Gram^+$ bacteria (319), hence ensuring early recognition of the pathogen invasion and shaping the downstream immune response. While TLR-2-mediated pro-inflammatory responses are important for controlling microbial infection and clearing pathogens (320, 321) disproportional/excessive immune responses via TLR-2 can contribute to disease progression and tissue damage (322, 323). Upon activation by viral dsRNA present in the host cells during replication, TLR-3 signalling is actioned by the recruitment of the TRIF adaptor, which stimulates conserved signalling pathways. Subsequently, this up-regulates transcription factors NF-κB, activating protein 1 (AP-1) (324, 325) and interferon regulated genes through binding IRAK kinases (313, 326). LPS of $Gram^-$ bacteria induce TLR-4 mediated NF-κB signal transduction (327). Interestingly, unlike any other TLRs, TLR-4 can elicit two independent signalling pathways through the recruitments of four different adaptors. Utilisation of TIRAP/MyD88 would generally control inflammatory responses, regulating IECs proliferation and apoptosis through activating COX-2 (328). Some TLR-4 ligands can activate MyD88-independent signalling of TLR-4, through recruitment of the TRAF3 and IRF3 adaptors, activating a type I interferon response (329).

Colonisation of some bacteria, e.g. commensal *E. coli* or *Salmonella* involves expression of flagellin, which is polymerised to form a flagellum. The host immune system recognises flagellin of flagellated microbes through TLR-5. Ligation of IEC TLR-5 to flagellin has a cyto-protective role (330). However, if the GI barrier is compromised, flagellated bacteria can be pathogenic, activating basolateral TLR-5, inducing pro-inflammatory cytokine TNF-α (331). Furthermore, *Salmonella* flagellin-induced TLR-5 can activate pro-apoptotic pathways, such as caspase-8, signifying a role of programmed cell death in controlling enteric infections. TLR-7/8 recognises viral ssRNA (e.g. vesicular stomatitis and influenza viruses) and induces IFN-α signalling through activation of pDCs and Mφs (332) by way of MyD88 (333).

Bacteria can be sensed by the immune system through their DNA (which features unmethylated CpG islands), inducing a strong polarisation of T_{h1} cells by activated DCs (334). Contrary to the bacterial genome, host DNA has a lower prevalence of CpG sequences, which are generally methylated and recognised by immune cells as self-antigens (335), hence displaying a non-immune stimulatory phenotype, namely immune tolerance. Bacterial CpG

dinucleotides are recognised by TLR-9 in the endosomes, triggering MyD88-mediated induction of NF-κB signalling, and IL-8 production (336).

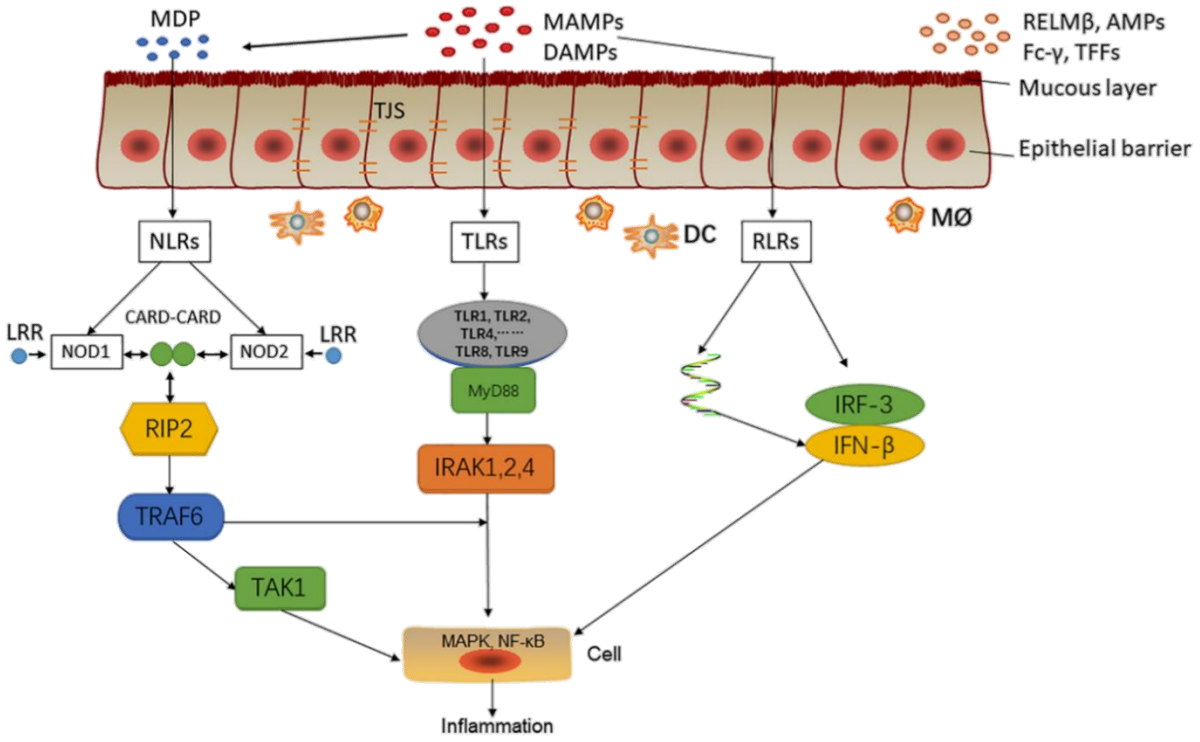


Fig. 11 | Signalling pathways triggered by PRRs upon activation. Microbial patterns are ligands for NLR, TLR and RLR members and prompt signalling pathways within the gut epithelium. Epithelial MAMPs and DAMPs activate signalling pathways through TLRs and RLRs. MAMPs, including MDP (muramyl dipeptide), which is found in the PG of both Gram$^-$ and Gram$^+$ bacteria and are also recognised by NLRs. NOD-1 and NOD-2 interact with RIP-2 via recognising caspase recruitment domains (CARD-CARD), stimulates TRAF-6, which subsequently recognises TAK-1, triggering NF-κB and MAPK. TLRs, TLR-1, 2 and 4, bind the MyD88 adapter which binds IRAK-1, 2, and 4 to activate pro-inflammatory NF-κB. RLRs can recognise viral RNA, releasing IFN-β, and which also activates NF-κB. Upon disruption, due to damage or pathogen invasion, the resulting chronic inflammation may lead to carcinogenic lesions. NLR, Nod-like receptor; TLR, toll-like receptor; RLR, retinoic acid-inducible gene I-like receptor; RIP-2, receptor-interacting protein 2. Adapted from (4).

Damaged IECs require a robust mechanism of self-destruction to prevent triggering an immune response. This involves self-DNA and RNA utilisation and avoidance, encoding of endogenous nucleases and compartmentalised and specialised innate PRRs (337). These features can occasionally be recognised by the immune system. If damaged epithelial cells release self-antigens (undigested DNA/RNA) this triggers the secretion of antibodies, underpinning autoimmune disorders (338). Hence, it is important to distinguish self-products of controlled degradation (apoptosis) or damage from *similar* molecular patterns of microbiota.

If colonic epithelial cells are infected (termed invasion) with pathogens and/or microbes that passed through the mucus barrier and interact with the epithelial membrane, PAMPs can be translocated to the cellular cytoplasm. Once PAMPs reach the intracellular compartment of IECs they are recognised by a cytoplasmic detection mechanism which is mediated by another set of receptors, termed nod-like receptors, NLRs initiating the intracellular defence response. The NLR-mediated cytoplasmic defence is likely involved in host immune defence wherever TLRs or their adaptors (e.g. MyD88) are minimally expressed or lacking entirely (339, 340). Intracellular signalling in response to cytoplasmic PAMPs and endogenous tissue damage products, DAMPs, is mediated by 23 NLR genes in human (341). NLRs detect cytosolic bacterial peptidoglycan (PG, a motif characteristic of γ-Proteobacteria) of invasive *Shigella flexneri* (342), *E. coli*, *Pseudomonas aeruginosa* or non-invasive *H. pylori* bacteria. Recognition of signals by NLR initiates a number of cellular defence mechanisms, including NF-κB, AP-1, type I IFN and inflammatory caspases (343).

NLRs are comprised of a central oligomerisation domain, a C-terminal LRR domain, NODs, within the family of NLRs, contain an N-terminal effector domain with caspase recruitment domains (CARDs), which is important for the function of inflammasomes, and NLRPs contain a pyrin domain which aids in the identification of flagellin in the cytosol through formation of inflammasomes (341). NOD-1, ubiquitously expressed in all adult tissues, recognises a dipeptide, γ-D-meso-DAP, iE-DAP and a tripeptide signature GlcNAc-MurNAc found in $Gram^-$ (SFB, *Bacteroides* spp., Clostridiales and Enterobacteriaceae) bacterial peptidoglycan (344). NOD-2 is expressed in a variety of cells, DCs, Mφs, Paneth, and epithelial cells, in response to recognition of bacterial PG degradation products and muropeptides, e.g. muramyl dipeptide (MDP), the common peptidoglycan motif of all bacteria (345). NOD-2 is important in maintaining bacterial tolerance of IECs and controls mucosal colonisation by specific bacteria, mutations in the *NOD-2* gene (e.g. L1007fsinsC) leads to increased Bacteroidaceae mucosal localisation (346). Consistent with this, NOD-2 activation is required for expression of AMPs and defensisns in Paneth cells (347). Hence, NOD-2 appears to be a general sensor for both $Gram^+$ and $Gram^-$ bacteria and is important for prevention of pathogen localisation and thus, dysbiosis. MDP-bound NODs are localised to

the plasma membrane and activate pro-inflammatory immune responses (NF-κB and MAPK) through recruitment of receptor-interacting proteins (e.g. RIP-2). For example, in response to *Listeria monocytogenes* IL-6, TNF, and IL-1β cytokine expression is induced, but in particular IL-8, in both monocytes and DCs, whereas the NOD-1 ligand appears to be less potent (348). Mutations in *NOD2* that result in the lack of recognition of MDP (3020insC, G908R, R702W, and L1007fsinsC) have been shown to be associated with auto-inflammatory Crohn's disease (349, 350). This implies that a lack of bacterial sensing through the loss of interaction between mutant NOD-2 and MDP contributes to the pathology of disease.

NOD-1 and NOD-2 modulate the activity of proteins that inhibit apoptosis, such as cIAP1 and cIAP2 (cellular inhibitor of apoptosis proteins containing C-terminal RING-finger domains with E3 ligase activity) (351). cIAP1 or cIAP2 binds and ubiquitinate RIP-2, adding K63-linked ubiquitin units to the kinase protein, ubiquitinated RIP-2 binds and activates TAK-1 (TGF-β-associated kinase 1) (352). Kinase actions of TAK-1 stimulates both MKK and IKKγ, which results in NF-κB and MAPK translocation to the nucleus, triggering upregulation of pro-inflammatory and defensin gene transcription, thereby mediating antibacterial effects. Viral ssRNAs can also be sensed by NOD-2, activating a mitochondrial antiviral membrane signalling protein (MAVS) (353) leading to dimerisation and activation of IRF-3 which induces the IFN-β response (354) as NOD2-deficient mice are highly susceptible to viral infections (355). Double knockout of NOD-1 and 2 in mice decreased expression of E-cadherin, causing increased gut permeability and impaired antimicrobial function (356). NLRPs are important for tissue damage repair through the production of IL-18 and block the expression of IL-22BP, mediating the activity of T_{h17} cells as discussed above (Fig. 6B).

TLR-3, 7/8 and 9 which recognise viral and bacterial nucleic acids in the endosomes cannot detect viral dsRNAs, a signature of actively replicating viruses. Two cytoplasmic RLRs, RIG-I (retinoic acid inducible gene I) (357) and MDA-5 (358) mediate dsRNA-induced antiviral responses. They promote phosphorylation and activation of IRF-3 followed by the induction of the IFN-β antiviral defence (359). RIG-I and MDA-5 contain CARD-like structures that are required for signalling. Similar to NOD-2-ssRNA signalling, CARDs interact with a mitochondrial adapter MAVS and initiate intracellular signalling pathways through IRF-3 to NF-κB and AP-1 via TBK-1/IKKi, IKKα/IKKβ and MAPKs, respectively. While RIG-5 targets 5'-triphosphate of dsRNA <1 kb, MDA-5 recognises dsRNA longer than 2 kb, ensuring that different lengths of replication products can be detected by host defence mechanisms. dsDNA derived from pathogens or apoptotic cells are degraded by the Mφs phagosome through the activity of DNase II, a primary lysosomal DNase (360). If phagosomes fail to clear DNA, Mφs will induce IFN-β, resulting in a TLR-9 independent inflammatory response (361) and evoke IFNs signalling in pDCs independently of TLRs (362). dsDNA activated IRF-3 and the *Ifnb*

promoter is recognised by through IKKi and TBK-1 but not RIG-1 or TLR-9. Intracellular viral dsDNA and possibly RNA can be recognised by a nucleic acid binding protein HMGB-1, high-mobility group box. HMGB1-3 serve as initial sensors of dsDNA in the cytoplasm and activate the type-I interferon response via IRF-3 and IRF-7 (363). Depletion of all three HMGBs impairs induction of TLR-9, TLR-7/8 and TLR-3 by their nucleic acid antigens. Alternatively, immunogenic dsDNA can be transcribed by RNA Polymerase III into dsRNA which in turn is recognised by cytoplasmic RIG-I/MAD-5 receptors, inducing type I IFN signalling. These data become relevant to CRC as recent studies have found that the intestinal virome specifically concerning bacteriophages are enriched and more diverse during disease and associated with a decreased diversity of bacteria within the gut. Not only is this process of viral expansion correlated with CRC progression from early to late stages, but this enrichment also has a distinct negative effect on patient prognoses. This may be due to bacteriophages influencing the behaviour of tumour cells through either altering the characteristics of specific species or the composition of gut bacteria as a whole, thus affecting bacterial invasiveness, and ultimately tumour progression (364).

Other constituents of the gut microbiota are fungi, while their metabolic contribution to the host may be minimal, there are potentially pathogenic fungi present, which may take advantage of a susceptible immune system (365). CLRs, C-type lectin receptors recognise carbohydrates on microorganism surfaces such as β-glucans present on fungi (366). Dectin-1 and -2 are immune-receptors, which are responsible for activating DCs to instruct T cells to confer protective immunity with respect to e.g. *Candida albicans*, an opportunistic pathogenic yeast, and a common member of the human gut microbiota. The Mφ and phagocytic Mincle (Clec4e and Clecsf9) C-type lectin receptors recognise colonisation and infection by pathogenic fungi such as *Malassezia* (via sensing alpha-mannosyl residues) as well as *C. albicans* (367) along with some self-ligands. These fungal signalling pathways induce Th_{17} responses and the induction of pro-inflammatory cytokines, such as IL-6 and TNF (367). Therefore, elevated fungal levels within the GI lumen may affect the inflammatory state alongside bacterial dysbiosis-associated inflammation, contributing to inflammation-mediated disease e.g. IBD and ultimately CRC.

2.8 Microbiota-Host Gastrointestinal Metabolism

The composition and activity of gut microbiota co-evolved with the host and is subject to diverse interactions that depend upon the host genome, nutrition and environment. The gut microbiota are critical for host homeostasis, underpinning collaborative host-microbiota metabolic, signalling, and immune-dependent inflammatory processes. The mechanisms by which host-microbe driven homeostasis and disease are determined involve a number of specific responses of the host epithelium to the availability of bacterial metabolites. The

microbes are capable of generating metabolites linked to disease, or beneficial compounds which can protect against disease in a concentration-dependent manner, such as e.g. SCFAs (368). However, even “health promoting” metabolites can be harmful at too high or low levels. Homeostasis of the human colonic epithelium in turn provides the microbiota with a stable environment. Recent data have shown that changes in the composition of the gut microbiota are associated with a number of diseases while the metabolic state of the epithelial cells dictates the overall microbial configuration both taxonomically and metabolically (369). However, it is this symbiotic relationship and its importance in dictating human health that drives our need to understand the mechanisms involved in their maintenance, and the factors that may disrupt them individually or collectively ultimately dictating the health status of an individual.

2.9 SCFA in the Intestinal Mucosa

SCFAs, mainly referring to *n*-butyrate, propionate and acetate, are key bacterial-derived metabolites, which are integral to human health as immune and metabolic mediators. *n*-Butyrate is the major energy currency for GI microbiota and colonocytes while propionate is mostly metabolised by the liver and acetate is transferred into the peripheral blood (370). They are also known to be anti-inflammatory by supressing NF-κB signalling (371), possessing anti-oncogenic properties (372), regulating gut barrier functions through tight junction integrity (373), mediating permeability of the epithelium and mucus production, hence protecting the colon against pathogen colonisation and influencing bacterial adherence (374). Luminal pH is lowered by production of SCFAs and this environmental factor alone can limit pathogenic colonisation as well as increase the absorption of dietary compounds (375). SCFAs are also implicated in immunomodulation and epigenetic functions through signalling via interaction with G-coupled receptor proteins, GCPRs (GPR109A, GPR41 and GPR43) activating anti-inflammatory responses (376, 377) and intracellular inhibition of histone deacetylation in a concentration-dependent manner (e.g. HDAC9) (378) in order to modulate immunological signalling cascades. Absorption of SCFAs is facilitated by transporters such as *SLC16A1*, the gene encoding MCT1, the proton-coupled mono-carboxylate transporter-1, and *SLC5A8*, encoding SMCT1, the Na^+-coupled mono-carboxylate transporter-1, promoting cellular metabolism (379, 380).

SCFAs, mainly *n*-butyrate, that are not absorbed within the intestine will be transported to the liver where they are metabolised and stored (182). SCFAs serve as substrates in carbohydrate and lipid metabolism, propionate is the primary substrate in gluconeogenesis with *n*-butyrate and acetate primarily involved in biosynthesis of lipids. An increase in *n*-butyrate has also been shown to decrease the incidence, with high fat diets (HFD), of obesity and insulin resistance in animal models (381). The protective effect of SCFAs against HFD-

induced metabolic shifts appears to depend upon peroxisome proliferator-activated receptor gamma (PPAR-γ), the intracellular butyrate sensor, down-regulation of which promotes a change from synthesis to oxidation of lipids (382). Signalling pathway activation mediated through protein kinases such as AMP-activated protein kinase or mitogen-activated protein kinases (MAPK) are commonly observed during these shifts in metabolism. The anti-inflammatory actions of *n*-butyrate work through signalling for the production of p25 and p72, cytoprotective heat-shock proteins acting to supress pro-inflammatory MAPK and NF-κB (383). Propionate and *n*-butyrate have also been shown to be involved in central nervous system (CNS) signalling through the induction of hormone synthesis and therefore reducing food intake (384), while acetate can interact with the CNS to reduce appetite (385). Therefore, not only can SCFA availability have profound implications for human metabolism directly but also for modulating behaviours in order to regulate consumption of dietary substrates like NDCs, non-digestible carbohydrates (described below).

SCFAs provide a protective environment against CRC in a dose-dependent manner, with most studies focussing on *n*-butyrate, the major energy source for the colon (386, 387). β-Oxidation of *n*-butyrate by the mitochondria consumes more than 70% of available O_2, thus promoting an anaerobic colonic environment which selects for the growth of health-promoting anaerobes in the gut (388), many of which aid in the prevention of CRC. Cancerous colonocytes undergo metabolic transformation, including enhanced anaerobic glycolysis, termed the Warburg effect (389) (characterised by increased glucose uptake and fermentation of glucose to lactate (390), causing an increase in O_2 due to its low consumption and acidification of the gut). This phenomenon leads to the accumulation of *n*-butyrate in the nucleus, increasing inhibition of histone deacetylation, causing cancerous colonocyte apoptosis. It has recently been shown that *n*-butyrate, secreted by *Holdemanella biformis in vitro*, can exert an anti-proliferative effect on human tumour-cells without compromising their viability, in a dose-dependent manner, 1-2mM (391). *n*-Butyrate blocks proliferation by enhancing histone H3 acetylation (H3K27ac), confirming its inhibition of histone deacetylase activity. This in turn led to downregulation of calcineurin, hence repressing NFATc3 (nuclear factor of activated T cells cytoplasmic-3) activation. While these data shows anti-tumorigenic properties of pure *H. biformis* cultures, it is not known whether this or other butyrate-producing bacteria can produce sufficient SCFAs in the gut and if this efficacy is mediated by the activities of other microbial species. In immune cells, *n*-butyrate and propionate but not acetate through inhibiting deacetylation of histones induce $CD4^+$ T_{reg} ($Foxp3^+$) cell differentiation, exposure of DCs to these SCFAs triggers *Foxp3* and *IL-10* expression, aiding in the control of GI inflammation (Fig. 12) (167, 392). Enhanced acetylation of H3 histone in the promoter at the *Foxp3* locus in a *n*-butyrate-dependent manner promotes expression of the *Foxp3* gene (168),

as described above. Levels of mucosal and faecal SCFAs have been found to be reduced in patients with inflammatory bowel diseases, UC and CD (393, 394). Therefore, chronic inflammatory responses that underpin some GI diseases may be caused by SCFA deficiencies brought about by or causing dysbiosis (395–398).

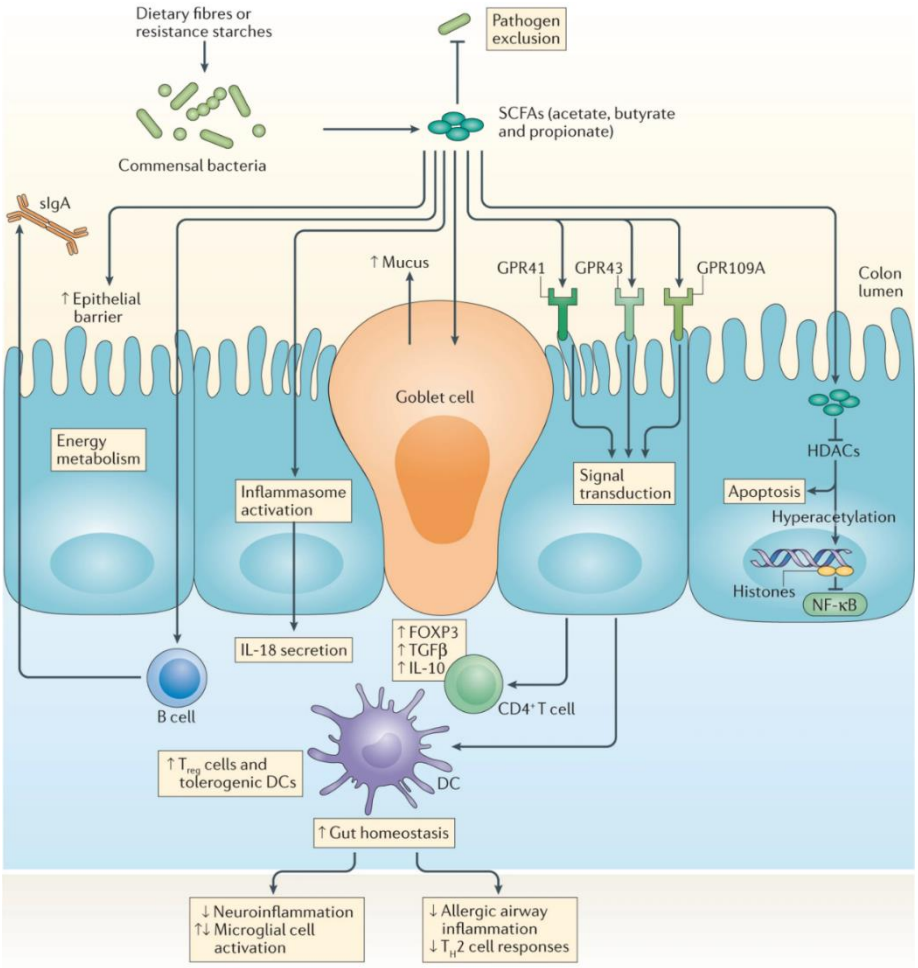


Fig. 12 | SCFAs, major bacterial-derived fermentation products. Luminal *n*-butyrate, propionate and acetate are sensed by the epithelial surface and immune cells, Mφs, neutrophils and possibly T cells (not shown) G protein-coupled receptors, GPR109A (butyrate), GPR41 (propionate) and GPR43 (propionate, acetate and to a lesser extent for *n*-butyrate) (392). *n*-Butyrate and propionate inhibits HDACs, leading to histone hyper-acetylation, apoptosis of CRC cells and prevention of the NF-κB nuclear localisation. Activation of SCFA receptors and inhibition of HDAC-9 exerts anti-inflammatory effects via enhancing polarisation of colonic $CD4^+FoxP3^+T_{reg}$ cells and production of anti-inflammatory IL-10 and TGF-β as well. SCFA-dependent activation of inflammasomes triggers secretion of IL-18, hence promoting production of AMPs and boosting the epithelial barrier function. GPR109A on mucosal $CD103^+$DCs upon engagement with *n*-butyrate controls expression of the retinal gene *Aldh1a2*, encoding for RALDH2 dehydrogenase, the tolerogenic acid-producing enzyme resulting in differentiation of $CD4^+$T cells into FoxP3-producing T_{regs} (399). Acetate-dependent activation of GPR43 on DCs also stimulates production of retinoic acid, which induces production of sIgA by B cells (398). *n*-Butyrate promotes the epithelial barrier function through up-regulation of *muc2* and *muc3* genes expression by Goblet cells, increasing mucus production, as well as depletion of intracellular O_2 (due to a high level of oxygen consumption by β-oxidation) and activation of HIF-1, stimulating TJ function. Adapted from (400).

2.10 Gastrointestinal Environment, Diet and Bacterial SCFA production

The environment of the colon itself has a profound effect on the native microbial populations, influencing bacterial selection through pressures such as pH, intraluminal gasses O_2 and N_2 and dietary components forming an ideal anaerobic environment for the breakdown of complex dietary fibres. Selecting for the growth or reduction of SCFA-producing bacteria upon environmental pressures may cause shifts in the microbial population that alters the levels of SCFAs (401). Therefore, environmental factors through modulating the microbial community composition may influence its capacity for carbohydrate fermentation, affecting energy harvesting and storage. Gut pH plays a major role in this bacterial selection, for example, mildly acidic pH (~5.5) was shown to inhibit growth of *Bacteroides* spp. and *E. coli* relative to Actinobacteria and Firmicutes in the presence of SCFAs (33 mM acetate, 9 mM propionate and 1 mM of iso-valerate, valerate and iso-butyrate) due to the lack of tolerance of *Bacteroides* spp. to SCFAs at pH 5.5 (402, 403). Shifts such as these enhance *n*-butyrate production while curtailing propionate when compared to pH 6.5-6.8. Bacteria which utilise the butyryl-CoA:acetate CoA-transferase route, *n*-butyrate production and acetate consumption increase at a mildly acidic pH when compared to pH ~6.7 (404). It is worthy of note that *Roseburia* spp. and *F. prausnitzii*, major *n*-butyrate producers, do not consume acetate, while *n*-butyrate production typically requires acetate consumption (405). Ultimately, acidic pH (~5.5) leads to an increase in acetate consumption and *n*-butyrate production, hence maintaining gut homeostasis while near neutral pH (~6.7) displays the inverse outcome. This shows the importance of pH to the GI environment in modulating the SCFA metabolism capacity of the microbiota.

O_2 levels within the healthy colon are minimal due to high oxygen consumption via host mitochondrial oxidative phosphorylation during energy production, this strictly anaerobic environment maintains the selection of *Bacteroides* and *Clostridia*, thus sustaining an *n*-butyrate-producing population (406). This is consistent with antibiotic-induced oxygenation leading to expansion of opportunistic pathogens, *E. coli* and *Salmonella* and supressing the growth of SCFA-producing bacteria (407). Depletion of SCFAs therefore decreases activities of SCFA receptors, GCPRs, silences PPAR-γ signalling and reduces the T_{reg} cell population, hence promoting inflammation of the colon (408, 409). Increases in Enterobacteriaceae populations are associated with reduced *n*-butyrate/PPAR-γ signalling, promoting an aerobic environment (increasing aerobic bacteria e.g. *Salmonella typhimurium*) and increasing iNOS, *i*nducible *n*itric *o*xide *s*ynthase, and nitrate levels further favouring Enterobacteriaceae expansion which in turn enhancing the colonial level of O_2 (407, 410).

2.11 SCFA Signalling and Transport in the Intestinal Mucosa

Propionate, *n*-butyrate and acetate all exert beneficial effects on the host by interacting with colonocytes and immune cells. They regulate their own absorption, intracellularly or extracellularly via inhibition of histone deacetylases (HDACs) and SCFA-specific GCPRs. SCFAs activate signalling pathways through at least three different G protein-coupled receptors, GPR109A (hydroxycarboxylic acid receptor 2, HCAR2), GPR43 (free fatty acid receptor 2, FFAR2) and GPR41 (free fatty acid receptor 3, FFAR3) found on many host cells, including colonocytes and immune cells (Fig. 12 and Fig. 13). GPR43-mediated repression of inflammation occurs through activation of FFAR2-expressing neutrophils which supress expression of inflammatory receptors, CXCR2 and C5aR (411). GPR43 may be activated by propionate and acetate at an effective half-maximal concentration (EC_{50}) of 50-200 μM, while propionate can activate GPR41 at an EC_{50} of only 10 μM. As these concentrations are found to be maintained during homeostasis, it is likely that both GPR43 and GPR41 are constitutively activated. GPR109A engages with *n*-butyrate, however, is less potent relative to acetate and propionate, responding to a concentration of ~1 mM, this concentration is only located within terminal ileum lumen and throughout the colon. MCT1 (*SLC16A1*), the primary transporter for *n*-butyrate uptake and MCT4 (*SLC16A3*), another H^+-coupled low-affinity monocarboxylate (e.g. lactate and pyruvate) transporter, are expressed in the apical and basolateral membranes of the colonic epithelium (412, 413). SMCT1 (*SLC5A8*) and SMCT2 (*SLC5A12*), Na^+-coupled monocarboxylate transporters are expressed exclusively in the apical membrane, in the crypts of proximal and distal colonocytes (414).

An anion-exchange mechanism of SCFA entry in anionic forms coupled to bicarbonate efflux is an alternate pathway of absorption in this membrane. All mechanisms of entry are regulated through the substrate themselves, making SCFAs key signalling molecules for their own translocation. For example, *n*-butyrate enhances transcription of *MCT1*, increasing its abundance and the SCFAs transport. Interestingly, the microbiota appear to modulate *n*-butyrate uptake by regulating transporter availability, *E. coli*-dependent endocytosis is attenuated by *Lactobacilli acidophilus*, enhancing MCT1 function (415). In UC and CRC, it has been observed that the expression of SCFA transporters is significantly altered. Specifically, the carcinogenic effects occur through promoter methylation and transcriptional modulation via p53 and HIF-1α (416). The promoter of *MCT1* for example contains binding sites for various pro-inflammatory transcription factors, NF-κB (MAPK), AP-1 and AP-2, Sp1, USF, TCF/LEF and c-MYC, thus potentially reduces the IECs ability to uptake beneficial bacterial-derived metabolites (417–420). Enhanced expression of *SLC16A1* coincides with the induction of IL-8, TNF-α and IFN-γ expression, suggesting that inflammation can influence the expression of the *n*-butyrate transporter MCT1. Similar to SCFA-mediated transporter

expression, epithelial GPR109A and GPR43 production in both human and murine intestinal mucosa is intrinsically correlated with colonic bacterial production of metabolites (421–423). These findings suggest that expression of these receptors is mediated by their respective substrates. Since GPR109A levels reduce in GF mice and are restored upon bacterial recolonisation, this argues that a decrease in SCFA levels resulting from deficiency in carbohydrate-fermenting bacteria causes down-regulation of SCFA-sensitive GCPRs. However, mucosal *FFAR3* and *FFAR2* mRNA levels (respectively encoding GPR41 and GPR43) were shown to be consistent irrespective of diet (424), suggesting expression of its metabolite-sensing receptors genes are not regulated by dietary fibre in these animals. Bacteria-produced SCFAs are essential in maintaining colonic homeostasis, via regulating GCPRs and transporters as well as certain immune functions, their regulation is critical in determining colonic function and therefore in human health and disease.

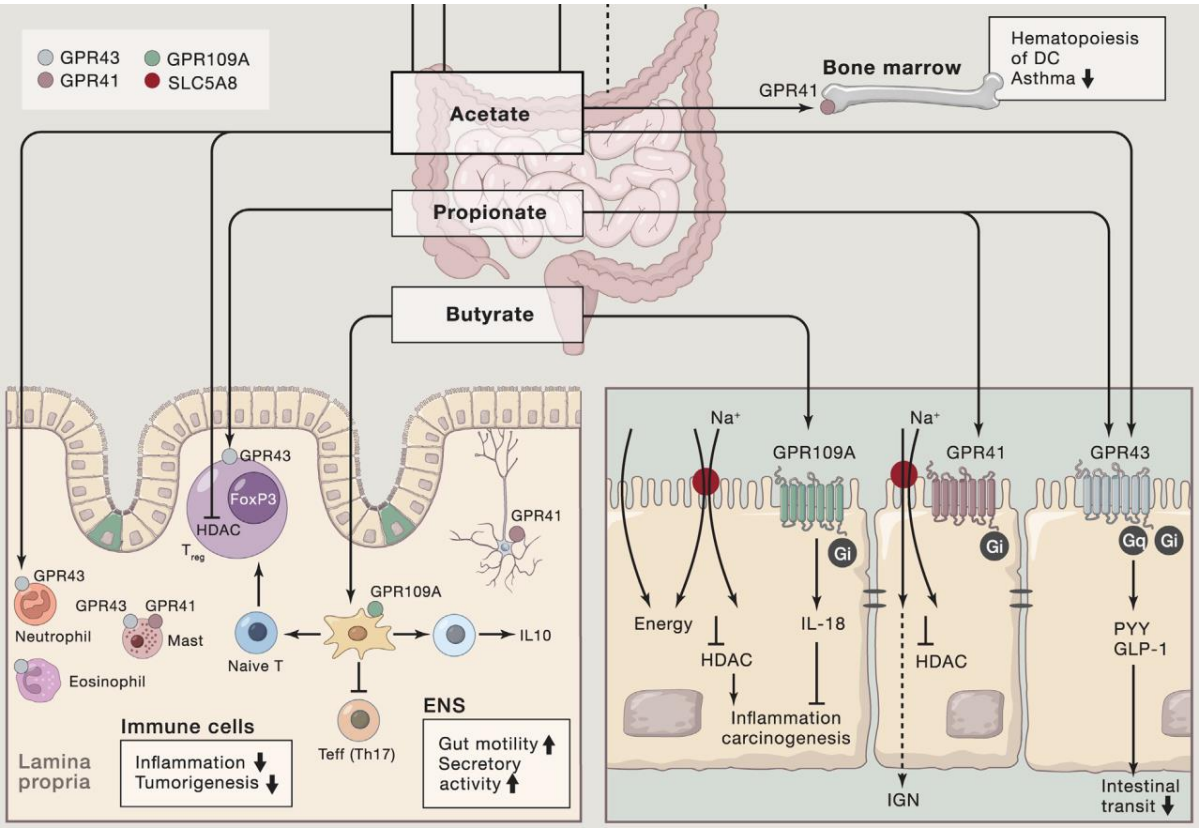


Fig. 13 | Microbial SCFAs are important for controlling inflammation and homeostasis. Within the distal colon, SCFAs are transported into cells by either passive (diffusion) or active (SLC5A8-mediated) mechanisms, serving as energy sources or an epigenetic regulator of HDACs, histone deacetylases. Propionate or acetate within the gut lumen via GPR43 and GPR41 releases GLP-1 and PYY, influencing transit. GPR109A senses luminal *n*-butyrate and mediates anti-inflammatory properties and inhibition of HDACs. Moreover, IGN can convert propionate to glucose, causing satiety (feeling of fullness) and reduction in glucose production by the liver. SCFAs can also act by stimulating secretory and activity motility, or stimulate Lamina propria immune cells, to produce anti-inflammatory signals and reduce tumorigenesis. PYY, peptide YY; GLP-1, glucagon like peptide-1; IGN, intestinal gluconeogenesis. Adapted from (425).

2.12 SCFA Regulation of Intestinal Barrier Function

SCFAs hold many crucial roles within the gastrointestinal tract, regulating intestinal barrier function is fundamental being the first line of defence of the innate immune system, separating bacteria from the mucosa and protecting against pathogen colonisation, thus minimising inflammatory responses. The GI tract has a single epithelial cell layer containing Goblet cells (Fig. 12) responsible for the secretion of mucin, charged glycoproteins which form mucus, a physical barrier between microorganisms and the epithelium (426). The mucus through its viscosity traps microbes and their products, restricting contact to the epithelium itself. The mucus inner layer is a scaffold for biomolecules, secretory IgA (produced by plasma cells) and AMPs (produced by Paneth cells), hence restricting bacterial adhesion to epithelium (Fig. 6A) (427). Tight junctions, responsible for the networking of colonocytes, govern the movement of water, nutrients and ions, while restricting the access of pathogens to the epithelium, thus controlling permeability of the intestinal mucosal barrier, cell polarity and maintaining homeostasis (32). Therefore, the mucus and epithelial integrity must be maintained at a high level with sufficient mucus and tightly interconnected colonocytes. Transmembrane proteins, such as occludin, claudin-1 and claudin-4 make up TJs and are responsible for forming a paracellular barrier via the sealing of intercellular spaces, and plaque proteins e.g. ZO-1/TJP1 situated intracellularly form an anchor point for transmembrane proteins (373). *n*-Butyrate at a low concentration of 2 mM has been shown to significantly improve intestinal barrier function of Caco-2 epithelial cells through decreasing permeability, however at higher concentration (8 mM) impairs barrier function, inhibits proliferation and promotes apoptosis (368).

Changes in the thickness or permeability of the mucus layer (428) can lead to inflammation due to direct contact between microbiota and epithelium as has been observed in *Muc2*$^{-}$ mice (429). GF mice when compared to conventionally raised mice have easily penetrable mucus in their large intestine which can be restored by transplantation of healthy gut microbes, arguing the barrier function is directly regulated by the GI microbiota (33). Although the majority of bacteria metabolise NDCs, some microbes, e.g. propionate producer *Akkermansia muciniphila* can metabolise mucin glycans as their source of energy (303) or modify (*Peptostreptococcus russellii*) it to enhance epithelial protection (430). Facultative anaerobe *Lactobacillus rhamnosus* promotes IEC renewal, hence promoting wound healing (157) and *Lactobacillus plantarum* are also implicated in enhancing IEC integrity (158, 431). The level of the mucus layer is modulated by the GI microbiota through exposure to bacterial peptidoglycan or lipopolysaccharides. Commensal species, such as *Bacteroides thetaiotaomicron* and *F. prausnitzii* coordinate mucus glycan production (189). *Lactobacillus casei*, *Ruminococcus gnavus*, and *B. thetaiotaomicron* are important for mucin expression,

remodelling its glycosylation and galactosylation via modulating the expression of modification genes (153, 155). Fucosylation of mucin and increased production of Goblet cells which is modulated by *Peptostreptococcus russellii* supresses inflammation through protecting from colitis induced by dextran sodium sulphate (DSS) (430). Additionally, colonised *Clostridia* promote differentiation of T_{reg} cells, hence can suppress inflammation, however, this is achieved through a yet unknown mechanism (432). Hence, epithelial homeostasis is maintained by the gut microbiota through supressing inflammation and mediating the ability of commensals and pathogens to adhere and colonise the colon.

The microbiota-driven increase in acetate production and consumption which occurs as a consequence of a HFD may modulate a microbiota-parasympathetic nervous system (PSNS) feedback loop, promoting hyperphagia (increased appetite) due to an increase in ghrelin (hormone controlling hunger and stimulating fat storage) secretion, and increased energy storage (433). The loci of SCFA-stimulus appear vital for the outcome of these SCFA-mediated feedback effects (434, 435), this argues a need for further investigation into the specific activity of SCFA-producing bacteria at the community level and how they influence energy usage and storage in obesity, a risk factor for CRC.

2.13 Host-Bacteria Quorum Sensing Signals

Trillions of microbes within the GI tract reside in a dynamic, highly dense and complex community (436). Such a high cell density environment allows bacteria to coordinate their behaviour, including forming fruiting bodies or spores, genetic competence (437), production of virulence factors, biofilm formation and secondary metabolites, facilitating colonisation and invasion of the host (438, 439). Like mammalian cells which use hormones and cytokines to communicate to each other, bacteria also regulate their cell-cell communication, termed quorum sensing (QS), through production of, auto-inducers the concentration of which depends on cell density, inducing group-wise chemotaxis. QS regulatory mechanisms are widespread in bacteria and may act as anti-immune mediators of host-bacteria interactions (440).

Production of auto-inducers by a bacterial population starts with synthesis at a low concentration by a small subpopulation. Those bacteria which sensed these signalling molecules also started producing their auto-inducers, gradually increasing its concentration within the community. Once a threshold is reached, inducers interact with the bacterial receptors, triggering a coordinated, community-wide alteration in gene expression. $Gram^-$ quorum sensing bacteria (e.g. *Pseudomonas aeruginosa*) utilise acylated homoserine lactones (AHL, long- and middle-chain fatty acids) as ligands for partner receptors, transmembrane cytoplasmic transcription factors and two-component His-sensor kinases

(441). Gram$^+$ (e.g. *Bacillus subtilis*) use modified oligo-peptides as signals to interact with the cognate receptor, transmembrane two-component His-sensor kinases (442). Direct interaction of auto-inducers with receptors often triggers expression of the auto-inducer synthase which in turn enhances the concentration of the signal as microbes are transformed by QS, promoting synchronisation of bacterial behaviours across the community (443).

QS auto-inducers often display immunosuppressive properties in order to evade host immune responses. *P. aeruginosa*-derived AHL (3O-C_{12}-HSL) appears to selectively deregulate NF-κB signalling through preventing activation of NF-κB-dependent inflammatory cytokines (e.g. IL-6) expression in macrophages (444). AHLs can promote chemotaxis in neutrophils, facilitating the localisation of inflammatory cells (e.g. neutrophils and Mφs) to infection sites (445). Gram$^+$ bacteria-produced QS signals have been shown to reduce intestinal inflammation by stimulating protein kinase B (Akt) and p38 MAPK survival pathways in host IECs (446). *Bacillus subtilis* produce competence and sporulation factor (CSF), a QS signal which induces anti-inflammatory IL-10 while repressing the secretion of IL-4, IL-6 and CXCL-1, pro-inflammatory agents in Caco-2 cells. Furthermore, CSF can induce HSPs that are critical for protecting IECs against oxidative stress, hence maintaining barrier function. Therefore, due to these anti-inflammatory properties, *B. subtilis*-produced CSF may represent a potential therapeutic target for controlling intestinal barrier function and inflammation. Gut-residing non-pathogenic *E. coli* can produce auto-inducer 2 (AI-2), a QS signal which upregulates the expression of IL-8 in human HCT-8 cells, is important for attracting neutrophils (447). This global-transcriptome study of bacterial-epithelial cell interactions revealed that secreted bacterial small QS signals from non-pathogenic microbes are critical for host-bacteria communication. This promotes activation of immune-related pathways that potentially interfere with the host immune response. For an effective immune response to occur TGF-β-dependent mucosal CD103$^+$ DCs-mediated generation of Foxp3$^+$ T_{reg} cells is required for adaptive immune responses to control any invading pathogens (448). Hence, the richness of the bacterial population and the production of bacterial-derived small chemicals (AHL and oligopeptides) as a result, is crucial for controlling the host immune system and may affect the integrity of IECs (448).

2.14 Roles of Bacterial Metabolites, Phenolic Compounds and Polyamines in GI homeostasis

Aside from the well-characterised bacterial metabolites, e.g. SCFAs, a number of bacteria-produced metabolites such as phenolic, benzoyl, phenyl, vitamins and polyamines hold important roles in maintaining gastrointestinal homeostasis. These small metabolites mediate host-microbiota symbiosis as well as interactions between bacterial species within the community.

Phenolic, benzoyl, and phenyl acids, secondary plant-derived metabolites, make up a subgroup of phenolic compounds, which contain one or more methoxy or hydroxyl groups along with aromatic rings, including hydroxycinnamic acids (C_6-C_3, caffeic and ferulic acids) and hydroxybenzoic acids (C_6-C_1, gallic acid) (449). Phenolics can act as non-enzymatic antioxidants (450), which delay or inhibit oxidation through the scavenging of free radicals (ROS and RNS), chelating metals (Fe^{2+} or Cu^{2+}) or inhibiting enzymes used in formation of free radicals, upregulating antioxidant defence mechanisms (449), hence reducing the risk of inflammation and DNA damage. Additionally, these aromatic compounds can modulate transcriptional factors such as PPARγ and NF-κB (451). These oxidative stresses and their corresponding responses are central among the processes driving degenerative chronic diseases such as cancer. During CRC levels of these aromatic phenolic, phenyl and benzoyl compounds and their derivatives e.g. phenylacetate, 4-cresol and 4-hydroxyphenylacetate have been shown to increase (452). Phenylacetate aerobic catabolism establishes an amalgam aerobic pathway wherein the primary stage is phenylacetate conversion to phenylacetyl-coenzyme A (phenylacetate-CoA) (453). Phenylacetate-CoA is later converted to the intermediate oxepin-CoA, a molecule further cleaved and ultimately, by way of a β-oxidation-like pathway, forms succinyl-CoA and acetyl-CoA, which are critical for SCFAs production and therefore colon homeostasis (454, 455). The *paaZ* operon encodes for enzymes responsible for phenylacetate catabolism are found in around 16% of bacterial genomes sequenced (456). The phenylacetate catabolising capacity of the gut bacteria appears to be reduced during CRC, this correlates strongly with the depletion of bacteria e.g. *Bifidobacterium, Clostridium difficile*, *F. prausnitzii*, *Lactobacillus* and *Subdoligranulum* capable of metabolising these aromatic compounds. Concentrations of these aromatic metabolites and their derivatives can act as a marker of dysbiosis, a consequence of, or caused by and driving CRC.

Polyamines (PAs) are polycationic molecules which are required for IEC growth, and in bacteria, typically include spermidine ($H_2N–(CH_2)_3–NH–(CH_2)_4–NH_2$), cadaverine ($H_2N–(CH_2)_5–NH_2$), putrescine ($H_2N–(CH_2)_4–NH_2$), and spermine ($H_2N–(CH_2)_3–NH–(CH_2)_4–NH–(CH_2)_3–NH_2$) (Fig. 14).

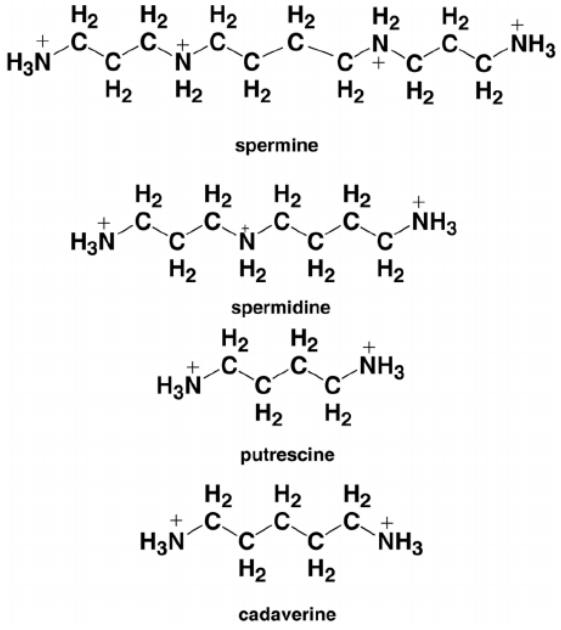


Fig. 14 | Polyamines are positively charged molecules. The chemical structure of the major native PAs. Adapted from (457).

These common small positively charged cellular metabolites are synthesised through amino acid decarboxylation and play a vital role in acid defence mechanisms, including decarboxylation of Lys and Arg amino acids. These microbial acid resistance mechanisms produce basic compounds and consume protons, hence increasing cytoplasmic pH (458). L-lysine is the precursor to cadaverine (via *cadA,* encoding for lysine decarboxylase), while L-arginine and L-ornithine are converted to putrescine either directly or through an intermediate, agmatine (459–461). Putrescine through e.g. *SpeE*, spermidine synthase is converted to spermidine, which subsequently either by spermine oxidase (SMO) or spermine synthase (SMS) becomes spermine. Orthologues encoding carboxyspermidine dehydrogenase (CASDH) and carboxyspermidine decarboxylase (CASDC) are present in a diverse range of biosynthetic putrescine genes throughout several bacterial clades. Bacteria that encoding PA biosynthetic pathway genes include *Alistipes putredinis, Bacteroides capillosus* (*Pseudoflavonifractor capillosus*), *Bacteroides eggerthii, Bacteroides uniformis, E. coli, Eubacterium siraeum, Roseburia intestinalis* M50/1 and *Ruminococcus torques* L2-14 (462). Bacteria however often possess genes of the operon *potABCD* encoding different transporters, which bind and uptake PAs from the surrounding environment. Two antiporters that exchange putrescine for ornithine and lysine for cadaverine (463), facilitate the trafficking of environmental PAs. These and other transport pathways are prevalent and highly conserved between both, $Gram^+$ and $Gram^-$ microorganisms such as *E. faecalis* and *E. coli*.

Positively charged PAs are involved in a multitude of cellular functions including the creation and maintenance of negatively charged DNA/RNA and proteins alongside membrane phospholipids (464), as well as in cell proliferation and differentiation while also regulating enzymatic activity, motility, apoptosis and resistance to oxidative and other stresses (465). In addition to these physiological functions PAs are also vital pathogen virulence determinants and are involved in QS and hence biofilm formation. At the level of RNA modification, several PA-mediated mechanisms exist modulating the translation of PA modulon factors via altering RNA structure and hence function (466). Firstly, PAs can stimulate translation when the mRNA Shine-Dalgarno (SD) sequence is obscured or is distant from the AUG translation start codon. Through causing structural changes of the 5’ untranslated region which includes the SD and AUG start site, the formation of an initiation complex can occur (for transcripts *Fis, FecI, OppA, H-NS* and *RpoN*). Secondly, in the *cya* (encoding for adenylate cyclase) and *cra* (encoding for a transcription factor that regulates expression of genes involved in glycolysis) mRNAs, PAs can enhance the initiation of translation from non-canonical UUG or GUG codons which are relatively inefficient initiators of translation. PAs can also facilitate the read-through of stop codons through promoting the interaction of UAG (amber termination codon) with Gln-$tRNA^{SupE}$ in the A-site as seen for *rpoS* mRNA. Alternatively PA modification of mRNA can

stimulate a frameshift (+1) at the UGA (opal or umber termination codon mutation) codon of *prfB* mRNA (467). These data argue for a unified PA-dependent molecular mechanism by which the PA modulon controls the expression of a large array of genes and conserves the optimal conditions to promote cell growth.

It is also established that PAs influence host cellular processes at the DNA level, such as transcription. PAs facilitate the *in vitro* oligomerisation of nucleosomal arrays, consisting of octamer-DNA-core histone complexes (spaced at 200 bp), the platform for transcription which forms the fundamental unit of chromosomal superstructures (468). PAs can control gene expression via modulation of DNA modification, hyperacetylation of histones inhibits chromosome condensation mediated by PAs (469). These results suggest that PAs are *in vivo* transcriptional suppressors, and that histone hyperacetylation plays a role in antagonising PAs capacity to stabilise condensed chromosomal structures. In fact, the TGF-β/Smad signalling pathway is induced through the depletion of PAs, which has been associated with a significant increase in transcription in IECs (470). These data demonstrate the integral effects PAs exert in cellular activity under physiological, elevated and depleted levels, illustrating the importance of their metabolism for the host and microbiome.

The microbiota that either synthesise PAs intracellularly through amino acid decarboxylation or transport PAs from the extracellular environment, can modulate their virulence through these metabolites. A primary example of bacterial utilisation of PA metabolism is genus *Shigella* enhancing their own survival by utilising the oxidative stress resistance offered by spermidine, due it's lack of the *speG* gene encoding spermidine acetyltransferase, the enzyme that converts spermidine into its inert form, acetylspermidine (471). This resistance to oxidative stress is crucial, as the exposure to H_2O_2 during macrophage responses could destroy the pathogen. Moreover, *Shigella* silence *cad* expression, which is crucial in its pathogenicity, this is due to the protective function of cadaverine for the colonic mucosa against enterotoxins ShET1 and ShET2. This negatively affects host immune response against *Shigella* by inhibiting polymorphonuclear neutrophil migration to the site of infection. Additionally, cadaverine blocks the cytoplasmic release of *Shigella* by stabilising the infected cells endosomal membrane (472). A similar gene deletion enhancing virulence has also been observed for enteropathogenic *E. coli.* Ultimately, the absence of the *cad* and *speG* genes critical in PA metabolism, demonstrates an effective strategy to improve virulence, via immune evasion to facilitate survival as well as invasion. This is one of many examples of pathogens utilising the PA metabolic pathways to their advantage, further highlighting the importance of metabolite levels either produced by host or microbiota, and how this affects symbiosis.

Multiple studies have shown that PA concentrations are elevated during CRC (473–475). The increase is triggered by the loss of polyamine homeostasis due to cell proliferation dysregulation (476). Data have shown evidence of upregulated PA biosynthesis, an elevated uptake (477) and decreased catabolism (478). PA biosynthesis and CRC are associated in a KRAS-dependent and p53-independent manner (479). As a result, PAs or PA-metabolising bacteria can be employed as biomarkers of neoplasia (475). Notably, data have shown that spermine levels in tissue may be employed as a reliable marker of prognosis and relapse in CRC patients (480). Dysbiosis caused by unbalanced dietary metabolite intake can further disrupt the microbial composition and most importantly, their metabolic activity, while the same is true for bacterial metabolism disrupting host homeostasis through mediating pro-inflammatory signalling and disrupting cell proliferation and apoptosis through their role in DNA and RNA modification.

3.0 Aims

Abundance of specific microbes is altered by the CRC gut therefore I aimed to i) define and characterise the metabolically active microbial species/groups in the CRC gut; ii) compare the profiles of microbial activity and abundance and identify potential microbial markers of the disease based on their regulated metabolic activity; ii) investigate the regulated CRC-specific patterns of microbial gene/pathway expression in the CRC gut and align this to the underlying mechanisms of disease pathology.

4.0 Objectives

- establish whether CRC is the primary/sole factor influencing changes in the microbial metatranscriptome by investigating any co-founder effects from patient metadata using Principal Component and Co-ordinate Analyses;

- identify differentially regulated microbial activities and taxonomic structure by metatranscriptome (mRNA-) and metagenome (rDNA-sequencing) to

    - propose the underlying mechanisms by which the microbiota influences human gut health in the context of CRC by aligning the regulated microbial metatranscriptome to specific disease features
    - establish the dormant and hyper-active sup-populations of the microbiota through aligning activity and abundance levels of its constituent species
- investigate regulation of gene expression (antibiotic- and acid-resistance genes) to deduce CRC-specific environmental pressures (low pH, high salinity and oxidative pressures)

5.0 Materials & Methods

5.1 Sample and patient data collection

Faecal samples from CRC patients and volunteers collected under the auspices of the Famished study at the University Hospital Coventry and Warwickshire NHS Trust (UHCW), UK Ethic certificate No: 09/H1211/38. UHCW NHS trust recruited 10 CRC patients and 10 non-CRC participants (Table 1) to ensure adequate statistical power of 20 samples for statistical analyses. Immediately upon collection samples were stored at -80°C to maintain nucleic acid integrity prior to processing. All patients are 18+ years old, excluding participants with previous cancer treatment (including chemotherapy) and those who had received any antibiotic treatment or surgery over the past 12 months. The health of the patients was monitored by Prof Arasaradnam's clinic over the duration of the study, patient metadata were also collected at UHCW including tumour site and stage, diet, height, weight, BMI, smoking status, medical conditions, and drug treatments as well as CEA and *KRAS*, *BRAF* and *MHL1* status (Table 1, not all metadata shown, only those used for PCA and PCoA analyses). Changes in diet were not recorded, only diet on the day of sample collection, this is likely to have remained stable due to the short period between diagnosis and sample collection.

5.2 Purification of microbiota from faecal samples

1 g of faecal sample was resuspended in 25 ml of ice cold sterile 1X Phosphate-buffered Saline (PBS) and centrifuged at 300 x g for 10' at 4°C to pull down debris. The supernatant was transferred to a new 50 ml falcon tube and centrifuged at 3,000 x g for 30' at 4°C to form a microbial pellet. The supernatant was discarded, and the pellet was resuspended again in 25 ml ice cold 1X PBS for washing, followed by a repetition of the previous centrifugation step. Finally, the washed pellet was resuspended in ice cold 1X PBS containing 20% glycerol in a volume of 6 ml, filtered through Miracloth (Calibochem®) to remove any traces of potential aggregates. Aliquots of equal volume of purified meta-microbiota from each sample of the cohort were pooled to make 2 ml of total CRC and total control meta-microbiota stocks were frozen in liquid nitrogen and stored at -80°C.

5.3 RNA isolation and in-column DNase I treatment

The RNeasy PowerMicrobiome kit (Qiagen, 26000-50) was used to purify total microbial RNA for further analysis from faecal samples. Briefly, 0.25-0.30 g of each faecal sample was added into a 55°C pre-heated PowerBead Tube (glass 0.1mm) containing 650 µl Solution PM1, 6.5 µl of β-mercaptoethanol and 100 µl phenol-chloroform (1:1, v/v) for deproteinisation. Samples were thoroughly vortexed horizontally for 10' to break the cell walls followed by centrifugation at 13,000 x g for 1' at 21°C (T_r, room temperature). The upper aqueous layer was mixed with 150 µl of Solution IRS and incubated on ice for 5'. The

supernatant after centrifugation at 13,000 x g for 1’ was mixed with 650 µl of Solution PM3 and 650 µl 70% ethanol to purify total RNA with length of >200 nts. The supernatant was loaded on the MB-RNA Spin Column via centrifugation at 13,000 x g for 1’, flow-through was discarded, this step was repeated until all the supernatant has been processed through the Spin Column (3 x 650 µl). 650 µl of Solution PM5 was added to the column and centrifuged at 13,000 x g for 1’. The flow-through was discarded and the column was centrifuged at 13,000 x g for 1’ to remove residual wash. 50 µl of DNase I (1 µg/µl) was added to the centre of the spin column and incubated at 21°C for 15’. 400 µl of Solution PM7 was added followed by centrifugation at 13,000 x g for 1’, flow-through discarded and 650 µl of Solution PM5 was added and centrifuged at 13,000 x g for 1’. This process was repeated once again with Solution PM4. The residual Solution PM4 was removed by centrifugation at 13,000 x g for 2’. The clean column-bound RNA was eluted with 100 µl DEPC-treated H_2O (incubated at 21°C for 5’ and RNA was eluted by centrifugation at 13,000 x g for 1’).

Study No.	Age	Sex	Smoker	Height & Weight	BMI	CEA range <6 ug/L	CRC Stage	PMH	Medication	Sample collection
CRC275	64	M	ex 20 years	180cm 85kg	26.2	3	T3 N1a M0	None	None	02/08/2019
CRC276	69	M	ex 30 years	166cm 108kg	39.2	4	T3 N2a M0	Type 2 Diabetes, 2 Hip Replacements, Sleep Apnoea, Hypertension	Empagliflozin, Losartan, Amlodipine, Co-Dydramol, Atorvastatin, omeprazole, Aspirin, Metformin,	04/08/2019
CRC277	67	M	Stopped 2007	182cm 85.5kg	25.6	4	T3 N1a M0	Emphysema, Aortic aneuryism, Hypothyroid, arthritis, anaemia	Levothyroxine, Atorvastatin, Omeprazole, Buspirone, Citalopram, Tiotropium, Zapain, Salamol, Duoresp	29/07/2019
CRC278	80	F	Never smoked	140cm 41.7kg	26.2	3	T3 N0 M0	Osteoarthritis, Breast CA 30 years ago, Anaemia	Adcal, Furosemide	11/08/2019
CRC279	52	M	Never smoked	173.5cm 62.2kg	20.7	2	T2 N0 M0	IBS	Mebeverine	18/08/2019
CRC280	72	M	Stopped 1994	172.5cm 90.4kg	30.4	4	T4a N2b M0	Hypertension, Asthma, Bilateral Knee replacements	Irbestartan, Ventolin	23/09/2019
CRC281	80	M	Stopped for 50 years	179cm 82.5kg	25.7	3	T4a N0 M0	Previous Squamous cell carcinoma skin, Myasthenia Gravis	Amitriplyline, Pednisolone, Lansoprazole, Thealoz eye drops	15/09/2019
CRC282	81	F	Never smoked	164cm 89.8kg	33.4	2	T3 N0 M0	T2DM, Stroke 30 years ago, Hypertension, Cholecystectomy, Hysterectomy	Ramipril, aspirin, Metformin, Co-amiloruse, ferrous fumarate, simvastatin	25/09/2019
CRC283	65	M	Stopped 1975	182.5cm 80.2kg	24.1	11	T4 N0 M0	T2DM, Anaemia	Atorvastatin, Dapagliflozin/Metformin, Gliclazide, Sitagliptin, Hydroxycobalamin	01/11/2019
CRC284	76	M	Stopped 15 years ago	176cm 97.8kg	31.3	3	T4 N1b M0	AF, COPD	Digoxin, Rivaroxiban, Lansoprazole, Ventolin inhaler	22/10/2019
V293	28	F	4 cigarettes a day	164cm 55.6kg	20.7	n/a	n/a	nil	nil	09/10/2019
V294	51	M	Never smoked	186cm 112.6kg	32.5	n/a	n/a	nil	nil	10/11/2019
V295	31	M	Never smoked	174cm 69kg	22.8	n/a	n/a	nil	nil	08/10/2019
V296	24	M	Never smoked	173cm 146kg	48.8	n/a	n/a	nil	nil	09/10/2019

V297	48	F	Never smoked	170cm 72kg	25	n/a	n/a	nil	nil	17/10/2019
V298	39	F	6 cigarettes a day	162cm 64kg	24.4	n/a	n/a	nil	nil	18/10/2019
V299	25	F	Never smoked	172cm 58.4kg	19.7	n/a	n/a	nil	nil	15/10/2019
V300	32	F	Never smoked	182cm 82.5kg	24.9	n/a	n/a	nil	nil	20/11/2019
V301	31	F	Never smoked	162cm 64kg	24.3	n/a	n/a	nil	nil	05/12/2019
V302	40	M	Never smoked	180cm 71kg	21.9	n/a	n/a	nil	nil	18/12/2019

Table 1 | Patient metadata. Tumour stages, T(1-4) denotes the size and progression of the tumour with a greater number following the T, representing more growth into surrounding tissues. T4a means the tumour has perforated and spread beyond the bowel wall. N represents nodal involvement, namely the number of surrounding lymph nodes (defined by the number and letter after the N) to which the cancer has spread. All faecal samples were collected in Falcon 50 ml tubes and frozen in dry ice followed by storage at -80˚C prior to microbial purification as described in the section 5.2. M denotes the state of metastasis, namely the spread of the tumour from the primary site to other tissues. M0, the cancer has not spread to other tissues. CRC, colorectal cancer patients; V, non-CRC volunteers; M, Male; F, Female; PMH, past medical history; CEA, carcinoembryonic antigen; BMI, body mass index.

5.4 DNase turbo treatment and ethanol precipitation

100 µl of total RNA was further treated with 5 µl (2 U/µl) of DNase Turbo (possessing greater affinity (K_m for DNA is a 6-fold lower) to lower DNA concentrations and is active in high salt comparing with DNase I due to a mutation in the DNA-binding pocket of the original DNase I) in 15 µl of 10X Turbo Buffer and 30 µl water for 30' at 37°C followed by chloroform (1:1) deproteinisation and ethanol precipitation with 1/10 volume of 3 M NaAc pH 4.5 (15 µl) and 3 volumes of ethanol (450 µl). The RNA samples were incubated at -20°C for at least 2 hours (or overnight) before collecting them by centrifugation at 21,000 x g for 20' at 21°C. The RNA pellets were washed with 70% ethanol twice, air-dried, dissolved in 50 µl of DEPC-treated H_2O and stored at -80°C for downstream applications.

5.5 Prokaryotic gDNA isolation

Using the DNeasy PowerSoil Pro Kit (Qiagen, 47014), 0.25-0.30 g of faecal sample (which were used for RNA purification) or a pellet from 200 µl of overnight aerobically gown bacteria in LB were added into a PowerBead Tube (glass 0.1 mm) containing 800 µl of Solution CD1 and vortexed horizontally for 10' before centrifugation at 15,000 x g for 1', T_r. The resultant supernatant was transferred to a new tube containing 200 µl of Solution CD2 at 4°C, vortexed and spun down for 1' at T_r. 700 µl of the sample was mixed with 600 µl of Solution CD3, vortexed and loaded onto an MB Spin Column and centrifuged for 1' (repeated until the entire sample, around 650 µl has passed through the column). 500 µl EA Solution was added to the column and centrifuged followed by the addition of 500 µl of Solution CM5, centrifuged and followed by a dry spin at 16,000 x g for 2'. DNA was eluted using 100 µl of sterile H_2O at 15,000 x g for 1' and stored at -20°C for downstream use.

5.6 Spectrophotometry of nucleic acids

Concentration and potential contamination of microbial gDNA extracted from faecal samples were measured using spectrophotometry. Nucleic acid concentrations are based on absorbance at A_{260} nm, for quantification a variation on the Beer-Lambert equation is used, $A = \varepsilon l c$ where c is nucleic acid concentration in ng/µl, A is the absorbance in AU (absorbance units), ε is the molar wavelength-dependent extinction coefficient in ng-cm/µl and l is the path-length in cm. One A_{260} unit is equal to 50 µg/ml of dsDNA, 33 µg/ml of ssDNA and 40 µg/ml of RNA.

5.7 Bioanalyzer and spectrophotometry of RNA

Concentration of the total RNA samples were also estimated via spectrophotometry using a NanoDrop UV-Vis Spectrophotometer in order to use an optimal amount of RNA for analysis of the concentration and quality by Bioanalyzer for RNA-seq. RNA was analysed through the Agilent Technologies 2100 Bioanalyzer systems Prokaryote total RNA Nano

assay via capillary gel-electrophoresis (using Agilent RNA-6000 Nano kit, 5067-1511) with a quantitative range of 25-500 ng. Briefly, the RNA ladder was prepared by denaturing at 70°C for 2’ and immediately placed on ice, aliquots were stored at -80°C. All reagents were equilibrated to room temperature for 30’ prior to use. The gel-matrix was prepared by adding 550 µl of Agilent RNA 6000 Nanogel Matrix (containing 100 mM Tris-HCl pH 7.5 @ 25°C and 0.1 mM EDTA) into the spin filter and centrifuged at 1,500 x g for 10’, aliquoted and stored at 4°C to be used over 4 weeks. The Gel-Dye Matrix was prepared by adding 1 µl of RNA-6000 Nano dye concentrate to 65 µl of filtered gel and spinning at 13,000 x g for 10’, used within one day of preparation. 5 µl of RNA 6000 Nano Marker was added in each well followed by 1 µl of RNA diluted in water to achieve the optimal amount of RNA within the range of 200-400 ng. The RNA ladder was added and the chip vortexed for 1’ prior to commencing the assay.

5.8 RNA and DNA high-throughput sequencing

Total RNA quality and concentration was analysed using the Agilent Technologies 2100 Bioanalyzer capillary gel electrophoresis system. RNA-seq was carried out by Vertis Biotechnologie AG, Germany, including depletion of rRNA, preparation of cDNA and Illumina NextSeq 500 sequencing (2 x 150 bp paired-end sequencing to produce 2 x 420 M reads). The cDNA inserts were flanked with the following adapter sequences, TruSeq_Sense_primer, i5 Barcode 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACN NNN NNN NAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T-3’ and TruSeq_Antisense_primer, i7 Index 5'-CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN NNG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT-3'. Blank extractions using 300 µl of H_2O were carried out to assess the quality of the DNA and RNA extraction kits and extractions did not yield any detectable nucleic acids. 16S rDNA V3-V4 regions were sequenced by Novogene Co., Ltd on Illumina (NovaSeq 6000 PE150) paired-end platform (100K tags of raw data per sample) to generate 250 bp paired-end raw reads (Raw PE), merged and pre-treated to obtain clean tags. Clean tags were removed to obtain the effective tags. Operational Taxonomic Units (OTUs) were obtained by clustering with >97% identity on the Effective Tags of all samples, taxonomic annotation was made for the representative sequence of each OTU to obtain the corresponding taxa information and taxa-based abundance distribution using MetaPhlAn (481) v3.0.11.

5.9 16S rRNA gene data processing and analysis

Paired-end reads were assigned to samples based on their unique barcodes and truncated by removing the barcode and primer sequences. Read pairs were merged using FLASH (482) v1.2.7. Quality filtering on the raw tags were performed according to QIIME (483) v1.7.0. Sequences were aligned to the SILVA database using UCHIME (484) to detect and remove chimeric sequences (485). Non-chimeric sequences with ≥97% similarity were

clustered into operational taxonomic units (OTUs) using UPARSE (486) v7.0.1090. Representative sequences for each OTU were mapped against the SILVA (487, 488) SSUrRNA database for taxonomic assignment, with a threshold of 80% using QIIME (483) v1.7.0. OTU abundances were normalized using a standard of sequence number corresponding to the sample with the fewest sequences.

5.10 Taxonomic metatranscriptome data processing and analysis

Raw reads were processed following the steps of the SAMSA2 (489) v2.2.0 pipeline. First, read pairs were trimmed to remove low quality bases using Trimmomatic (118) v0.36, then overlapping read pairs were merged into single sequences using PEAR (490) v0.9.11. Sequences were 'ribodepleted' *in silico* using SortMeRNA (491) v2.1 to remove those representing ribosomal RNA. The remaining sequences were translated and assigned to a database of 68,433,538 protein sequences from RefSeq (492) with protein names and taxonomic information, using DIAMOND (493, 494) v0.8.38.

5.11 Functional metatranscriptome data processing and analysis

Raw reads were processed following the steps of the SAMSA2 (489) v2.2.0 pipeline. First, read pairs were trimmed to remove low quality bases using Trimmomatic (118) v0.36 followed by overlapping read pairs were merged into single sequences using PEAR (490) v0.9.11. These sequences were 'ribodepleted' *in silico* using SortMeRNA (491) v2.1 to identify and remove sequences representing ribosomal RNA. These ribodepleted sequences were translated and assigned to functional classes of the SEED Subsystems hierarchical database (495) using DIAMOND (493, 494) v0.8.38 to align reads against a database of 7,939,855 protein sequences. Sequences assigned to each functional class were aggregated to give raw abundance count data for each class. These counts were used to determine statistically significant differential abundance of transcripts (496) between conditions using DESeq2 (121) v1.26.0 with *P*-values adjusted via Benjamini-Hochberg False Discovery Rate (497) (FDR<0.1).

5.12 Microbial diversity and statistical analyses

To compare the transcriptome-based taxonomic data to those of 16S rRNA-marker data, 16S data were re-processed using QIIME2 (498), denoising and merging reads using DADA2 (499) (assigning taxonomy using a Naïve Bayes classifier trained on V3-V4 SILVA (500) and quantifying reads assigned to taxa at taxonomic level 7, namely species. Per-taxon read counts for both transcriptome and 16S data were normalised using scaling with ranked subsampling using the SRS (501) R package. Normalised counts were used for ordination by Principal Coordinate Analysis (PCoA), to calculate α- and β-diversities and to perform ANOSIM and PERMANOVA analyses, all using the VEGAN (502) v2.6-2 R package. For α-

diversity the Shannon (503) and the Gini-Simpson (504) diversity (1-Simpson (inverse) diversity) indices were calculated for each sample. Shannon diversity index is a quantitative index accounting for species evenness and abundance, it measures both the number of species and the inequality between species abundance within a certain sample group or samples bound by a commonality, defined as, $H' = -\sum_{i=1}^{S}(p_i ln(p_i))$. Inverse Simpson index measures species diversity independent of richness using, $1 - D = \sum_{i=1}^{S} p_i^2$, where p_i is the proportion of total species contained within species i. For β-diversity, the Bray-Curtis distance (505) was calculated. Bray-Curtis dissimilarity/distance is defined as $BC_{ij} = \frac{S_i+S_j-2C_{ij}}{S_i+S_j}$, where S_i and S_j are the amount of species in both populations i and j, and C_{ij} is the fewest number of common species from a specific group. Bray-Curtis distance was used in PERMANOVA and ANOSIM analyses. PERMANOVA was applied to assess the influence of clinical and lifestyle factors of the study participants on taxonomic distributions. ANOSIM was applied to identify taxonomic dissimilarly between healthy control and CRC samples. The LEfSe (linear discriminant analysis (LDA) Effect Size) package was used to establish the significant differences of microbial community composition between CRC and control groups for 16S taxonomy. The significance of observed differences in microbial 16S abundance among groups was evaluated by multiple hypothesis-test for sparsely-sampled features and false discovery rate (FDR) through MetaStat. *P*-values were obtained by permutation test and q-value was calculated by Benjamini-Hochberg False Discovery Rate (497). Differences in microbial activity were measured through a generalised linear model, Wald T tests and q-values calculated with Benjamini-Hochberg False Discovery Rate (FDR<0.1) for differential taxonomic activity through DESeq2 (121) v1.26.0.

Organism name	5'- Forward primer -3'	5'- Reverse primer -3'	Gene name	Annealing Temperature
Faecalibacterium prausnitzii	CACCGGCAAACTGCACAC	GGCCGTCAAGACCGACAA	*argS*	56°C
Enterococcus faecalis	TGCTTTGTTCGTTGCCGAC	CGGGATTTAGCTGCTGCG	*argS*	56°C
Prevotella copri	GGGCTGACCTCACCAACG	GAAACGGAGGTCGGCAGT	*argS*	56°C
Fusobacterium nucleatum	TGCTTGGTTGGGACATTGA	AGCACGGCTATGAGCTTC	*argS*	56°C
Bacteroides fragilis	ACCAACATCTCACGCGCT	GCAAGGGTGAAACTCCCGA	*argS*	56°C
Escherichia coli	GAGCAGGTGCTGACTCAT	TTCAAGGTCAGCCAGCTC	*argS*	56°C
Bacterial universal 16S	CGGTGAATACGTTCYCGG	TACGGCTACCTTGTTACGACTT	16S	56°C

Table 2 | Primers to validate species activity and abundance. qPCR primers for arginyl-tRNA synthetase (ligase) (*argS*) genes.

5.13 PCR

PCR reactions contained 2X *Taq* PCR MasterMix (Qiagen, 201443) (*Taq* polymerase with a concentration of 5 U/µl, 5' –> 3' exonuclease activity, polymerisation rate of at least 2 kb/min at 72°C (as per the manufacturers specifications), containing substrate analogues dNTP, ddNTP, dUTP, biotin-11-dUTP, DIG-11-dUTP, fluorescent-dNTP/ddNTP for enzyme stabilisation during storage, 50 mM KCl/$(NH_4)_2SO_4$ and 2 mM $MgSO_4$). Primers at a final concentration of 500 nM and DNA (or RNA to test gDNA contamination) at 10 ng/µl were used. 16S rRNA gene was amplified with either universal forward 27F-primer, 5'-AGAGTTTGATCCTGGCTCAG-3' and reverse 534R-primer, 5'-ATTACCGCGGCTGCTGG-3' to produce a fragment of 507 nucleotide long of the V1-V3 region to test the quality of purified gDNA and/or for Sanger sequencing of the fragments using forward primer BACT1369-F, 5'-CGGTGAATACGTTCYCGG-3' and reverse primer BACT1492-R, 5'-TACGGCTACCTTGTTACGACTT-3' for quantitative analyses by qRT-PCR and qPCR (Table 2) (506). Hot start for polymerase activation was carried out at 95°C for 3' followed by 15 cycles of 1' at 95°C, 45'' at 56°C and 1' at 72°C followed by analysis on 1% TBE agarose gels. Independent triplicate samples were analysed to establish statistical significance. PCR of total RNA samples using the qPCR set of primers for 16S rDNA sequences as the template tested DNA contamination of the total RNA samples, performed as describe in (108).

5.14 cDNA synthesis

1.5 µl of mix[1] (a mixture of 2 µl 100 µM d(N_6) random primer (5'-d(N_6)-3' [N=A, C, G, T]), 1 µl 25 mM dNTPs and 3 µl H_2O) was added to 5 µl of total RNA (0.5-1 µg/µl) and incubated at 65°C for 5' to melt the local RNA secondary structure followed by incubation on ice for 2'. 3 µl of mix[2] (a mixture of 2 µl 5X RT buffer and 1 µl of 100 mM DTT) was added and incubated at 25°C for 2' for efficient primer annealing. RTase Superscript-II (Invitrogen, 18064-022) at a volume of 1 µl (200 U) was added and samples were incubated at 42°C for 90' followed by 5' at 75°C to terminate the reaction. A 50-fold dilution of cDNA was used for quantitative PCR. The SYBR® Green iTaq (Bio-Rad) qRT-PCR system was tested with the 16S rDNA primers for contamination (water used as the template) and DNA contamination of the RNA samples (proportionally to the amount of cDNA used for amplification, diluted RNA samples were added as templates for PCR). No amplification was observed for all the control samples.

5.15 Bacterial taxonomic profiling q(reverse transcription-based) PCR

To confirm taxonomic profiling conducted through metatranscriptome sequencing analysis, species-specific qRT-PCR was conducted for *E. coli* and *F. prausnitzii*. 1 µl of 20 ng/µl gDNA/cDNA, 5 µl of 2X iTaq Universal SYBR Green Supermix (antibody-mediated hot-start iTaq DNA polymerase, SYBR® Green I dye, $MgCl_2$, passive ROX and fluorescein dyes and dNTPs) (BioRad, 1725124), 1 µl of each gene specific and housekeeping bacterial 16S

rRNA gene forward BACT1369-F and reverse BACT1492-R primers (506) at a final concentration of 500 nM (Table 2, Table 3 and Table 4) and 2 µl H_2O were used. Hot start for polymerase activation was carried out at 95°C for 2’ followed by 35 cycles of 15’’ at 95°C, 30’’ at 56°C and 45’’ at 72°C. Melting curve was conducted from 65°C to 95°C with an increment of 0.5°C being held for 5’’. The housekeeping gene PCR was used for normalisation with a minimum of five technical repeats, and at least three technical repeats being conducted in parallel for each sample taxonomic analysis. Gene expression analysis was conducted using non-parametric two-sample Mann-Whitney U (Wilcoxon rank-sum) tests on normalised (to 16S rRNA gene levels) C_t values to establish significance of changes to gene expression.

Gene	5'- PCR Primers -3'	size (nts)	Resistance with perfect matches, species of origin	AB resistance	Resistance with sequence variants
cls	F_GATCACCGGAAAATTGTTG R_AAGAGACGTTCCAATCCAT	181	*Enterococcus faecalis*	Daptomycin	N/A
bla2	F_ GAAGCAGTTCCTTCGAAC R_ ATCAGCGTGTGCATGTGT	162	*Bacillus cereus,* *Bacillus thuringiensis,* *Bacillus toyonensis*	B1 metallo-β-lactamase, penicillin, cephalosporin, carbapenem	*Bacillus anthracis,* *Vibrio cholerae*
mdtO	F_ ATGCTCGACTATCCGGAA R_ GCATTTGCGAAATGGCAC	127	*Escherichia coli,* *Shigella spp.*	Puromycin, acriflavine, nucleoside ABs	*ESKAPE spp.,* *Citrobacter spp.,* *Salmonella spp.,* *Serratia marcescens*
phoP	F_CTGTCGGTGAATGACCAG R_CGTCGATGGTGTGGCTTT	154	*Klebsiella pneumoniae,* *Escherichia coli*	Colistin, Macrolides, Peptide ABs	N/A
eptB	F_ CCTTCTTCCTGTTACGTC R_ GATATCGGTGGTCATCAC	161	*Klebsiella pneumoniae,* *Escherichia coli*	Peptide ABs	*Acinetobacter baumannii,* *Citrobacter spp.,* *Enterobacter spp.,* *Enterococcus faecium,* *Pseudomonas aeruginosa,* *Mycobacterium tuberculosis,* *Serratia marcescens,* *Shigella spp.*
catA	F_ CAGACCGTTCAGCTGGAT R_ TATCACCAGCTCACCGTC	150	*Acinetobacter baumannii,* *Citrobacter spp.,* *Proteus mirabilis,* *Serratia marcescens,* *Shigella flexneri,* *Klebsiella pneumoniae*	Chloramphenicol	*Alcaligenes faecalis,* *Chlamydia trachomatisa,* *Citrobacter spp.,* *Enterobacter spp.,* *Helicobacter pylori,* *Neisseria gonorrhoeae,* *Pseudomonas aeruginosa,* *Proteus spp.,* *Serratia marcescens,* *Shigella spp.*
bla_{CMY}	F_ GCTGCTGACAGCCTCTTT R_ TGCGTGACTGGGTGGTTA	198	*Citrobacter freundii,* *Providencia rettgeri,* *Klebsiella pneumoniae,* *Escherichia coli*	CMY-type β-lactamase, cephamycins	N/A
nfsA	F_ TCCATTCGCCATTTCACT R_ TAATGCTACTGCACTGCA	109	*Escherichia coli,* *Shigella spp.*	Nitrofurantoin	*Citrobacte. Freundii,* *Enterobacter cloacae*
marA	F_ GGACTGGATCGAGGACAA R_ CTGCGGATGTATTGGCCT	135	*Escherichia coli,* *Shigella flexneri*	Multi-drug efflux pump	*Klebsiella pneumoniae,* *Shigella spp.*
gadX	F_ TGTCAAGGGACACGCTTT R_ GATAGTTGCGCAACTTCC	142	*Escherichia coli,* *Salmonella spp.,* *Shigella spp.*	Penam, fluoroquinolone macrolide ABs	N/A

Table 3 | Antibiotic resistance gene primers used for qPCR. All primers are sourced from the comprehensive antibiotic resistance database (CARD) (507). *cls*, cardiolipin synthase (required for membrane synthesis; specific mutations of *cls* in *Enterococcus* confer resistance to daptomycin); *bla2*, Zn dependent β-lactamase type-II with a broad spectrum of activity; *mdtO*, multidrug (major facilitator superfamily) efflux transporter permease subunit; *phoP*, virulence transcriptional repressor of the *macAB* efflux genes; *eptB*, kdo(2)-lipid A phosphoethanolamine 7''-transferase; *catA*, type A-1 chloramphenicol O-acetyltransferase; bla_{CMY}, class C β-lactamase; *nfsA*, oxygen-insensitive nitroreductase; *marA*, multidrug resistant efflux pump AcrAB transcriptional activator; *gadX*, acid

resistance transcriptional activator which enhances expression of *mdtEF*, RND efflux pump; ESKAPE, six highly pathogenic and antibiotic resistant bacterial species, including *E. faecium, S. aureus, K. pneumoniae, A. baumannii, P. aeruginosa* and *Enterobacter spp*. AB, antibiotic.

Gene	5'- Primers -3'	Product size (bp)	Function
speA	F_GGTGTACTACGCTCCATG R_TAATGTGGCCCAGCTCGT	124	Biosynthetic Arg-decarboxylase involved in putrescine synthesis; pH independent
adiA	F_CTCCATCAAGACACCTGG R_AGGCAGTCAATGGCTTCG	140	Degradative Arg-decarboxylase, inducible by low pH in rich media anaerobically
cadA	F_CCATCCGTGAACTTCATC R_ATTTCTTCGCACAGCTCG	157	Inducible Lys-decarboxylase, producing cadaverine, a superoxide radical scavenger

Table 4 | Primers for Arg- and Lys-decarboxylase genes for qPCR. Expression of Arg- and Lys-decarboxylase genes which are part of *E. coli* amino acid-dependent acid resistance mechanisms were tested. Total microbiota purified from CRC and control meta-samples were grown aerobically for 24 hours at 37°C in LB. The forward and reverse gene specific primers for PCR with annealing temperature of 56°C and the size of the amplicons are shown. PCR fragments were cloned into the TA-pGEM vector (Promega) and 10 randomly selected clones for each gene after transformation were Sanger sequenced and BLAST against the nucleotide collection database for confirmation of target amplification.

5.16 pGEM®-T Easy vector ligation with Taq-derived PCR products

1 µl (50 ng) pGEM®-T Easy Vector in 5 µl 2x Rapid Ligation Buffer (1x 30 mM Tris-HCl pH 7.8 @ 25°C, 10 mM $MgCl_2$, 1m M ATP, 10 mM DTT and 5% PEG-4000) (Promega, A1380), with 1 µl 2 U/µl T4 DNA ligase and 3 µl of purified PCR fragment were incubated overnight at 21°C.

5.17 Transformation of *E. coli* with pGEM®-T easy ligation of PCR fragments

Chemically competent *E. coli* (*hsdR*, *lac*ZΔM15, *mcrA*, *recA* and *endA*) DH10B (>1x10^9 cfu/µg) in a volume of 50 µl were thawed on ice for 20' and 5 µl of ligation reaction was added and incubated on ice for 30'. Samples were subjected to heat-shock at 42°C for 20'' to induce uptake of DNA fragments and immediately placed on ice for 2' to allow cells to recover. Following this, 250 µl of S.O.C medium (Thermo Fisher, 15544034) was added to each sample and incubated for 1 hr at 37 °C with agitation at 100 x rpm to activate expression of β-lactamase, the ampicillin resistance gene acting as a selectable marker.

5.18 Blue-white screening for recombinant pGEM®-T easy plasmids

LB (Luria-Bertani)-ampicillin agar plates containing IPTG (1 mM) and X-Gal (20 mg/ml) were plated, in duplicate, with 150 µl of transformation reaction and incubated overnight at 37°C. White or pale blue colonies were selected (re-plated onto fresh LB-amp$^+$ IPTG/X-Gal$^+$ plates) for further analysis to confirm the presence (and the size of the insertion) of the recombinant plasmid (508).

5.19 pGEM®-T easy vector EcoRI digest

To assess the length of insertion of PCR fragments that would be suitable for Sanger sequencing 2 µg pGEM®-T Easy recombinant plasmids extracted (QIAprep Spin Miniprep Kit, Qiagen following the manufacturers protocol) from either white or pale blue colonies post-transformation in 13 µl were mixed with 1 µl EcoRI restriction endonuclease (Thermo Fisher, ER0271), 2 µl 10X Buffer EcoRI and 2 µl water in a final volume of 20 µl and incubated for 60' at 37°C. Restriction digest was analysed by 1% TBE-agarose gel-electrophoresis and plasmids that carry PCR fragments of expected sizes were selected for Sanger sequencing by Eurofins with M13uni-21 primer 5'-TGTAAAACGACGGCCAGT-3'.

5.20 Expression of antibiotic resistance determinants and amino acid resistance genes under different growth conditions *in vitro*

Meta-microbial cultures were exposed to two acid stress conditions, (HCl and DL-lactate pH adjusted media), alongside osmotic stress (NaCl treated media) and oxidative stress (H_2O_2 treated media), before assessment of differential gene expression. 50 µl of purified meta-microbiota from both cohorts, CRC (n=3) and control (n=3) were cultured in 50 ml LB (pH 6.8) at 37°C until stationary phase (pH 8.4-8.6) for 24 hrs at 200 rpm. Optical density

of the 1,000-fold diluted overnight cultures were measured at A_{600} and appropriate adjustments were made to make microbial concentrations equal in all cultures. 100 µl of control and CRC overnight cultures were transferred into 50 ml of fresh LB pH 5.8 (acid adaptation, n=12) and fresh LB pH 6.8 ("untreated", n=4) for 2 hours as standing cultures at 37°C. After two hours of acid adaptation the pH of CRC and control cultures were adjusted to pH 3.5 with either HCl (n=6, CRC culture: n=3 and control culture: n=3) or DL-lactate (n=6, CRC culture: n=3 and control culture: n=3). The untreated CRC (n=2), control (n=2) pH 6.8 cultures and HCl- pH 3.5 and DL-lactate-adjusted pH 3.5 cultures were further incubated at 37°C as standing cultures. After 2 hours of incubation microbes were collected by centrifugation at 3,000 x g for 30' at 4°C. The pellets were washed with 5 ml cold 1X PBS followed by centrifugation using the same conditions as above and pellets were immediately frozen in liquid nitrogen and stored at -80°C for RNA and DNA extractions. The above overnight cultures were used for NaCl osmotic stress (n=6) and 1.5 mM H_2O_2 oxidative stress (n=6) and untreated CRC (n=2) and control (n=2) cultures. 100 µl of the overnight cultures were transferred into 50 ml of fresh LB pH 6.8 adjusted to 5% NaCl: CRC cultures n=3 and control cultures n=3 and oxidative stress: CRC culture n=3 and control culture n=3 at 37°C as standing cultures. After 2 hours the cultures were treated as above for RNA and DNA purification. Microbial gDNA from overnight cultures (CRC and control, 200 µl each) were purified using the DNeasy PowerSoil Pro Kit as described above. Purified gDNA was used for 16S rDNA sequencing by Novogene to establish the microbial composition of the CRC and non-cancerous aerobic cultures. qRT-PCR was used to quantify the level of expression of specific AB genes (Table 1), using 16S rDNA primers for normalisation as described above.

5.21 Targeted antibiotic resistance gene expression analysis

A more focused analysis of antimicrobial resistance genes was undertaken using the Resistance Gene Identifier (RGI) software (509). The 'rgi bwt' pipeline was run to map the meta-transcriptome data to the curated Comprehensive Antimicrobial Resistance Database (CARD) (507) v3.1.4 using bwa (510). It is noteworthy that the current CARD is a poor predicter of the AB resistance for *P. aeruginosa* and may give a high rate of false-positive predictions in a Global Isolate Dataset (511). Read counts per gene were extracted and used to compare conditions ('healthy' against 'CRC') using DESeq2 (121) to identify differentially abundant/expressed AMR-associated genes.

To validate antibiotic resistance gene primers DNA was isolated from the untreated control and CRC cultures as performed previously (see Genomic (g)DNA Isolation) (Table 3 and Table 4). PCR products were ligated into cloning vectors (see pGEM®-T Easy Vector Ligation with PCR Products), digested with EcoRI (see pGEM®-T Easy Vector EcoRI Digest) and gel electrophoresis was conducted to validate insertion and ligated plasmids were used for

sequencing by Eurofins using M13uni-21 5'-TGTAAAACGACGGCCAGT-3' primers to validate antibiotic resistance gene primer specificity. RNA was isolated from all cultures in parallel as previously carried out (see RNA Isolation) for conversion to cDNA, as previously conducted (see cDNA Synthesis (Invitrogen)) for qPCR. For each biological sample (n=3) 3 technical repeats of qPCR were conducted as previous (see Bacterial Taxonomic Profiling q(reverse transcriptase) PCR). Gene expression analysis was conducted using 16S rRNA gene expression to normalise data, non-parametric two-sample Mann-Whitney U (Wilcoxon rank-sum) tests were conducted on normalised C_t values to establish significance of changes to gene expression.

6.0 Results

6.1 Sample and metadata collection, DNA and RNA extraction and quality control

University Hospital Coventry and Warwickshire (UHCW) recruited 10 CRC patients and 10 healthy participants (Table 1), subject to the selection criteria previously outlined. Faecal samples (10-50 g per sample) were collected, and total bacteria was purified from all participants. Total DNA and RNA (25-30 µg) were purified. The quality and concentration of the extracted DNA and RNA were tested by spectrophotometry at 260-280 nm (Table 5), bioanalyser and PCR/RT-PCR for suitability of the samples for sequencing. Testing for any potential DNA contamination of RNA was also carried out. RNA samples were treated with DNase Turbo post-extraction and have shown no DNA contamination, to confirm this, 12 ng (400-fold diluted) of total RNA was used as template for 16S rDNA gene PCR and revealed no products in all samples (data not shown). RT-PCR was used to test RNA sample suitability for sequencing, clear 500 nts 16S rRNA bands were observed for all samples. The bioanalyser assays did not identify the presence of eukaryotic (human and fungal) rRNAs, demonstrating that the total RNA samples contain only prokaryotic and archaeal rRNA and do not require depletion for eukaryotic rRNA before cDNA library preparation (Fig. 15). The absence of eukaryotic RNAs is evident through the lack of ribosomal rRNA, 1869 nts and ~5070 nts, while bands at 1.5 kb and 2.9 kb indicate prokaryotic RNA. These data confirm that all RNA samples are suitable for RNA-sequencing, despite some CRC samples showing visibly higher levels of degradation, confirmed by VERTIS Biotechnologie AG using a Shimadzu MultiNA microchip (Fig. 16).

Sample	Concentration (ng/µl)		Volume (µl)		Amount (µg)	
CRC275	290	822	39	50	1.1	41.1
CRC276	396.4	744	19	50	7.5	37.2
CRC277	129.6	353	29	100	3.8	35.3
CRC278	492.6	571	19	100	9.4	57.1
CRC279	865	468	19	100	16.4	46.8
CRC280	600.3	471	19	100	11.4	47.1
CRC281	921.5	407	19	100	17.5	40.7
CRC282	721	364	19	100	13.7	36.4
CRC283	617.7	346	19	100	11.7	34.6
CRC284	668.2	291	19	100	12.7	29.1
V293	710.9	417	19	50	13.5	41.7
V294	153.4	456	19	50	2.9	45.6
V295	1014.5	1318	19	50	19.3	131.8
V296	694	953	19	50	13.2	95.3
V297	315.2	838	19	50	6	83.8
V298	929.3	592	19	50	17.7	59.2
V299	550.6	920	19	50	10.5	92
V300	645.4	433	34	100	21.9	43.3
V301	656.7	461	39	100	25.6	46.1
V302	717	355	29	100	20.8	35.5
	DNA	RNA	DNA	RNA	DNA	RNA

Table 5 | In-house analysis of bacterial DNA and RNA concentrations, volumes, and total amounts. 5 µg of RNA per samples was used for metatranscriptome sequencing, the same amount for 16S rRNA gene sequencing.

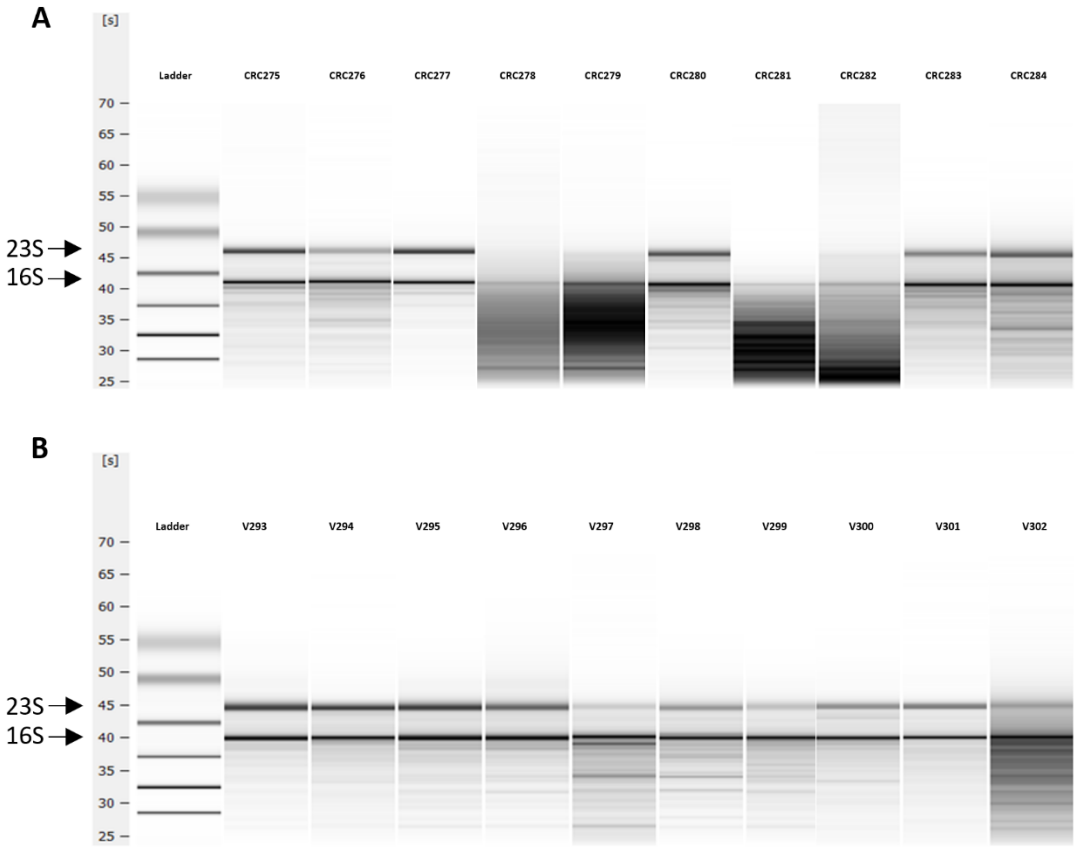


Fig. 15 | Bioanalyser analysis of RNA isolated from CRC and non-CRC samples. RNA samples analysed by bioanalyser (Agilent Technologies) using capillary gel-electrophoresis. **A.** Prokaryotic RNA from 10 CRC pre-surgery faecal samples. Samples CRC278, CRC279, CRC281 and CRC282 appear to be more degraded than other CRC RNA samples, however of sufficient integrity for sequencing and downstream analysis. **B.** Prokaryotic RNA from 10 non-CRC faecal samples. 23S, 23S rRNA;16S, 16S rRNA, [s] (y axis) denotes time in seconds.

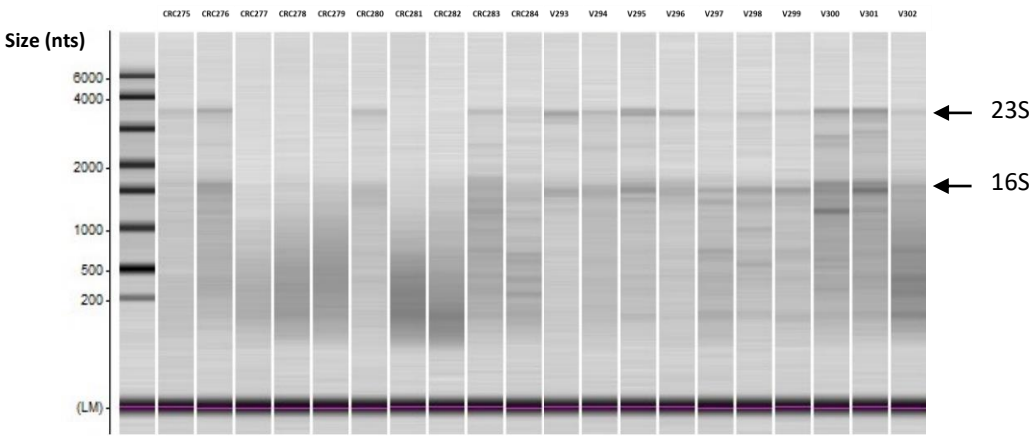


Fig. 16 | Capillary gel-electrophoresis of sample total RNA analysed by Bioanalyser: Prokaryotic rRNA subunits, 16S rRNA (16S) and 23S rRNA (23S) indicated with arrows for all samples provided by VERTIS by Shimadzu MultiNA microchip. Results confirmed in-house RNA integrity analysis (Fig. 15).

6.2 Sequencing and community diversity metrics

Metatranscriptomic sequencing of the 20 samples produced 909,748,013 read pairs, which were processed to output 693,090,228 sequences (18,510,095-38,871,022 per sample, median 21,946,018), a 76% rate of retention consistent with the expected Illumina Q30 score (a metric indicating accuracy of base calling) of ≥75% (512). 16S rDNA sequencing of the 20 samples produced 3,603,795 read pairs, which were processed to produce 3,303,648 sequences. (All metatranscriptome and 16S rRNA gene sequencing data were submitted to the European Nucleotide Archive (ENA) under the project accession PRJEB53891 and are accessible from https://www.ebi.ac.uk/ena/browser/view/PRJEB53891). Alpha-diversity, as assessed by Gini-Simpson indices, measured for transcriptomic and for 16S rDNA-seq data correlated (Spearman's ρ=0.67, *P*=0.0017; Fig. 17A) indicating a similar signal in the data despite the different nature. PERMANOVA analyses indicated the only significant grouping factor was healthy *vs*. CRC samples (and not age, sex, smoking status, or BMI) (Table S1). This was reflected in the ordination (PCoA) plots, that indicated some separation between healthy and CRC samples (Fig. 17B and Fig. 17C). Alpha-diversity was greater in healthy samples than in CRC samples for both transcriptomic (healthy=0.0226; CRC=0.0170) and 16S (healthy=0.0336; CRC=0.0234) data (Fig. 17D and Fig. 17E). ANOSIM analyses indicated that CRC samples were significantly more divergent from one another than healthy samples for both metatranscriptomic (*P*=0.043) and 16S rDNA (*P*=0.002) data (Fig. 17F and Fig. 17G).

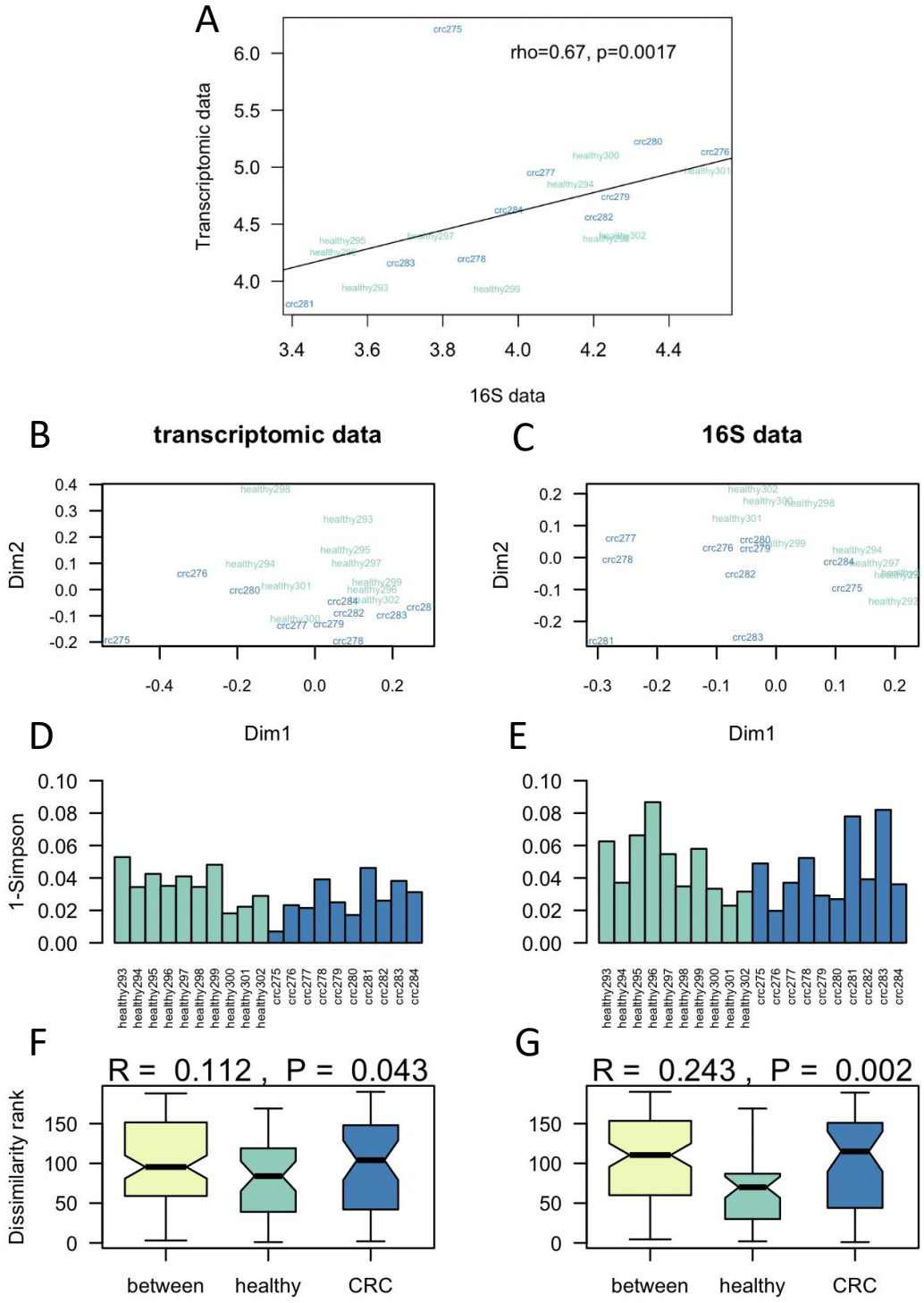


Fig. 17 | Community diversity of gut microbiota based on meta-transcriptomic sequencing and on amplified 16S markers. A. Spearman's correlation between 16S rRNA gene and metatranscriptome datasets. **B.** PCoA ordination plot of non-CRC- (healthy) and CRC-associated active taxonomy. **C.** PCoA ordination plot of non-CRC- and CRC-associated microbial 16S abundance-based taxonomy. **D.** The Gini-Simpson α-diversity index (1-Simpson) of each sample's active taxonomy. **E.** The Gini-Simpson α-diversity index (1-Simpson) of each sample's 16S abundance-based taxonomy. **F.** ANOSIM results, based on Bray-Curtis dissimilarity of samples, showing the distribution of ranks of pairwise dissimilarities *between* non-CRC and CRC-associated active taxonomy and *among* non-CRC

and CRC-associated active taxonomy. **G.** ANOSIM results, showing the distribution of ranks of pairwise dissimilarities *between* non-CRC and CRC-associated 16S taxonomy and *among* non-CRC and CRC-associated 16S taxonomy. The plots shows that CRC-associated samples are significantly more dissimilar from one another than the healthy microbiota in both metatranscriptome (*P*=0.043) and 16S (*P*=0.002) datasets. Principal Coordinate Analysis; PCoA. Analysis of similarity; ANOSIM.

6.3 Profiles of the most abundant and most active species in both health and colorectal cancer groups are distinct from one another

It is well established that bacteria can exist in different metabolic states to adapt to adverse environmental conditions, including dormant and persistence phenotypes (518). To characterise different subpopulations of specific bacteria in the gut in both groups we compared the 16S rDNA-based taxonomic abundance with the metatranscriptome-based relative activity of bacterial species. We identified 240 species whose activities (based on total transcript abundance) were significantly different in CRC (Fig. 18, Dataset 1, Tab 2 of (513)) and 15 species the presence of which was altered in the disease (Dataset 1, Tab 3 of (513)). Based on 16S rDNA analysis we identified the 20 most abundant bacteria in the healthy cohort (Fig. 19A) represented 38.70% of the total microbial population. Of those, the abundance of *Faecalibacterium* sp., *Faecalibacterium prausnitzii*, *Ruminococcus bicirculans* and *Lactobacillus amylovorus* was significantly reduced in CRC. The combined relative activity of these species in the healthy group accounted for 26.39% of the total expressed microbiome and reveals several interesting observations (Fig. 19C). Five species, *F. prausnitzii*, a major butyrate-producing bacterium (514), *Bacteroides vulgatus* and *B. stercoris* (members of the pro-inflammatory *Bacteroides fragilis* group (515)), *B. uniformis* and *Parabacteroides merdae* (commensal saccharolytic species), also resided within the profile of the most active species in the control cohort. The activity of only *F. prausnitzii* was significantly altered in the cancerous gut. Activity of only two of the most abundant Firmicutes bacteria, *F. prausnitzii* and *R. bicirculans* (a 'secondary-degrading' cellulolytic, hemicelluloses, bacterium (516)), was reduced in CRC but not of the lactic acid bacterium *L. amylovorus*. These data suggest that the most prevalent microorganisms of the non-cancerous gut (a small proportion of species), as averaged across all samples, perform more than a quarter of all microbiome activity, providing that the RNA level from extracts accurately represent *in vivo* transcript abundance. Critically, these profiles are largely distinct from their corresponding DNA-based signatures.

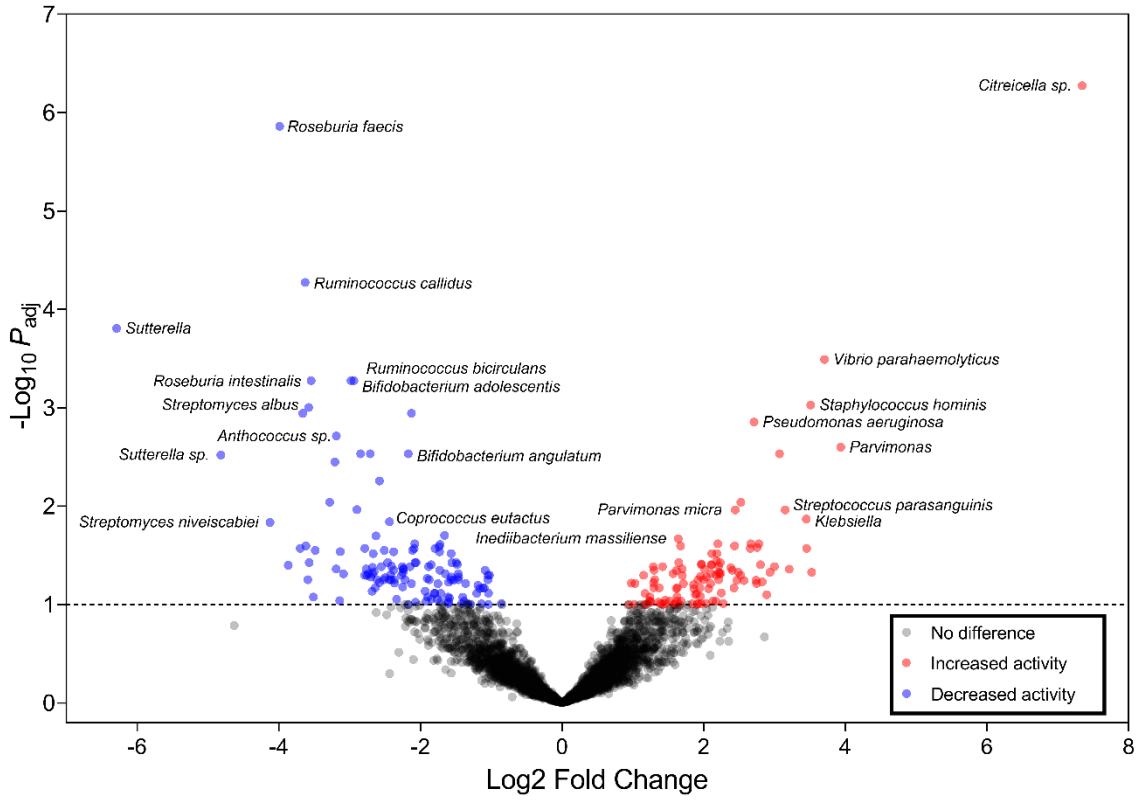


Fig. 18 | Activity of microbial species during colorectal cancer. The volcano plot indicates -Log_{10} BH-adjusted *P*-values for transcriptome-wide gene expression per species against their respective Log_2 fold changes in activity during CRC compared to control. Blue and red dots represent significantly increased and decreased activity respectively between CRC and control groups. Grey dots represent species with no changes in activity between the two groups (-Log_{10} BH-adjusted P-values <1). 240 of 8739 (3667 shown) species identified show significantly altered activity during CRC.

In CRC, *Escherichia coli* and *Bacteroides gallinaceum* of the top 20 most abundant species were significantly enriched as opposed to *Faecalibacterium* sp. and *F. prausnitzii* (Fig. 19B). Six of the most prevalent species, *B. uniformis*, *F. prausnitzii*, *B. vulgatus*, *Parabacteroides distasonis*, *E. coli* and *P. merdae* displayed similar levels of activity in CRC (Fig. 19D). Interestingly, the level of transcripts of four species, *Methanobrevibacter smithii* (a methanogenic archaea), *B. gallinaceum*, *Dialister sp*. and *Romboutsia ilealis* were not detected in either cohort. It is noteworthy that the archaeal genome is poorly represented in the SAMSA2 database, but four bacteria that belong to the genus *Dialister* were identified in both groups with no significant changes in their activities across conditions.

The signature of the 20 most active bacteria in the non-cancerous gut was compared with their corresponding abundance and activity in CRC. They shared almost half of the total microbial activity, 47.86% while their abundance made up only 16.76% to the total population (Fig. 19C). The corresponding profile of most active bacteria in CRC was 48.80% (Fig. 19D). The abundance of these most active bacteria constituted 21.22% of the total microbial population in CRC. Interestingly, 19 of those microbes had unchanged levels of transcription, with only activity of *F. prausnitzii* altered in CRC. The presence of the most active species, *B. dorei*, *Duodenibacillus massiliensis* (a gram-positive Betaproteobacteria), *Alistipes shahii* (a gram-negative strict anaerobe), *B. cellulosilyticus* (a cellulolytic bacterium), *Acinetobacter baumannii* (a gram-negative aerobic opportunistic pathogen belonging to the clinically relevant multi-drug resistant ESKAPE group (517)) and *Bacteroides timonensis* had no detectable rDNA. The metabolic activity of just 20 bacteria displays almost a half of the total microbial activity of the gut in both cohorts but only made up around one fifth of the population. These findings suggest that bacteria in the gut may exist in a dormant state (such as *R. ilealis*) and *vice versa*, actively transcribing while their DNA level is low or undetectable (see below). The presented data indicate that members of the gut microbiota who constitute the most abundant and most active profiles in each condition are distinct from one another, and the majority of transcription is carried out by a minority of the population, disproportionate to their presence.

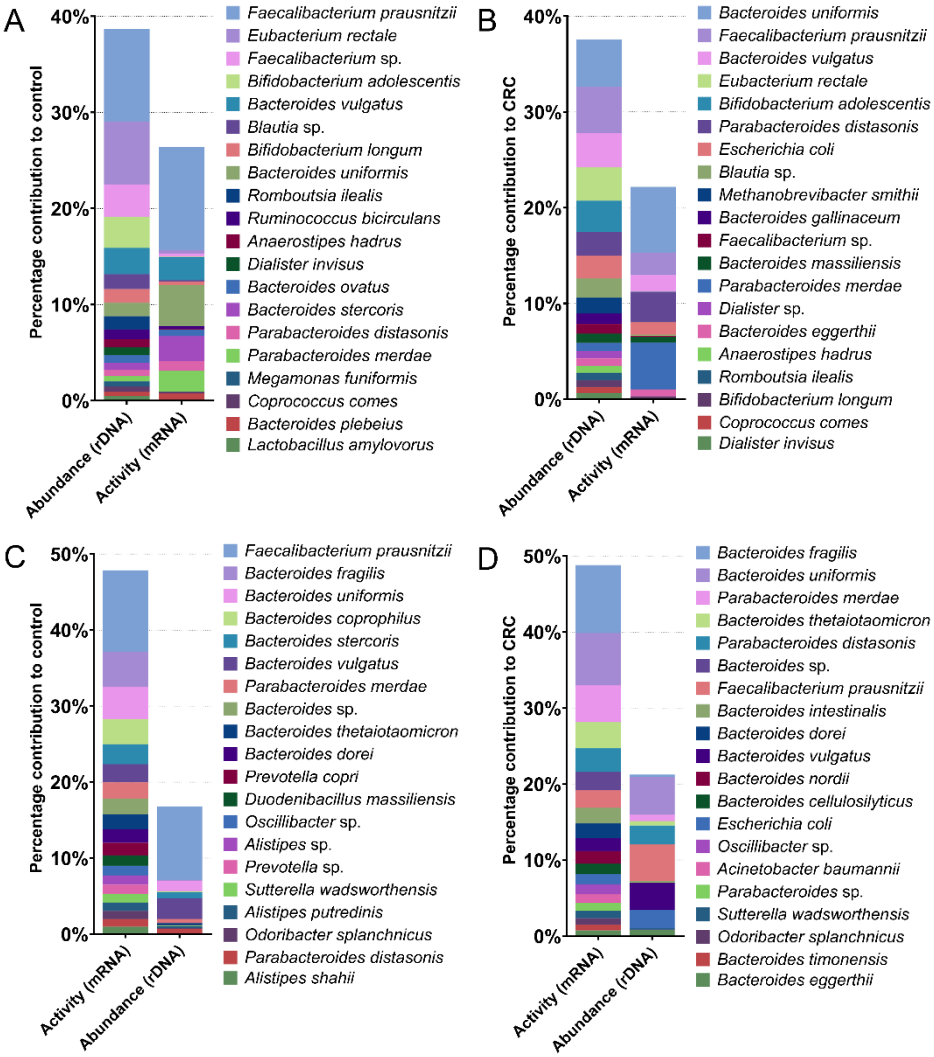


Fig. 19 | Profiles of the most abundant and most active species in both health and colorectal cancer groups are distinct from one another. A. The profile of the most abundant (38.70% of total microbial population) control cohort species and their activity (26.39% of total expressed microbiome). *Faecalibacterium prausnitzii*, *Bacteroides vulgatus*, *Bacteroides uniformis*, *Bacteroides stercoris* and *Parabacteroides merdae* were present in the profile of most active control group species (Fig. 19C). **B.** The profile of most abundant (37.59% of total microbial population) CRC cohort species and their comparative activity (22.16% of total expressed microbiome). *B. uniformis*, *F. prausnitzii*, *B. vulgatus*, *Parabacteroides distasonis*, *Escherichia coli* and *P. merdae* were also represented in the profile of most active CRC bacteria (Fig. 19D). **C.** The profile of most active (47.86% of total expressed microbiome) control group species and their comparative abundance (16.76% of total microbial population). *F. prausnitzii*, *B. uniformis*, *B. stercoris*, *B. vulgatus*, *P. merdae* and *P. distasonis* were also represented in the profile of most abundant species for the control cohort (Fig. 19A). **D.** The profile of most active (48.80% of total expressed microbiome) CRC group species and their comparative abundance (21.22% of total microbial population). *B. uniformis*, *P. merdae*, *P.s distasonis*, *F. prausnitzii*, *B. vulgatus*, *E. coli* and *Bacteroides eggerthii* were also represented in the profile of most abundant CRC species (Fig. 19B).

6.4 Members of the gut microbiota display divergent levels of abundance and transcriptional activity in CRC and non-cancerous environments

To further characterise taxonomic differences between the abundance and activity of bacteria in the gut we investigated whether DNA-based undetectable species display any detectable transcriptional activity and *vice versa*. In this study 34 species were identified at the rDNA level with no traces of transcriptional activity, 8 of those were present in the CRC gut and absent in the control group (Table S2). *Bacteroides gallinaceum*, *Helicobacter canadensis*, *Corynebacterium durum* and *Dialister* sp. were among species common to both groups with undetectable expression. *Corynebacterium amycolatum*, *Mesorhizobium plurifarium*, *Sphingopyxis terrae* and *Promicromonospora* sp. etc. were unique to the CRC gut. All microbes compared at the DNA and RNA levels were present in both databases (489, 518) (including previous taxonomic classifications). Although it is possible that the lack of transcripts may be due to faster mRNA degradation in some species relative to others, and we conclude that those organisms are transcriptionally inactive or dead.

A further subset of species showed a detectible transcriptome level while their relative abundance was greater by at least an order of magnitude (Fig. 20A), suggesting a dormant-like phenotype. A dormant microbial cell will still have low mRNA levels which can be used for housekeeping or be translationally silent (519). The low presence of a dormant species would very likely result in failure to detect its transcripts. For instance, *Chlamydia trachomatis* was represented in the CRC population more than 430-fold compared to that of their relative activity, but this DNA:RNA ratio was down 8.3-fold in the control group. Members of this sub-set of microbes, strikingly, includes species that hold significant probiotic potential. *Bifidobacterium* species *B. adolescentis*, *B. animalis* and *B. pullorum* have relative abundances 10-60-fold greater than their relative activity in both groups (excluding *B. pullorum* in CRC whose activity is greater relative to its abundance). *B. adolescentis* (a bacterium which maintains the gut-brain axis through production of the primary inhibitory neurotransmitter γ-aminobutyrate) (520, 521) made up 3.25% of the microbiome in both disease and control groups. However, the proportion of their RNA levels was at least 100-fold less in CRC. This further supports the idea that the CRC environment can regulate activity of microbes without changing their abundance. This trend is even more pronounced for several *Lactobacillus* species, e.g. *L. amylovorus*, *L. reuteri* and *L. mucosae*. *Bifidobacterium* spp. and *Lactobacillus* spp. share many phenotypic characteristics and are often considered for probiotic administration due to their acid fermenting capabilities (522). However, the realisation of their functional potential (based on DNA level alone) appears to be limited, thus their protective role may be overestimated. Consistently, *E. rectale* (inflammation-causing species), *Streptococcus salivarius* and *D. invisus* follow this dormant-like pattern. Although roles of certain species may

appear significant to health and disease due to their genome abundance, they may be overestimated in their contribution.

Contrarily, we found many transcriptionally active microorganisms in both cohorts that were largely absent (genome level) in at least one group (Table S2). The active species, *B. dorei*, *Duodenibacillus massiliensis*, *Alistipes shahii*, *B. cellulosilyticus* and a member of the ESKAPE group *Acinetobacter baumannii* (517) had no detectable rDNA. Metabolically active pro-inflammatory *A. shahii*, associated with optimal responses to cancer immunotherapy (131), and *Acinetobacter junii*, linked to nosocomial pneumonia, were also absent at 16S levels. These data further evidence the importance of studying activity of gut microbes to avoid overlooking small populations which are in fact transcriptionally active, which in this instance appear to harbour clinically relevant pathogens.

Additionally, several species were identified with relative activity far greater than their respective representation in the community, indicating hyper-active species relative to population size (Fig. 20B). For instance, *Ruminococcus torques*, a CRC-associated bacterium (222), displayed the most divergent relationship between relative presence and activity, with an average of ~$1.3x10^3$-fold difference between the DNA and RNA levels, in both health and CRC. *Bacteroides nordii*, *Bacillus thuringiensis*, *F. nucleatum*, *Prevotella timonensis*, *Megamonas* sp., *P. copri*, *Flavonifractor plautii*, *B. fragilis* and *Collinsella aerofaciens* constitute further examples of bacteria, including established CRC-associated species, where relative transcriptional activity was disproportionately greater than abundance within the community. Moreover, ecologically under-characterised species within this subset, such as *Candidatus soleaferrea* (523), warrant more detailed functional analysis. These data highlight that genome abundance is a poor indicator of microbial activity in the community and may lead to the underestimation of the roles of certain potentially relevant pathogens in CRC.

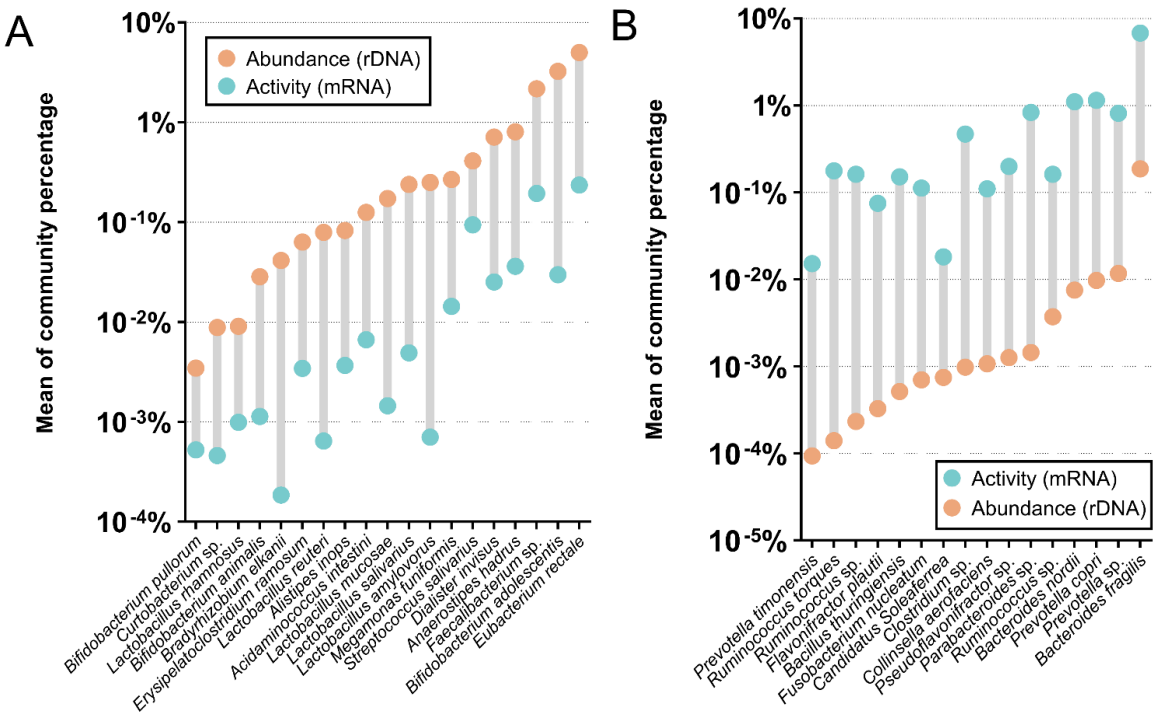


Fig. 20 | Relative abundance and activity are affected by the CRC gut niche independently. A Species, the abundance (level of 16S rDNA) of which overrepresents their corresponding activity (level of expressed mRNA). **B** Species, the activity of which overrepresents their corresponding abundance. All data presented are as a mean percentage of total microbial population (abundance) or transcriptome (activity) across both cohorts.

6.5 Activity of the microbiome in CRC shifts away from beneficial species towards a diverse range of pathogens

Intestinal microbiota provide the host with metabolites critical to the maintenance of gut homeostasis, including short-chain fatty acids (SCFAs), such as *n*-butyrate (524). *n*-butyrate, the preferred energy source for colonocytes, fortifies the epithelial barrier and suppresses inflammation (371) and is mainly produced by *Clostridium* clusters IV, XIV and XVI, abundances of which have been shown to negatively correlate with CRC (177). We observed, with the exception of *C. perfringens* (a common cause of food poisoning (525)) (Table 6), enhanced activity of *Clostridium kluyveri* (synthesising *n*-butyrate from ethanol and lactate) and *Inediibacterium massiliense* (encoding butyrate kinase), and significant loss of activity of 22 major *n*-butyrate producing species (Fig. 21). This is in line with transcription of *n*-butyrate-metabolising genes/pathways (526). We also found that activity of several *Streptomyces* species (major antibiotic-producing bacteria which control microbial community composition (527)) were diminished in CRC. Interestingly, only three species, *S. albus*, *S. mangrovisoli* and one unknown species (out of ten identified) contributed almost half (48%) of all *Streptomyces* activity in the control group, with undetectable levels of rDNA, and proportional levels of transcript 39% lower than the CRC group (down to 9%). This suggests that specific *Streptomyces* species may be important for the health of the gut.

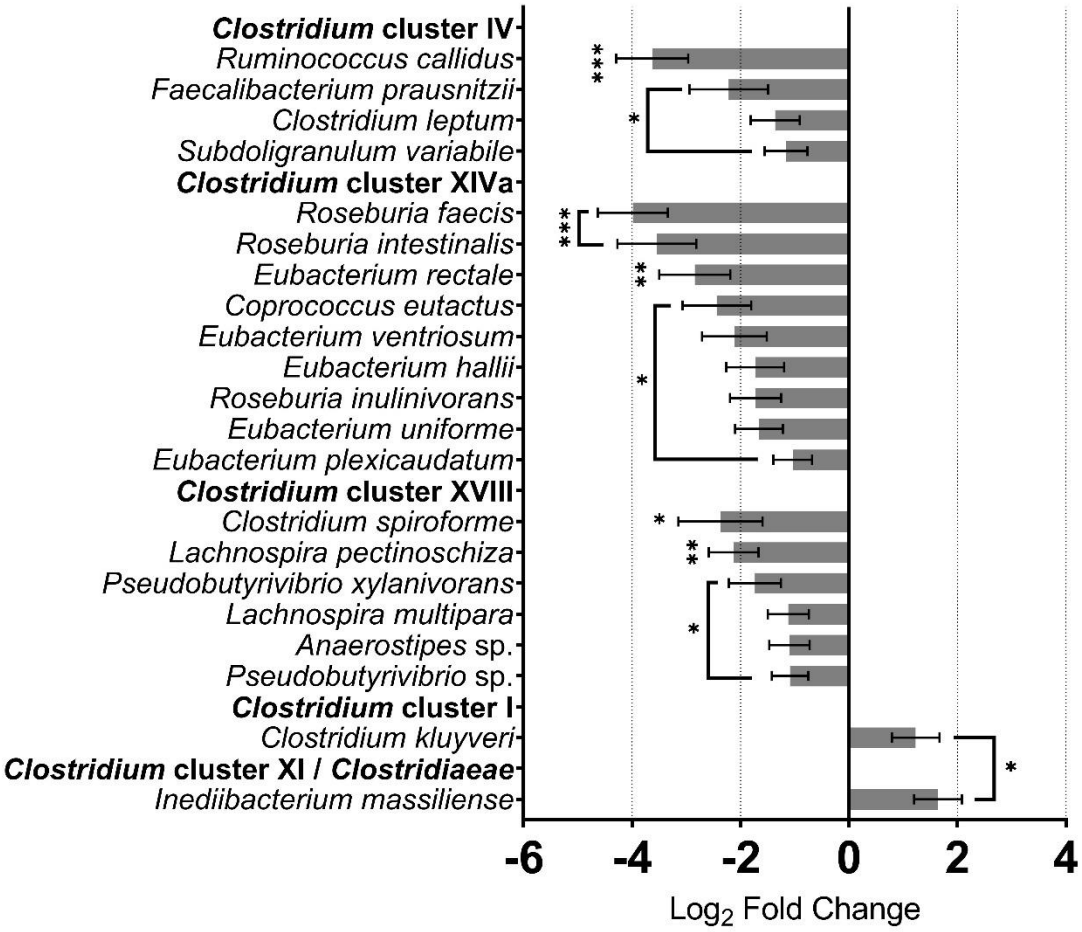


Fig. 21 | Beneficial butyrate-producing species are less active in CRC while some pathogenic butyrate-producers gain activity. Species which possess butyrate-synthesising genes activity are widely affected in the CRC gut. The beneficial *n*-butyrate-producing species with significantly altered activity in CRC belong to *Clostridium* clusters IV, XIVa and XVIII. Pathogenic *n*-butyrate-producing *Clostridium* cluster I, *Clostridium kluyveri,* and *Clostridium* cluster XI/*Clostridiaceae*, *Inediibacterium massiliense* (528), are also shown, however with elevated activity in the malignancy. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

Organism Name	Log2 Fold Change	lfcSE	P_{adj}
Upregulated in CRC			
Actinomyces cardiffensis	1.92	0.70	0.0972
Actinomyces dentalis	1.95	0.66	0.0676
Actinomyces graevenitzii	1.90	0.66	0.0773
Actinomyces israelii	1.86	0.62	0.0629
Actinotignum urinale	1.86	0.63	0.0686
Clostridium perfringens	1.84	0.67	0.0980
Dolosigranulum pigrum	1.18	0.43	0.0999
Enterobacter lignolyticus	2.47	0.81	0.0565
Gemella sanguinis	2.44	0.67	0.0253
Gemella sp.	1.91	0.70	0.0995
Klebsiella aerogenes	2.43	0.82	0.0676
Klebsiella michiganensis	1.95	0.68	0.0827
Klebsiella quasipneumoniae	2.18	0.78	0.0924
Klebsiella variicola	2.74	0.90	0.0565
Mycoplasma pneumoniae	2.83	0.93	0.0588
Parvimonas micra	2.45	0.62	0.0109
Staphylococcus hominis	3.51	0.74	0.0009
Staphylococcus pasteuri	1.62	0.58	0.0924
Staphylococcus saprophyticus	2.25	0.78	0.0761
Streptococcus agalactiae	1.68	0.60	0.0924
Streptococcus cristatus	1.96	0.63	0.0515
Streptococcus dysgalactiae	1.61	0.50	0.0452
Streptococcus gordonii	2.20	0.60	0.0240
Streptococcus infantis	2.40	0.74	0.0447
Streptococcus mitis	2.11	0.63	0.0395
Streptococcus parasanguinis	3.15	0.80	0.0109
Streptococcus pneumoniae	2.11	0.63	0.0393
Streptococcus porci	2.78	0.76	0.0240
Streptococcus pyogenes	1.58	0.58	0.0995
Streptococcus sanguinis	1.45	0.53	0.0975
Vibrio parahaemolyticus	3.71	0.73	0.0003
Downregulated in CRC			

Aggregatibacter aphrophilus	-2.24	0.70	0.0469
Anaerobiospirillum succiniciproducens	-1.33	0.48	0.0975
Campylobacter insulaenigrae	-2.76	0.88	0.0505
Chryseobacterium gleum	-1.80	0.63	0.0768
Clostridium neonatale	-1.02	0.33	0.0502
Haemophilus influenzae	-2.25	0.76	0.0660
Haemophilus parainfluenzae	-2.68	0.81	0.0418
Legionella steigerwaltii	-2.27	0.72	0.0502
Prevotella bivia	-2.33	0.83	0.0882
Pseudomonas citronellolis	-3.57	1.05	0.0373
Sphingomonas sp.	-1.65	0.52	0.0487
Eubacterium hallii	-1.73	0.53	0.0447
Eubacterium plexicaudatum	-1.04	0.36	0.0763
Eubacterium rectale	-2.84	0.65	0.0029
Eubacterium uniforme	-1.66	0.44	0.0198
Eubacterium ventriosum	-2.11	0.60	0.0281
Streptomyces albus	-3.58	1.07	0.0001
Streptomyces mangrovisoli	-3.86	1.16	0.0398
Streptomyces niveiscabiei	-4.12	1.07	0.0146
ESKAPE Pathogens			
Enterococcus faecium	0.15	0.40	0.8466
Staphylococcus aureus	1.63	0.55	0.0674
Klebsiella pneumoniae	1.56	0.63	0.1392
Acinetobacter baumannii	1.45	0.69	0.2145
Pseudomonas aeruginosa	2.71	0.59	0.0014
Enterobacter lignolyticus	2.47	0.81	0.0565

Table 6 | Activity of clinically relevant species, including ESKAPE pathogens, is positively regulated by CRC. Opportunistic pathogens which may cause infections in immunocompromised individuals and the activity of which are regulated (significantly different transcript levels) in CRC are presented. Transcriptome of three ESKAPE bacteria, *E. faecium*, *K. pneumoniae* and *A. baumannii*: lfcSE, log2 fold change standard error. P_{adj}, *P*-value adjusted with Benjamini-Hochberg.

Metagenome studies have found many lactic acid-producing bacteria, such as *Bifidobacterium* and *Streptococcus thermophilus*, are underrepresented in CRC (529). We identified that functional activity of six members of the *Bifidobacterium* genus was downregulated in CRC while activity of a probiotic *S. thermophilus* was enhanced (Fig. 22). Notably, it has been suggested certain probiotics (naturally present GI species known to produce beneficial metabolites, often consumed in large quantity, from certain foods or dietary supplements, to maintain or restore their population) pose a risk to immunocompromised individuals (530). Interestingly, activity of four *Lactobacillus* bacteria was stimulated in cancer while the activity of another probiotic, *Lactobacillus ruminis* (531) was reduced. Despite sharing characteristics, namely lactic acid producing genes, total activities of these species are differentially regulated under the same CRC conditions.

There are a number of pathogens present in the human gut that upon gaining access to the bloodstream may cause life-threatening infections (532), many of which carry multi-drug resistance, such as Enterobacteriaceae (533). We identified members of three genera *Klebsiella*, *Enterobacter* and *Kluyvera* with significantly enhanced activity in CRC (Fig. 23), yet surprisingly activity of *E. coli* was not modulated by CRC (Dataset 1, Tab 2 of (513)) and this was confirmed by qRT-PCR (Fig. 24). This aligns, at least in-part with our functional metatranscriptome analyses (presented below), that the gut microbiome displayed greater expression of antibiotic resistance determinants in CRC, including β-lactams and efflux pumps. The factors responsible for enhancing Enterobacteriaceae activity, such as nutrient availability or colonocyte invasion, are yet to be established.

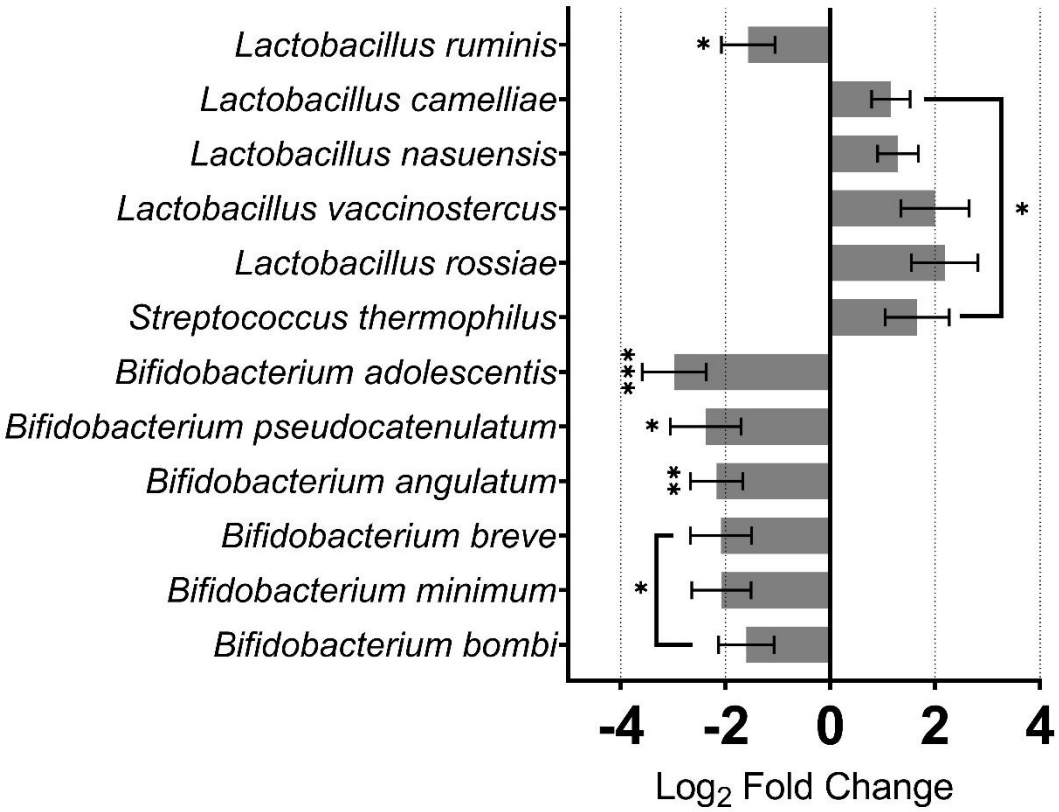


Fig. 22 | Probiotic genera, including *Lactobacillus* and *Bifidobacterium* show varied alterations in activity in CRC. Species with probiotic capacity from three different genera, *Lactobacillus*, *Bifidobacterium* and *Streptococcus* exhibit different patterns of expression during malignancy. *Lactobacillus* displayed species-specific regulation of transcription during CRC, *L. ruminis* (beneficial) alongside the *Bifidobacterium* species was less active while other *Lactobacillus* and *Streptococcus* spp. were more active. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

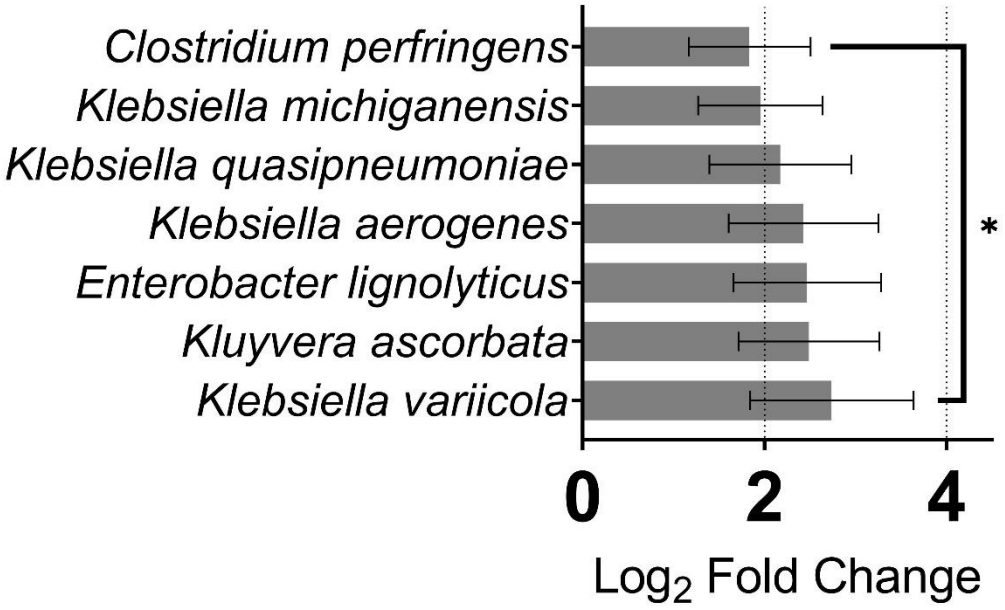


Fig. 23 | Enterobacteriaceae family pathogens have increased activity in the CRC niche. Four genera of the Enterobacteriaceae family, *Clostridium*, *Klebsiella*, *Enterobacter* and *Kluyvera* have member species with augmented transcriptional activity during CRC. * $P \leq 0.05$.

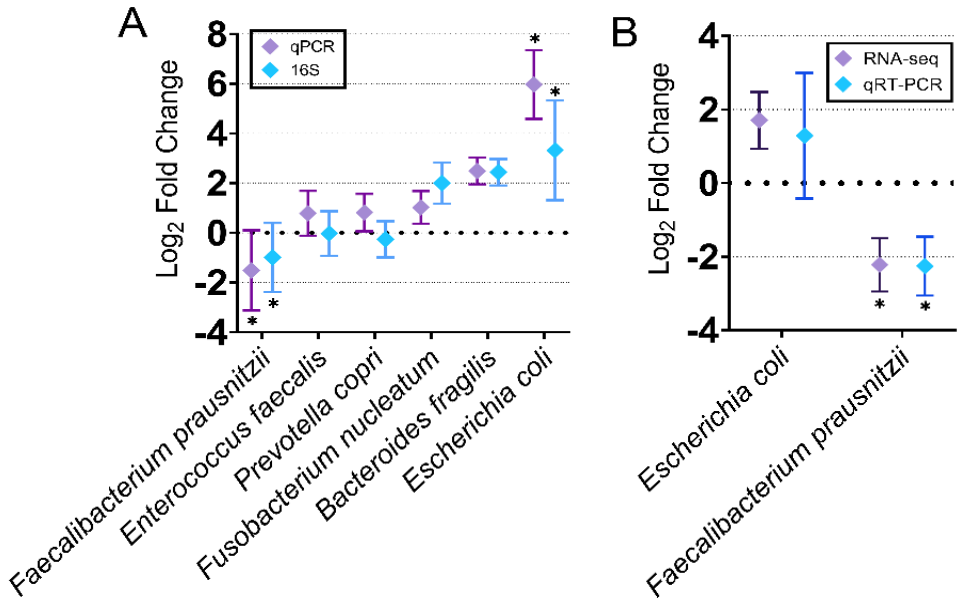


Fig. 24 | qPCR confirmation of 16S sequencing analysis and qRT-PCR confirmation of metatranscriptome sequencing analysis. A. Patterns of abundance for six CRC-associated species established by 16S rDNA sequencing confirmed by *argS* targeted qPCR. *F. prausnitzii* (*P*=0.022) and *E. coli* (*P*=0.015) both show significantly altered abundance in CRC, underrepresented and enriched respectively. All data are presented as Log_2 fold change of abundance between CRC and control cohorts. **B.** Transcriptome profiling confirmed through qRT-PCR. Levels of expression and their fold differences between conditions established through metatranscriptome sequencing were confirmed by qRT-PCR for *Escherichia coli* and *Faecalibacterium prausnitzii*. Both species were selected for having high levels of transcription and differential activity between conditions. * $P \leq 0.05$.

We identified a set of 31 pathogens, which can cause severe invasive infections particularly in vulnerable immunocompromised individuals (532), with enhanced activity in CRC (Table 5). Among them were *Staphylococcus aureus*, *Pseudomonas aeruginosa* and *Mycoplasma pneumoniae* that usually in combination cause pneumonia, sinusitis, and diarrhoea, in addition to *Staphylococcus saprophyticus*. Furthermore, 11 *Streptococcus* species (belonging to pathogenic groups A, B, C and G, e.g. *S. pneumoniae*, *S. pyogenes*, *S. dysgalactiae* and *S. agalactiae*) were among these clinically important pathogens.

We also observed repressed activity of 16 pathogens, including *Haemophilus* sp., *Haemophilus parainfluenzae* and *Sphingomonas* sp. (Table 5). Interestingly, this included five *Eubacterium spp.*, opportunistic pathogens that can produce SCFAs. These five species constituted 54% of the *Eubacterium* genus's transcription in health and only 21% in CRC, suggesting these bacteria may play more health-related, protective roles than pathogenic. ESKAPE pathogens, as recognised by the World Health Organisation, are exhausting treatment options as they carry multidrug resistance, including extended spectrum β-lactamase, vancomycin and methicillin resistances (534). All ESKAPE bacteria, including *Enterobacter lignolyticus*, were active in both cohorts (Table 5). *E. faecium* (0.009% of total metatranscriptome activity), *Klebsiella pneumoniae* (0.136%) and *A. baumannii* (1.028%) showed no significant difference in their activity. We have identified two members of the *K. pneumoniae* phylogroup, *K. quasipneumoniae* (KpII subgroup (535)) and *K. variicola* (KpIII subgroup) the activities of which were significantly upregulated in CRC. Metatranscriptome levels of *S. aureus*, *P. aeruginosa* and *E. lignolyticus* were also significantly higher. These data argue the activity of ESKAPE pathogens is regulated in the CRC gut niche.

6.6 Sulphate-Reducing Bacteria and Thermophiles are differentially active in the CRC gut

Expansion of sulphate-reducing bacteria (SRB) which produce hydrogen sulphide (H_2S), a by-product of anaerobic respiration, has been linked to the development of CRC (536). SRB reduce inorganic sulphate SO_4^{2-} to H_2S which, although do not damage DNA directly, exhibits pro-cancerous effect via RAS-MEK signalling. We found the activity of five SRB was differentially regulated (Fig. 25), three species displayed enhanced and two reduced activities. Our functional analysis of microbial pathway and gene expression indicated that the CRC gut environment is highly acidic (see below), which can be due to a high level of organic acids, such as lactate, an electron donor for SRB. Hence, alteration in gut acidity e.g. due to changes in metabolism of members of the microbiota, or the Warburg effect in cancer cells may influence metabolism of other groups of microorganisms, such as SRB that could enact a genotoxic stress on colonocytes.

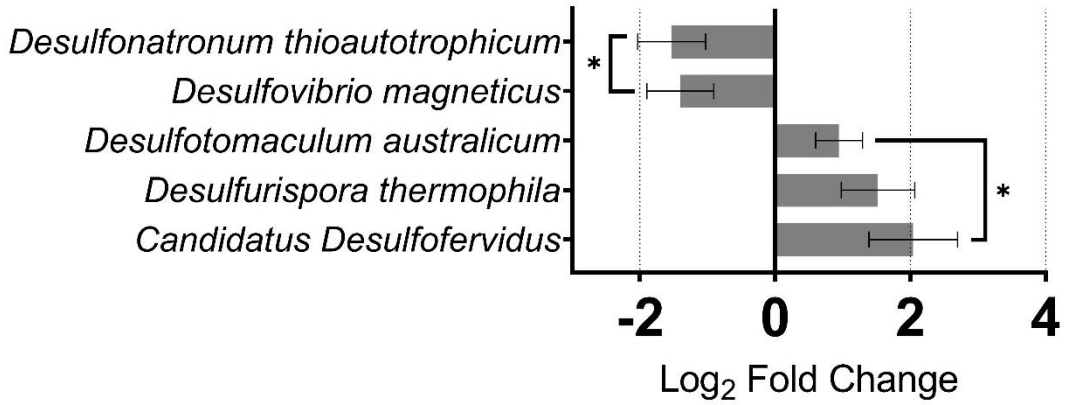


Fig. 25 | Sulphate-reducing bacteria display species-specific CRC regulated activity. Bacteria with the capacity to reduce environmental sulphate to potentially carcinogenic hydrogen sulphide (537, 538), display differing patters of activity, dependent on the species, during CRC. * $P \leq 0.05$.

We found 15 thermophilic species with significantly higher transcriptional activity in CRC and *Desulfovibrio magneticus* (a SRB) and *Treponema caldarium* (oral cavity bacteria) displayed attenuated activity (Fig. 26). The majority of thermophiles grow optimally at a basic pH between 7 and 11 despite the CRC gut likely being highly acidic (as later finding of this work). Depending upon microbial loci, activity of these thermophiles may be regulated by dynamic microenvironments and/or transient stimuli.

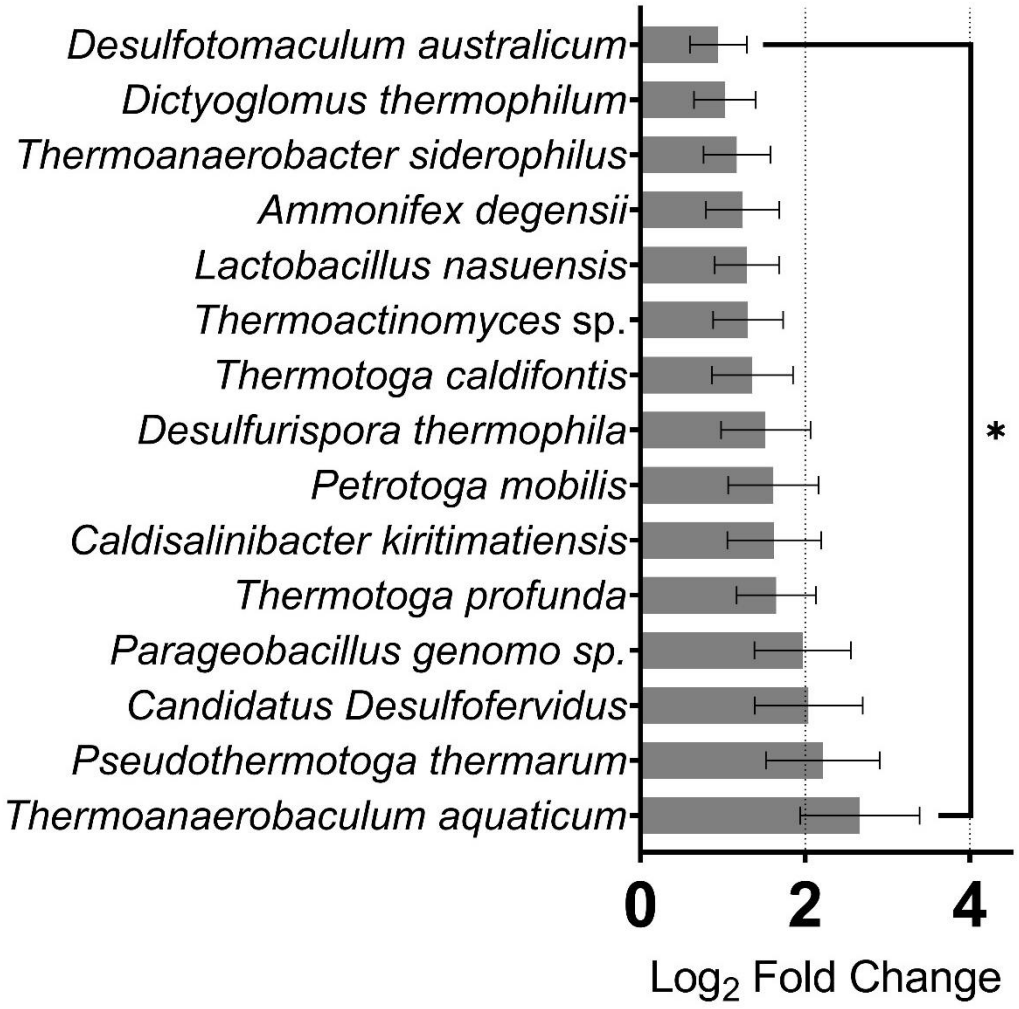


Fig. 26 | Despite the potentially high acidity of the CRC gut, thermophiles can thrive. 15 species of thermophilic (grow optimally at temperatures >41°C and pH 7-11) bacteria possess elevated activity in the CRC gut niche. * $P \leq 0.05$.

6.7 Gut microbiota express antibiotic resistance genes without antibiotic treatment and transcription of many resistance genes is upregulated by the gut environment

Based on extensive DNA-based analyses it has been established that the gut microbiota is a reservoir of antibiotic resistance determinants. However, expression of antibiotic resistance determinant genes in the gut has not been investigated so far. We applied the RGI approach to the metatranscriptomic data to analyse the level of and any differences in expression of antibiotic resistance determinants across groups, as described in 5.21. The surprising data of enhanced AB (antibiotic) resistance determinant expression by the CRC microbiota without external AB pressure prompted us to investigate differential AB resistome profile in greater detail (Dataset 1, Tab 4 of (513)). A high level of resistance genes expressed in a number of clinically relevant microorganisms was observed for, e.g. *S. aureus*, *C. difficile*, *Salmonella enterica serovar Typhimurium*, *S. pneumoniae*, *Neisseria gonorrhoeae* and *N. meningitidis*, *Campylobacter jejuni*, *H. pylori*, *E. coli*, *Mycobacterium tuberculosis* etc, to a wide spectrum of antibiotic classes. Critically, these AB resistance determinants were expressed irrespectively of the health status of the gut. This suggests that the human gut environment supports expression of AB resistances as a part of microbial adaptation. Comparison of expression of AB resistance genes by the CRC and control gut microbiota revealed differential expression of 45 resistance genes ($P<0.05$) with a further 71 genes differentially transcribed with 90% confidence. Observed AB resistance determinants belonged to more than twelve different families (Fig. 27A). Among them, CRC-dependent overexpression of 52 AB resistance genes was observed for all ESKAPE species (Fig. 27B). Interestingly, expression of only 9 AB resistance genes were downregulated in the CRC microbiome, transcription of all other AB resistance determinants was significantly increased in cancer. Strikingly, EF-Tu dependent elfamycin and *rpoB* mediated rifamycin resistances were inhibited. This could be due to a “collateral” effect of acidity of the CRC gut, low pH may affect DNA replication, transcription and translation (539), hence expression of AB resistances that target translation and transcription may be affected by downregulation of targeted genes.

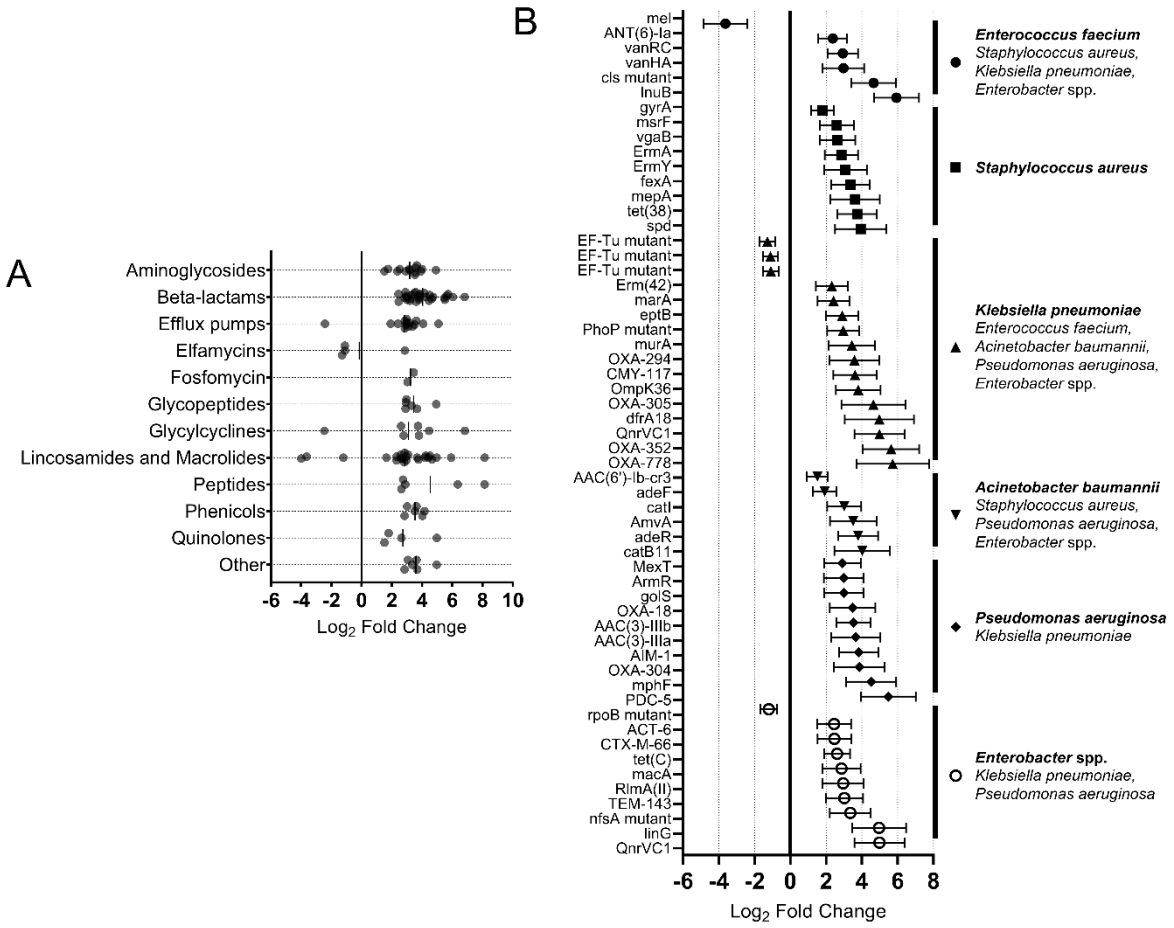


Fig. 27 | Resistance genes to over 12 classes of antibiotics as well as multidrug resistance are found differentially expressed in CRC including by ESKAPE pathogens. **A.** CRC microbiota upregulates expression of AB resistance determinants of 12 major families of antibiotics, including aminoglycosides, β-lactams, lincosamides and macrolides. Other ABs/AB families include synthetic oxazolidinone, rifamycin, streptogramins, pleuromutilin, nitrofuran, isoniazid and diaminopyrimidine. **B.** ESKAPE pathogens of the CRC gut upregulate expression of a gamut of AB resistance determinants. Bold species denote primary expressor of genes, listed ESKAPE pathogens below bold species can also express genes displayed for that group. All data points plotted are statistically significant (either >90% or >95% confidence). Solid black circle: *E. faecium*; Black square: *S. aureus*; Upward pointing black triangle: *K. pneumoniae*; Downward pointing black triangle: *A. baumannii*; Black diamond: *P. aeruginosa*; Hollow black circle: *Enterobacter* spp.

One of the striking features of the CRC gut environment is that it differs from the control counterpart with respect to environmental acidic, osmotic and oxidative pressures. We proposed that such factors may in part regulate expression of AB resistance genes. Hence, we tested the influence of these environmental factors *in vitro* on microbial expression of AB resistance genes (Table 3). Purified metabacteria were cultured under distinct stress conditions which based on our metatranscriptome analysis we believe the CRC gut microbiota are subjected to, in the presence of H_2O_2, NaCl, Lactate or HCl. Expression of AB resistance determinants (identified as differentially expressed through targeted RNA sequence data analysis) were investigated through qRT-PCR using primers from the CARD database. Non-treated cultures acted as controls, and housekeeping 16S rRNA was used for normalisation, when conducting statistical analyses.

6.8 H_2O_2 represses and enhances expression of AB resistance determinates depending upon the health status of the gut

The unexpected discovery of a high level of expression of a significant proportion of AB resistance determinants in the gut microbiota regardless of the health status of the host and significantly enhanced expression of >200 of AB resistance determinants in CRC led us to test if their expression by metabacteria *in vitro*. We hypothesised their expression may be regulated by specific environmental stresses of the CRC gut, as suggested via our metatranscriptomic analysis.

Our metatranscriptome analysis revealed differential expression of several bacterial genes that respond to the presence of ROS, and specifically the availability of H_2O_2. This suggests that availability of H_2O_2 in the gut depends upon the health status (6.13). Based on these findings we decided to test whether H_2O_2 has an effect on expression of specific AB resistance determinants *in vitro* using the gut metabacteria purified from CRC and control faecal samples. Hydrogen peroxide simultaneously upregulated expression of *nsfA*, *mdtO*, *catA*, and *gadX* antibiotic resistance genes of the control microbes while repressing their transcription in the CRC derived microbes (Fig. 28). Expression of *phoP*, and bla_{CMY} genes was also inhibited in the CRC microbes with no difference in its expression in microbes derived from the control samples. Expression of the *bla2* gene was significantly attenuated in both cultures. Metatranscriptomic analyses for these genes showed their transcription was significantly greater in CRC. This suggests that these antibiotic resistance determinants are not regulated by hydrogen peroxide, and this is consistent with our observation that H_2O_2 oxidative pressure is not a major feature of CRC in contrast to the control gut environment. Expression of *cls* and *marA* was enhanced in CRC microbes in response to H_2O_2. Transcription by *E. coli* in the control cultures was elevated for *marA* and repressed for *bla2*. Expression of *eptB* was not changed by the CRC *E. coli* but switched on in the control. This suggests that oxidative

pressure is a potential regulator of antibiotic resistance of the gut microbiome in a health-dependent manner.

6.9 Osmotic pressure inhibits AB resistance irrespectively of the health of the host but may activate certain resistance mechanisms

Our metatranscriptomic analysis of the CRC microbiome showed enhanced expression of several genes which are involved in the osmotic stress response. This suggests that microbiota in CRC may be under greater osmotic pressure. Hence, to test if osmotic stress is a potential regulator of AB resistance determinant expression, we applied NaCl to the growth media of the gut metabacteria and monitored the level of expression of specific AB determinants, the differential expression of which was observed via the metatranscriptomics. Osmotic pressure downregulated expression of *cls*, *bla2*, *gadX*, bla_{CMY}, *mdtO*, *phoP* and *catA* genes in CRC and control bacteria (Fig. 28). Expression of *nsfA* was not affected by NaCl in either culture while expression of *eptB* and *marA* was upregulated by the CRC microbes. The control culture did not change expression of *marA* while expression of *eptB* mirrored the phenotype for oxidative pressure. These data argue that osmotic pressure suppresses antibiotic resistance gene expression in a health-independent manner while it may activate expression of specific AB resistance genes depending upon the health status of the host.

6.10 Organic and inorganic acids control AB resistance gene expression differently

The CRC gut microenvironment is well known to be featured with highly acidic conditions due to Warburg metabolism of the tumour, leading to the secretion of lactic acid to the surrounding environment (556). Our metatranscriptomic analysis of differential gene expression also revealed expression of genes which are involved in acid stress responses (6.15). To test if acidity influences the expression of specific AB resistance determinants *in vitro,* we subjected metabacterial cultures to organic (lactic acid) and non-organic (HCl) acids and measured expression of these resistance genes. Expression of *nsfA*, *eptB*, *marA* and *phoP* was regulated by acidity in a health dependent manner (Fig. 28), both acid conditions upregulated the gene expression by CRC aerobes but not in the control, except for *marA* expression in response to inorganic acid (HCl). DL-lactate promoted expression of *mdtO*, *catA* and *bla2* genes in the CRC culture but had no effect of their expression in the control. Transcription of *cls* and *gadX* genes was repressed by lactate in the control cultures and their expression was enhanced and non-changed in the CRC cultures, respectively. Expression of bla_{CMY} was not regulated by lactate in either culture but inorganic acid supressed its expression in both. Inorganic acid prompted no changes in expression of *gadX*, *mdtO* or *catA* in the CRC aerobes while attenuated their expression in the control. Furthermore, HCl adjusted acidity significantly downregulated transcription of *cls* and *bla2* genes irrespective of the health status origin of bacteria. Lactate appears to be a major positive regulator of AB resistance gene

expression in CRC-derived aerobes and affects resistance gene expression generally in a health-dependent manner. Inorganic acid, in contrast, variably regulated resistance gene expression. HCl-dependent AB resistance gene expression by health-associated aerobes differs to that of lactate however, there is limited overlap in the patterns of gene expression following exposure to the two acids.

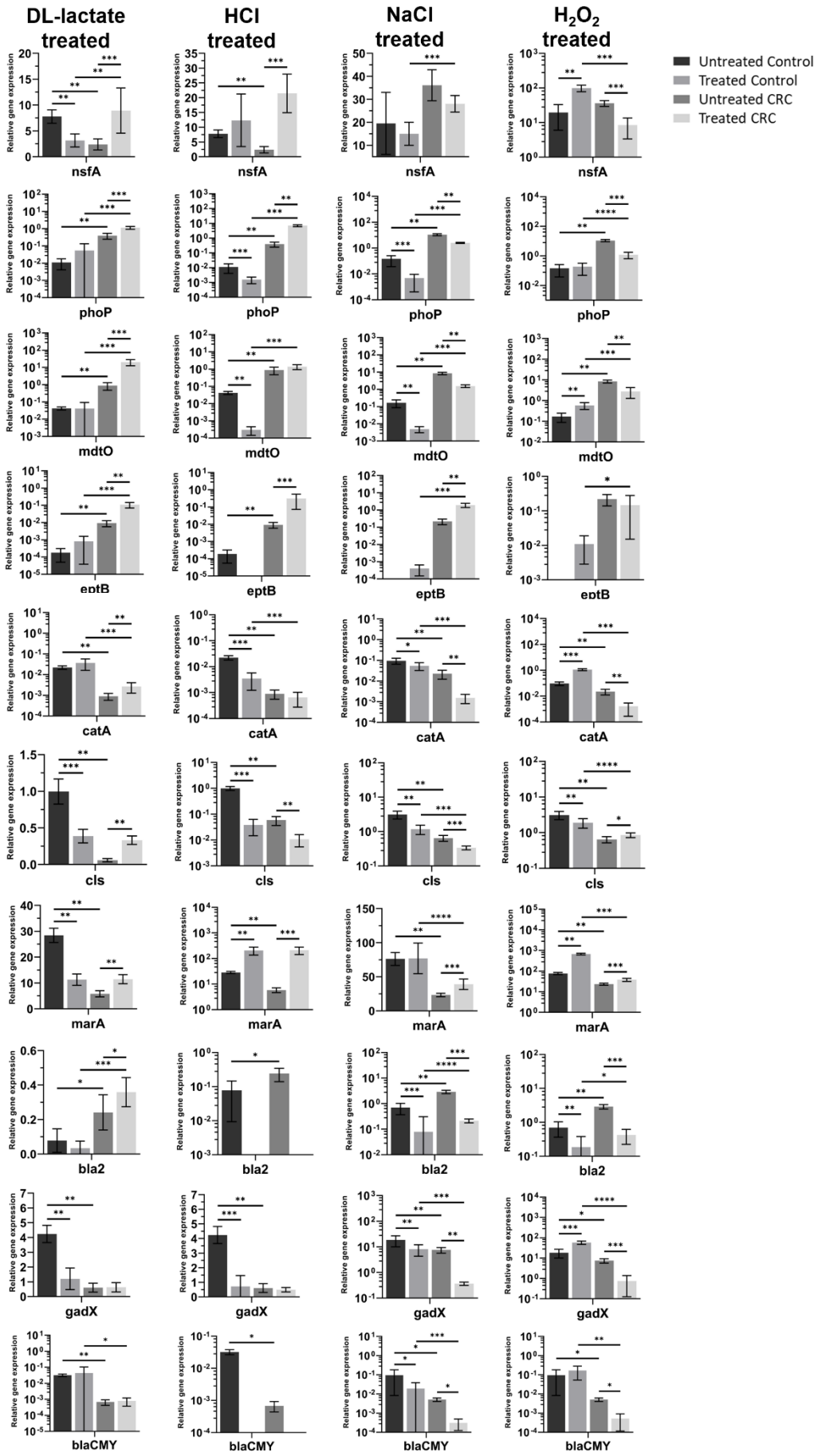

DL-lactate treated
HCl treated
NaCl treated
H_2O_2 treated
Untreated Control
Treated Control
Untreated CRC
Treated CRC
Relative gene expression
nsfA
phoP
mdtO
eptB
catA
cls
marA
bla2
gadX
blaCMY

Fig. 28 | Environmental pressures regulate expression of multidrug resistance determinants. Expression of 10 antibiotic resistance genes was quantified by qRT-PCR conducted following growth of meta-microbiota isolated from both CRC and control cohorts under four different CRC-related pressures, DL-lactate (pH 3.5), hydrogen chloride, HCl (pH 3.5), sodium chloride, NaCl (5%) and hydrogen peroxide, H_2O_2 (1.5mM). Levels of expression are shown in arbitrary relative units. Error bars denote standard deviation (n=9), asterisks represent statistically significance * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$.

6.11 Oral cavity and biofilm forming bacteria are more active in the CRC gut

Several reports have associated CRC with oral cavity microbiota (524), such as *Parvimonas micra* and *Streptococcus spp*., suggesting the use of oral bacteria as CRC markers. We observed an extensive collection of oral cavity species which were differentially active between conditions (Fig. 29A). Later colonisers of oral biofilms *Parvimonas micra*, a core CRC-enriched oral pathogen identified through metagenomic analysis (540), and *Veillonella magna* were among those with elevated activity. Early-stage plaque and biofilm forming species, including *Rothia* sp., *Gemella* sp., four *Actinomyces* pathogens, including *A. dentalis*, *Streptococcus pyogenes* among other *Streptococcus* species, all which cause periodontal diseases, were more active in the CRC gut. This suggests these pathogens which initiate oral biofilm formation may act with a similar *modus operandi*. Consistent with this, we found that activity of *Anaerostipes* sp. and *Roseburia* are diminished in cancer (Fig. 29B), the increased abundance of which was negatively associated with colonocyte colonisation by oral bacteria (223). Furthermore, among bacteria that lost activity in CRC, we observed several oral pathogens which are known for being secondary/later colonisers, such as *Aggregatibacter* sp., pro-inflammatory *Tannerella* sp., *Porphyromonas* and *Prevotella* (541). Interestingly, their, and *Haemophilus* spp. genome enrichment in CRC has been reported, further emphasising the importance of functional analysis. This also suggests that activity of secondary colonisers may be regulated by their ability to adhere to other periodontal pathogens while they may lose out to gut bacteria for secondary colonisation.

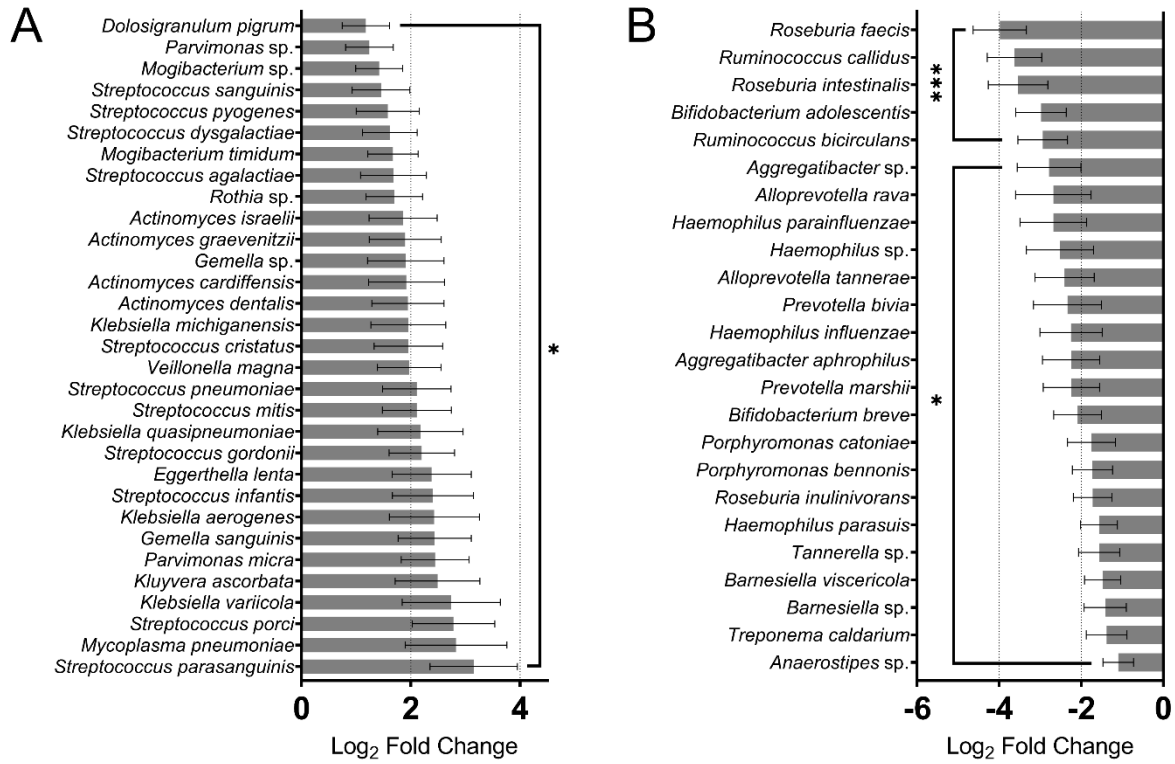


Fig. 29 | Known early- and late-stage coloniser oral cavity bacteria show CRC-regulated activity. **A.** Oral cavity species with significantly elevated activity in the CRC gut. **B.** Oral cavity species with significantly diminished activity in the CRC gut. * $P \leq 0.05$, *** $P \leq 0.001$.

6.12 Cofounder effect exclusion and overall regulated genes and pathways reported

Through principal component analysis (PCA) we established that only 'health condition', namely CRC, had any effect on global transcription by the gut microbiome, other factors included in patient metadata (e.g. age, sex and BMI) had little to no influence over transcriptome composition (Fig. 30 and Fig. 31A). Calculation of α-diversity showed consistency between CRC and non-CRC groups (Fig. 31B). However, ANOSIM (analysis of similarities) based on Bray-Curtis β-diversity revealed inter-sample diversity was significantly greater than intra-sample dissimilarity between CRC and control (Fig. 31C). Sequences were mapped to annotated gene sequences and assigned to curated subsystems of functional roles (SEED 'Subsystems' hierarchy level 3 in MG-RAST). The differential relative transcript level of these subsystems was compared between control and CRC samples to characterise the CRC-associated functional transcriptome. Of the 1361 curated subsystems, 901 were identified in this analysis (Dataset 1, Tab 1 of (526)). 49 subsystems were significantly over-represented and 24 significantly under-represented across all samples with 261 genes out of 6495 were differentially expressed, 182 up- and 79 downregulated in CRC (Fig. 32 and Dataset 1, Tab 2 of (526)). These differentially expressed subsystems and genes of the gut microbiota represent a CRC-specific transcriptional signature.

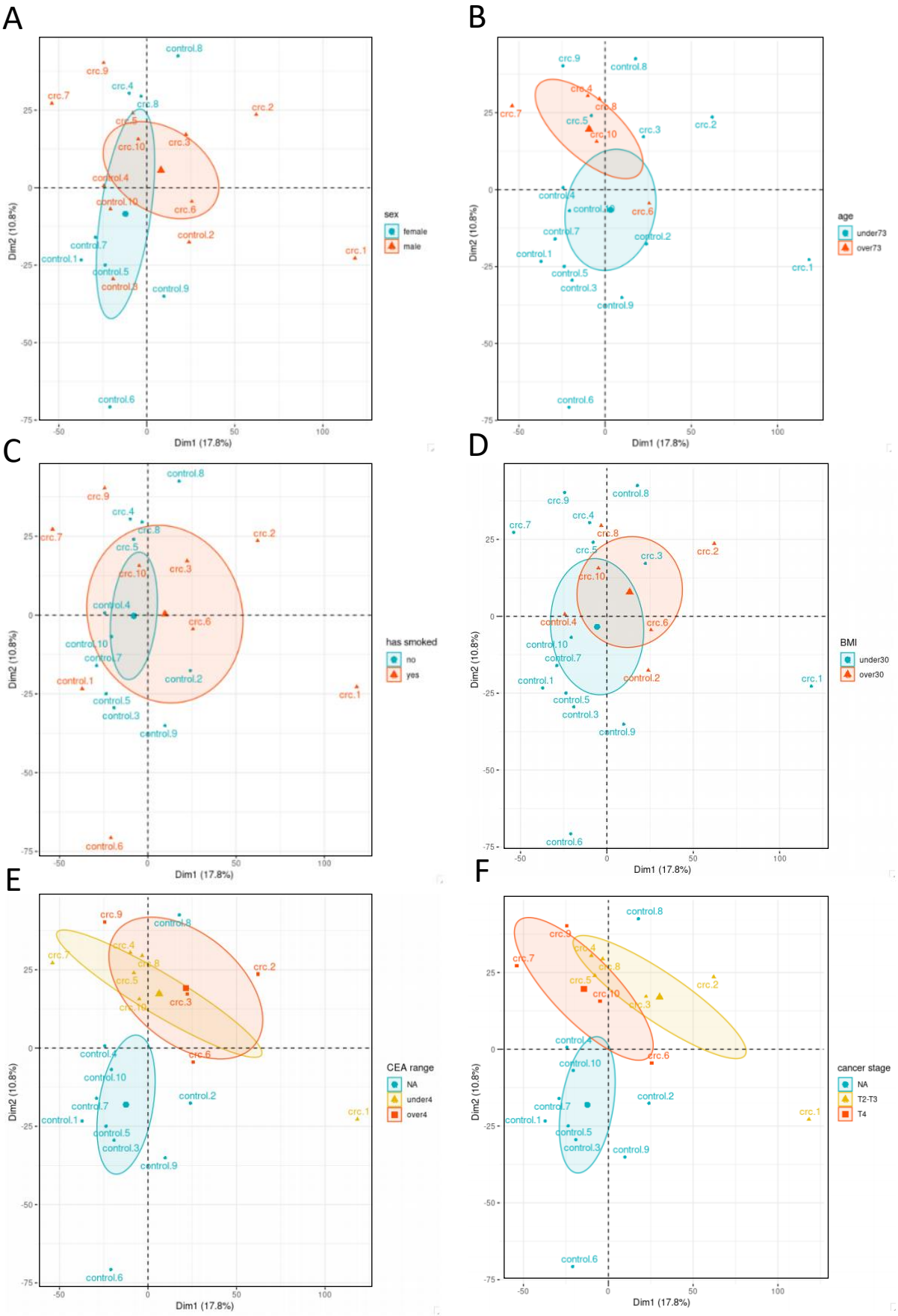


Fig. 30 | Factors included in patient metadata show limited to no effect on microbial gene expression during colorectal cancer. Scatter plots of variation across Dim1 and Dim2 of microbial

transcriptome-wide gene expression. **A.** Gender. Microbial gene expression was not affected by participant sex. **B.** Age (above and below 73 years). Microbial gene expression shows clustering in the over 73 group, no such clustering under 73. **C.** Smoking status. Microbial gene expression was not affected by participant smoking status. **D.** Body mass index, BMI (obese, >30 and non-obese, <30). Microbial gene expression was not affected by participant obesity status. **E.** Carcinoembryonic Antigen (CEA) range <6 µg/L. Microbial gene expression was not affected by patient CEA range (grouped above and below 4 µg/L). **F.** Cancer stage (early stage, T2-T3 and late stage T4). Cancer stage may affect microbial gene expression, CRC patient groups showed weak clustering. NA, not applicable.

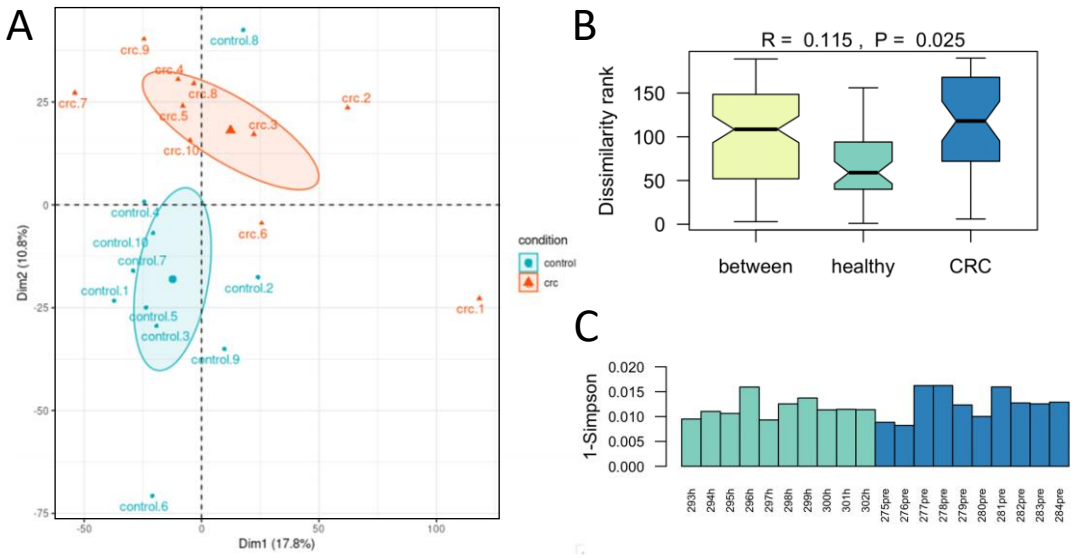


Fig. 31 | Colorectal cancer influences the microbial transcriptome. A. PCoA scatter plot of Bray-Curtis dissimilarity (beta-diversity) across dimension (Dim)1 and Dim2 of microbial transcriptome-wide gene expression. Blue circles represent control samples and red triangles represent CRC samples. Samples demonstrate clustering with respect to origin as CRC influences variation in microbial gene expression. **B.** ANOSIM of Bray-Curtis dissimilarity revealed intra-sample diversity is significantly lower than inter-sample dissimilarity between CRC and control. **C** Species alpha-diversities (a measure of biodiversity within a single sample, which quantifies the richness and evenness of different taxa present) for all 20 samples as assessed by Gini-Simpson index.

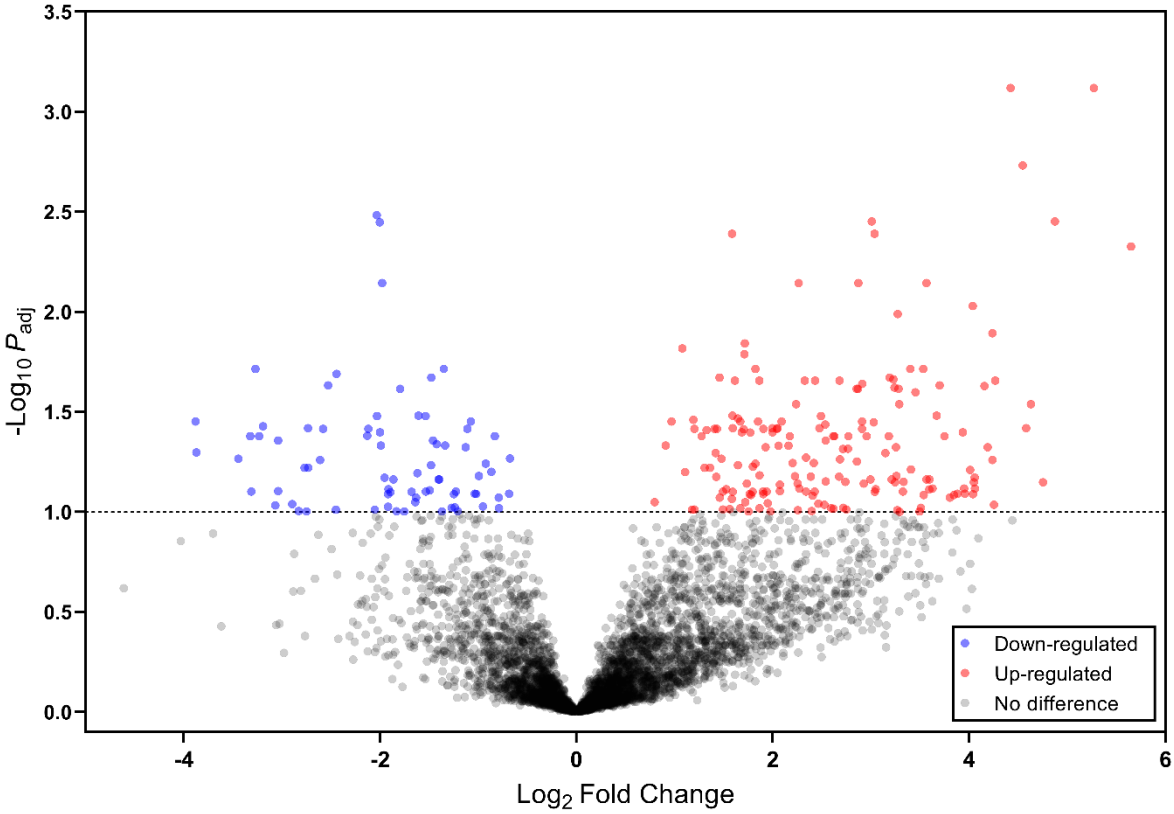


Fig. 32 | Microbial transcriptome-wide changes in gene expression during colorectal cancer. The volcano plot of -Log_{10} BH-adjusted (FDR <0.1) *P*-values for transcriptome-wide gene expression against their respective Log_2 fold changes. Blue and red dots represent significantly down and up regulated genes between CRC and control groups, respectively. Grey dots represent genes with no changes in expression. 261 of 6495 (4856 shown) genes have significantly altered expression during CRC.

6.13 Oxidative stress responses are housekeeping functions of the microbiome irrespective of gut health status

The housekeeping activity of the human gut microbiome has been studied at the genomic and transcriptomic levels in healthy adults (125), however it is yet to be elucidated in CRC. The most active subsystems, the core transcriptome (the most prevalently expressed subsystems, each constituting >1% of the total transcriptome) accounted for ~40% of total microbial activity (Fig. 33A), only one of these, pyruvate:ferrodoxin oxidoreductase that decarboxylates pyruvate to acetyl-CoA in anaerobes, showed a significant transcriptional reduction in CRC (Fig. 33B). This 'core' transcriptome appears to be responsible for housekeeping activities, biosynthesis and energy production. Interestingly, oxidative stress responses dominate, despite inflammation/oxidative stress being long considered a disease-specific phenotype. This indicates that, despite previous knowledge regarding the role of the microbiome in mediating ROS through influencing antioxidant production (542), the microbiome may play a more crucial role in mediating the level of ROS within the gut than previously thought, which has not been found previously.

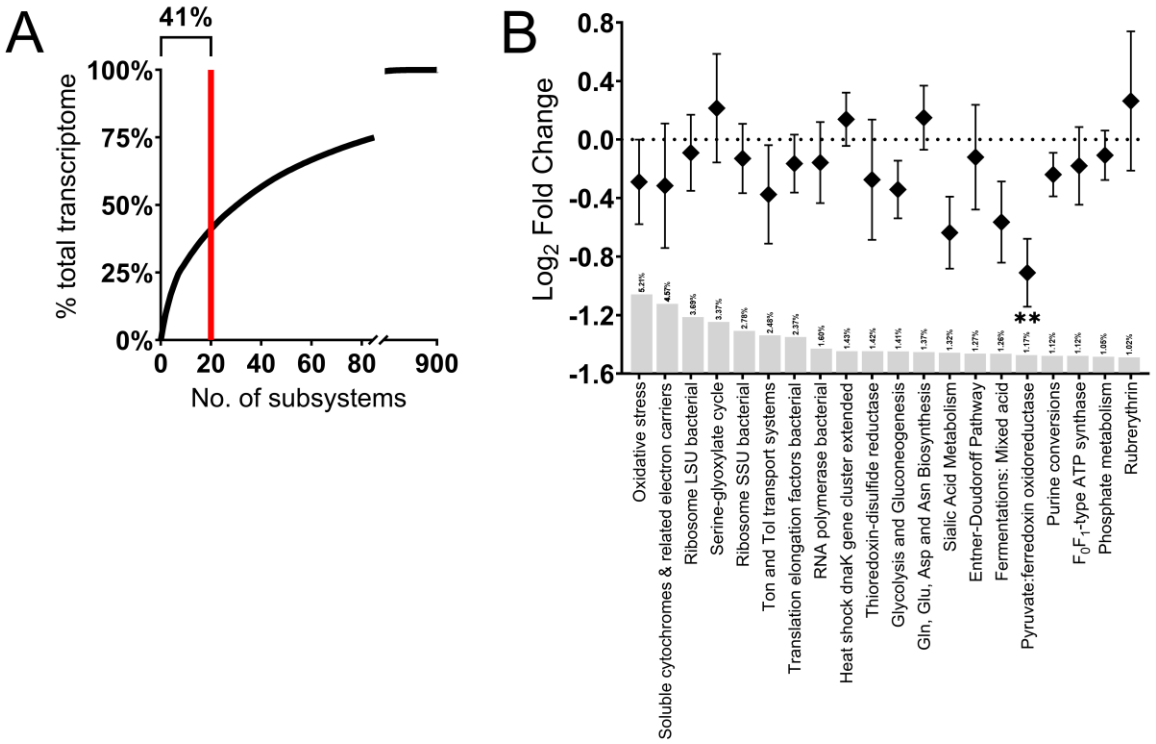


Fig. 33 | The "core" transcriptome of the gut microbiota is mostly maintained in colorectal cancer. A. Threshold of subsystems considered as the core, 20 subsystems of 900 identified contribute 41% of total transcriptome activity. Asterisks denote statistically significant differences between the health and CRC cohorts (** *P*=0.008). **B.** Metatranscriptional profile of the most prevalently expressed, 'core' subsystems which were mostly unchanged across all samples in both CRC and non-CRC cohorts. The gut microbiota generate biomass primarily through glycolysis-gluconeogenesis, the serine-glyoxylate cycle, purine metabolism, amino acids (Gln/Glu and Asn/Asp) biosynthesis and ions, vitamins and iron transport. Microbial metabolism of sialic acid, a terminal modification of host colonocytes and mucus, also appears to be a common housekeeping activity of the human gut microbiome. We also observed that oxidative stress responses (Ton and Tol transport systems, thioredoxin reduction, heat shock *dnaK* gene cluster subsystems) featured within the core transcriptome of both healthy and CRC-associated microbiota. Individual subsystem contribution to the overall transcriptome displayed as percentage above grey bars.

6.14 Gut microbiota alters the level of enzymatic and non-enzymatic antioxidative activities in CRC

The majority of CRC cases have a sporadic origin and result from gradual accumulation of somatic mutations in glandular epithelial cell DNA (543). This is attributed to the deleterious effects of ROS and reactive nitrogen species (RNS) on DNA integrity and repair. We found that microbial ROS/RNS-scavenging activities were altered in CRC. Unexpectedly, several ROS-reducing subsystems were significantly repressed in CRC (Fig. 34A). Alkyl hydroperoxide reductase and thiol oxidoreductase scavenge H_2O_2, the most potent DNA damaging agent (non-charged H_2O_2 is easily taken up by colonocytes) (228). Transcription of nine genes alongside genes involved in oxidative DNA damage responses was significantly downregulated (Fig. 34B).

Bacteria can also produce and utilise protective non-enzymatic antioxidants. We found that the ectoine biosynthesis and regulatory subsystem, which scavenges hydroxyl radicals and has anti-inflammatory activities (544) was downregulated in CRC (Fig. 34C). The L-ascorbate utilisation subsystem displayed the opposite pattern of activity, suggesting the microbiome may deplete the host of L-ascorbate. We observed upregulated transcription of the multicopper oxidase gene, involved in oxidation of different antioxidants, such as polyphenols, L-ascorbate, aromatic polyamines and metal ions. Expression of the monoamine oxidase gene, the product of which is required for oxidative deamination of monoamines such as serotonin, a neurotransmitter present in the gastrointestinal mucosa (545) was also increased. This suggests the gut microbiota can deplete and/or are depleted for secondary antioxidants during the cancer.

Significantly higher levels of glutaredoxin and glutathione reductase expression in CRC demonstrates the significant role the microbiota plays in maintaining the redox status of the cell (Fig. 34D). Additionally, expression of several reactive species scavenging genes was significantly upregulated, suggesting the CRC gut is featured with elevated $O_2^{\cdot -}$ and NO· levels. Consistent with NO· being a major RNS in CRC, primarily produced by neutrophils, expression of genes encoding for the glutathione redox cycle pathway, which senses NO· levels and some universal stress proteins (546) was increased (Fig. 34E). In the CRC gut it would appear NO· and $O_2^{\cdot -}$ are the primary radicals to which the microbiota responds to different extents.

Unexpectedly, expression of genes involved in multiple ROS reduction pathways showed equally high levels of expression in both groups (Fig. 34F). Overall, these data showed the microbial responses to $O_2^{\cdot -}$ was largely unchanged, H_2O_2 were lessened, and NO· were enhanced during CRC. This strongly implies the microbiome differentially responds

depending on the nature of the ROS/RNS as a result of the gut health status. While a high level of background ROS reduction appears to be a housekeeping feature of the gut microbiome, fluctuations in compound-specific responses may mediate potential damaging effects over time.

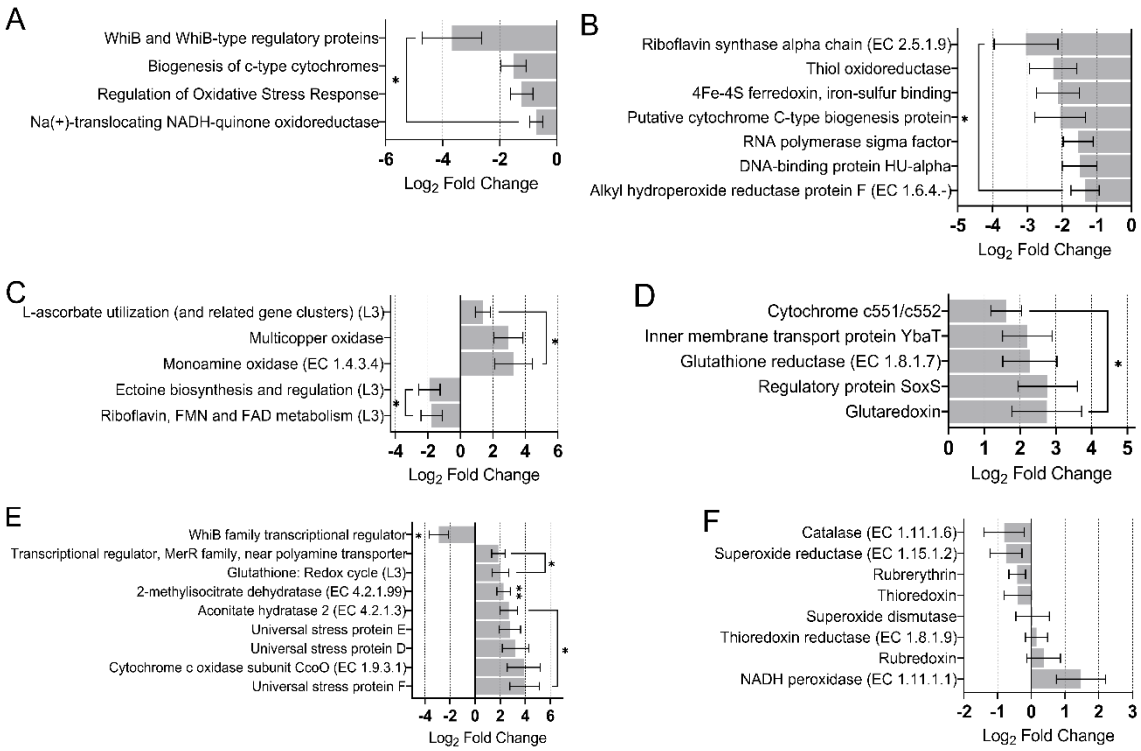


Fig. 34 | The microbiome response to H_2O_2 is diminished and to NO is increased in colorectal cancer despite high background levels of oxidative stress activities in health and disease. A. Activity of subsystems involved in modulation of oxidant levels are repressed in CRC. These subsystems involve sensors of oxidative stress (547), reduction of quinones (548), *c*-type cytochrome and the antioxidant riboflavin (vitamin B_2) synthesis (549). **B.** Expression of specific genes related to oxidative damage in CRC: expression of RNA polymerase sigma factor, a universal regulator of microbial oxidative stress response, the DNA-binding protein HU-α, a bacterial histone-like protein which displays high affinity to damaged DNA and plays a part in the oxidative DNA damage response (550), is also significantly downregulated. Expression of 4Fe-4S ferredoxin, thiol oxidoreductase and putative cytochrome *c*-type biogenesis protein genes, prominent regulators of redox status and global nitrogen and sulfur cycles, was also significantly diminished. Transcription of the riboflavin synthase and alkyl hydroperoxide reductase genes was also downregulated. **C.** The CRC gut microbiota express genes for the utilisation and oxidation of several non-enzymatic antioxidants such as ectoine and L-ascorbate. **D.** Microbiota in CRC maintains a reduced gut environment. Expression of cytochrome c551/c552 and regulatory protein SoxS, superoxide response regulon transcriptional regulator (551) was upregulated. The CRC microbiota showed a high uptake of Se (selenate and selenite), an essential element that is critical for production and activity of antioxidative selenoproteins. Selenoproteins are vital for host immunity and antiviral defence which, enhanced levels of the inner membrane transport protein YbaT and selenoproteins O synthesis have been observed (552), correlating with higher Se uptake. **E.** The CRC gut contains elevated $O_2^{\cdot-}$ and NO levels, expression of genes the activity of which is implicated in their removal was elevated. Transcription of cytochrome *c* oxidase, CcoO subunit, with high NO· reductase activity and *MerR*, a transcriptional factor which regulates NO defense (553) was significantly overactive in CRC. Synthesis of NO-induced universal stress proteins D, E and F (554)

was significantly enhanced. Aconitate hydratase 2 and 2-methylisocitrate dehydratase, the expressions of which are negatively regulated by NO· are also transcribed to a higher degree. **F.** High level of ROS-reducing activity appears to be a housekeeping characteristic of the gut microbiome. Expression of major ROS-reducing genes was maintained in a health-status independent manner. Data are normalised to gene copy number and are proportional. (L3) denotes a subsystem. * $P \leq 0.05$, ** $P \leq 0.01$.

6.15 CRC-associated microbiota depletes the host of beneficial metabolites and responds to the acidic gut environment

It has long been known that the pH of the colorectum can drop to levels as low as 2.3-3.4 during severe disease (555). Yet, the impact on microbial physiology remained unknown. We observed evidence of microbial adaptation to highly acidic conditions during CRC, at the molecular (Fig. 35A) and phylogenetic (as described above) levels. The Na^+-H^+ antiporter subsystem, which modulates H^+ potential across the bacterial membrane was downregulated, implying high extracellular proton concentrations, low pH. A gamut of 19 differentially expressed genes support this assertion. We also observed evidence that bacteria and archaea attempt to protect their cell membrane against H^+ permeability. They may reinforce it with more cyclopropane fatty acids, over-expressing S-adenosyl-L-methionine-dependent methyltransferase (SAM MTase) (Fig. 35A) and unsaturated fatty acids through 3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase (Fig. 36) found in CRC.

Iron availability and uptake has been associated with bacterial pathogenicity and is often linked to low environmental pH. Expression of the ferrous iron (Fe^{2+}) transporter EfeUOB operon, which allows uptake of the relatively soluble Fe^{2+} was elevated by the cancerous gut microbiota (Fig. 35B). However, the community downregulates its prominent non-chelating ferric iron uptake mechanism, the ferric iron ABC transporter. Despite iron uptake being conducted by a core member of the housekeeping transcriptome (Ton-Tol transport systems) certain iron acquisition mechanisms are more active in CRC, suggesting greater access to Fe^{2+}.

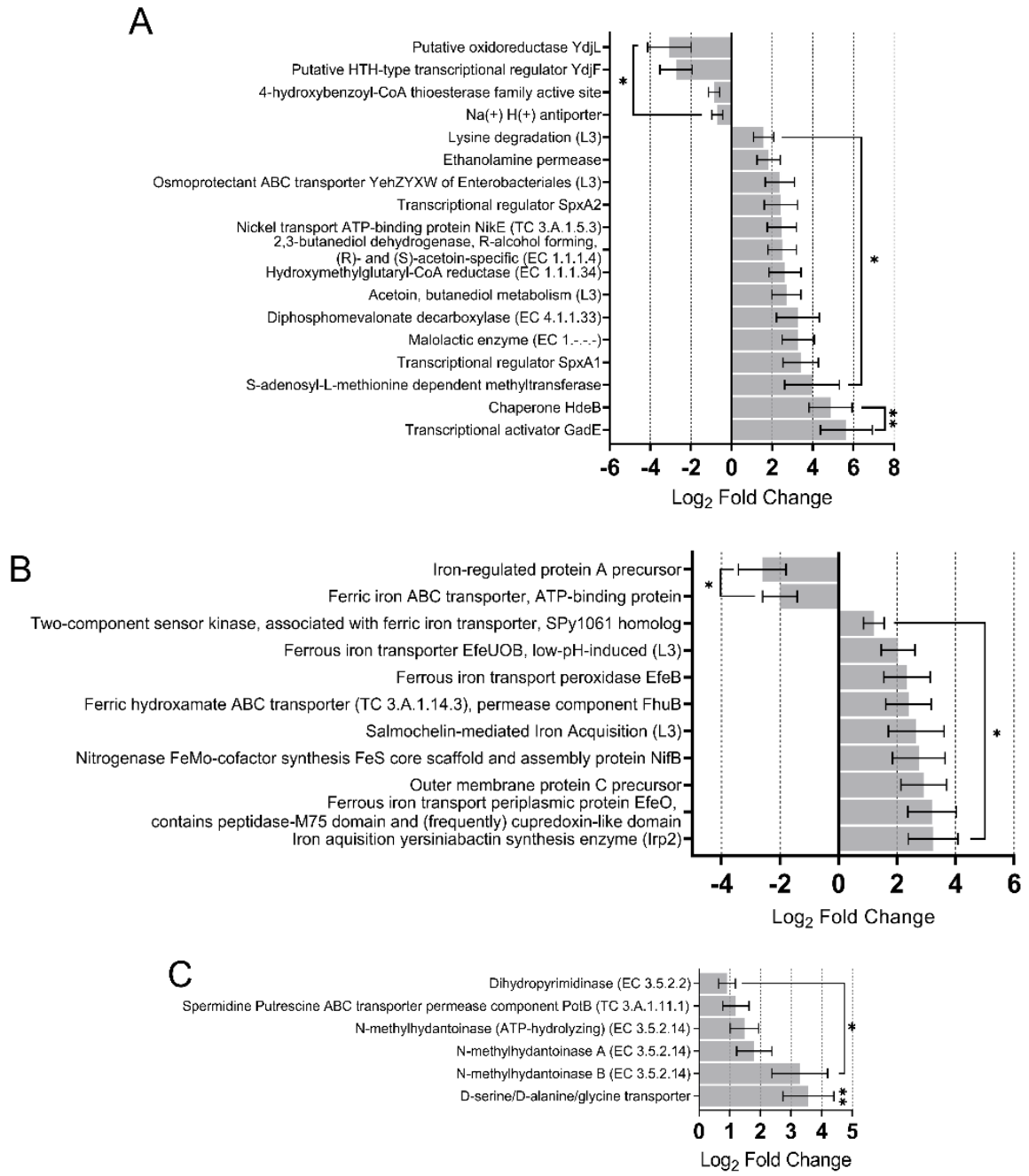


Fig. 35 | The CRC microbiome is adapted to the high acidity of the gut and metabolises host-required metabolites more readily. A. Activity of glutamate-dependent acid resistance mechanisms through transcriptional activator GadE, glutamate transport membrane-spanning protein and inner membrane transport protein YbaT (Fig. 34D), were all enhanced in CRC alongside the acid stress chaperone HdeB. Basic compounds such as ammonia (NH_3^+) can be produced by bacteria to offset low cellular pH, particularly from urea (557), the higher transcription of nickel transport ATP-binding protein NikE observed may be critical in providing the nickel for the activity of ureases that catalyse this conversion. Production of L-malate via expression of malate synthase and its conversion to L-lactate and CO_2 by malolactic enzyme was also a prominent feature of the CRC microbiome, the activity of which is triggered at pH <2.3. Levels of ethanolamine permease transcription and acid stress-induced transcriptional regulators *SpxA1* and *SpxA2*, which are virulence determinants in pathogens, were over-

represented. Conversely, alkali pH-induced genes 4-hydroxybenzoyl-CoA thioesterase and putative HTH-type transcriptional regulator *YdjF* and *YdjL* oxidoreductase genes exhibited lower expression during cancer. **B.** Iron uptake and transport related genes are upregulated by the gut microbiota in CRC. Expression of *EfeO* and *EfeB*, iron acquisition yersiniabactin synthesis enzyme, outer membrane protein C precursor, ferric hydroxamate ABC transporter (a chelating mechanism of ferric iron (Fe^{3+}) uptake) and two-component sensor kinase SPy1061 homolog that respond to iron availability and acid stress was more active. **C.** The CRC gut microbiota actively metabolise exogenous DNA. Transcription of dihydropyrimidinase, N-methylhydantoinases A and B, guanine-hypoxanthine permease, D-serine/D-alanine/glycine transporter, phage-associated cell wall hydrolase and *PotB* genes was increased in CRC. * $P \leq 0.05$, ** $P \leq 0.01$.

6.16 The CRC-associated microbiota depletes the supply of *n*-butyrate to the host

It is well established that the gut microbiota produces *n*-butyrate, a major SCFA, by fermentation of non-digestible dietary carbohydrates. This microbial-derived metabolite is the primary energy source of colonocytes which catabolise it in the mitochondria via β-oxidation, hence helping to maintain the strict anaerobic environment of the gut and promote growth of beneficial anaerobic bacteria. Unlike healthy epithelial cells, cancerous colonocytes do not utilise *n*-butyrate for their growth, instead preferring other carbon and energy sources (137). *n*-Butyrate is produced by two major bacterial metabolic pathways, via either phosphorylation of butyryl-CoA to *n*-butyrate by butyrate kinase or the butyryl-CoA:acetate CoA-transferase (acetyl-CoA:acetoacetyl-CoA transferase) route (558). The activity of butyrate kinase was comparable across cohorts (data not shown) while transcription of the acetyl-CoA:acetoacetyl-CoA transferase (β-subunit) gene was significantly increased in CRC (Fig. 36), suggesting the synthesis of *n*-butyrate from butyryl-CoA and acetate is enhanced. However, synthesis of crotonyl-CoA (from 3-hydroxybutyryl-CoA), the substrate for acetyl-CoA:acetoacetyl-CoA transferase, was significantly reduced due to repressed expression of the crotonase gene. Despite the enhanced activity of acetyl-CoA:acetoacetyl-CoA transferase, the potential depletion of crotonyl-CoA and hence, butyryl-CoA may switch the substrate specificity from butyrate production towards interconversion of acetoacetate and acetoacetyl-CoA for acetone and/or acetyl-CoA, the latter appears to be in sufficient supply. This is consistent with our recent observation that activity of 22 major butyrate producing bacteria in CRC is significantly reduced. The observed metabolic activities of the gut microbiota in CRC, with respect to the supply of beneficial micronutrients, indicates significant overall reductions. Furthermore, we found a high level of carnitine utilisation (Fig. 36) by microbiota in CRC, a further strong indication of restricted supply of the major mitochondrial fatty acid carrier which is crucial for β-oxidation, hence limiting the ability of colonocytes to utilise *n*-butyrate in CRC.

6.17 Nucleic acids are metabolised to a greater degree by the CRC microbiome

Purine utilisation by the CRC gut microbiota appears to be more pronounced (Fig. 36). This includes enhanced expression of genes which are involved in (i) uptake of guanine and hypoxanthine bases, (ii) purine degradation, (iii) cleaning up of the nucleotide pool and (iv) regulation of purine metabolism (559) This suggests the CRC microbiome salvages xanthosine/inosine nucleosides as an energy-saving alternative to maintain the integrity of the genome (e.g. during replication) may use these purines as C- and N- sources to compensate for diminished carbohydrate metabolism. Consistent with the activation of the *PurR* operon, microbial purine nucleotide synthesis seemed to be repressed during CRC (purine nucleotide synthesis regulator subsystem). However, conversion of adenine to guanine nucleotides and production of adenosines via hydrolysis of AT(D/M)P and mRNA degradation, were

downregulated in CRC, suggesting that microbes do not use an adenine/adenosine salvage pathway as a C- and energy source in the cancerous gut.

The gut microbiota supplies a range of health-maintaining essential metabolites to the host. For example, carnitine, that delivers fatty acids for β-oxidation to the mitochondria for energy production. Increased transcription of microbial genes involved in carnitine and selenium (Se) uptake and catabolism was observed in CRC (Fig. 36). Moreover, the pattern of expression of *n*-butyrate synthesising genes indicates a switch in substrate specificity to favour acetone production from acetoacetyl-CoA suggesting limited supply for the host. We have shown the attenuated activity of 22 *n*-butyrate-producing species in the CRC microbiome (see above), which corroborates these data. Carnitine is also important for osmotic adaptation of the microbiota, suggesting they are under increased osmotic stress. Microbial uptake of Queuosine (Q), a precursor base of modified Q_{34}-tRNA in bacteria and eukaryotes, critical for translation fidelity, a contributor to human health (560) was elevated in CRC. The microbiota also reduced transcription of genes implicated in ceramide production, an apoptosis activator and enhanced vitamin B_2, B_6 and B_{12} uptake, further suggesting depletion of the host for beneficial compounds.

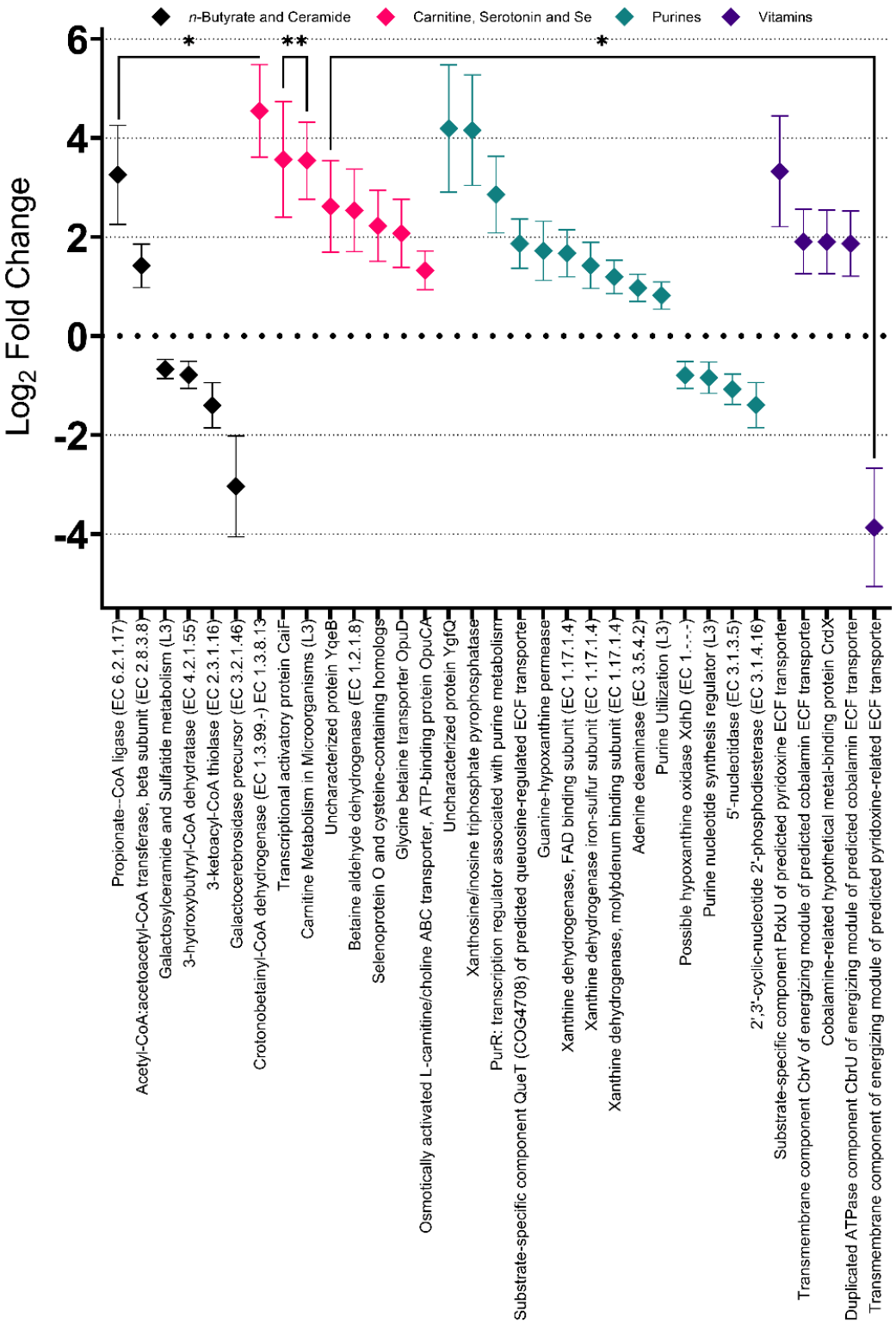


Fig. 36 CRC microbiota depletes the host of beneficial metabolites and micronutrients. The CRC-associated microbiota transcribes *n*-butyrate synthesising acetyl-CoA:acetoacetyl-CoA transferase more readily but 3-hydroxybutyryl-CoA dehydratase and 3-ketoacyl-CoA thiolase to a lesser extent. Betaine aldehyde dehydrogenase, glycine betaine specific transporter OpuD and the precursor for non-specific OmpC, which are preferentially expressed at low pH and in high osmolarity, were overexpressed alongside the subsystem for carnitine metabolism. Upregulated transcription of *CaiF*,

CaiA and *OpuCA* (the osmotically activated L-carnitine/choline ABC transporter) are evidence of higher carnitine metabolism rates by the microbiome. The biosynthesis and uptake of glycine betaine are also more pronounced in CRC. Enterobacteriales upregulated expression of the osmoprotectant ABC transporter *YehZYXW* (Fig. 35A), the pleotropic role of which is important for the accumulation of the glycine betaine osmolyte under hyperosmotic stress, nutrient starvation and acidic environments (561). The bacterial production of the sphingolipid, ceramide and the galactocerebrosidase precursor protein in CRC were downregulated. This may lead to a reduced supply of ceramide to the host, hence attenuating apoptosis, prolonging cell survival, and enhancing tumour growth and glucose metabolism (562). Transcription of the *Cbr* modules of the predicted cobalamin ECF transporter was more active in cancer along with expression of the *CrdX* gene, a cobalamine-related hypothetical metal-binding protein. Expression of genes which respond to purine availability to repress purine biosynthesis, was enhanced in CRC. Levels of hypoxanthine oxidase XdhD, 5'-nucleotidase and 2',3'-cyclic-nucleotide 2'-phosphodiesterase transcription, which produce adenosines were diminished. Genes implicated in regulating microbial vitamin B_2 (Fig. 34B and 34C), B_6 and B_{12} and their derivatives were differentially expressed. Three predicted pyridoxine and cobalamin ECF transporters and cobalamin metal-binding protein CrdX were expressed to a higher degree during CRC. Black diamonds, *n*-Butyrate and ceramide; Pink diamonds, Carnitine, serotonin and Se; Green diamonds, Purines; Purple diamonds, Vitamins.

The CRC-associated microbiota activated expression of hydantoin uptake and metabolism genes, the products of oxidation of cytosine and thymine bases of dead cell DNA (563) (Fig. 35C). Lysis of bacterial cells due to higher bacteriophages activity and biofilm formation (564) in CRC may in part explain the availability of exogenous hydantoins and purines. Cell-free (cf)DNA in the gut may also be available from the accelerated death of tumour and immune cells (565). Additionally, transport of spermidine and putrescine (biogenic amines, products of fatty and amino acid breakdown from decaying cells/tissues) was also significantly increased. We observed xanthine/xanthosine, inosine and guanine metabolising genes were upregulated by the microbiome, a source of microbial ROS, while adenine/adenosine salvage genes were downregulated (Fig. 36). We found expression of a number of genes involved in carbohydrate metabolism was diminished in CRC as opposed to activities for utilisation of amino acid and aromatic compounds (Fig. 37A and 37B). This suggests a switch from carbohydrates in health to amino acids and aromatic compounds metabolism in CRC. Archaeal methanogenesis activities and expression of microbial genes for biosynthetic pathways were enhanced in CRC (Fig. 38A and 38B).

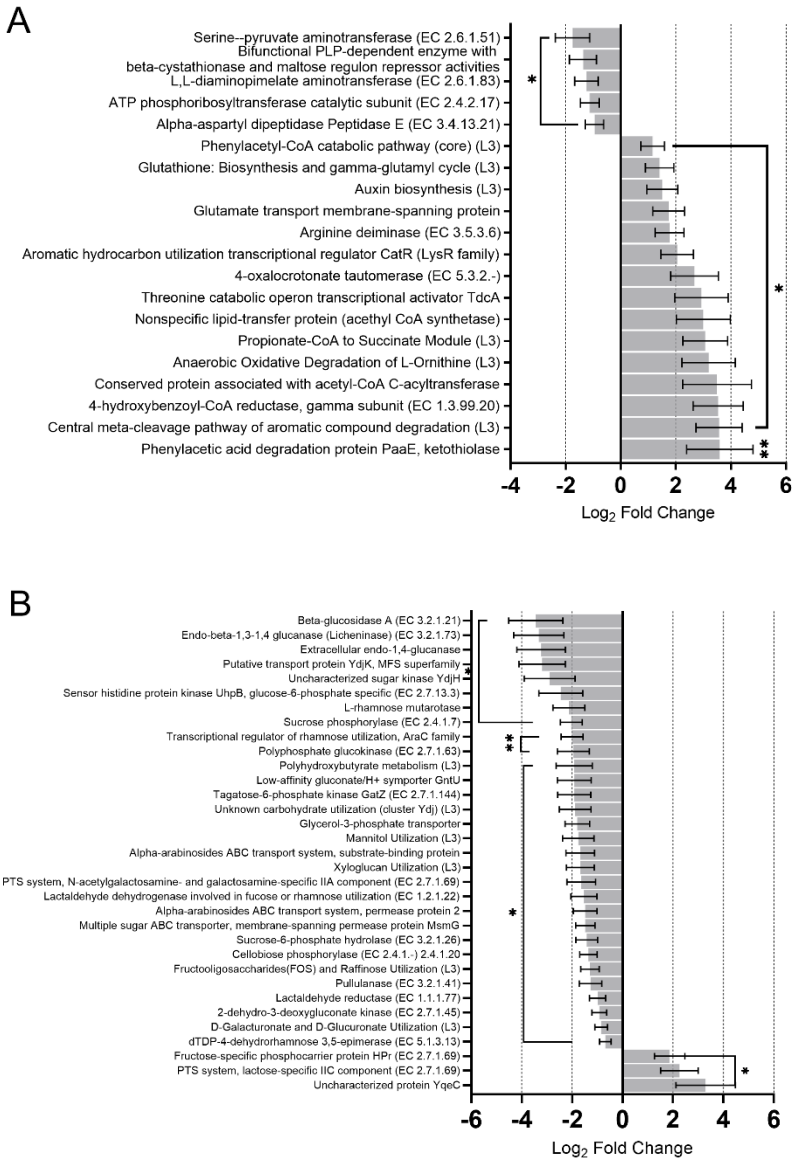


Fig. 37 | Microbial metabolic activity switches from carbohydrate utilisation in health to amino acids, aromatic compounds and nucleic acids metabolism in CRC. A. Anaerobic oxidative degradation of L-ornithine is upregulated in CRC alongside degradation of branched-chain amino acids (Iso, Val and Leu). Microbial expression of L-lysine catabolism, phenylacetyl-CoA catabolic pathway, auxin biosynthesis, glutathione: biosynthesis and gamma-glutamyl cycle subsystems was augmented. Transcription of glutamate transport membrane-spanning protein and genes involved in amino acid degradation, 2-oxoglutarate dehydrogenase, arginine deiminase and threonine catabolic operon transcriptional activator TdcA, S-adenosyl-L-methionine dependent methyltransferase was enhanced in cancer. PaaE ketothiolase (involved in phenylacetic acid degradation in aerobic conditions) was also expressed to a greater extent. **B.** Expression of 30 genes which are involved in metabolism of carbohydrates was significantly attenuated in cancer while expression of fructose metabolising and transporting genes was elevated.

6.18 The CRC microbiota adjusts their respiratory pathways to utilise amino acids and aromatic compounds, alters H^+ transport in the low pH milieu and diverts resources towards biosynthesis

Microbes have evolved several mechanisms to utilise a wide variety of carbon and nitrogen sources, often cross-feeding on the products of metabolism of other species. Our analysis shows dynamic metabolism of amino acids and carbohydrates. Anaerobic oxidative degradation of L-ornithine, producing acetate, alanine and tumour-promoting ammonia (important for bacterial colonisation) (566) appears to be a signature of the CRC microbiota (Fig. 40A). However, anaerobic metabolism of complex carbohydrates such as D-galacturonate and D-glucuronate derived from pectin and other heteropolysaccharide was repressed. Upregulation of archaeal methanogenesis genes, the anaerobic endpoint of fermentation, the acetoin, butanediol metabolism module suggests that some members of the microbial community grow under hypoxia and use fermentation for energy production to a greater degree. We found that expression of several genes that support aerobic metabolism under oxygen-limited settings, in other words, associated with growth under microaerobic conditions, was enhanced in CRC.

A recent large meta-analysis of CRC gut metagenomes has revealed a significant enrichment in pathways of amino acid degradation alongside a depletion of genes for carbohydrate utilisation (222) We found that the activity of these pathways reflected the reported altered metagenome abundances (Fig. 40, Fig. 37A and Fig. 37B). Interestingly, enhanced degradation of some amino acids may serve another adaptive purpose, namely protection of microorganisms against low environmental pH. An acidic environment can favour the growth of some archaea, and we saw a host of pathways involved in the biosynthesis of archaeal ribosomes, coenzyme M and B and methanopterins which were more active in the CRC gut (Fig. 38A). These data coupled with methanogenesis from methylated compounds and formaldehyde assimilation subsystems and archaeal vitamin B_2 synthesis upregulation, all point to a cancerous gut with a flourishing archaeome. Methanogenesis, the anaerobic endpoint of fermentation, the process of reducing CO_2 into CH_4, methane, in the presence of H_2 for the purpose of energy production is a prominent CRC signature. We also observed enhanced utilisation of aromatic compounds as the primary C and N sources, shown through CRC-dependent expression of e.g. 4-oxalocrotonate tautomerase involved in the central meta-cleavage pathway for aromatic compound degradation, a subsystem with augmented activity alongside the phenylacetyl-CoA catabolic pathway (Fig. 37A). Degradation of branched amino acids (as showed above) to propionate-CoA and L-lysine catabolism was also a distinct signature of the CRC microbiota as seen through upregulation of both the propionate-CoA to succinate module (Fig. 37A) and lysine degradation (Fig. 35A) subsystems.

Transcription of glutathione metabolising genes, involved in amino acid transport as part of the γ-glutamyl cycle (567) were amplified. Notably, enhanced amino acid uptake by CRC-associated microbiota coincided with a high level of acidity. Expression of acid-induced amino acid transport genes and inner membrane transport protein YbaT (Fig. 34D) and glutamate/aspartate uptake were all up-regulated (Fig. 40A). Transcription of genes needed for degradation of Arg, Ser/Thr metabolism and Phe catabolism were also elevated. Trp degradation also appeared to be promoted, as expression of the auxin biosynthesis subsystem and monoamine oxidase gene was greater in CRC. Contrarily, activity of His, Lys and D-amino acid biosynthesis genes and α-aspartyl dipeptidase peptidase E was significantly reduced in CRC.

Such high catabolic activity of amino acids in CRC may lead to the limitation of e.g. Ser availability. Transcription of the serine-pyruvate aminotransferase gene was decreased (Fig. 37A) while expression of the GTP pyrophosphokinase protein which catalyses formation of pppGpp, the precursor of ppGpp, an alarmone active in response to amino acid starvation, (568) was enhanced (Fig. 38B). The observed amino acid starvation in turn may trigger the elevated ribosomal hibernation response observed, through production of membrane protein YqjD and ribosome hibernation protein YfiA. This starvation response may also be, at least in part a cause of the dormancy of certain species.

Interestingly, microorganisms in CRC activated expression of glutamine synthetase, a member of the ubiquitous pathway for ammonium assimilation (569). Downregulation of the structural gene for nitrogen regulatory protein P-II, that regulates the catalytic activity and transcription of glutamine synthetase was observed in CRC (Fig. 38B). This suggests that glutamine synthetase-dependent N-assimilation from ammonium is a preferred nitrogen source in CRC to support high cell growth under nitrogen-limited conditions. By contrast, the CRC gut appeared enriched with phosphate as expression of *PstS* (a phosphate ABC transporter) was repressed, its transcription is known to be negatively regulated by the availability of phosphate.

Genes encoding multiple sugar ABC transporters, as well as lactaldehyde reductase and dehydrogenase responsible for L-fucose and L-rhamnose degradation were expressed to a lesser extent in CRC. We found that expression of 24 genes, activities of which are required for carbohydrate transport and metabolism, was negatively correlated with CRC (Fig. 37B). Interestingly however, genes involved in fructose metabolism (described above) exhibit enhanced transcription, required for nucleic acid biosynthesis, consistent with increased biosynthetic metabolism.

Transcription of a key β-oxidation of fatty acids enzyme 3-ketoacyl-CoA thiolase was downregulated by the CRC microbiota (Fig. 36), suggesting the catabolism of amino acids and not β-oxidation of fatty acids was a primary source of acetyl/propionate-CoA. The two decarboxylating steps (isocitrate → 2-oxoglutarate → succinyl-CoA) of the TCA cycle can be circumvented via the glyoxylate bypass, this shunt is preferred upon a metabolic switch from energy to biomass producing activity and was more active in CRC (Fig. 38B). The preferred glyoxylate bypass in accordance with the anabolic pentose phosphate pathway argues that the microbiota has high biosynthetic demands during colorectal cancer, potentially facilitated by their access to colonocytes triggering microbial proliferation.

Succinate, the major substrate for oxidative phosphorylation during respiration can be produced not only through the TCA cycle and glyoxylate bypass, but also through the γ-aminobutyrate (GABA) shunt, converting 2-oxoglutarate to succinate, bypassing succinyl-CoA and NADH production. This TCA cycle re-entry point was potentially utilised more readily by the gut microbiota in cancer, via lactam utilisation protein, LamB (Fig. 38B). Once more, the propionate-CoA to succinate module demonstrates elevated activity, supplying succinate. However, dietary carboxylic sugar, *myo*-inositol, can also be catabolised to e.g. propionate, acetate, CO_2, acetaldehyde and succinate by commensal microbes (570) which also displayed upregulation through diminished expression of the transcriptional repressor of the *myo*-inositol catabolic operon *DeoR*. These data argues that there is an abundance of succinate, supplied more actively by four routes, GABA and glyoxylate shunts, the propionate-CoA to succinate module and *myo*-inositol catabolism, which the CRC microbiota can utilise to provide electrons to the electron transport chain in respiration during cancer and satiate their biosynthetic requirements.

The observed upregulation of the glyoxylate bypass would result in less NADH, one of two electron donors required for Complex I to catalyse electron transfer to ubiquinone. This, coupled with Na(+)-translocating NADH-quinone reductase subunit E and the FMN-producing pathway under-transcription (Fig. 38B) by the cancerous microbiome, suggests Complex I of the electron transport chain has a diminished role in cellular respiration during cancer. However, transcription of genes encoding for enzymes which carry out electron transfer from Complex II through to Complex V was higher in CRC. Complex II, where multiple component-encoding genes display elevated expression, alongside ample succinate availability suggests Complex II is the entry point for the generally more active electron transport chain in CRC.

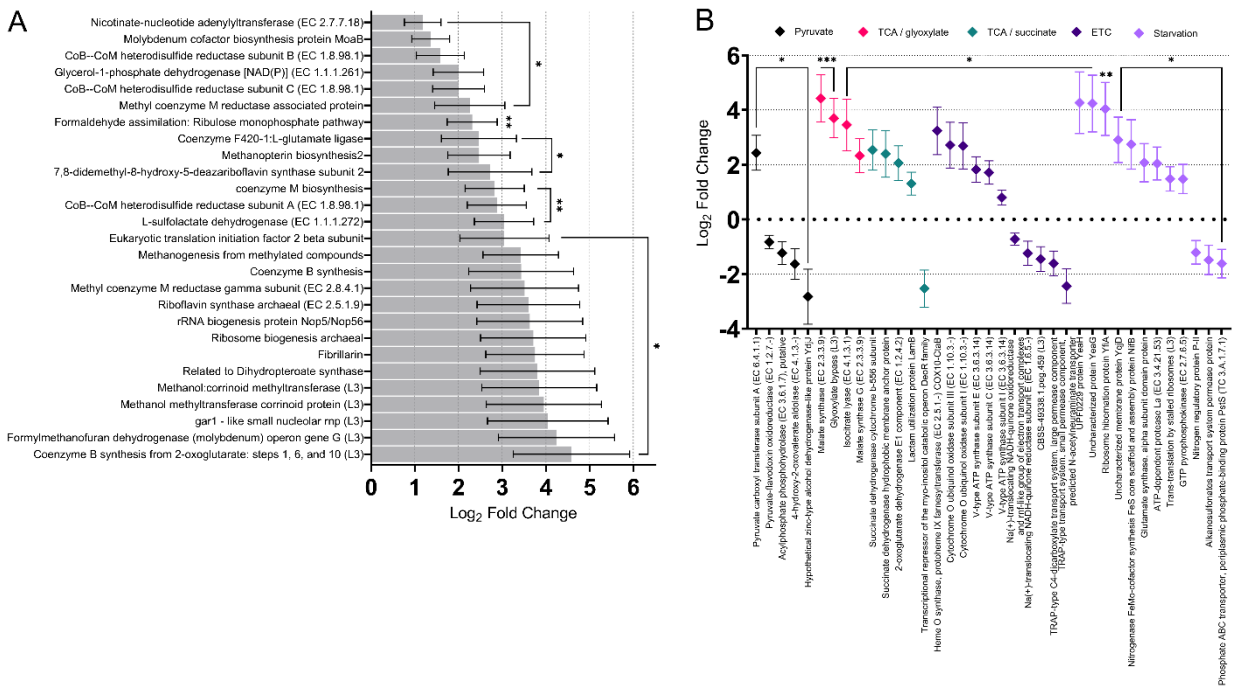


Fig. 38 | Methanogenic and biosynthetic metabolic features of the CRC microbiome coincide with local starvation. A. Archaeal growth and methanogenesis related gene expression in the CRC gut was enhanced. Coenzymes M, B and F_{420} synthesis displayed amplified transcription along with methylotrophic pathway components methanol:corrinoid methyltransferase and methanol methyltransferase corrinoid protein genes, all crucial components in CH_4 production by methanogenic microorganisms. Fibrillarin, Nop5/Nop56 and translation initiation factor 2β, glycerol-1-phosphate dehydrogenase [NAD(P)], hydroxymethylglutaryl-CoA reductase, and archaeal riboflavin synthase genes, which synthesise cell wall components, all displayed increased expression in the cancerous gut. **B.** Pyruvate metabolism associated gene expression was repressed through weakened microbial transcription of the pyruvate:ferredoxin subsystem (Fig. 33A) and pyruvate-flavodoxin oxidoreductase contained therein, required for aerobic acetyl-CoA production. Transcription of biosynthetic glyoxylate bypass genes, circumventing the decarboxylating steps of the tricarboxylate cycle, and the whole subsystem itself was upregulated, including, aconitate hydratase 2 (Fig. 34E), isocitrate lyase and malate synthases. Succinate dehydrogenase subunits of respiratory complex II that supply electrons to the respiratory chain, 2-oxoglutarate dehydrogenase and LamB all responsible for succinate metabolism were upregulated by the microbiome, in conjunction with downregulation of the transcriptional repressor of the *myo*-inositol catabolic operon which can also be utilised in succinate synthesis. Genes encoding complex III- (cytochrome o ubiquinol oxidase, subunits I and III), IV- (cytochrome c oxidase and heme O synthase) and V- (V-type ATP synthase, subunits C, E, I) related components to reduce oxygen to H_2O were upregulated. Complex I, Na(+)-translocating NADH-quinone reductase subunit E and the FMN-producing pathway were less active in CRC. Gene expression in response to nutrient starvation, e.g. amino acid, such as ribosome hibernation protein YfiA was a feature of the CRC microbiome. The microbiota also enhanced expression of the *YeaH* and *YeaG*, which are implicated in nitrogen starvation responses. The level of transcripts for aromatic hydrocarbon utilisation transcriptional regulator CatR, hypothetical nudix hydrolase YeaB, *La* protease and glutamate

synthase, alpha subunit were increased in malignancy. Black diamonds, Pyruvate; Pink diamonds, TCA/glyoxylate; Green diamonds, TCA/succinate; Purple diamonds, Electron transport chain (ETC); Lilac diamonds, Starvation.

6.19 Amino acid-dependent acid resistance mechanisms of aerobic microbes derived from CRC and healthy guts are regulated differently

Microbial RNA-seq data argues that the CRC gut environment is more acidic (section 6.15), fluctuates in osmotic potential (section 6.17) and is less saturated with H_2O_2 (section 6.13) compared to the control gut. The aerobic microbial populations of both conditions, grown in LB over 24 hours until stationary phase, were enriched with 60-70% of *E. coli* based on rDNA-seq profiling (Dataset 1, Tab 3 of (526)) and they are known to be highly resistant to acidic conditions (571) and can survive in the mammalian stomach (572). Hence, we tested whether acidity and other environmental factors (osmotic and oxidative pressures) regulate *E. coli* Arg- and Lys-dependent acid resistance systems by quantifying the expression of amino acid decarboxylases, *speA* (459), *adiA* (460) and *cadA* (461) (Table 4 and Fig. 39).

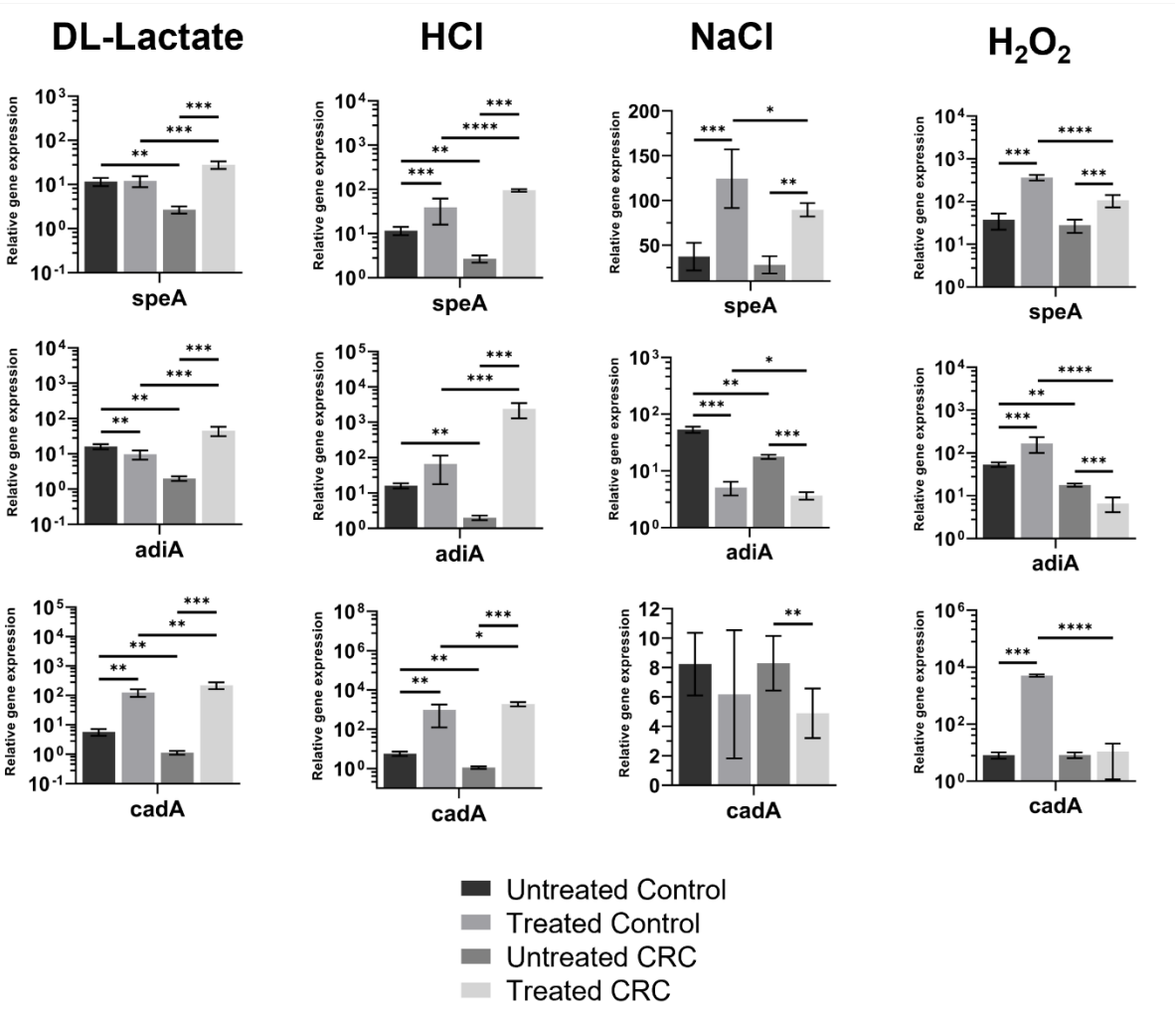


Fig. 39 | Amino acid Arg- and Lys-dependent acid defense mechanisms in *E. coli* are regulated by environmental factors and the health status origin of bacteria. The level of expression of acid resistance genes, *speA* (Arg-decarboxylase), *adiA* (biosynthetic Arg-decarboxylase) and *cadA* (Lys-decarboxylase) of *E. coli* grown at 37°C in LB overnight before 4 hours of exposure to environmental stresses, acid (pH 5.8 to 3.5 adjusted by either DL-lactate or HCl), osmotic (5% NaCl) and oxidative (1.5 mM H_2O_2) pressures. Microbiota were purified from n=10 CRC (Untreated CRC and Treated CRC) and n=10 control (Untreated Control and Treated Control) samples and pooled (in equal volume) with respect to group. Gene expression was quantified by qRT-PCR. Error bars denote standard deviation (Treated n=9 and Untreated n=6), asterisks represent statistically significance, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$.

Expression of *speA* (pH independent Arg-decarboxylase) was positively regulated by *E. coli* in response to all four growth conditions irrespectively of the health status of the host, except for DL-lactate in the non-cancerous samples. This indicates *E. coli* of the gut microbiota activate the SpeA resistance pathway in response to acid and non-acid pressures primarily in a health independent manner. This i) shows that *speA* transcription is activated in response to salt, oxidative and inorganic acid pressures irrespectively of health status but to lactate-based acidity in a health-dependent way and ii) suggests that gene expression of individual bacteria (such as *E. coli*) of the gut microbiota is regulated by specific pressures and dependent on the health status of the host. Together, this suggests that the SpeA response represents a broad-spectrum stress protection mechanism of the aerobic gut microbiota.

Expression of *adiA* (biosynthetic Arg-decarboxylase) was enhanced in CRC-derived *E. coli* regardless of the nature of acidity as opposed to both osmotic and oxidative pressures that downregulated *adiA* expression. In contrast, a mixed picture (downregulation by lactate and salt, upregulation by H_2O_2 and no effect by HCl) was observed for control cultures on transcription of *adiA*. This is consistent with the health status of the host being a key mediator of the AdiA-dependent acid stress response mechanism to all but osmotic pressure. Thus, expression of the *E. coli* AdiA Arg-dependent acid resistance system is differentially regulated by different environmental factors in a host health status-dependent manner.

Expression of *cadA* (pH inducible Lys-decarboxylase) was positively regulated by either acidic condition regardless of health status, showing that the CadA Lys-dependent acid resistance mechanism operates independently of host health. Osmotic pressure however inhibited its expression in only CRC *E. coli* while this enhanced expression of the gene in the control culture in the presence of H_2O_2, demonstrating that salt and oxidative pressures regulate *cadA* transcription in a manner influenced by host health. Hence, the CadA Lys-dependent acid resistance mechanism is activated in response to acidity in a health independent manner while playing a role in protection of bacteria of the healthy but not cancerous gut against oxidative stress.

Both Arg and Lys-dependent acid resistance mechanisms were positively regulated under acid conditions in CRC-derived *E. coli* regardless of the nature of the acid. However, the Lys-dependent acid resistance mechanism, unlike the Arg-dependent systems, responded in a health-independent manner. It appears that *E. coli* originating from different microbiomes responds differentially to the same acid stresses. Both amino acid-dependent acid defence systems sensed the oxidative pressure in a health dependent manner, while the SpeA Arg-dependent sub-system responded irrespectively of the origin of *E. coli*. Osmotic stress elicited

opposite patterns of Arg-dependent system regulation and was not influenced by health status, while the Lys-dependent mechanism displayed health-dependent regulation. *In fine*, this argues that the gut microbiome, at least its aerobic population, respond to the same environmental pressures in a unique fashion depending upon their native gut environment, be it CRC affected or healthy.

6.20 CRC-associated microbiome expresses a plethora of virulence and colonisation factors

The gut microbiota is a known complex microbial community containing commensal and potential pathogenic microbes, many of which favour colonisation as an adaptation to the gut environment. This requires bacteria to express numerous factors, including those important for adhesion and virulence to ensure colonisation and survival. However, it is not known if the health status of the gut can influence activity of potential pathogens and regulate expression of adhesion and virulence genes.

Taxonomic analysis of this microbiome revealed elevated activity of oral cavity, Enterobacteriaceae, ESKAPE and other clinically relevant pathogenic species (see above). The same communities displayed enhanced activity of numerous specific virulence determinants. The CRC microbiota transcribed exopolysaccharide, heteropolysaccharide and capsular polysaccharide biosynthesis genes more readily (Fig. 40A and 40C). This suggests Gram-positive microbes in CRC can colonise the mucosal surface and evade opsonophagocytosis more effectively (573). Furthermore, expression of lipoteichoic acid polymer forming genes (574) were also promoted in CRC, the produced polymer is anchored to the bacterial membrane and involved in adherence of e.g. group A streptococci (575). Improved adhesion properties of the microbiota in CRC were evidenced by enhanced activity of type-1 pili and the adherence of *Enterobacteria* subsystems and a number of adhesion-related genes (Fig. 40A). Interestingly, expressions of the *VgrG* gene, a component of the type VI secretion system (576) and *YdjG*, hypothetical oxidoreductase, which is required for *E. coli* colonisation (577) was downregulated. This suggests that adhesion/colonisation of some pathogenic Proteobacteria, such as *Pseudomonas*, *Escherichia, Klebsiella, Burkholderia* and *Acinetobacter*, does not involve a phage-like secretion mechanism.

We observed enhanced activity of salmochelin-mediated iron acquisition subsystem (Fig. 35B). Salmochelin has been shown to promote both pathogenic *E. coli* colonisation and biofilm formation *in vivo* (578). Increased production of lipopolysaccharide by Gram-negative bacteria in CRC was evident through an increase of pseudaminic acid biosynthesis gene expression (Fig. 40C), that helps microorganisms evade the host immune system (579). Consistent with both enhanced adhesive properties, biofilm formation associated gene

transcription is also promoted. Upregulation of curli production (Fig. 40A), amyloid fibres which form the extracellular biofilm matrix is a signature of the CRC-associated microbiota. It appears that sporulation activities of microbes in CRC were diminished while the early-stage germination seemed to be increased (Fig. 42).

The bacterial ability to perform curli-mediated adherence is inversely coordinated with their motility (580) required for movement and adhesion to the mucosa or epithelium (581). Motility and chemotaxis of the microbiota in CRC was repressed (Fig. 40B). These data argue that colonised and clustered microorganisms are a potential signature of CRC. A key microbial gene, β-galactosidase which is involved in degradation of mucus was downregulated in CRC (Fig. 40A), consistent with the notion that during CRC the microbiota has already colonised the host epithelium and to a lesser extent the mucus.

One of many mechanisms which allows the microbiome to adapt to environmental changes (15) is the HT of genetic information. HT facilitates the creation of a diverse and fluctuating array of genetic combinations often enforced by selective pressures. Conjugation, which requires cell-to-cell interaction (582), in *Enterococcus* and other Firmicutes was upregulated (Fig. 40E). HT in Gram-positive bacteria (competence in *Streptococci* and sex pheromones in *Enterococcus faecalis* and other Firmicutes) (Fig. 40B) and DNA repair (CBSS-214092.1.peg.3450 and EC699-706) were all enhanced in CRC (Fig. 40D). Two antiviral defence mechanisms, CRISPR-Cas (the adaptive microbial immune system, CBSS-216592.1.peg.3534) and group II intron-associated genes (preventing phage propagation through the microbial population at the expense of infected microbes, termed abortive infection) (583) were upregulated in CRC (Fig. 40D). At the transcriptional level, CRISPR Type III system was upregulated while the *E. coli* CRISPR subtype I-E was downregulated. We found that DNA repair was augmented in CRC, including ATP-dependent DNA ligase, also crucial for DNA replication and recombination.

Antimicrobial resistance has been a major health-related concern for decades and subjecting microbial communities to antibiotic pressure plays a major role in the development and spread of these determinants (108). Surprisingly, we found that the gut microbiota of the CRC cohort (who were not subjected to antibiotic treatment in at least the two months preceding sample collection) displayed the potential for a multi-drug resistant phenotype (Fig. 40E) via induced expression of *MarB* (584). We also observed enhanced expression of genes conferring resistance to vancomycin and β-lactams. These data demonstrates that the CRC gut can promote expression of antibiotic resistance determinants, this may be due to the enhanced activity of microbes carrying antibiotic resistance, including ESKAPE and Enterobacteriaceae species.

Furthermore, the CRC microbiota showed a significantly higher transcription of bacitracin transport genes, cyclic peptide antibiotics that disrupt Gram-positive cell wall synthesis. Production of microcin B17, a peptide toxin that causes microbial dsDNA breaks (585), and fosfomycin, which interferes with cell wall biosynthesis was also a CRC signature. Interestingly, fosfomycin acts against methicillin- and vancomycin-resistant Enterobacteriaceae pathogens with increased susceptibility to the antibiotic in an acidic environment (586), a feature of the malignant gut. The higher production of microbiota-derived antimicrobials suggests increased competition between microorganisms in CRC.

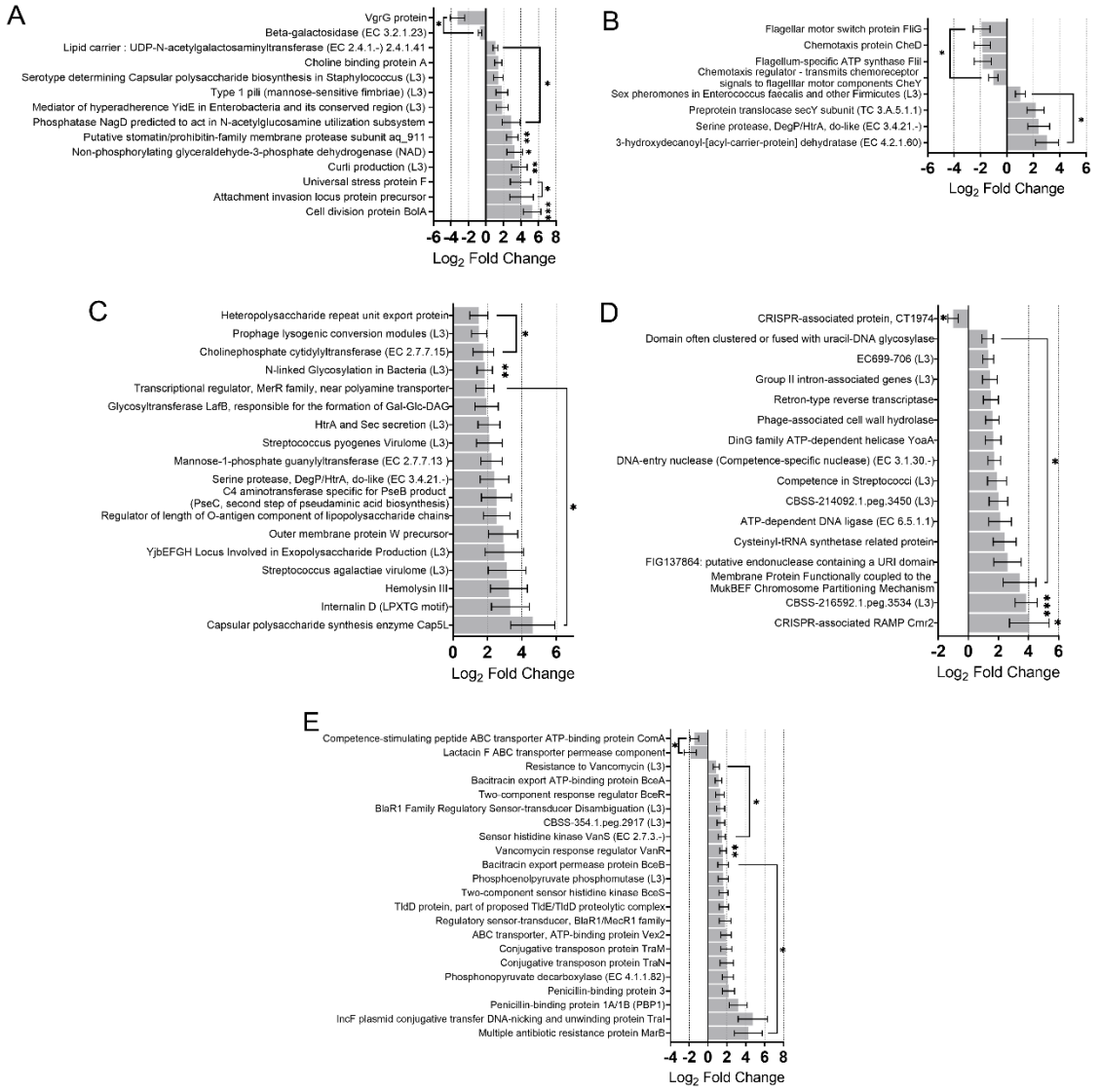


Fig. 40 | The microbiome in colorectal cancer colonises the host and form biofilms, exchanges DNA and overexpresses numerous virulence determinants. A. Transcription of genes which are important for colonisation, flagellin and pilin modifications and the formation/remodelling of the cell wall (587) was elevated in the CRC microbiome. Higher transcription of *BolA* and the curli production subsystem (which play roles in biofilm formation) and lower transcription of the possible hypoxanthine oxidase *XdhD* and the bifunctional PLP-dependent enzyme with β-cystathionase and maltose regulon repressor activities (which facilitate biofilm disassembly) suggest increased biofilm formation in the CRC-associated microbiome. **B.** Quorum sensing (QS) and motility were differentially regulated in CRC. Gram-negative QS-associated genes were overrepresented in CRC, expression of the *secY* gene, translocase and *DegP*/*HtrA* serine proteases were higher in CRC. Gram-positive QS mechanisms were however attenuated in cancer. Transcription of several chemotaxis and flagellar production/function genes (*CheY*, *FliI*, *FliG* and *CheD*) was reduced in the CRC niche. **C**. The CRC microbiome activate expression of virulence factors. Production of capsular polysaccharide synthesis enzyme Cap5L, heteropolysaccharide repeat unit export protein, *Irp2* which encodes the iron acquisition yersiniabactin

synthesis enzyme (Fig. 35B), hemolysin III and the LPXTG-containing motif Internalin D was increased. Expression of R-alcohol forming, (R)- and (S)-acetoin-specific 2,3-butanediol dehydrogenase (Fig. 35A), which reduces acetoin to 2,3-butanediol, was enhanced in CRC, suggesting a potentially high supply of acetoin, promoting a pro-cancerous phenotype of the CRC-specific microbiota. **D.** The CRC gut microbiota are prone to the exchange of genetic information, protective against pervasive bacteriophages and repair errors in their genome. Transcription of a DNA-entry nuclease (a competence-specific nuclease) was increased in CRC. Expression of the CRISPR-associated RAMP *Cmr2* gene, a part of the Type III system, and retron-type reverse transcriptase was amplified. Yet, transcription of the CRISPR-associated protein CT1974, a member of the CRISPR subtype I-E of *E. coli* (588) was decreased. There was increase in transcription of genes for helicase YoaA (involved in the repair of replication forks), domain clustered with uracil-DNA glycosylase and FIG137864:putative endonuclease domain (involved in releasing damaged pyrimidines from dsDNA). Higher expression of cysteinyl-tRNA synthetase related protein in CRC suggests that the RecA-mediated recombinational repair mechanism and hence the SOS response was increased under cancerous conditions. **E.** Antibiotic resistance activities of the microbiome are positively regulated in CRC. Increased transcription of the two-component regulatory system VanR/VanS (589), which senses either the presence of extracellular vancomycin and/or cell wall disruption by e.g. bacitracin, was observed. *Vex2*, encoding an ATP transporter which is important for a vancomycin-tolerant phenotype was overexpressed. The CRC gut microbiota showed an enhanced expression of *MarB*, a periplasmic protein which may indirectly repress the expression of *MarA*, a trigger of bacterial response to different toxic compounds, including antibiotics (590). β-Lactam resistance of the CRC microbiome appears to be significantly enhanced, as it is seen via greater activity of the BlaR1 family regulatory sensor-transducer disambiguation subsystem. The expression of *BlaR1*/*MecR1* family genes (591) that sense β-lactams and activate expression of β-lactamase PC1/*blaZ* and penicillin-binding proteins 1A/1B and 3 (poorly acylated by β-lactam antibiotics) that confer resistance to the antibiotic, was elevated. Activity of the subsystem, phosphoenolpyruvate phosphomutase and expression of the phosphonopyruvate decarboxylase gene, involved in biosynthesis of fosfomycin, were increased in CRC. Lactacin F ABC transporter permease component, a bacteriocin, was transcribed less in CRC. Horizontal gene transfer facilitated through expression of *ComA*, a member of bacteriocin-associated ATP-binding transporter family was repressed. However, higher conjugative activity was likely a feature of the microbiome through enhanced transcription of *TraM* and *TraN* genes as well as the *TraI* gene, encoding for *IncF* plasmid conjugative transfer DNA-nicking and unwinding protein. This would enhance genome plasticity and confer more adaptive traits to the microbiota in CRC. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

The cancer-associated microbiota was also more active with respect to translation, cell division/replication, cleansing the nucleotide pool, folding of bacterial chromosomes, and daughter cell separation (Fig. 41). Together with more pronounced gluconeogenesis and biosynthetic metabolism in CRC, an enhanced level of microbial growth in the cancerous gut is also supported by our findings. This includes enhanced biosynthesis of NAD(+) cofactor, folate (vitamin B_9) and nucleotides, membrane phospholipids and their carriers and peptidoglycan as well as methylation (SAm MTase) in CRC. Overexpression of some serine protease and hydrolase genes which are implicated in the prevention of misfolded proteins and potentially toxic nucleoside diphosphate derivative accumulation in fast growing cells was observed.

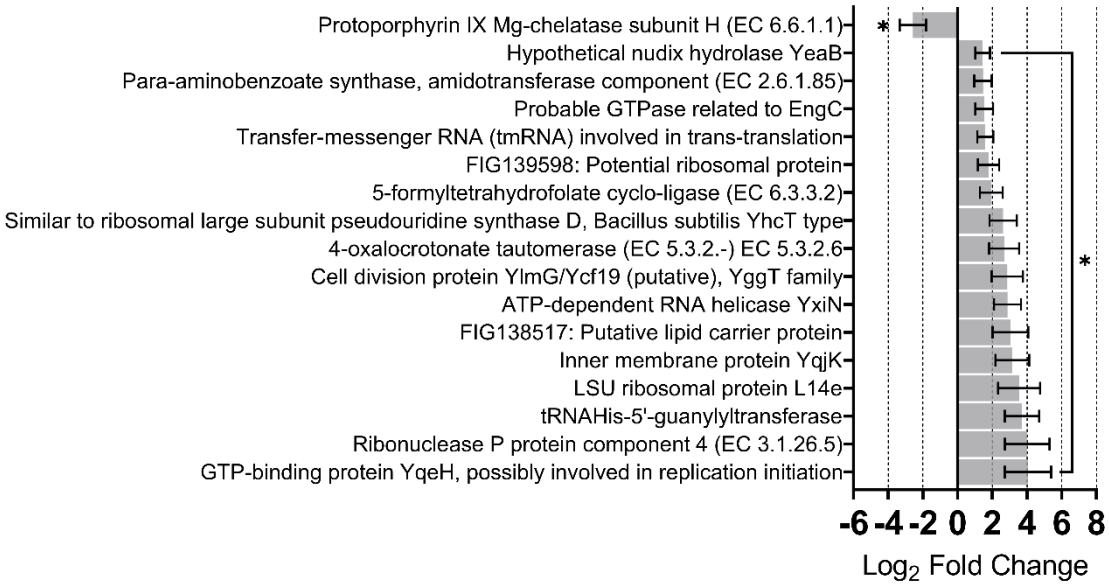


Fig. 41 | The CRC microbiota exhibit an enhanced rate of growth. Genes required for global processes such as translation, replication, cell division, membrane and cell wall biosynthesis displayed enhanced transcript levels under CRC conditions. Anaerobic photosynthetic growth in the CRC gut may be diminished as photoporphyrin IX Mg-chelatase subunit H expression, required for growth of genera such as *Xenococcus*, was downregulated, a possible reflection of less active Cyanobacteria in the CRC gut. * $P \leq 0.05$.

6.21 Quorum sensing activities of Gram-positive and Gram-negative bacteria are diminished and elevated, respectively, and less spore formation occurs during CRC

Microbes of complex communities can communicate to each other via production of specific signalling molecules, the concentration of which is critical for a proper microbial response. The signalling molecules, signal members of the microbial community about the surrounding conditions, including nutrient availability and stresses. This in turn may affect specific bacteria to adjust their metabolism to particular conditions, involving germination or sporulation.

We observed significant changes in transcription of a small group of genes required for sporulation and germination (Fig. 42). The CRC gut environment may suppress sporulation and support germination of sporulated bacteria under certain conditions. This may be a consequence of the local environment and signals therein. Quorum sensing (QS) activity which responds to growth conditions (592) was affected. Two lines of evidence argue that the Gram-negative QS machinery was upregulated in CRC. Primarily, enhanced transcription of SAM MTase (Fig. 35A) which suggests a greater supply of SAM in CRC and is indirect evidence that excess SAM in the cancerous gut can also stimulate production of homoserine lactone by acyl-homoserine lactone (AHL) pheromone synthases. Secondarily, expression of 3-hydroxydecanoyl-[*a*cyl-*c*arrier-*p*rotein, ACP] dehydratase (Fig. 40B) implicated in C_{10}-AHL production was also augmented. The CRC-associated microbiota further displayed QS-mediated regulation of gene expression, such as enhanced production of proteases, e.g. HtrA and Sec secretion subsystem (Fig. 40C). Contrarily, the peptide-based QS system of Gram-positive bacteria was repressed in CRC.

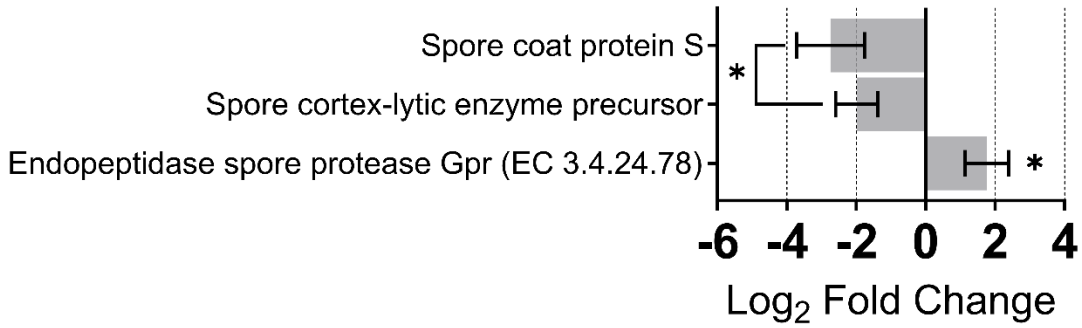


Fig. 42 | Microbial sporulation is diminished while germination is dynamic in the CRC gut niche. Expression of WhiB-like transcription regulator (Fig. 34E) (necessary for sporulation in Actinomycetes), spore cortex-lytic enzyme precursor (required for completion of germination) and spore coat protein S (a major spore coat protein produced early in sporulation) are repressed while transcription of the *Gpr* gene, which encodes for a spore protease required for early-stage germination of spores was increased in CRC. * $P \leq 0.05$.

6.22 A range of virulence determinants are overexpressed by a more inflammatory yet immune evasive microbial population in the CRC gut

The CRC-associated microbiota upregulated production of a number of virulence factors as described above (section 6.20). Expression of the *S. pyogenes* and *S. agalactiae* virulomes, prophage lysogenic conversion modules (genes of prophage origin which can enhance pathogen virulence) (593) and auxin biosynthesis genes (a virulence factor which is implicated in suppression of plant defence) (Fig. 37A) were more pronounced in CRC. Microorganisms enhanced expression of genes whose products synthesise and transport inflammation-promoting intermediates. Greater microbial transport of polyamines, via upregulation of spermidine putrescine ABC transporter permease component PotB (Fig. 35C) and transcriptional regulator, MerR family, near polyamine transporter (Fig. 34E), suggests an amplified abundance of inflammatory polyamines (594). This has been shown to facilitate both biofilm formation in the colonic mucosa and cancer cell growth, invasion and metastasis (218).

Overproduction of the proinflammatory O-antigen of LPS in Gram-negative bacteria, via upregulation of mannose-1-phosphate guanylyltransferase and a regulator of O-antigen component length (595), is consistent with a more inflammatory microbiome in CRC, or a reflection of greater Gram-negative species activity. Accelerated degradation of phenylacetic acid, an anti-inflammatory catabolite of phenylalanine, through phenylacetic acid degradation protein PaaE (Fig. 37A) may further increase the pro-inflammatory environment of the gut. Acetoin, a potential pro-inflammatory intermediate of bacterial fermentation, the availability of which in the CRC gut appeared to be higher (Fig. 35A), can induce IL-8, IL-6 (pro-inflammatory cytokines interleukins 8 and 6) and TNF-α production, causing significant loss of epithelial barrier function (596). Acetoin can also promote colonisation/biofilm formation of environmental microbes in the lungs, including *Pseudomonas aeruginosa* (597) (due to acetoin cross-feeding by other bacteria). It can be reduced to 2,3-butanediol under oxygen limited conditions. Gram-negative facultative anaerobes, such as *Klebsiella*, *Pseudomonas* and *Enterobacter* activities of which have been significantly enhanced in CRC (556) under anaerobic conditions can divert carbon flux from production of organic acids to neutral products, such as 2,3-butanediol. This suggests that acidification of the gut during CRC has a pleotropic effect on microbial metabolism, from enhanced amino acid catabolism to supporting a pro-cancerous gut environment. By contrast, butanediol fermentation in Gram-positive bacteria was attenuated, transcription of *YdjL* (Fig. 35A), putative oxidoreductase, a *Bacillus subtilis* 2,3-butanediol dehydrogenase homologue (598) was repressed. This suggests that major 2,3-butanediol-producing Gram-positive bacteria like *Bacillus* spp. and potentially *Streptococcus* spp. are more tolerant to low pH and still can produce mixed acids from pyruvate during cancer. These data *in fine* strongly suggest that microbiota of the gut

community in CRC interacts with host cells and expresses diverse virulence factors to facilitate their colonisation.

Elevated production of microbial virulence factors may be a consequence of a higher level of microbial access to and colonisation of the epithelium. The observed overexpression of hemolysin III (Fig. 40C), a virulence determinant, may help microorganisms, e.g. members of *Bacteroides fragilis* group, to colonise the CRC gut (599). Hemolysins are known to be central in developing systemic infections by opportunistic pathogens, weakening the immune system and gaining advantages in a competitive niche. Critically, hemolysins lyse can kill host cells, and hence provide a supply of nutrients, e.g. nucleic acids. This is consistent with our finding that the cell death rate in the CRC gut is elevated (Fig. 35C). Induced production of capsular polysaccharides and EPS results in more hydrophilic properties of the cell wall surface which was shown to be less susceptible to phagocytosis by neutrophils (600). facilitating microbial immune evasion. Furthermore, biosynthesis of pseudaminic acid (579) (Fig. 40C) supports the assertion that an array of adaptative approaches are employed by the microbiota in CRC to evade host immune cells.

7.0 Discussion

Here for the first time the CRC gut microbiota were analysed through metatranscriptomics to assess health-dependent changes to their specific (gene and species) and wider (pathway and microbial sub-population) activity. Through profiling the active CRC microbiome, we identify activity-based marker species of the disease (Fig. 21-23, Fig 25-26, Fig. 29 and Table 6). We also discovered that expression of antibiotic resistance determinant genes can also be regulated (Fig. 27 and Fig. 40E), at least in part by specific gut environmental factors without external antibiotic pressures (Fig. 28). Intriguingly, activity of non-cancerous and CRC gut derived *E. coli* grown as a part of the community and in pure culture is fundamentally different, indicative of microbial transcriptional memory. Importantly functional dependency of the microbial community on the health status of the host was uncovered (Fig. 32). Inherently different regulated responses of gut microbes were found to diverse environmental factors depending upon if bacteria were health or disease associated (Fig. 39). This work reveals distinct mechanisms by which gene expression of the microbiota responds to the malignant state of the gut. The findings herein are aligned to the potential mechanisms by which the CRC microenvironment is maintained.

To ensure a high quality of RNA was isolated from the faecal microbiota without external RNA contamination, blank RNA extractions using water were carried out prior to bioanalyser analysis and PCR screening. Blank extractions ensured no foreign RNA was being introduced during isolation, while the bioanalyser allowed for the visualisation of

prokaryotic 23S and 16S ribosomal (r)RNA as well as assessment of RNA degradation. No eukaryotic 28S and 18S rRNA was observed and RNA degradation was limited (Fig. 15). PCR was performed using the RNA isolate as template to confirm the absence of DNA contamination (data not shown). Quantity of RNA was established through spectrophotometry (Table 5). All samples passed these quality control criteria and were deemed suitable for quantitative high-throughput RNA sequencing, confirmed by bioanalyser analysis conducted by VERTIS biotechnologie, AG (Fig. 16). Following ribodepletion and cDNA synthesis an average of 40M read pairs were produced per sample, over 75% of reads were retained following trimming and removal of short, below 50 bp reads. This depth of sequencing and retention of data are desired to perform a high-fidelity analysis of bacterial composition and downstream functional analysis of the microbiome. Microbial DNA extractions were subject to similar quality control procedures, namely blank extraction, PCR assessment of rDNA quality and spectrophotometry. The PCR revealed no eukaryotic DNA had been isolated alongside the prokaryotic target DNA and following 16S rDNA V3-V4 sequencing, between 120-140k reads were produced per sample, surpassing the 100k reads suggested for sufficient taxonomic survey of microbial communities, these outputs were similar for both meta faecal microbial DNA and later DNA of *in vitro* cultures.

It has long been reported that microbiome diversity, the metric by which ecosystem (or sample) species richness and evenness of distribution is assessed, is lower in those affected by gastrointestinal pathologies, including CRC (601). Diversity can be investigated between different groups, or within an individual sample, termed β-diversity, and α-diversity respectively (504, 602). Through comparing the metatranscriptomes of the two sample groups (CRC and control) using Bray-Curtis dissimilarity, a β-diversity metric, intra-sample diversity was recorded as significantly lower than inter-sample dissimilarity (Fig. 17). While it is also noteworthy that the intra-sample diversity for the CRC group appears to be more variable than that of the controls, implying the heterogenicity of the microbiome is more pronounced in the colorectal cancer gut. However, DNA-level abundance analysis appears to overestimate sample α-diversity, this is consistent with and reflected in the greater significance recorded between sample dissimilarities (Fig. 17 and Fig. 31). This overestimation is likely due to species who, while having a detectable population at the DNA level, may be transcriptionally silent (dormant) or dead. This suggests that genomic abundance alone is a less informative tool for assessing microbial diversity than transcript abundance, or at the very least the two methods should be used in combination. The pattern of lower α-diversity in the guts of those with gut pathologies has overlooked the potential of significant microbial dormancy within the population. This ‘shutdown’ of metabolic activity of bacteria is becoming more well understood

and is even hypothesised to be the default mode of life for microbes due to the dynamic and potentially high-stress environment of the gut particularly during disease (603).

Changes in the composition of the gut microbiota in CRC have long been reported (540, 604), common CRC-specific patterns of taxonomy have emerged as potential disease markers. In this work we show that the abundance and activity of the microbiota in the control and CRC groups do not always correlate (Fig. 20), in line with metatranscriptome and metagenome data reported for IBD microbiota (113). It has been previously found that the metatranscriptome of the human gut microbiome is functionally stable, meaning average transcript levels remain consistent when compared to the taxonomic composition (125). However, prior to this study, the share of total microbiome activity across its constituents had not been reported. We found that, despite the presence of around 1000 distinct species, almost half of all transcripts identified belonged to only 20 species (Fig. 19). This substantially uneven weighting of transcript contribution across the microbiome highlights the disproportional relationship between abundance and activity. This finding becomes particularly pertinent when assessing the profile of species with traditionally altered (particularly enriched) abundance in CRC, such as *Fusobacterium nucleatum* (222), that exhibits great divergence between activity and abundance, calling into question the relevance of previously reported changes in presence associated with disease. We found that several species exist in a dormant (or dormant-like) state, with levels of activity at least an order of magnitude lower than their abundance, if detectible at all (Fig. 20A). The opposite was also observed, several, including previously CRC-linked species (e.g. *F. nucleatum*) appeared hyper-active, with total transcript level at least an order of magnitude greater than genome level (Fig. 20B). This shows that activity of specific microorganisms is regulated independently of abundance in CRC. However, despite substantial differences in microbial abundance and activity, this does not rule out that a species, such as *Eubacterium rectale* (Fig. 20A), with low activity relative to abundance, may still significantly contribute to the onset of CRC while being less active during the disease, as is proposed (605). Metatranscriptome levels of many clinically relevant bacteria, such as ESKAPE, Enterobacteriaceae and oral pathogens was significantly elevated in CRC, consistent with enhanced β-lactam and vancomycin resistances of this microbiome. Responses of aerobes of the CRC and control gut microbiota *in vitro* differed to environmental pressures, particularly to acid pressures. Additionally, multiple commensal (e.g. *n*-butyrate producing) bacteria lose significant activity in the CRC niche.

Many microorganisms were identified in one or both groups that exhibited a dormant, dormant-like or hyper-active phenotype (Fig. 20), potentially common features of microbial life in the diseased gut (113) by which bacteria can adapt to survive different environmental stresses. The CRC gut features loci of varied oxygen saturations (606) and easier access to

colonocytes (607). Such a diseased niche may cause the observed switch from active to latent forms of metabolism and *vice versa*. Members of the probiotic *Bifidobacterium* and *Lactobacillus* genera were highly abundant across both cohorts while transcriptome levels were underrepresented by at least 10-fold in comparison (Fig. 20A). It is noteworthy that CRC patients often present as folate (vitamin B_9, required for DNA synthesis, methylation, and metabolism) deficient (608). The diminished activity and dormancy of major *de novo* folate producers, *Bifidobacterium,* and *Lactobacillus spp*., in the gut niche during cancer may at least in-part, contribute to this. A consequence of this depletion may be the eventual hypomethylation of host DNA, a phenotype associated with CRC through the synthesis of 5-methyltetrahydrofolate, which donates a methyl group to SAM (609). The link between depletion of folate-producing bacteria and host DNA hypomethylation has previously been reported (610), however there is now substantial evidence that this mechanism exists in the colorectal cancer gut. The gut environment appears to control activity of probiotic species in a health status-independent manner. Once probiotic administration is ceased any measurable health benefit quickly diminishes (611) as the conditions of the gut likely controls their overall and specific activity. Therefore, it would appear manipulating microbial activity, either directly or indirectly through altering the intestinal environment, such as through diet and/or physical activity, could be a more efficient therapeutic approach than probiotic administration to gain any lasting potential health benefits (612). This is a concept which is already being explored, as well as microbiome transplant/transfer elsewhere (613), often through faecal transplant. However, while this has proved to be an effective therapeutic option, it also has significant limitations. These include lack of sustainability (donor availability), reproducibility (standardisation), and knowledge regarding under-characterised and potentially harmful constituent species with the potential to cause infections in the immunocompromised. Critically, these potentially infectious species are common in the faecal microbiome which could be dormant upon collection and become active post-transplantation. Hence it is imperative to avoid transplantation of these bacteria that may be pathology-causing in the long-term. These findings bring insight into the gut niche dynamically altering, either through amplifying or muting, the metabolic states of different microbes which may have been historically overlooked. Additionally, these findings highlight the importance of characterising activity of species to provide context to associative studies based on genomic abundance before moving forward with developing strategies to manipulate the human gut microbiome.

Enhanced activity of Proteobacteria was also found in CRC alongside genome enrichment. Over-representation of this phylum in the gut has been widely reported, particularly via faecal samples of human and animals during disease, including CRC (177) and in response to antibiotic treatments (107, 108). However, it has been found that

Proteobacteria were under-represented in the microbiota of CRC tissues, particularly *E. coli* (614, 615), but their activity had not been investigated previously (616). No significant differences in the activity of this species in CRC were found, confirmed via qRT-PCR (Fig. 24). Instead, higher metatranscriptome levels of other Enterobacteriaceae family pathogens were observed, such as *Klebsiella*, *Enterobacter* and *Kluyvera* species (Table 6), often resistant to antibiotics, and naturally competent to HT (617). Proteobacteria do not generally specialise in digesting complex carbohydrates and utilising fermentation products (618), they can however cross-feed on simple sugars made available through the saccharolytic activity of other microbes, such as Bacteroidetes (619). A reduction in expression of carbohydrate and increase in amino acid metabolising genes coincides with an increase in activity of these pathogens, likely to concurrently augment their virulence (620). These findings suggest a link between carbon source utilisation in the CRC gut and enhanced Enterobactiaceae activity and potentially virulence.

Both aerobic and anaerobic respiration entails glycolysis, pyruvate metabolism, tricarboxylic acid production and oxidative phosphorylation using either oxygen as the terminal electron acceptor or less energetic molecules, e.g. nitrate, sulphate and fumarate. These together enable the production of ATP, the reducing agent NADPH, lipids, nucleic and amino acids. The TCA cycle is fed with a supply of pyruvate from various sources, the glycolytic Embden-Meyerhof-Parnas (EMP) pathway chief among them. However, glucose availability appears limited during CRC, reflected in attenuated microbial glycolysis (Fig. 37). The gluconeogenic pyruvate carboxylase can be activated by the allosteric regulator acetyl-CoA and high pyruvate concentrations. This initiates production of glucose from non-carbohydrates, such as lactate and amino acids and converts pyruvate to oxaloacetate instead of acetyl-CoA, the respective genes of these alternate pathways displayed a marked rise in expression (Fig. 37A), again suggesting lack of glucose and potentially glutamate (the allosteric inhibitor of the enzyme) availability (621). Interestingly, production of glyceraldehyde-3-phosphate (GADP) in both ED and EMP glycolytic pathways was augmented (Fig. 40A), perhaps facilitating host mucin adhesion, if it is cell wall-associated (622), by the microbiota rather than playing a crucial role in energy generation. Expression of pyruvate producing oxo-acid lyase, 4-hydroxy-2-oxovalerate aldolase, was down-regulated in CRC (Fig. 38B), hence the production of acetaldehyde, an inhibitor of gluconeogenesis, was suppressed likely alongside the supply of pyruvate. Further evidence of enhanced gluconeogenesis, and thus higher activity of biosynthetic pathways, was supported by upregulation of the propionate-CoA to succinate module (Fig. 38B). Propionate can feed gluconeogenesis though the TCA cycle after conversion to succinyl-CoA followed by oxidation to glucose via pyruvate and oxaloacetate. β-Oxidation of odd carbon fatty acids and the catabolism of Iso, Val, Met and

Thr amino acids are major sources of propionate-CoA, supporting our findings of increased amino acid catabolism by the CRC microbiota. The microbiota favoured oxidation of the limited glucose for anabolic purposes through the pentose phosphate pathway as seen through elevated transcription of *YqeC* (Fig. 37B), implicated in production of the second NAD(P)H cofactor within the pathway, necessary for reductive biosynthetic reactions in fatty acid, aromatic amino and nucleic acid production. These data overall shows that the CRC microbiota may deploy gluconeogenesis, amino acid catabolism and specific nucleotides to offset the deficiency of simple sugar availability.

Several members of the ESKAPE group as well as a sub-set of oral pathogens gained activity in the CRC gut (Table 6 and Fig. 29). ESKAPE pathogens, clinically important multi-drug resistance carrying bacteria can acquire said resistance determinants through HT of mobile genetic elements (MGEs) (623). The dissemination of MGEs is predominantly carried out by conjugation and requires expression of the *tra* operon (624). Activity of two ESKAPE pathogens, *K. pneumoniae* and *A. baumannii*, was unchanged but their metatranscriptome substantially contributed to the total activity of the microbiome at 0.14% and 1% correspondently (data not shown). The observed elevated activity of other ESKAPE pathogens in CRC potentially poses a high risk of infection and the dissemination of MGEs. *Acinetobacter*, *Pseudomonas*, and *Staphylococcus* species can uptake DNA via natural transformation as well (625). Enhanced expression of genes involved in conjugative activity were shown for this CRC microbiome, including overexpression of *TraM* (DNA transfer) and *TraN* (mating pair stabilisation) genes (Fig. 40D). While it is currently difficult to assign these activities to specific species, it is feasible that these pathogens may increase their HT-specific gene expression, further enhancing their genetic plasticity (626). These findings suggest a potential risk of disseminating resistance determinants and highlight the importance of mapping antibiotic susceptibility patterns of CRC patients to afford appropriate treatment options.

A dramatic change in microbiome metabolism from utilisation of carbohydrates to amino acid catabolism supposes two important physiological characteristics of the cancerous gut. Firstly, tumour cells sustain uncontrolled growth through enhancing glycolytic activity, namely upregulating expression of glycolytic enzymes, glucose- and sodium glucose transporters (627, 628). This is consistent with the suggested lower availability of simple carbohydrates in the gut for microbial metabolism. Secondly, it is very unlikely that the diet of patients would be significantly altered compared with the control group (patients were scheduled for emergency surgery, therefore would not alter long-term dietary habits). Therefore, supply of carbohydrates should remain. However, it cannot be ruled out that a short-term change in diet between diagnosis and sample collection (around 4 weeks) may

have influenced the availability of simple carbohydrates, affecting the observed modes of microbial metabolism. However, during cancer, pathogens such as Enterobacteriaceae and ESKAPE pathogens that do not metabolise fibre, are more active in CRC and can increase their colonisation efficiency due to an expansion of the microaerobic niche, which may in-part lead to the decreased activity of saccharolytic *Bacteroides* and some *Clostridium* (629). Genomes of saccharolytic species are enriched with hydrolases which degrade complex carbohydrates (630), hence supplying carbon sources to other microbes (cross-feeding) as a product of anaerobic fermentation (619). These fermentation end products include *n*-butyrate, propionate, acetate, and lactate (631). Increased activity of pathogenic bacteria at the expense of carbohydrate hydrolysing species may also contribute to a deficiency of simple sugar metabolism and force microbiota to utilise amino acids, nucleotides, and aromatic compounds as alternative carbon sources. Moreover, extracellular (e)DNA can also be utilised as a nutrient source to offset the lack of simple carbohydrates in the cancerous gut (632). Biofilms, prevalent within the CRC gut (217, 218), constitute a rich source of eDNA, the primary component of the matrix, forming a nucleotide pool which is renewed through bacterial killing due to competition or immune activation (633). Hence, we cannot rule out that inflammation, nutrient depletion and/or biofilm formation in the cancerous gut (179) can also be a further factor(s) that induces DNA uptake.

Additionally, *S. aureus*, *A. baumannii* and *P. aeruginosa* can form biofilms, the expression of these determinants is elevated in the CRC microbiome (Fig. 40A), hence facilitating immune evasion, persistence, and antimicrobial resistance (634). Several oral cavity pathogens also exhibited an elevated level of transcriptional activity (Fig. 29A), including primary and secondary tissue colonisers, consistent with the observed augmented expression of colonisation factors by this microbiome. Additionally, significant overexpression of *Streptococcus agalactiae* and *Streptococcus pyogenes* virulomes (Fig. 40C) aligns with the enhanced levels of biofilm-associated gene expression in CRC. The CRC environment, such as transiently enhanced levels of oxygen and acidity (130, 136), mucus depletion (33, 189, 309), and potentially altered cross-feeding may trigger significant changes in the metabolism of opportunistic oral cavity, ESKAPE and Enterobacteriaceae pathogens. Therefore, immunocompromised individuals who are subject to recurring infections by these pathogens (including periodontal infections) should be considered vulnerable to dys/neoplasia and offered earlier CRC screening.

It has been found that levels of faecal hydrogen sulphide (H_2S) are greater in those suffering from CRC, meaning the capacity of SRB to reduce sulphate to H_2S is likely enhanced (635). Interestingly, reports surrounding the roles of H_2S within the gut vary, some suggest an anti-cancerous effect while others claim the opposite, a more commonly held view. Promoting

inflammation and genotoxicity (537), alongside *in vitro* promotion of colonocyte proliferation and attenuation of oxidative phosphorylation of *n*-butyrate have been observed in response to H_2S treatment (636). However, some works have suggested that H_2S protects the mucus layer by promoting mucin production and therefore decreases inflammation (538). The findings from the metatranscriptome analysis of this work shows that three SRB species gain in activity during CRC (Fig. 25), providing potential candidate species in the explanation for the elevated H_2S levels previously reported. While two other SRB species lost activity in CRC, these data provide a precise and previously unknown cohort of SRB who may be the driving force behind H_2S-mediated damage, or in fact highlight specific SRB whose potential protective role through H_2S is lost during CRC, depending on the prevailing nature of the contribution of H_2S to gut health which is yet to be fully clarified.

It has been suggested that thermophilic bacteria, which have an optimal growth temperature of >41°C and pH of 7-11, can confer certain health benefits, e.g. stimulation of host immunoglobulin A, IgA when introduced to the gut, shown through oral administration of compost (or compost extract and pure cultures of identified thermophiles) to gnotobiotic mice, flatfish, and rats (637–639). Yet, it must be noted that this is not a unique host response to thermophilic species, many conserved antigens of commensal microbes such as SFB *Candidatus arthromitus* and *Mucispirillum* can also trigger IgA secretion (640). By metatranscriptome profiling, a group of 15 thermophilic gut species were found to exhibit elevated activity during CRC (Fig. 26), this is despite the substantial evidence of a highly acidic gastrointestinal environment, this is consistent with findings that although thermophiles prefer basic pH conditions, they are highly adaptable and capable of becoming acid tolerant (641). This either suggests that the acid gut environment causes physiological changes to the microbiota allowing certain sub-groups to thrive (consistent with what is reported later in this work), or that local areas of higher pH exist within the large intestine, and perhaps the heat-producing fermentative metabolism that allows these species to grow is facilitated in specific loci of the diseased gut. This could be consistent with the observation that probiotic introduction of thermophilic bacteria corresponds with higher energy harvesting of gut microbes (greater proportion of calories harvested from food) and more rapid gain of host biomass (639), a trait more commonly associated with Proteobacteria. However, these microbes are not well characterised and their contribution to host health needs to be established in greater detail, they may indeed increase IgA production by the host, however whether this is sufficient to elicit a beneficial reduction in pathogenic microbes in the complex gut microbiome is not known, particularly as certain microbes have evolved strategies to evade host immune activity (642).

The low pH environment of the diseased gut (643, 644) may provide favourable conditions for the growth of methanogenic bacteria and archaea that are naturally more acid tolerant. Archaea, while being prokaryotic like bacteria possess analogues of eukaryotic genes and fermentative metabolic pathways (645). However, while there is little known about the contribution of the archaeome to host health, enrichment in abundance of these methanogens and halophiles (species with a preference for high salt conditions) has been reported during disease (646). Intriguingly, archaea have been found to induce strong pro-inflammatory host responses by DCs (647) and have been associated with, as for thermophiles, greater energy harvesting within the gut (648). At the functional level it was observed in this metatranscriptome work that expression of transcripts specific to the synthesis of archaeal enzymes (including methanogenic co-factors) was augmented during CRC (Fig. 38A), perhaps alluding to a pro-inflammatory role of this microbial population in the progression and/or onset of the malignancy. Taxonomic analysis of the archaeome at the metatranscriptome level would help in uncovering the specific species which exhibit altered metabolic activity within the population during disease, however the archaeome is currently not well represented in the available databases.

The human intestinal microbiota represent a dense microbial population and are natural reservoirs of antibiotic resistance determinant genes (649). AB genes are stable in such bacterial communities regardless of the presence of antibiotic pressure (e.g. (650)) and therefore the potential for AB gene expression and horizontal transfer within such a community is very high (651). RNA-seq analysis revealed enhanced expression of genes involved in HT and resistance determinants to twelve AB families in CRC in the absence of AB pressure (Fig. 40E). It was found that many AB resistance determinants are active in the gut microbiome irrespective of the health status of the host with expression of a sub-set of AB resistance determinants significantly upregulated by the CRC microbiota (Fig. 27B). This leads to two important conclusions. Firstly, a significant proportion of over-expressed AB resistance determinants in CRC were encoded by Enterobacteriaceae. It is known that gut inflammation, a feature of the CRC gut, promotes HT between pathogens and *Enterobacteria* (652), suggesting that CRC microbes are prone to HT and it would be feasible to expect this is happening in the non-cancerous gut but to a lesser degree. Secondly, expression of numerous AB resistance genes by the gut microbiota strongly argues that the environment of the gut is able to induce AB resistance determinant expression. However, the reason for this is not abundantly clear. Furthermore, the CRC gut environment induces expression of a number of resistance determinants which belong to different AB classes and expression of only a handful of resistance genes was attenuated (Fig. 27A). This shows that specific CRC gut (micro)environments may trigger such responses in an AB-independent manner. Enhanced

expression of some AB resistance determinants, such as *bla*$_{CMY}$ and *gadX*, found through RNA-seq in CRC was not determined by a single environmental factor *in vitro* (Fig. 28). These resistances further reduce therapeutic options, particularly for methicillin-resistant *Staphylococcus aureus*, MRSA, coagulase negative staphylococci and other Gram-positive infections in penicillin allergic individuals (653). These data argue that bacterial competition may be enhanced in CRC through production of antimicrobials, bacitracin (654), microcin (585) and fosfomycin (655), and appears to be a primary feature of their co-habitation. Higher activity of the multiple antibiotic resistance phenotype and some efflux transporters is consistent with the CRC gut exerting more stresses to the microbiota. However, it cannot be ruled out that endogenous production of antimicrobial peptides, such as defensins, by colonocytes (656) may play a role in the induction of microbial AB resistance determinant expression and HT. In CRC the microbiome may be in close proximity to or colonising the epithelium due to compromised epithelial barrier function, hence promoting endogenous defensin production.

AB resistome data revealed that ESKAPE pathogens display a multi-drug resistant phenotype (upregulating AB resistance in CRC) (Fig. 27B), and expression of *catA* by *A. baumannii* and *phoP* by *K. pneumoniae* was confirmed *in vitro* to be regulated by environmental factors (Fig. 28). For example, overexpression of *phoP* and *eptB* genes in *K. pneumoniae* in the CRC gut was observed in response to acids, while *phoP*, *eptB* and *bla*$_{CMY}$ transcription was repressed by HCl but not lactate. A similar pattern was observed for expression of *mdtO*, *nfsA*, *marA* and *gadX* by *E. coli*. This shows that multi-drug resistant bacteria express their AB resistance genes in response to environmental pressures in a gene specific manner and this often depends upon the health status of the host. Hence, other factors, such as antimicrobials and/or cell-cell interaction of bacteria and host may be required for inducing expression of specific AB resistance determinants *in vivo*. It is known that the *Salmonella* PhoP/PhoQ two-component system, which is critical for its virulence (657) and confers resistance to colistin, macrolides and peptide ABs, responds to the host environment and enhances modification of the bacterial envelope, hence reducing membrane permeability (658–661). Exposure to antimicrobial peptides, such as defensins and polymyxin, has been shown to increase expression of *phoP*/*phoQ* and resistance to antimicrobial peptides (662). The findings from this work show that non-acid pressures generally inhibit *phoP* expression, but acidity promotes its transcription in a health-dependent manner in CRC derived *K. pneumoniae*. This shows that *phoP* can be activated by acidity which in turn may lead to solidification of the cell membrane by enriching the proportion of unsaturated fatty acids (663, 664). It is reasonable to propose that gut bacteria could deploy putative “long-term memory” (665–669) of acid-dependent membrane solidification upon repeat exposure to these conditions. Similar transcriptional memory had been observed in *E. coli* in regulating the

expression of the *lac* operon in response to changing environmental nutrient availability (665). Another possible explanation of CRC-dependent gene overexpression in response to a pressure, including *phoP* is that the control culture produces an unknown inhibitor which prevents the acid-dependent activation of transcription. Interestingly, 16S rDNA-sequencing failed to detect *K. pneumoniae* in culture, highlighting further the importance of studies of active microbes rather than their abundance.

The *in vitro* data shows that expression of AB resistance genes can be regulated by different environmental factors, acidity, osmotic potential, and oxidative stress (Fig. 28). Interestingly, acidification of the gut appears to be a potentially key factor in activation of AB resistance gene expression in CRC. The CRC gut features excessive lactate availability due to anaerobic glycolysis (Warburg metabolism) of tumour cells (389). Enhanced transcription of the gene encoding malolactic enzyme (Fig. 35A), which converts malate to less acidic lactate, by the CRC gut microbiota, is consistent with increased intestinal acidification during CRC (643). Consistent with this, lactate seems to upregulate expression of a wider population of resistance genes *in vitro* compared to HCl, however a larger array of genes should be tested to validate this. This argues that acid stress, during gastrointestinal malignancy, could propagate an antibiotic resistance phenotype of gut microbes. While osmotic pressure in the CRC culture represses expression of many AB resistance genes *in vitro* (Fig. 28), this can also activate transcription of other genes, including *catA* in *A. baumannii*, an ESKAPE pathogen (Fig. 27B). Oxidative pressure, like osmotic stress, inhibits expression of many AB resistance determinants in CRC-derived culture while upregulating expression of certain genes (Fig. 28). In the control-derived culture however, H_2O_2 enhanced expression of more than half of the AB resistance genes (6 out of 10) tested, showing that oxidative pressure may activate AB resistance mechanisms, primarily in a health-dependent manner. This also argues that gut bacteria possess unidentified long-term transcriptional memory mechanisms with respect to AB resistance that provide cells with a rapid response to previously encountered pressures, such as acidity for expression of the *phoP* gene in *K. pneumoniae* (Fig. 28). These findings highlight the notion that microbiota of the healthy gut are under severe pressure from H_2O_2 compared to the CRC gut and the *in vitro* transcriptional data supports this assertion.

Through profiling the active CRC microbiome and investigating the control of its activity we showed global CRC-specific transcriptional regulation (Fig. 32), particularly for species associated with recurrent nosocomial infections which gained in activity (Table 6). Repeat infections caused by those bacteria can be an early marker of suppressed immunity or gastrointestinal vulnerability which poses a high risk for onset of colorectal cancer. Here it was also shown for the first time that expression of antibiotic resistance determinant genes can also be regulated by specific gut environmental factors without external antibiotic pressures

and for specific resistances the health status of the host plays a key role (Fig. 28). Distinct responses of CRC- and control-derived aerobes to each pressure suggests that i) the health status of the gut environment influences expression of AB resistance genes in the absence of AB treatment and ii) activity of the same non-cancerous and CRC gut species is fundamentally different and can be modulated by external factors.

The gut microbiota, the 'germ organ' of the host, is a unique microbial community as it develops with the host from birth. It is well known that, despite the constant interaction of the microbiota with the colonic mucosa, no general inflammation of the gut is observed day to day (193). The observed core metabolic functions of the microbiota across cohorts are in concordance with general housekeeping activities which allow microbes to co-exist with their host despite the inflammatory potential of the community and its members (Fig. 33). The downregulation of the pyruvate:ferredoxin/flavodoxin core subsystem indicates the microbiome is conducting less anaerobic respiration as a whole in CRC. However, certain pathway activities which occur in the absence of oxygen are still observed, and in some cases are even upregulated (see below, TME), indicating local areas of hypoxia in a potentially more oxygen rich CRC niche.

The stepwise accumulation of sporadic genetic lesions causing CRC has been attributed to the damaging effects of ROS (670). Oxidative stress response constitutes pathways which reduce ROS, such as O_2^- and H_2O_2 to protect membranes, proteins and DNA from damage (671). If the damage exceeds the capacity of host DNA repair mechanisms, genetic mutations may occur. H_2O_2-dependent dysregulation of epithelial barrier function would facilitate microbial colonisation and invasion, promoting inflammation and ROS production. However, the findings presented argue that inflammation-derived ROS in the gut appears to be only half the picture. Several unexpected lines of evidence strongly posit that the gut microbiota is a crucial mediator of ROS levels through their ability to scavenge and reduce ROS (Fig. 33 and Fig. 34). If the capacity of the microbiota to control the level of physiological ROS (during mitochondrial oxidative phosphorylation) is reduced, even temporarily, the cumulative effects over 15-30 years may facilitate the accumulation of ROS-induced damage (genetic and epithelial barrier function) and hence onset of CRC (672). Expression of the oxidative stress subsystem appears to be *the* core housekeeping function of the microbiome (Fig. 33). In response to enhanced ROS availability, the microbiota is to control this by increased ROS reduction, to lower ROS to physiological levels. Conversely, over-induced activity of microbial anti-oxidative mechanisms may lead to diminished ROS levels, also causing gut pathology through compromising epithelial barrier integrity or reducing the efficacy of host-produced ROS as a means of controlling the microbial population (673). Through these findings, two modes of ROS-mediated genetic damage are proposed. i)

Chronic or intermittent inflammation-dependent ROS accumulation due to e.g. IBD or antibiotic treatment (172, 187, 674). ii) Inflammation-independent, where compromised ROS-reducing microbiome functions leads to excess or diminished ROS (674). However, these modes may together form one continuous cyclic pathology wherein the epithelium can be compromised in an inflammation-independent manner, leading to inflammation-mediated damage (32, 309, 675).

Osmotic pressure which in-part regulates non-enzymatic antioxidants, such as ectoine (544) synthesis appears to be a regulated factor which may be associated with elevated levels of these antioxidants in the gut. Interestingly, a major response by the CRC gut microbiota appears to be to RNS, specifically nitric oxide (Fig. 34E). This is in concordance with high levels of inflammation that occur in the cancerous gut, including elevated levels of microbial colonisation, resulting in activation of host inducible nitric oxide synthase, *iNOS* (676). This also occurs during Warburg metabolism, saturating the lumen with nitrate (NO_3^-) and O_2 from the reduction in oxidative phosphorylation, facilitating the reduction in population of commensal anaerobes and expansion of potentially damaging facultative anaerobes (629).

Iron uptake and transport by the CRC microbiota appears to be enhanced (Fig. 35B). It was shown that tumour cells accumulate iron while blocking its export (677), a likely cause of the deficiency common to the condition, alongside inflammation via the hepcidin pathway and chronic blood loss (678). Elevated microbial uptake of iron should be considered, alongside inflammation and blood loss, as a mechanism by which the host becomes deficient. Iron supplementation, therefore, may have adverse effects by feeding tumour growth and pathogen virulence, hence posing a greater risk of infection and further inflammation. Intravenous iron supplementation, which has become more common practice in the NHS (particularly post-surgery) may subvert this issue by bypassing the gut microbiota, however this is not the case for other health conditions and the reason for this deficiency was not well understood. Many cancer patients also suffer from carnitine deficiency, 75% of which is derived from the diet (679). The observed enhanced catabolism of carnitine by microbiota in CRC (Fig. 36) may explain, at least in part, this phenomenon. Carnitine, the transporter of *n*-butyrate across the mitochondrial membrane, is the rate-limiting step of mitochondrial oxidative phosphorylation (680). The depletion of this metabolite may limit the beneficial O_2-consuming SCFA metabolism of healthy colonocytes during disease, further disrupting the activity and composition of the gut microbiota. Uncontrolled growth of cancerous colonocytes is underpinned by dysregulation and reprogramming of gene expression, including translation (681). Humans rely on dietary scavenging and the gut microbiota for their supply of queuosine (Q, a hyper-modified guanosine analogue), necessary for tRNA (Q_{34}tRNA) to ensure translation fidelity (682). Elevated microbial transport of Q (Fig. 36) suggests they are

assimilating the modified guanosine at a higher level and potentially depleting the host of this vital molecule, thus decreasing the accuracy of host protein synthesis. This could have far-reaching implications for host health, however no evidence or suggested mechanisms surrounding bacterial regulation of host translation fidelity exist as of now.

The data presented argue for local regions of hypoxia and O_2 saturation, as both oxygen-dependent and anaerobic metabolic processes are concurrently differentially active, consistent with the known architecture and metabolism of the TME (683) and resident biofilms (564). These conditions facilitate sequential colonisation of oxygen-respiring microbes in proximity to tumour blood vessels (and local areas of inflammation) and facultative and obligate anaerobes further from the O_2 supply (683). The TME in close proximity to the vasculature is less acidic (CO_2, H^+ and lactate are vented into the bloodstream) and more saturated with oxygen. Further from the vasculature the TME becomes more hypoxic and acidic. Under this O_2 gradient cancer cells become more glycolytic and release lactate and protons into the surrounding lumen (684), forcing anaerobes to modify their membrane structure with unsaturated fatty acids to decrease H^+ permeability (572), a trait observed being adopted by the microbiota (Fig. 36). This suggests close interaction of microbial sub-populations co-inhabiting specific niches which cannot support growth of anaerobes and aerobes simultaneously. However, under oxygen rich conditions anaerobes can still thrive in CRC via formation of biofilms with obligate anaerobes being the primary colonisers, forming the inner biofilm layers, which become hypoxic following colonisation of other bacteria (217).

The presented analysis of microbial RNA-seq data revealed evidence that the CRC gut environment, compared to its healthy counterpart, is more acidic (Fig. 35A). This can be due, in part, to the altered metabolism of cancerous colonocytes, which excessively produce lactate even in the presence of oxygen, namely aerobic glycolysis or the Warburg effect (389). CRC gut microbiota were observed exhibiting enhanced expression of GAD, of the glutamate decarboxylase acid defence mechanism (685). This defence is like other microbial acid resistance mechanisms, such as Arg- and Lys-decarboxylase systems which produce basic compounds and consume protons, hence increasing cytoplasmic pH (458). *E. coli* strains, whether pathogenic or not, abundant constituents of the gut microbiota, are remarkably well equipped with acid resistance mechanisms and can cause different diseases, including infections born from contaminated acidic food (461).

The Lys-dependent acid-resistance mechanism appears to be a universal acid defence system which protects gut bacteria against acid irrespective of host health status or nature of the acid (organic or inorganic). CadA decarboxylates Lys to cadaverine, a superoxide antioxidant, which is exported from cells in exchange for extracellular lysine and thus

alkalinises the cytoplasm by consuming a proton. However, the CadA system in the non-acid adapted microbiota is either unresponsive to or repressed by high salinity, while playing an adaptive role in response to oxidative stress in a health-dependent manner (Fig. 39). This suggests that the Lys-dependent resistance mechanism may provide the gut microbiota with additional non-enzymatic protection against ROS in response to high acidity and hydrogen peroxide availability, while activation by the latter is health status-dependent. In contrast, both Arg-dependent acid protective systems responded differently to both acid pressures and in a health-dependent manner (Fig. 39). Both subsystems appear to sense low pH (H^+ and lactate) only if they originated from the CRC microbiome, confirming the Arg-dependent systems are important for maintaining pH-homeostasis of aerobic microbiota in CRC. It has been shown that expression of *adiA* is triggered only under anaerobic growth at low pH (686). Our *in vitro* data clearly showed *adiA* is expressed under aerobic conditions by *E. coli* of gut microbiota in a pH-independent manner regardless of the health status of origin (Fig. 39). One possible explanation could be that growth of a complex mixture of aerobic microbes reduces the level of oxygen in the medium resulting in de-repression of the *E. coli* AdiA system. Additionally, low pH is not required for de-repression of transcription of Arg- and Lys-acid dependent systems in aerobic conditions. A potential cross-communication of microbiota and/or acidification of the medium due to CO_2 production (e.g. by the activity of pH independent SpeA) may be sufficient to maintain a constitutive level of expression of these amino acid defence mechanisms in an aerobic environment *in vitro*. Expression of *adiA* appears to be also regulated by salt and oxidative pressures. Interestingly, up-regulation was observed only in microbiota derived from the control samples and in response to H_2O_2 (Fig. 39). These data further suggest that microbiota adapt to the environment of the gut and can exhibit these “inherited” properties later by regulating their patterns of gene expression as part of survival strategies, perhaps through transcriptional memory (665). The expression of *speA* appears to be a broad-spectrum stress defence mechanism, at least in *E. coli* under aerobic conditions. However, the health status of the host also affects the ability of the SpeA-mediated mechanism to maintain pH homeostasis of the cell, supporting the view that the microbiota of the CRC and non-cancerous guts are fundamentally different. Acids, salt and reactive oxygen species trigger *in vitro* amino acid-dependent acid resistance mechanisms in a health-dependent manner (Fig. 39), suggesting these factors are features of the CRC human gut which in-turn direct microbial acid tolerance, consistent with metatranscriptome analysis. It cannot be ruled out however, that despite the overall similarities in taxonomic composition between each cohorts cultures (60-70% *E. coli*) (Dataset 1, Tab 3 of (527)), potential differences in the strains of *E. coli* grown may have contributed to the observed health status-specific differences in gene expression *in vitro*.

Gut microbiota have evolved numerous adaptive mechanisms by which they can exchange and expand their genetic information, termed genetic plasticity. One of these mechanisms includes bacteriophage infection through lysogeny facilitated by their co-habitation (687). It was also observed that the microbiota in the cancerous gut promote HT, a prominent feature of biofilms (564). Improved competence may be a result of either transient or enduring pressures on the gut, such as antibiotics/drug treatments or dsDNA breaks (688). It is also known that bacteriophages facilitate HT (689) and findings of enhanced anti-viral mechanisms (Fig. 40D) suggests this may be the case to a greater degree in the cancerous gut. Activities of two major anti-viral defence mechanisms, that can trigger abortive infection, namely group II intron-associated genes (retron-type reverse transcriptase) and the broad range (naïve) CRISPR-Cas Type III system (690) were upregulated coinciding with phage lysogeny/prophage overexpression. It is also known that bacteriophages facilitate HT (691), and the findings of this work suggests this may be the case to a greater degree in the cancerous gut. Interestingly, the CRC-associated gut microbiota downregulates CRISPR-Cas Type I (protein CT1974, a homolog of Cse3/CasE, one of the three endo-RNases implicated in processing of short CRISPR pre-RNA (692)) while enhancing activity of CRISPR-Cas Type III (RAMP *Cmr2* gene, a member of the Cas10-Cmr CRISPR Type III system which can target both foreign ssDNAs and ssRNAs) (Fig. 40D). CRISPR Type I can exhibit active primed adaptation which targets invasive mobile genetic element (iMGEs) sequences (693). Due to the presence of adjacent PAM (protospacer adjacent motif) and seed sequences which ensure precise target specificity, the Type I CRISPR system would fail to recognise mutated sequences, allowing accumulation of invader escapers. To minimize this, the Type I system promotes primed adaptation to target the same mutated invader via multiple spacers by rapidly acquiring new spacers complementary to other non-mutated regions of the invader DNA. Primed adaptation is the biased acquisition of new spacers derived from sequences that carry the targets of pre-existing spacers. Interestingly, while it has been shown that PAM facilitates prevention of autoimmunity (694), recent bioinformatic analysis of Type I targets of CRISPR spacers has revealed strong self-targeting potential with the core function to regulate HT (695). However, the Type I CRISPR system is known to fail in recognition of new and mutated sequences, allowing accumulation of invader escapers like prophages (690). CRISPR Type I is likely to respond to a symbiotic co-existence between phages and microbes. Greater uptake of extracellular DNA by the gut microbiota and higher activity of a wide range of bacteriophages in CRC would be consistent with enhanced activity of the Type III CRISPR-Cas system. Type III CRISPR is less stringent than Type I and targets both DNA and RNA, specifically highly transcribed plasmid and phage DNA. This can be beneficial in protecting cells against a variety of active phages (696). Finally, upregulation of the retron module, which uses the second element of the retron, non-coding RNA to generate a covalently linked

RNA/DNA molecule to initiate the major abortive infection defence mechanism (697), could correlate with the observed enhanced rate of microbial death in the CRC gut.

8.0 Future Prospects

As ROS are a primary trigger of CRC development (698–700), the ability of the microbiota to modulate ROS levels in the gut poses some important questions. For example, what are the environmental signals which can regulate microbial antioxidative activities, e.g. diet, antibiotics, toxins or other pressures and how are these signals modulated by the gut microbiota. Another important task is to characterise the specific pressure(s) that promote the enhanced antibiotic resistance phenotype displayed during disease. This is critical for patients who require surgery to prevent post-operative infection. Understanding and subsequently manipulating such adaptive mechanisms which the microbiota uses to compete for nutrients, exchange genetic material and control prevalence and activity of other gut species can be a useful tool in developing bacteria-based therapy.

Metatranscriptomics can identify disease-related changes in activity of individual microbes, functionally related groups and even at the community level, here shown at the community level for the CRC microbiome. The altered microbial active taxonomy and community characteristics reported sets up two major questions. Firstly, what are the specific patterns of gene expression of the identified species dominating the microbial transcriptome in both health and CRC and which are the marker species with dynamic CRC-specific activity. Secondly, what are the CRC-specific environmental factors that drive activity of these species *in vivo* and expression and dissemination of genetic elements. Antibiotic resistance appears to be a mechanism which provides the cell with the capacity for adaptation to environmental pressures in the gut. This in turn poses a high risk to health services to treat infections. Different non-antibiotic related factors differentially regulate AB resistance mechanisms. Thus, it would be of a great importance to investigate the combined effects of different pressures on a given resistance phenotype, especially for those which conversely regulate AB resistance determinant gene expression. Effects of colonocyte metabolism as well as other human tissues on the regulation of AB resistance gene expression should also be a high priority topic for antibiotic resistance research.

This work provides direct links between specific adaptive responses of the gut microbiome in the colorectal cancer gut via metatranscriptomics. These findings reveal important insights into the protective role of gut microbiome against developing cancer and its adaptive responses to the tumour environment. A striking example is the high background level of microbial-mediated ROS reduction activities by both the CRC and healthy gut microbiomes, an apparent “core housekeeping” role of the gut community, protecting

colonocytes against ROS-induced DNA damage and promoting epithelial integrity (Fig. 33 and Fig. 34F). These data shows that the CRC and control gut microbiota adapt through inherently different mechanisms to environmental pressures of the gut *in vitro*. This suggests that the health status of the host influences adaptive responses of the microbiota to specific stresses, laying the foundations for investigation into effective strategies for microbial manipulation. Depletion of the gut for beneficial metabolites in combination with enhanced genetic exchange, virulence, host colonisation, antibiotic and acid resistance in colorectal cancer make the microbiome more pathogenic and less protective. Therefore, it would be of a great importance to establish environmental factors that modulate the ROS-reducing capacity of the gut microbiota with the aim of protecting colonocytes from ROS-mediated DNA damage.

Moreover, from this work more specific questions may be proposed regarding the precise molecular mechanisms by which the microbiota and their secretome regulate colonocyte function. There is a wealth of evidence which shows that the gut microbiota and the level of their specific metabolic products such as SCFAs, polyamines and tryptophan intermediates have a profound effect on host cell metabolism and function (138, 182, 183, 218, 452, 701–706). Hence, the question emerges, what is the molecular mechanism(s) by which the microbiota may modulate colonocyte cancer hallmarks such as Warburg metabolism, the shift from oxidative phosphorylation to anaerobic glycolysis and associated cellular redox status through e.g. affecting the $NADP^+$:NADPH ratio. Answering this question would have broad implications for the understanding of how colonocytes become susceptible to ROS and hence DNA damage in an otherwise healthy individual. Similarly, what are the microbially secreted metabolites underpinning regulation of host cancer-driving epigenetics such as the CpG island hypermethylation and histone modification phenotypes, building on the knowledge gained through *in vitro* investigation of the ability of SCFA to inhibit histone deacetylases. Research regarding these topics would potentially identify novel therapeutic targets of the microbiota and/or its secreted compounds to prevent the progression of, or relapse to CRC. Furthermore, discovering the molecular mechanisms by which the complex microbiome regulates colonocyte metabolism and epigenetics could aid in the development of earlier diagnostic markers and intervention strategies, ultimately significantly lowering the mortality rate of the disease, particularly if these mechanisms are features of precursor conditions.

9.0 References

1. A. J. M. Watson, P. D. Collins, Colon cancer: A civilization disorder in *Digestive Diseases*, (2011), pp. 222–228.

2. I. Nishisho, *et al.*, Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science (1979)* **253**, 665–669 (1991).

3. G. Poulogiannis, I. M. Frayling, M. J. Arends, DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. *Histopathology* **56**, 167–179 (2010).

4. Z. Dai, *et al.*, The role of microbiota in the development of colorectal cancer. *Int J Cancer* **145**, 2032–2041 (2019).

5. M. X. Byndloss, A. J. Bäumler, The germ-organ theory of non-communicable diseases. *Nat Rev Microbiol* **16**, 103–110 (2018).

6. J. Gagnière, *et al.*, Gut microbiota imbalance and colorectal cancer. *World J Gastroenterol* **22**, 501–518 (2016).

7. C. A. Lopez, D. D. Kingsbury, E. M. Velazquez, A. J. Bäumler, Collateral damage: microbiota-derived metabolites and immune function in the antibiotic era. *Cell Host Microbe* **16**, 156–163 (2014).

8. I. Sobhani, *et al.*, Microbial Dysbiosis in Colorectal Cancer (CRC) Patients. *PLoS One* **6**, e16393 (2011).

9. I. Koliarakis, *et al.*, Oral Bacteria and Intestinal Dysbiosis in Colorectal Cancer. *Int J Mol Sci* **20**, 4146 (2019).

10. Y. Lu, *et al.*, Mucosal adherent bacterial dysbiosis in patients with colorectal adenomas. *Sci Rep* **6**, 26337 (2016).

11. F. Genua, V. Raghunathan, M. Jenab, W. M. Gallagher, D. J. Hughes, The Role of Gut Barrier Dysfunction and Microbiome Dysbiosis in Colorectal Cancer Development. *Front Oncol* **11** (2021).

12. W. B. Whitman, D. C. Coleman, W. J. Wiebe, Prokaryotes: The unseen majority. *Proc Natl Acad Sci U S A* **95**, 6578–6583 (1998).

13. J. Lederberg, A. T. Mccray, "COMMENTARY 'Ome Sweet 'Omics-A Genealogical Treasury of Words."

14. G. Bordenave, Louis Pasteur (1822-1895). *Microbes Infect* **5**, 553–60 (2003).

15. V. D'Argenio, F. Salvatore, The role of the gut microbiome in the healthy adult status. *Clinica Chimica Acta* **451**, 97–102 (2015).

16. P. B. Eckburg, *et al.*, Microbiology: Diversity of the human intestinal microbial flora. *Science (1979)* **308**, 1635–1638 (2005).

17. L. Li, N. Mendis, H. Triguri, J. D. Oliver, S. Faucher, The importance of VBNC state in human bacterial pathogens. *Front Microbiol* **5**, 1–19 (2014).

18. W. Manz, R. Amann, W. Ludwig, M. Vancanneyt, K. H. Schleifer, Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum

cytophaga-flavobacter-bacteroides in the natural environment. *Microbiology (N Y)* **142**, 1097–1106 (1996).

19. E. Birney, *et al.*, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).

20. O. Mizrahi-Man, E. R. Davenport, Y. Gilad, Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. *PLoS One* **8** (2013).

21. C. R. Woese, *et al.*, Conservation of primary structure in 16S ribosomal RNA. *Nature* **254**, 83–86 (1975).

22. Z. Liu, T. Z. DeSantis, G. L. Andersen, R. Knight, Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* **36**, e120 (2008).

23. J. Kuczynski, *et al.*, Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet* **13**, 47–58 (2012).

24. D. A. Peterson, D. N. Frank, N. R. Pace, J. I. Gordon, Metagenomic Approaches for Defining the Pathogenesis of Inflammatory Bowel Diseases. *Cell Host Microbe* **3**, 417–427 (2008).

25. H. Sung, *et al.*, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).

26. K. W. Jasperson, T. M. Tuohy, D. W. Neklason, R. W. Burt, Hereditary and Familial Colon Cancer. *Gastroenterology* **138**, 2044–2058 (2010).

27. R. E. Graff, *et al.*, Familial Risk and Heritability of Colorectal Cancer in the Nordic Twin Study of Cancer. *Clinical Gastroenterology and Hepatology* **15**, 1256–1264 (2017).

28. P. Lichtenstein, *et al.*, Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *New England Journal of Medicine* **343**, 78–85 (2000).

29. I. Mármol, C. Sánchez-de-Diego, A. P. Dieste, E. Cerrada, M. J. R. Yoldi, Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. *Int J Mol Sci* **18** (2017).

30. K. D. Mills, D. O. Ferguson, F. W. Alt, The role of DNA breaks in genomic instability and tumorigenesis. *Immunol Rev* **194**, 77–95 (2003).

31. A. J. Schetter, N. H. H. Heegaard, C. C. Harris, Inflammation and cancer: Interweaving microRNA, free radical, cytokine and p53 pathways. *Carcinogenesis* **31**, 37–49 (2009).

32. L. W. Peterson, D. Artis, Intestinal epithelial cells: Regulators of barrier function and immune homeostasis. *Nat Rev Immunol* **14**, 141–153 (2014).

33. H. E. Jakobsson, *et al.*, The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Rep* **16**, 164–177 (2015).

34. K. E. B. Knudsen, *et al.*, Impact of diet-modulated butyrate production on intestinal barrier function and inflammation. *Nutrients* **10** (2018).

35. J. Petersson, *et al.*, Importance and regulation of the colonic mucus barrier in a mouse model of colitis. *Am J Physiol Gastrointest Liver Physiol* **300** (2011).

36. I. mi Okazaki, A. Kotani, T. Honjo, Role of AID in Tumorigenesis. *Adv Immunol* **94**, 245–273 (2007).

37. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

38. L. Li, *et al.*, Activation-induced cytidine deaminase expression in colorectal cancer. *Int J Clin Exp Pathol* **12**, 4119–4124 (2019).

39. S. Gemble, *et al.*, Pyrimidine Pool Disequilibrium Induced by a Cytidine Deaminase Deficiency Inhibits PARP-1 Activity, Leading to the Under Replication of DNA. *PLoS Genet* **11**, e1005384 (2015).

40. W. M. Grady, J. M. Carethers, Genomic and Epigenetic Instability in Colorectal Cancer Pathogenesis. *Gastroenterology* **135**, 1079–1099 (2008).

41. S. S. Zeki, T. A. Graham, N. A. Wright, Stem cells and their implications for colorectal cancer. *Nat Rev Gastroenterol Hepatol* **8**, 90–100 (2011).

42. E. J. Kuipers, *et al.*, Colorectal cancer. *Nat Rev Dis Primers* **1**, 15065 (2015).

43. N. N. Keum, E. Giovannucci, Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol* **16**, 713–732 (2019).

44. D. J. Weisenberger, *et al.*, CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787–793 (2006).

45. M. C. Wong, H. Ding, J. Wang, P. S. Chan, J. Huang, Prevalence and risk factors of colorectal cancer in Asia. *Intest Res* **17**, 317–329 (2019).

46. J. Guinney, *et al.*, The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350–1356 (2015).

47. T. E. Raevaara, *et al.*, Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. *Gastroenterology* **129**, 537–549 (2005).

48. M. Toyota, M. Ohe-Toyota, N. Ahuja, J. P. J. Issa, Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype. *Proc Natl Acad Sci U S A* **97**, 710–715 (2000).

49. C. Lengauer, K. W. Kinzler, B. Vogelstein, Genetic instability in colorectal cancers. *Nature* **386**, 623–627 (1997).

50. C. KR, V. B, Genetic alterations in the adenoma--carcinoma sequence. *Cancer* **70** (1992).

51. S. L. Thompson, S. F. Bakhoum, D. A. Compton, Mechanisms of Chromosomal Instability. *Current Biology* **20** (2010).

52. C. Z. Zhang, *et al.*, Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).

53. M. S. Pino, D. C. Chung, The Chromosomal Instability Pathway in Colon Cancer. *Gastroenterology* **138**, 2059–2072 (2010).

54. N. R. Rodrigues, *et al.*, p53 mutations in colorectal cancer. *Proc Natl Acad Sci U S A* **87**, 7555–7559 (1990).

55. T. W. Chittenden, *et al.*, Functional classification analysis of somatically mutated genes in human breast and colorectal cancers. *Genomics* **91**, 508–511 (2008).

56. A. L. Hsieh, Z. E. Walton, B. J. Altman, Z. E. Stine, C. V. Dang, MYC and metabolism on the path to cancer. *Semin Cell Dev Biol* **43**, 11–21 (2015).

57. M. Brocardo, B. R. Henderson, APC shuttling to the membrane, nucleus and beyond. *Trends Cell Biol* **18**, 587–596 (2008).

58. C. Artus, *et al.*, The Wnt/planar cell polarity signaling pathway contributes to the integrity of tight junctions in brain endothelial cells. *Journal of Cerebral Blood Flow and Metabolism* **34**, 433–440 (2014).

59. O. J. Sansom, *et al.*, Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev* **18**, 1385–1390 (2004).

60. D. M. Muzny, *et al.*, Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

61. A. Lasry, A. Zinger, Y. Ben-Neriah, Inflammatory networks underlying colorectal cancer. *Nat Immunol* **17**, 230–240 (2016).

62. J. Behrens, *et al.*, Functional interaction of β-catenin with the transcription factor LEF- 1. *Nature* **382**, 638–642 (1996).

63. F. T. Kolligs, G. Bommer, B. Göke, Wnt/Beta-Catenin/Tcf Signaling: A Critical Pathway in Gastrointestinal Tumorigenesis. *Digestion* **66**, 131–144 (2002).

64. W.-L. He, *et al.*, Association Between c-Myc and Colorectal Cancer Prognosis: A Meta-Analysis. *Front Physiol* **9** (2018).

65. W. ZHANG, H. T. LIU, MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res* **12**, 9–18 (2002).

66. L. C. Cantley, The Phosphoinositide 3-Kinase Pathway. *Science (1979)* **296**, 1655–1657 (2002).

67. T. Sjöblom, *et al.*, The consensus coding sequences of human breast and colorectal cancers. *Science (1979)* **314**, 268–274 (2006).

68. W.-J. Jeong, E. J. Ro, K.-Y. Choi, Interaction between Wnt/β-catenin and RAS-ERK pathways and an anti-cancer strategy via degradations of β-catenin and RAS by targeting the Wnt/β-catenin pathway. *NPJ Precis Oncol* **2**, 1–10 (2018).

69. A. J. L. Clark, S. Ishii, N. Richert, G. T. Merlino, I. Pastan, Epidermal growth factor regulates the expression of its own receptor. *Proc Natl Acad Sci U S A* **82**, 8374–8378 (1985).

70. P. Wee, Z. Wang, Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers (Basel)* **9** (2017).

71. K. Inamura, Colorectal cancers: An update on their molecular pathology. *Cancers (Basel)* **10** (2018).

72. M. G. Summers, *et al.*, BRAF and NRAS locus-specific variants have different outcomes on survival to colorectal cancer. *Clinical Cancer Research* **23**, 2742–2749 (2017).

73. A. Russo, *et al.*, The TP53 colorectal cancer international collaborative study on the prognostic and predictive significance of p53 mutation: Influence of tumor site, type of mutation, and adjuvant treatment. *Journal of Clinical Oncology* **23**, 7518–7528 (2005).

74. T. Takayama, K. Miyanishi, T. Hayashi, Y. Sato, Y. Niitsu, Colorectal cancer: genetics of development and metastasis. *J Gastroenterol* **41**, 185–192 (2006).

75. M. A. Hahn, *et al.*, Methylation of polycomb target genes in intestinal cancer is mediated by inflammation. *Cancer Res* **68**, 10280–10289 (2008).

76. T. Cooks, *et al.*, Mutant p53 Prolongs NF-κB Activation and Promotes Chronic Inflammation and Inflammation-Associated Colorectal Cancer. *Cancer Cell* **23**, 634–646 (2013).

77. K. T. Bieging, S. S. Mello, L. D. Attardi, Unravelling mechanisms of p53-mediated tumour suppression. *Nat Rev Cancer* **14**, 359–370 (2014).

78. E. Yonish-Rouach, *et al.*, Wild-type p53 induces apoptosis of myeloid leukaemic cells that is inhibited by interleukin-6. *Nature* **352**, 345–347 (1991).

79. I. López, *et al.*, Different mutation profiles associated to P53 accumulation in colorectal cancer. *Gene* **499**, 81–87 (2012).

80. S. Schwitalla, *et al.*, Loss of p53 in Enterocytes Generates an Inflammatory Microenvironment Enabling Invasion and Lymph Node Metastasis of Carcinogen-Induced Colorectal Tumors. *Cancer Cell* **23**, 93–106 (2013).

81. N. R. West, S. Mccuaig, F. Franchini, F. Powrie, Emerging cytokine networks in colorectal cancer. *Nat Rev Immunol* **15**, 615–629 (2015).

82. S. Ambs, *et al.*, p53 and vascular endothelial growth factor regulate tumor growth of NOS2-expressing human carcinoma cells. *Nat Med* **4**, 1371–1376 (1998).

83. J. Bergounioux, *et al.*, Calpain activation by the Shigella flexneri effector VirA regulates key steps in the formation and life of the bacterium's epithelial niche. *Cell Host Microbe* **11**, 240–252 (2012).

84. B. Iacopetta, TP53 mutation in colorectal cancer. *Hum Mutat* **21**, 271–276 (2003).

85. D. A. Ahlquist, *et al.*, Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Ann Intern Med* **149**, 441–450 (2008).

86. T. F. Imperiale, *et al.*, Multitarget stool DNA testing for colorectal-cancer screening. *New England Journal of Medicine* **370**, 1287–1297 (2014).

87. G. E. Fox, K. R. Pechman, C. R. Woese, "Comparative Cataloging of 16s Ribosomal Ribonucleic Acid: Molecular Approach to Procaryotic Systematics" (1977).

88. R. J. Case, *et al.*, Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies. *Appl Environ Microbiol* **73**, 278–288 (2007).

89. A. Suau, *et al.*, Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol* **65**, 4799–4807 (1999).

90. R. Poretsky, *et al.*, Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS One* **9**, 93827 (2014).

91. A. J. Pinto, L. Raskin, PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets. *PLoS One* **7**, e43093 (2012).

92. D. L. Church, *et al.*, Performance and Application of 16S rRNA Gene Cycle Sequencing for Routine Identification of Bacteria in the Clinical Microbiology Laboratory. *Clin Microbiol Rev* **33** (2020).

93. S. A. Barghouthi, A Universal Method for the Identification of Bacteria Based on General PCR Primers. *Indian J Microbiol* **51**, 430–444 (2011).

94. N. Youssef, *et al.*, Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Appl Environ Microbiol* **75**, 5227–36 (2009).

95. S. M. Huse, D. M. Welch, H. G. Morrison, M. L. Sogin, Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12**, 1889–98 (2010).

96. R. Rosselló-Mora, R. Amann, The species concept for prokaryotes. *FEMS Microbiol Rev* **25**, 39–67 (2001).

97. J. C. Wooley, A. Godzik, I. Friedberg, PLoS Computational Biology: A Primer on Metagenomics. *PLoS Comput Biol* **6**, e1000667 (2010).

98. R. P. Smith, *et al.*, Gut microbiome diversity is associated with sleep physiology in humans. *PLoS One* **14**, e0222394 (2019).

99. S. C. Campbell, *et al.*, The effect of diet and exercise on intestinal integrity and microbial diversity in mice. *PLoS One* **11** (2016).

100. H. M. P. Consortium, Structure, Function and Diversity of the Healthy Human Microbiome The Human Microbiome Project Consortium HHS Public Access. *Nature* **486**, 207–214 (2012).

101. A. Zhernakova, *et al.*, Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science (1979)* **352**, 565–569 (2016).

102. C. Xu, H. Zhu, P. Qiu, Aging progression of human gut microbiota. *BMC Microbiol* **19**, 236 (2019).

103. M. J. Claesson, *et al.*, Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proceedings of the National Academy of Sciences* **108**, 4586–4591 (2011).

104. M. J. Hopkins, Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. *Gut* **48**, 198–205 (2001).

105. N. Li, W.-T. Ma, M. Pang, Q.-L. Fan, J.-L. Hua, The Commensal Microbiota and Viral Infection: A Comprehensive Review. *Front Immunol* **10** (2019).

106. M. T. Bailey, *et al.*, Exposure to a social stressor alters the structure of the intestinal microbiota: Implications for stressor-induced immunomodulation. *Brain Behav Immun* **25**, 397–407 (2011).

107. C. Jernberg, S. Löfmark, C. Edlund, J. K. Jansson, Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology (N Y)* **156**, 3216–3223 (2010).

108. D. A. Dobrzanska, *et al.*, Preventive antibiotic treatment of calves: emergence of dysbiosis causing propagation of obese state-associated and mobile multidrug resistance-carrying bacteria. *Microb Biotechnol* **13**, 669–682 (2020).

109. W. Chen, F. Liu, Z. Ling, X. Tong, C. Xiang, Human Intestinal Lumen and Mucosa-Associated Microbiota in Patients with Colorectal Cancer. *PLoS One* **7**, e39743 (2012).

110. X. C. Morgan, *et al.*, Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* **13**, R79 (2012).

111. M. Joossens, *et al.*, Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* **60**, 631–637 (2011).

112. J. A. Navas-Molina, *et al.*, "Advancing Our Understanding of the Human Microbiome Using QIIME" in (2013), pp. 371–444.

113. M. Schirmer, *et al.*, Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* **3** (2018).

114. B. Canard, R. S. Sarfati, DNA polymerase fluorescent substrates with reversible 3'-tags. *Gene* **148**, 1–6 (1994).

115. M. Margulies, *et al.*, Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).

116. M. J. Gosalbes, *et al.*, Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS One* **6** (2011).

117. I. Korf, Genomics: The state of the art in RNA-seq analysis. *Nat Methods* **10**, 1165–1166 (2013).

118. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

119. E. M. Glass, J. Wilkening, A. Wilke, D. Antonopoulos, F. Meyer, Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc* **2010**, pdb.prot5368 (2010).

120. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

121. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

122. R. L. Tatusov, The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**, 22–28 (2001).

123. C. von Mering, *et al.*, STRING 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35**, D358-62 (2007).

124. Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).

125. G. S. Abu-Ali, *et al.*, Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat Microbiol* **3** (2018).

126. A. Raj, A. van Oudenaarden, Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226 (2008).

127. K. A. Lema, *et al.*, Inter- and Intra-Specific Transcriptional and Phenotypic Responses of *Pseudo-nitzschia* under Different Nutrient Conditions. *Genome Biol Evol* **11**, 731–747 (2019).

128. G. D'Souza, C. Kost, Experimental Evolution of Metabolic Dependency in Bacteria. *PLoS Genet* **12**, e1006364 (2016).

129. T. Kovács, E. Mikó, G. Ujlaki, Z. Sári, P. Bai, "The Microbiome as a Component of the Tumor Microenvironment" in *Advances in Experimental Medicine and Biology*, (Springer, 2020), pp. 137–153.

130. V. G. Peddareddigari, D. Wang, R. N. DuBois, The Tumor Microenvironment in Colorectal Carcinogenesis. *Cancer Microenvironment* **3**, 149–166 (2010).

131. N. Iida, *et al.*, Commensal Bacteria Control Cancer Response to Therapy by Modulating the Tumor Microenvironment. *Science (1979)* **342**, 967–970 (2013).

132. A. El Kaoutari, F. Armougom, J. I. Gordon, D. Raoult, B. Henrissat, The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* **11**, 497–504 (2013).

133. S. J. Kallus, L. J. Brandt, The intestinal microbiota and obesity. *J Clin Gastroenterol* **46**, 16–24 (2012).

134. Y. Litvak, M. X. Byndloss, A. J. Bäumler, Colonocyte metabolism shapes the gut microbiota. *Science (1979)* **362**, eaat9076 (2018).

135. V. Ganapathy, M. Thangaraju, P. D. Prasad, Nutrient transporters in cancer: Relevance to Warburg hypothesis and beyond. *Pharmacol Ther* **121**, 29–40 (2009).

136. X. Zhong, *et al.*, Warburg effect in colorectal cancer: the emerging roles in tumor microenvironment and therapeutic implications. *J Hematol Oncol* **15**, 160 (2022).

137. D. R. Donohoe, *et al.*, The Warburg Effect Dictates the Mechanism of Butyrate-Mediated Histone Acetylation and Cell Proliferation. *Mol Cell* **48**, 612–626 (2012).

138. D. J. Morrison, T. Preston, Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism (Taylor and Francis Inc., 2016).

139. H.-J. Wu, E. Wu, The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes* **3**, 4–14 (2012).

140. M. Yamamoto, *et al.*, A microarray analysis of gnotobiotic mice indicating that microbial exposure during the neonatal period plays an essential role in immune system development. *BMC Genomics* **13**, 335 (2012).

141. N. Kamada, G. Y. Chen, N. Inohara, G. Núñez, Control of pathogens and pathobionts by the gut microbiota. *Nat Immunol* **14**, 685–690 (2013).

142. J. P. Ouwerkerk, W. M. De Vos, C. Belzer, Glycobiome: Bacteria and mucus at the epithelial interface. *Best Pract Res Clin Gastroenterol* **27**, 25–38 (2013).

143. A. J. Baümler, V. Sperandio, Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**, 85–93 (2016).

144. Y. Momose, K. Hirayama, K. Itoh, Competition for proline between indigenous Escherichia coli and E. coli O157:H7 in gnotobiotic mice associated with infant intestinal microbiota and its contribution to the colonization resistance against E. coli O157:H7. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **94**, 165–171 (2008).

145. G. P. Schamberger, F. Diez-Gonzalez, Selection of recently isolated colicinogenic Escherichia coli strains inhibitory to Escherichia coli O157:H7. *J Food Prot* **65**, 1381–1387 (2002).

146. J. L. Giel, J. A. Sorg, A. L. Sonenshein, J. Zhu, Metabolism of bile salts in mice influences spore germination in clostridium difficile. *PLoS One* **5** (2010).

147. I. Sekirov, S. L. Russell, L. Caetano M Antunes, B. B. Finlay, Gut microbiota in health and disease. *Physiol Rev* **90**, 859–904 (2010).

148. D. A. O'Neil, *et al.*, Expression and regulation of the human beta-defensins hBD-1 and hBD-2 in intestinal epithelium. *J Immunol* **163**, 6718–24 (1999).

149. A. Mathias, B. Pais, L. Favre, J. Benyacoub, B. Corthésy, Role of secretory IgA in the mucosal sensing of commensal bacteria. *Gut Microbes* **5**, 688–695 (2015).

150. E. W. Rogier, A. L. Frantz, M. E. C. Bruno, C. S. Kaetzel, Secretory IgA is concentrated in the outer layer of colonic mucus along with gut bacteria. *Pathogens* **3**, 390–403 (2014).

151. R. Randal Bollinger, *et al.*, Human secretory immunoglobulin A may contribute to biofilm formation in the gut. *Immunology* **109**, 580–587 (2003).

152. P. Schnupf, *et al.*, Growth and host interaction of mouse segmented filamentous bacteria in vitro. *Nature* **520**, 99–103 (2015).

153. F. Graziani, *et al.*, Ruminococcus gnavus E1 modulates mucin expression and intestinal glycosylation. *J Appl Microbiol* **120**, 1403–1417 (2016).

154. S. Varyukhina, *et al.*, Glycan-modifying bacteria-derived soluble factors from Bacteroides thetaiotaomicron and Lactobacillus casei inhibit rotavirus infection in human intestinal cells. *Microbes Infect* **14**, 273–278 (2012).

155. M. Freitas, *et al.*, A heat labile soluble factor from Bacteroides thetaiotaomicron VPI-5482 specifically increases the galactosylation pattern of HT29-MTX cells. *Cell Microbiol* **3**, 289–300 (2001).

156. E. Thursby, N. Juge, Introduction to the human gut microbiota. *Biochemical Journal* **474**, 1823–1836 (2017).

157. P. A. Swanson, *et al.*, Enteric commensal bacteria potentiate epithelial restitution via reactive oxygen species-mediated inactivation of focal adhesion kinase phosphatases. *Proc Natl Acad Sci U S A* **108**, 8803–8808 (2011).

158. H. Q. Chen, *et al.*, Lactobacillus plantarum ameliorates colonic epithelial barrier dysfunction by modulating the apical junctional complex and PepT1 in IL-10 knockout mice. *Am J Physiol Gastrointest Liver Physiol* **299** (2010).

159. A. S. Neish, *et al.*, Prokaryotic regulation of epithelial responses by inhibition of IκB-α ubiquitination. *Science (1979)* **289**, 1560–1563 (2000).

160. M. Linsalata, R. Giannini, M. Notarnicola, A. Cavallini, Peroxisome proliferator-activated receptor gamma and spermidine/spermine N1-acetyltransferase gene expressions are significantly correlated in human colorectal cancer. *BMC Cancer* **6**, 191 (2006).

161. D. Kelly, *et al.*, Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shutting of PPAR-γ and ReIA. *Nat Immunol* **5**, 104–112 (2004).

162. X. Xu, *et al.*, Madecassic acid, the contributor to the anti-colitis effect of madecassoside, enhances the shift of Th17 toward Treg cells via the PPARγ/AMPK/ACC1 pathway. *Cell Death Dis* **8** (2017).

163. G. Pascual, *et al.*, A SUMOylation-dependent pathway mediates transrepression of inflammatory response genes by PPAR-γ. *Nature* **437**, 759–763 (2005).

164. S. Alex, *et al.*, Short-Chain Fatty Acids Stimulate Angiopoietin-Like 4 Synthesis in Human Colon Adenocarcinoma Cells by Activating Peroxisome Proliferator-Activated Receptor . *Mol Cell Biol* **33**, 1303–1316 (2013).

165. L. Dubuquoy, *et al.*, Impaired expression of peroxisome proliferator-activated receptor γin ulcerative colitis. *Gastroenterology* **124**, 1265–1276 (2003).

166. A. Szanto, *et al.*, STAT6 transcription factor is a facilitator of the nuclear receptor PPARγ-regulated gene expression in macrophages and dendritic cells. *Immunity* **33**, 699–712 (2010).

167. N. Arpaia, *et al.*, Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451–455 (2013).

168. Y. Furusawa, *et al.*, Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446–450 (2013).

169. N. Singh, *et al.*, Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* **40**, 128–139 (2014).

170. M. C. Becerra, I. Albesa, Oxidative stress induced by ciprofloxacin in Staphylococcus aureus. *Biochem Biophys Res Commun* **297**, 1003–1007 (2002).

171. X. Wang, X. Zhao, Contribution of Oxidative Damage to Antimicrobial Lethality. *Antimicrob Agents Chemother* **53**, 1395–1402 (2009).

172. M. A. Kohanski, D. J. Dwyer, B. Hayete, C. A. Lawrence, J. J. Collins, A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics. *Cell* **130**, 797–810 (2007).

173. S. Y. Kim, *et al.*, Antibacterial strategies inspired by the oxidative stress and response networks. *Journal of Microbiology* **57**, 203–212 (2019).

174. R. Mundade, T. F. Imperiale, L. Prabhu, P. J. Loehrer, T. Lu, Genetic pathways, prevention, and treatment of sporadic colorectal cancer. *Oncoscience* **1**, 400–406 (2014).

175. C. M. Johnson, *et al.*, Meta-analyses of colorectal cancer risk factors. *Cancer Causes & Control* **24** (2013).

176. F. Islami, *et al.*, Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA Cancer J Clin* **68** (2018).

177. C. H. Park, C. S. Eun, D. S. Han, Intestinal microbiota, chronic inflammation, and colorectal cancer. *Intest Res* **16**, 338–345 (2018).

178. S. K. Vanaja, V. A. K. Rathinam, K. A. Fitzgerald, Mechanisms of inflammasome activation: Recent advances and novel insights. *Trends Cell Biol* **25**, 308–315 (2015).

179. A. Lasry, A. Zinger, Y. Ben-Neriah, Inflammatory networks underlying colorectal cancer. *Nat Immunol* **17**, 230–240 (2016).

180. C. Lucas, N. Barnich, H. T. T. Nguyen, Microbiota, inflammation and colorectal cancer. *Int J Mol Sci* **18** (2017).

181. S. K. Mazmanian, J. L. Round, D. L. Kasper, A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**, 620–625 (2008).

182. G. Den Besten, *et al.*, The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res* **54**, 2325–2340 (2013).

183. J. G. LeBlanc, *et al.*, Beneficial effects on host energy metabolism of short-chain fatty acids and vitamins produced by commensal and probiotic bacteria. *Microb Cell Fact* **16**, 79 (2017).

184. J. K. Gustafsson, *et al.*, An ex vivo method for studying mucus formation, properties, and thickness in human colonic biopsies and mouse small and large intestinal explants. *Am J Physiol Gastrointest Liver Physiol* **302**, 430–438 (2012).

185. M. E. V. Johansson, J. M. Holmén Larsson, G. C. Hansson, The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A* **108**, 4659–4665 (2011).

186. K. Smith, K. D. McCoy, A. J. Macpherson, Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin Immunol* **19**, 59–69 (2007).

187. P. A. Swanson, *et al.*, Enteric commensal bacteria potentiate epithelial restitution via reactive oxygen species-mediated inactivation of focal adhesion kinase phosphatases. *Proc Natl Acad Sci U S A* **108**, 8803–8808 (2011).

188. H. Q. Chen, *et al.*, Lactobacillus plantarum ameliorates colonic epithelial barrier dysfunction by modulating the apical junctional complex and PepT1 in IL-10 knockout mice. *Am J Physiol Gastrointest Liver Physiol* **299** (2010).

189. L. Wrzosek, *et al.*, Bacteroides thetaiotaomicron and Faecalibacterium prausnitzii influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. *BMC Biol* **11** (2013).

190. M. A. Rahat, H. Bitterman, N. Lahat, Molecular mechanisms regulating macrophage response to hypoxia. *Front Immunol* **2** (2011).

191. S. Huber, *et al.*, Th17 Cells Express Interleukin-10 Receptor and Are Controlled by Foxp3- and Foxp3+ Regulatory CD4+ T Cells in an Interleukin-10-Dependent Manner. *Immunity* **34**, 554–565 (2011).

192. S. Huber, *et al.*, IL-22BP is regulated by the inflammasome and modulates tumorigenesis in the intestine. *Nature* **491**, 259–263 (2012).

193. N. Gagliani, B. Hu, S. Huber, E. Elinav, R. A. Flavell, The fire within: Microbes inflame tumors. *Cell* **157**, 776–783 (2014).

194. D. Hawiger, *et al.*, Dendritic cells induce peripheral T cell unresponsiveness under steady state conditions in vivo. *Journal of Experimental Medicine* **194**, 769–779 (2001).

195. P. A. Morel, T. B. Oriss, Crossregulation between Th1 and Th2 cells. *Crit Rev Immunol* **18**, 275–303 (1998).

196. A. Boonstra, *et al.*, Flexibility of mouse classical and plasmacytoid-derived dendritic cells in directing T helper type 1 and 2 cell development: Dependency on antigen dose and differential toll-like receptor ligation. *Journal of Experimental Medicine* **197**, 101–109 (2003).

197. E. Elinav, *et al.*, Inflammation-induced cancer: Crosstalk between tumours, immune cells and microorganisms. *Nat Rev Cancer* **13**, 759–771 (2013).

198. M. A. Kinnebrew, *et al.*, Interleukin 23 Production by Intestinal CD103 +CD11b + Dendritic Cells in Response to Bacterial Flagellin Enhances Mucosal Innate Immune Defense. *Immunity* **36**, 276–287 (2012).

199. J. A. Walker, J. L. Barlow, A. N. J. McKenzie, Innate lymphoid cells-how did we miss them? *Nat Rev Immunol* **13**, 75–87 (2013).

200. S. Huber, N. Gagliani, R. A. Flavell, Life, death, and miracles: Th17 cells in the intestine. *Eur J Immunol* **42**, 2238–2245 (2012).

201. T. Zelante, *et al.*, Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity* **39**, 372–385 (2013).

202. L. Yang, *et al.*, IL-21 and TGF-β are required for differentiation of human T H17 cells. *Nature* **454**, 350–352 (2008).

203. M. Nagai, Y. Obata, D. Takahashi, K. Hase, Fine-tuning of the mucosal barrier and metabolic systems using the diet-microbial metabolite axis. *Int Immunopharmacol* **37**, 79–86 (2016).

204. S. Hapfelmeier, *et al.*, Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science (1979)* **328**, 1705–1709 (2010).

205. N. W. Palm, *et al.*, Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000–1010 (2014).

206. Y. Zheng, *et al.*, Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nat Med* **14**, 282–289 (2008).

207. J. Chen, E. Pitmon, K. Wang, Microbiome, inflammation and colorectal cancer. *Semin Immunol* **32**, 43–53 (2017).

208. E. M. Shevach, *et al.*, The lifestyle of naturally occurring CD4 + CD25 + Foxp3 + regulatory T cells. *Immunol Rev* **212**, 60–73 (2006).

209. Y. Belkaid, Regulatory T cells and infection: A dangerous necessity. *Nat Rev Immunol* **7**, 875–888 (2007).

210. K. Honda, D. R. Littman, The Microbiome in Infectious Disease and Inflammation. *Annu Rev Immunol* **30**, 759–795 (2012).

211. A. Wächtershäuser, S. M. Loitsch, J. Stein, PPAR-γ is selectively upregulated in Caco-2 cells by butyrate. *Biochem Biophys Res Commun* **272**, 380–385 (2000).

212. L. Klotz, *et al.*, The nuclear receptor PPARγ selectively inhibits Th17 differentiation in a T cell-intrinsic fashion and suppresses CNS autoimmunity. *Journal of Experimental Medicine* **206**, 2079–2089 (2009).

213. D. Cipolletta, *et al.*, PPAR-γ is a major driver of the accumulation and phenotype of adipose tissue T reg cells. *Nature* **486**, 549–553 (2012).

214. S. G. Jeon, *et al.*, Probiotic Bifidobacterium breve induces IL-10-producing Tr1 cells in the colon. *PLoS Pathog* **8** (2012).

215. C. L. Sears, W. S. Garrett, Microbes, microbiota, and colon cancer. *Cell Host Microbe* **15**, 317–28 (2014).

216. C. M. Dejea, *et al.*, Microbiota organization is a distinct feature of proximal colorectal cancers. *Proceedings of the National Academy of Sciences* **111**, 18321–18326 (2014).

217. J.-P. Motta, J. L. Wallace, A. G. Buret, C. Deraison, N. Vergnolle, Gastrointestinal biofilms in health and disease. *Nat Rev Gastroenterol Hepatol* **18**, 314–334 (2021).

218. C. H. Johnson, *et al.*, Metabolism Links Bacterial Biofilms and Colon Carcinogenesis. *Cell Metab* **21**, 891–897 (2015).

219. C. M. Dejea, *et al.*, Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science (1979)* **359**, 592–597 (2018).

220. S. Tomkovich, *et al.*, Human colon mucosal biofilms from healthy or colon cancer hosts are carcinogenic. *Journal of Clinical Investigation* **129**, 1699–1712 (2019).

221. M. X. Chen, S.-Y. Wang, C.-H. Kuo, I.-L. Tsai, Metabolome analysis for investigating host-gut microbiota interactions. *Journal of the Formosan Medical Association* **118**, S10–S22 (2019).

222. J. Wirbel, *et al.*, Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* **25**, 679–689 (2019).

223. B. Flemer, *et al.*, The oral microbiota in colorectal cancer is distinctive and predictive. *Gut* **67**, 1454–1463 (2018).

224. A. Hartwich, *et al.*, Helicobacter pylori infection, gastrin, cyclooxygenase-2, and apoptosis in colorectal cancer. *Int J Colorectal Dis* **16**, 202–210 (2001).

225. Y. Yang, *et al.*, Fusobacterium nucleatum Increases Proliferation of Colorectal Cancer Cells and Tumor Development in Mice by Activating Toll-Like Receptor 4 Signaling to Nuclear Factor−κB, and Up-regulating Expression of MicroRNA-21. *Gastroenterology* **152**, 851-866.e24 (2017).

226. A. C. Goodwin, *et al.*, Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis. *Proc Natl Acad Sci U S A* **108**, 15354–15359 (2011).

227. C. Pleguezuelos-Manzano, *et al.*, Mutational signature in colorectal cancer caused by genotoxic pks + E. coli. *Nature* (2080) https:/doi.org/10.1038/s41586-020 (March 5, 2020).

228. D. R. M. Mark M. Huycke, Victoria Abrams, Enterococcus faecalis produces extracellular superoxide and hydrogen peroxide that damages colonic epithelial cell DNA. *Carcinogenesis* **23**, 529–536 (2002).

229. S. Wu, *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).

230. J. C. Arthur, *et al.*, Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science (1979)* **338**, 120–123 (2012).

231. M. M. Huycke, D. R. Moore, In vivo production of hydroxyl radical by Enterococcus faecalis colonizing the intestinal tract using aromatic hydroxylation. *Free Radic Biol Med* **33**, 818–826 (2002).

232. A. Boleij, H. Tjalsma, The itinerary of Streptococcus gallolyticus infection in patients with colonic malignant disease. *Lancet Infect Dis* **13**, 719–724 (2013).

233. A. N. McCoy, *et al.*, Fusobacterium Is Associated with Colorectal Adenomas. *PLoS One* **8** (2013).

234. X. J. Shen, *et al.*, Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. *Gut Microbes* **1**, 138–147 (2010).

235. S. Yachida, *et al.*, Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med* **25**, 968–976 (2019).

236. H. Sokol, *et al.*, Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* **105**, 16731–16736 (2008).

237. R. F. Benus, *et al.*, Impact of digestive and oropharyngeal decontamination on the intestinal microbiota in ICU patients. *Intensive Care Med* **36**, 1394–1402 (2010).

238. W. Chen, F. Liu, Z. Ling, X. Tong, C. Xiang, Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS One* **7** (2012).

239. S. Dulal, T. O. Keku, Gut Microbiome and Colorectal Adenomas. *The Cancer Journal* **20**, 225–231 (2014).

240. C. Sergaki, *et al.*, Developing whole cell standards for the microbiome field. *Microbiome* **10**, 123 (2022).

241. Q. R. Ducarmon, B. V. H. Hornung, A. R. Geelen, E. J. Kuijper, R. D. Zwittink, Toward Standards in Clinical Microbiota Studies: Comparison of Three DNA Extraction Methods and Two Bioinformatic Pipelines. *mSystems* **5** (2020).

242. C. Mirzayi, *et al.*, Reporting guidelines for human microbiome research: the STORMS checklist. *Nat Med* **27**, 1885–1892 (2021).

243. G. C. A. Amos, *et al.*, Developing standards for the microbiome field. *Microbiome* **8**, 98 (2020).

244. S. Suerbaum, P. Michetti, Helicobacter pylori infection. *New England Journal of Medicine* **347**, 1175–1186 (2002).

245. J. H. Nam, *et al.*, Helicobacter pylori infection is an independent risk factor for colonic adenomatous neoplasms. *Cancer Causes and Control* **28**, 107–115 (2017).

246. T. J. Koh, *et al.*, Overexpression of glycine-extended gastrin in transgenic mice results in increased colonic proliferation. *Journal of Clinical Investigation* **103**, 1119–1126 (1999).

247. S. F. Tatishchev, C. VanBeek, H. L. Wang, Helicobacter pylori infection and colorectal carcinoma: Is there a causal association? *J Gastrointest Oncol* **3**, 380–385 (2012).

248. H. J. Forman, M. Torres, Reactive oxygen species and cell signaling: Respiratory burst in macrophage signaling in *American Journal of Respiratory and Critical Care Medicine*, (American Lung Association, 2002).

249. O. Handa, Y. Naito, T. Yoshikawa, Helicobacter pylori: A ROS-inducing bacterial species in the stomach. *Inflammation Research* **59**, 997–1003 (2010).

250. D. J. Evans, *et al.*, Characterization of a Helicobacter pylori neutrophil-activating protein. *Infect Immun* **63**, 2213–2220 (1995).

251. B. E. Dunn, H. Cohen, M. J. Blaser, Helicobacter pylori. *Clin Microbiol Rev* **10**, 720–741 (1997).

252. H. Shmuely, *et al.*, Relationship between Helicobacter pylori CagA status and colorectal cancer. *Am J Gastroenterol* **96**, 3406–3410 (2001).

253. T. L. Cover, S. R. Blanke, Helicobacter pylori VacA, a paradigm for toxin multifunctionality. *Nat Rev Microbiol* **3**, 320–332 (2005).

254. M. Gerhard, *et al.*, Clinical relevance of the Helicobacter pylori gene for blood-group antigen-binding adhesin. *Proc Natl Acad Sci U S A* **96**, 12778–12783 (1999).

255. Y. Yamaoka, *et al.*, Helicobacter pylori outer membrane proteins and gastroduodenal disease. *Gut* **55**, 775–781 (2006).

256. O. Lee, *et al.*, Helicobacter pylori CagA phosphorylation status determines the gp130-activated SHP2/ERK and JAK/STAT signal transduction pathways in gastric epithelial cells. *Journal of Biological Chemistry* **285**, 16042–16050 (2010).

257. S. Wessler, L. M. Krisch, D. P. Elmer, F. Aberger, From inflammation to gastric cancer - The importance of Hedgehog/GLI signaling in Helicobacter pylori-induced chronic inflammatory and neoplastic diseases. *Cell Communication and Signaling* **15** (2017).

258. G. D. Eslick, Helicobacter pylori infection causes gastric cancer? A review of the epidemiological, meta-analytic, and experimental evidence. *World J Gastroenterol* **12**, 2991–2999 (2006).

259. B. J. Marshall, J. R. Warren, UNIDENTIFIED CURVED BACILLI IN THE STOMACH OF PATIENTS WITH GASTRITIS AND PEPTIC ULCERATION. *The Lancet* **323**, 1311–1315 (1984).

260. P. Escobar-Páramo, *et al.*, Large-scale population structure of human commensal Escherichia coli isolates. *Appl Environ Microbiol* **70**, 5698–5700 (2004).

261. B. Olesen, Characterization of four Escherichia coli clonal groups. *APMIS* **125**, 1–28 (2017).

262. A. Cougnoux, *et al.*, Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut* **63**, 1932–1942 (2014).

263. M. R. Wilson, *et al.*, The human gut bacterial genotoxin colibactin alkylates DNA. *Science (1979)* **363** (2019).

264. J. Raisch, N. Rolhion, A. DuBois, A. Darfeuille-Michaud, M. A. Bringer, Intracellular colon cancer-associated Escherichia coli promote protumoral activities of human macrophages by inducing sustained COX-2 expression. *Laboratory Investigation* **95**, 296–307 (2015).

265. O. Marchès, *et al.*, Enteropathogenic and enterohaemorrhagic Escherichia coli deliver a novel effector called Cif, which blocks cell cycle G2/M transition. *Mol Microbiol* **50**, 1553–1567 (2003).

266. G. Flatau, *et al.*, Toxin-induced activation of the G protein p21 Rho by deamidation of glutamine. *Nature* **387**, 729–733 (1997).

267. F. Taieb, C. Petit, J.-P. Nougayrède, E. Oswald, The Enterobacterial Genotoxins: Cytolethal Distending Toxin and Colibactin. *EcoSal Plus* **7** (2016).

268. O. D. K. Maddocks, A. J. Short, M. S. Donnenberg, S. Bader, D. J. Harrison, Attaching and effacing Escherichia coli downregulate DNA mismatch repair protein in vitro and are associated with colorectal adenocarcinomas in humans. *PLoS One* **4** (2009).

269. E. C. Wick, C. L. Sears, Bacteroides spp. and diarrhea. *Curr Opin Infect Dis* **23**, 470–474 (2010).

270. C. L. Sears, The toxins of Bacteroides fragilis. *Toxicon* **39**, 1737–1746 (2001).

271. R. J. Obiso, A. O. Azghani, T. D. Wilkins, The Bacteroides fragilis toxin fragilysin disrupts the paracellular barrier of epithelial cells. *Infect Immun* **65**, 1431–1439 (1997).

272. C. Ramakrishna, *et al.*, Bacteroides fragilis polysaccharide A induces IL-10 secreting B and T cells that prevent viral encephalitis. *Nature communication* **10** (2019).

273. C. M. Dejea, *et al.*, Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science (1979)* **359**, 592–597 (2018).

274. C. J. van der W. Yi Li, Jasper Deuring, Maikel P Peppelenbosch, Ernst J Kuipers, Colin de Haar, IL-6-induced DNMT1 activity mediates SOCS3 promoter hypermethylation in ulcerative colitis-related colorectal cancer. *Carcinogenesis* **33**, 1889–96 (2012).

275. G. Muyzer, A. J. M. Stams, The ecology and biotechnology of sulphate-reducing bacteria. *Nat Rev Microbiol* **6**, 441–454 (2008).

276. C. M. L. M. Coutinho, *et al.*, Sulphate-reducing bacteria from ulcerative colitis patients induce apoptosis of gastrointestinal epithelial cells. *Microb Pathog* **112**, 126–134 (2017).

277. C. Bernstein, *et al.*, Carcinogenicity of deoxycholate, a secondary bile acid. *Arch Toxicol* **85**, 863–871 (2011).

278. X. Wang, T. D. Allen, Y. Yang, D. R. Moore, M. M. Huycke, Cyclooxygenase-2 Generates the Endogenous Mutagen trans-4-Hydroxy-2-nonenal in Enterococcus faecalis–Infected Macrophages. *Cancer Prevention Research* **6**, 206–216 (2013).

279. Y. Zhou, *et al.*, Association of oncogenic bacteria with colorectal cancer in South China. *Oncotarget* **7**, 80794–80802 (2016).

280. B. J. Wallace, I. G. Young, Role of quinones in electron transport to oxygen and nitrate in Escherichia coli. Studies with a ubiA- menA- double quinone mutant. *BBA - Bioenergetics* **461**, 84–100 (1977).

281. P. A. Ruiz, A. Shkoda, S. C. Kim, R. B. Sartor, D. Haller, IL-10 Gene-Deficient Mice Lack TGF-β/Smad Signaling and Fail to Inhibit Proinflammatory Gene Expression in Intestinal Epithelial Cells after the Colonization with Colitogenic Enterococcus faecalis . *The Journal of Immunology* **174**, 2990–2999 (2005).

282. X. Wang, Y. Yang, M. M. Huycke, Commensal bacteria drive endogenous transformation and tumour stem cell marker expression through a bystander effect. *Gut* **64**, 459–468 (2015).

283. A. S. Abdulamir, R. R. Hafidh, F. A. Bakar, Molecular detection, quantification, and isolation of Streptococcus gallolyticus bacteria colonizing colorectal tumors: inflammation-driven potential of carcinogenesis via IL-1, COX-2, and IL-8. *Mol Cancer* **9**, 249 (2010).

284. H. T. Annemarie Boleij, Carla M. J. Muytjens, Sarah I. Bukhari, Nadège Cayet, Philippe Glaser, Peter W. M. Hermans, Dorine W. Swinkels, Albert Bolhuis, Novel Clues on the Specific Association of Streptococcus gallolyticus subsp gallolyticus With Colorectal Cancer. *J Infect Dis* **203**, 1101–1109 (2011).

285. Y. W. Han, Fusobacterium nucleatum: A commensal-turned pathogen. *Curr Opin Microbiol* **23**, 141–147 (2015).

286. M. Castellarin, *et al.*, Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res* **22**, 299–306 (2012).

287. A. D. Kostic, *et al.*, Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host Microbe* **14**, 207–215 (2013).

288. J. Yu, *et al.*, Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).

289. G. Zeller, *et al.*, Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* **10**, 766 (2014).

290. T. Wang, *et al.*, Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J* **6**, 320–9 (2012).

291. I. Allali, *et al.*, Gut microbiome compositional and functional differences between tumor and non-tumor adjacent tissues from cohorts from the US and Spain. *Gut Microbes* **6**, 161–172 (2015).

292. G. Nakatsu, *et al.*, Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun* **6** (2015).

293. B. Flemer, *et al.*, Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut* **66**, 633–643 (2017).

294. C. A. Brennan, W. S. Garrett, Fusobacterium nucleatum — symbiont, opportunist and oncobacterium. *Nat Rev Microbiol* **17**, 156–166 (2019).

295. M. S. Hayden, S. Ghosh, NF-κB, the first quarter-century: Remarkable progress and outstanding questions. *Genes Dev* **26**, 203–234 (2012).

296. Y. Wu, *et al.*, MicroRNA-21 (Mir-21) Promotes Cell Growth and Invasion by Repressing Tumor Suppressor PTEN in Colorectal Cancer. *Cellular Physiology and Biochemistry* **43**, 945–958 (2017).

297. M. R. Rubinstein, *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).

298. P. Lee, K. S. Tan, Fusobacterium nucleatum activates the immune response through retinoic acid-inducible gene i. *J Dent Res* **93**, 162–168 (2014).

299. J. L. Drewes, *et al.*, High-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. *NPJ Biofilms Microbiomes* **3**, 34 (2017).

300. B. van den Bergh, M. Fauvart, J. Michiels, Formation, physiology, ecology, evolution and clinical importance of bacterial persisters. *FEMS Microbiol Rev* **41**, 219–251 (2017).

301. C. M. Dejea, *et al.*, Microbiota organization is a distinct feature of proximal colorectal cancers. *Proceedings of the National Academy of Sciences* **111**, 18321–18326 (2014).

302. S. Tomkovich, *et al.*, Human colon mucosal biofilms from healthy or colon cancer hosts are carcinogenic. *Journal of Clinical Investigation* **129**, 1699–1712 (2019).

303. N. A. Pudlo, *et al.*, Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. *mBio* **6**, e01282--15--e01282--15 (2015).

304. O. Takeuchi, S. Akira, Pattern Recognition Receptors and Inflammation. *Cell* **140**, 805–820 (2010).

305. C. L. Bevins, N. H. Salzman, Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nat Rev Microbiol* **9**, 356–368 (2011).

306. E. Elinav, J. Henao-Mejia, R. A. Flavell, Integrative inflammasome activity in the regulation of intestinal mucosal immune responses. *Mucosal Immunol* **6**, 4–13 (2013).

307. N. J. Gay, M. F. Symmons, M. Gangloff, C. E. Bryant, Assembly and localization of Toll-like receptor signalling complexes. *Nat Rev Immunol* **14**, 546–558 (2014).

308. T. Kawai, S. Akira, TLR signaling. *Cell Death Differ* **13**, 816–825 (2006).

309. A. Swidsinski, *et al.*, Comparative study of the intestinal mucus barrier in normal and inflamed colon. *Gut* **56**, 343–350 (2007).

310. N. Matsushima, *et al.*, Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics* **8** (2007).

311. N. J. Gay, M. Gangloff, Structure and Function of Toll Receptors and Their Ligands. *Annu Rev Biochem* **76**, 141–165 (2007).

312. F. Sipos, I. Furi, M. Constantinovits, Z. Tulassay, G. Muzes, Contribution of TLR signaling to the pathogenesis of colitis-associated cancer in inflammatory bowel disease. *World J Gastroenterol* **20**, 12713–12721 (2014).

313. S. Akira, K. Takeda, Toll-like receptor signalling. *Nat Rev Immunol* **4**, 499–511 (2004).

314. L. A. J. O'Neill, A. G. Bowie, The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol* **7**, 353–364 (2007).

315. E. Cario, D. K. Podolsky, Differential alteration in intestinal epithelial cell expression of Toll-like receptor 3 (TLR3) and TLR4 in inflammatory bowel disease. *Infect Immun* **68**, 7010–7017 (2000).

316. S. Rakoff-Nahoum, J. Paglino, F. Eslami-Varzaneh, S. Edberg, R. Medzhitov, Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229–241 (2004).

317. A. O. Aliprantis, *et al.*, Cell activation and apoptosis by bacterial lipoproteins through Toll- like receptor-2. *Science (1979)* **285**, 736–739 (1999).

318. E. M. Creagh, L. A. J. O'Neill, TLRs, NLRs and RLRs: a trinity of pathogen sensors that co-operate in innate immunity. *Trends Immunol* **27**, 352–357 (2006).

319. T. L. Rothstein, *et al.*, Porcine B cell subset responses to Toll-like receptor ligands. **8**, 25 (2017).

320. S. M. Reba, *et al.*, TLR2 engagement on CD4+ T cells enhances effector functions and protective responses to Mycobacterium tuberculosis. *Eur J Immunol* **44**, 1410–1421 (2014).

321. Y.-G. Chen, *et al.*, Control of Methicillin-Resistant Staphylococcus aureus Pneumonia Utilizing TLR2 Agonist Pam3CSK4. *PLoS One* **11**, e0149233 (2016).

322. C. X. Lima, *et al.*, Therapeutic Effects of Treatment with Anti-TLR2 and Anti-TLR4 Monoclonal Antibodies in Polymicrobial Sepsis. *PLoS One* **10**, e0132336 (2015).

323. C. A. Cowardin, *et al.*, The binary toxin CDT enhances Clostridium difficile virulence by suppressing protective colonic eosinophilia. *Nat Microbiol* **1**, 16108 (2016).

324. S. Trop-Steinberg, Y. Azar, AP-1 Expression and its Clinical Relevance in Immune Disorders and Cancer. *American Journal of the Medical Sciences* **353**, 474–483 (2017).

325. L. Alexopoulou, A. C. Holt, R. Medzhitov, R. A. Flavell, Recognition of double-stranded RNA and activation of NF-κB by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).

326. E. Della Mina, *et al.*, Inherited human IRAK-1 deficiency selectively impairs TLR signaling in fibroblasts. *Proc Natl Acad Sci U S A* **114**, E514–E523 (2017).

327. J. C. Chow, D. W. Young, D. T. Golenbock, W. J. Christ, F. Gusovsky, Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *Journal of Biological Chemistry* **274**, 10689–10692 (1999).

328. M. Fukata, *et al.*, Cox-2 Is Regulated by Toll-Like Receptor-4 (TLR4) Signaling: Role in Proliferation and Apoptosis in the Intestine. *Gastroenterology* **131**, 862–877 (2006).

329. S. Basith, B. Manavalan, T. H. Yoo, S. G. Kim, S. Choi, Roles of toll-like receptors in cancer: A double-edged sword for defense and offense. *Arch Pharm Res* **35**, 1297–1316 (2012).

330. A. T. Gewirtz, T. A. Navas, S. Lyons, P. J. Godowski, J. L. Madara, Cutting Edge: Bacterial Flagellin Activates Basolaterally Expressed TLR5 to Induce Epithelial Proinflammatory Gene Expression. *The Journal of Immunology* **167**, 1882–1885 (2001).

331. H. Zeng, *et al.*, Flagellin/TLR5 responses in epithelia reveal intertwined activation of inflammatory and apoptotic pathways. *Am J Physiol Gastrointest Liver Physiol* **290** (2006).

332. F. Heil, *et al.*, Species-Specific Recognition of Single-Stranded RNA via Till-like Receptor 7 and 8. *Science (1979)* **303**, 1526–1529 (2004).

333. J. M. Lund, *et al.*, Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci U S A* **101**, 5598–5603 (2004).

334. G. Hartmann, G. J. Weiner, A. M. Krieg, CpG DNA: A potent signal for growth, activation, and maturation of human dendritic cells. *Proc Natl Acad Sci U S A* **96**, 9305–9310 (1999).

335. M. S. LAMPHIER, C. M. SIROIS, A. VERMA, D. T. GOLENBOCK, E. LATZ, TLR9 and the Recognition of Self and Non-Self Nucleic Acids. *Ann N Y Acad Sci* **1082**, 31–43 (2006).

336. H. Hemmi, *et al.*, A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–745 (2000).

337. T. J. D. Coutinho, G. R. Franco, F. P. Lobo, Homology-Independent Metrics for Comparative Genomics. *Comput Struct Biotechnol J* **13**, 352–357 (2015).

338. V. Bordignon, *et al.*, "Nucleic Acid Sensing Perturbation: How Aberrant Recognition of Self-Nucleic Acids May Contribute to Autoimmune and Autoinflammatory Diseases" in (2019), pp. 117–137.

339. F. Bäckhed, M. Hornef, Toll-like receptor 4-mediated signaling by epithelial surfaces: Necessity or threat? *Microbes Infect* **5**, 951–959 (2003).

340. D. J. Philpott, S. E. Girardin, P. J. Sansonetti, Innate immune responses of epithelial cells following infection with bacterial pathogens. *Curr Opin Immunol* **13**, 410–416 (2001).

341. M. H. Shaw, T. Reimer, Y. G. Kim, G. Nuñez, NOD-like receptors (NLRs): bona fide intracellular microbial sensors. *Curr Opin Immunol* **20**, 377–382 (2008).

342. D. J. Philpott, S. Yamaoka, A. Israël, P. J. Sansonetti, Invasive Shigella flexneri Activates NF-κB Through a Lipopolysaccharide-Dependent Innate Intracellular Response and Leads to IL-8 Expression in Epithelial Cells . *The Journal of Immunology* **165**, 903–914 (2000).

343. R. G. Correa, S. Milutinovic, J. C. Reed, Roles of NOD1 (NLRC1) and NOD2 (NLRC2) in innate immunity and inflammatory diseases. *Biosci Rep* **32**, 597–608 (2012).

344. M. Chamaillard, *et al.*, An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nat Immunol* **4**, 702–707 (2003).

345. S. E. Girardin, *et al.*, Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *Journal of Biological Chemistry* **278**, 8869–8872 (2003).

346. D. N. Frank, *et al.*, Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* **17**, 179–184 (2011).

347. R. J. Xavier, D. K. Podolsky, Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448**, 427–434 (2007).

348. J. H. Fritz, *et al.*, Synergistic stimulation of human monocytes and dendritic cells by Toll-like receptor 4 and NOD1- and NOD2- activating agonists. *Eur J Immunol* **35**, 2459–2470 (2005).

349. M. C. Grimm, P. Pavli, NOD2 mutations and Crohn's disease: Are Paneth cells and their antimicrobial peptides the link? *Gut* **53**, 1558–1560 (2004).

350. A. Biswas, T. Petnicki-Ocwieja, K. S. Kobayashi, Nod2: A key regulator linking microbiota to intestinal mucosal immunity. *J Mol Med* **90**, 15–24 (2012).

351. M. J. M. Bertrand, *et al.*, Cellular Inhibitors of Apoptosis cIAP1 and cIAP2 Are Required for Innate Immunity Signaling by the Pattern Recognition Receptors NOD1 and NOD2. *Immunity* **30**, 789–801 (2009).

352. J. Pedersen, E. C. LaCasse, J. B. Seidelin, M. Coskun, O. H. Nielsen, Inhibitors of apoptosis (IAPs) regulate intestinal immunity and inflammatory bowel disease (IBD) inflammation. *Trends Mol Med* **20**, 652–665 (2014).

353. R. B. Seth, L. Sun, C. K. Ea, Z. J. Chen, Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-κB and IRF3. *Cell* **122**, 669–682 (2005).

354. T. Kawai, *et al.*, IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat Immunol* **6**, 981–988 (2005).

355. A. Sabbah, *et al.*, Activation of innate immune antiviral responses by Nod2. *Nat Immunol* **10**, 1073–1080 (2009).

356. J. M. M. Natividad, *et al.*, Commensal and probiotic bacteria influence intestinal barrier function and susceptibility to colitis in Nod1-/-;Nod2-/- Mice. *Inflamm Bowel Dis* **18**, 1434–1446 (2012).

357. M. Yoneyama, *et al.*, The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol* **5**, 730–737 (2004).

358. H. Kato, *et al.*, Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441**, 101–105 (2006).

359. E. C. Lavelle, C. Murphy, L. A. J. O'Neill, E. M. Creagh, The role of TLRs, NLRs, and RLRs in mucosal innate immunity and homeostasis. *Mucosal Immunol* **3**, 17–28 (2010).

360. K. Kawane, *et al.*, Impaired thymic development in mouse embryos deficient in apoptotic DNA degradation. *Nat Immunol* **4**, 138–144 (2003).

361. H. Yoshida, Y. Okabe, K. Kawane, H. Fukuyama, S. Nagata, Lethal anemia caused by interferon-β produced in mouse embryos carrying undigested DNA. *Nat Immunol* **6**, 49–56 (2005).

362. K. J. Ishii, *et al.*, A toll-like receptor-independent antiviral response induced by double-stranded B-form DNA. *Nat Immunol* **7**, 40–48 (2006).

363. H. Yanai, *et al.*, HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. *Nature* **462**, 99–103 (2009).

364. G. Nakatsu, *et al.*, Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes. *Gastroenterology* **155**, 529-541.e5 (2018).

365. L. W. Lamps, K. K. T. Lai, D. A. Milner, Fungal Infections of the Gastrointestinal Tract in the Immunocompromised Host. *Adv Anat Pathol* **21**, 217–227 (2014).

366. M. J. Robinson, *et al.*, Dectin-2 is a Syk-coupled pattern recognition receptor crucial for Th17 responses to fungal infection. *Journal of Experimental Medicine* **206**, 2037–2051 (2009).

367. S. Yamasaki, *et al.*, C-type lectin Mincle is an activating receptor for pathogenic fungus, Malassezia. *Proc Natl Acad Sci U S A* **106**, 1897–1902 (2009).

368. L. Peng, Z. He, W. Chen, I. R. Holzman, J. Lin, Effects of butyrate on intestinal barrier function in a caco-2 cell monolayer model of intestinal barrier. *Pediatr Res* **61**, 37–41 (2007).

369. E. Holmes, J. V. Li, T. Athanasiou, H. Ashrafian, J. K. Nicholson, Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol* **19**, 349–359 (2011).

370. F. Guarner, J. R. Malagelada, Gut flora in health and disease. *Lancet* **361**, 512–519 (2003).

371. J. P. Segain, *et al.*, Butyrate inhibits inflammatory responses through NFκB inhibition: Implications for Crohn's disease. *Gut* **47**, 397–403 (2000).

372. W. Scheppach, H. P. Bartram, F. Richter, Role of short-chain fatty acids in the prevention of colorectal cancer. *Eur J Cancer* **31**, 1077–1080 (1995).

373. L. Peng, Z.-R. Li, R. S. Green, I. R. Holzman, J. Lin, Butyrate Enhances the Intestinal Barrier by Facilitating Tight Junction Assembly via Activation of AMP-Activated Protein Kinase in Caco-2 Cell Monolayers. *J Nutr* **139**, 1619–1625 (2009).

374. T. H. Jung, J. H. Park, W. M. Jeon, K. S. Han, Butyrate modulates bacterial adherence on LS174T human colorectal cells by stimulating mucin secretion and MAPK signaling pathway. *Nutr Res Pract* **9**, 343–349 (2015).

375. G. T. Macfarlane, S. Macfarlane, Bacteria, colonic fermentation, and gastrointestinal health. *J AOAC Int* **95**, 50–60 (2012).

376. A. J. Brown, *et al.*, The orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *Journal of Biological Chemistry* **278**, 11312–11319 (2003).

377. K. Ahmed, S. Tunaru, S. Offermanns, GPR109A, GPR109B and GPR81, a family of hydroxy-carboxylic acid receptors. *Trends Pharmacol Sci* **30**, 557–562 (2009).

378. A. Sekhavat, J. M. Sun, J. R. Davie, Competitive inhibition of histone deacetylase activity by trichostatin A and butyrate. *Biochemistry and Cell Biology* **85**, 751–758 (2007).

379. J. Stein, M. Zores, O. Schröder, Short-chain fatty acid (SCFA) uptake into Caco-2 cells, by a pH-dependent and carrier mediated transport mechanism. *Eur J Nutr* **39**, 121–125 (2000).

380. S. Miyauchi, E. Gopal, Y. J. Fei, V. Ganapathy, Functional Identification of SLC5A8, a Tumor Suppressor Down-regulated in Colon Cancer, as a Na+-coupled Transporter for Short-chain Fatty Acids. *Journal of Biological Chemistry* **279**, 13293–13296 (2004).

381. Z. Gao, *et al.*, Butyrate improves insulin sensitivity and increases energy expenditure in mice. *Diabetes* **58**, 1509–1517 (2009).

382. G. Den Besten, *et al.*, Short-chain fatty acids protect against high-fat diet-induced obesity via a pparg-dependent switch from lipogenesis to fat oxidation. *Diabetes* **64**, 2398–2408 (2015).

383. D. L. Arvans, *et al.*, Luminal bacterial flora determines physiological expression of intestinal epithelial cytoprotective heat shock proteins 25 and 72. *Am J Physiol Gastrointest Liver Physiol* **288** (2005).

384. H. V. Lin, *et al.*, Butyrate and propionate protect against diet-induced obesity and regulate gut hormones via free fatty acid receptor 3-independent mechanisms. *PLoS One* **7** (2012).

385. G. Frost, *et al.*, The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nat Commun* **5** (2014).

386. R. B. Canani, *et al.*, Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World J Gastroenterol* **17**, 1519–1528 (2011).

387. T. O. Keku, S. Dulal, A. Deveaux, B. Jovov, X. Han, The gastrointestinal microbiota and colorectal cancer. *Am J Physiol Gastrointest Liver Physiol* **308**, G351–G363 (2015).

388. W. E. W. Roediger, Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. *Gut* **21**, 793–798 (1980).

389. O. Warburg, F. Wind, E. Negelein, The metabolism of tumors in the body. *Journal of General Physiology* **8**, 519–530 (1927).

390. Y. Y. Fan, *et al.*, A bioassay to measure energy metabolism in mouse colonic crypts, organoids, and sorted stem cells. *Am J Physiol Gastrointest Liver Physiol* **309** (2015).

391. E. Zagato, *et al.*, Endogenous murine microbiota member Faecalibaculum rodentium and its human homologue protect from intestinal tumour growth. *Nat Microbiol* **5** (2020).

392. P. Louis, G. L. Hold, H. J. Flint, The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* **12**, 661–72 (2014).

393. W. Wang, *et al.*, Increased Proportions of Bifidobacterium and the Lactobacillus Group and Loss of Butyrate-Producing Bacteria in Inflammatory Bowel Disease. *J Clin Microbiol* **52**, 398–406 (2014).

394. K. Takahashi, *et al.*, Reduced Abundance of Butyrate-Producing Bacteria Species in the Fecal Microbial Community in Crohn's Disease. *Digestion* **93**, 59–65 (2016).

395. P. M. Smith, *et al.*, The microbial metabolites, short-chain fatty acids, regulate colonic T reg cell homeostasis. *Science (1979)* **341**, 569–573 (2013).

396. M. A. R. Vinolo, *et al.*, Suppressive effect of short-chain fatty acids on production of proinflammatory mediators by neutrophils. *Journal of Nutritional Biochemistry* **22**, 849–855 (2011).

397. C. J. Kelly, *et al.*, Crosstalk between microbiota-derived short-chain fatty acids and intestinal epithelial HIF augments tissue barrier function. *Cell Host Microbe* **17**, 662–671 (2015).

398. W. Wu, *et al.*, Microbiota metabolite short-chain fatty acid acetate promotes intestinal IgA response to microbiota which is mediated by GPR43. *Mucosal Immunol* **10**, 946–956 (2017).

399. J. Tan, *et al.*, Dietary Fiber and Bacterial SCFA Enhance Oral Tolerance and Protect against Food Allergy through Diverse Cellular Pathways. *Cell Rep* **15**, 2809–2824 (2016).

400. B. Hu, *et al.*, Microbiota-induced activation of epithelial IL-6 signaling links inflammasome-driven inflammation. *Proc Natl Acad Sci U S A* **110**, 9862–9867 (2013).

401. P. Louis, H. J. Flint, Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol* **19**, 29–41 (2017).

402. A. W. Walker, S. H. Duncan, E. Carol McWilliam Leitch, M. W. Child, H. J. Flint, pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Appl Environ Microbiol* **71**, 3692–3700 (2005).

403. S. H. Duncan, P. Louis, J. M. Thomson, H. J. Flint, The role of pH in determining the species composition of the human colonic microbiota. *Environ Microbiol* **11**, 2112–2122 (2009).

404. H. Kettle, P. Louis, G. Holtrop, S. H. Duncan, H. J. Flint, Modelling the emergent dynamics and major metabolites of the human colonic microbiota. *Environ Microbiol* **17**, 1615–1630 (2015).

405. M. Braun, S. Schoberth, G. Gottschalk, Enumeration of bacteria forming acetate from H2 and CO2 in anaerobic habitats. *Arch Microbiol* **120**, 201–4 (1979).

406. C. J. Kelly, S. P. Colgan, Breathless in the Gut: Implications of Luminal O2 for Microbial Pathogenicity. *Cell Host Microbe* **19**, 427–8 (2016).

407. M. X. Byndloss, *et al.*, Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science* **357**, 570–575 (2017).

408. K. Atarashi, *et al.*, Induction of Colonic Regulatory T Cells. *Science (1979)* **331**, 337–342 (2011).

409. P. M. Smith, *et al.*, The microbial metabolites, short chain fatty acids, regulate colonic Treg cell homeostasis. *Science (1979)* **341** (2013).

410. F. Rivera-Chávez, *et al.*, Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella. *Cell Host Microbe* **19**, 443–54 (2016).

411. K. M. Maslowski, *et al.*, Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–1286 (2009).

412. A. Ritzhaupt, I. S. Wood, A. Ellis, K. B. Hosie, S. P. Shirazi-Beechey, Identification and characterization of a monocarboxylate transporter (MCT1) in pig and human colon: its potential to transport L-lactate as well as butyrate. *J Physiol* **513 (Pt 3**, 719–32 (1998).

413. H. Al-mosauwi, *et al.*, Differential protein abundance of a basolateral MCT1 transporter in the human gastrointestinal tract. *Cell Biol Int* **40**, 1303–1312 (2016).

414. C. K. Garcia, J. L. Goldstein, R. K. Pathak, R. G. W. Anderson, M. S. Brown, Molecular characterization of a membrane transporter for lactate, pyruvate, and other monocarboxylates: Implications for the Cori cycle. *Cell* **76**, 865–873 (1994).

415. A. Kumar, W. A. Alrefai, A. Borthakur, P. K. Dudeja, Lactobacillus acidophilus counteracts enteropathogenic E. coli-induced inhibition of butyrate uptake in intestinal epithelial cells. *Am J Physiol Gastrointest Liver Physiol* **309**, G602–G607 (2015).

416. S. Sivaprakasam, Y. D. Bhutia, S. Yang, V. Ganapathy, Short-Chain Fatty Acid Transporters: Role in Colonic Homeostasis. *Compr Physiol* **8**, 299–314 (2018).

417. M. A. Cuff, S. P. Shirazi-Beechey, The human monocarboxylate transporter, MCT1: genomic organization and promoter analysis. *Biochem Biophys Res Commun* **292**, 1048–56 (2002).

418. J. R. Doherty, *et al.*, Blocking lactate export by inhibiting the Myc target MCT1 Disables glycolysis and glutathione synthesis. *Cancer Res* **74**, 908–20 (2014).

419. C. Hadjiagapiou, *et al.*, Role of USF1 and USF2 as potential repressor proteins for human intestinal monocarboxylate transporter 1 promoter. *Am J Physiol Gastrointest Liver Physiol* **288**, G1118-26 (2005).

420. S. Sprowl-Tanio, *et al.*, Lactate/pyruvate transporter MCT-1 is a direct Wnt target that confers sensitivity to 3-bromopyruvate in colon cancer. *Cancer Metab* **4** (2016).

421. N. Singh, *et al.*, Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* **40**, 128–139 (2014).

422. M. Thangaraju, *et al.*, GPFM 09A 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).

423. A. Agus, *et al.*, Western diet induces a shift in microbiota composition enhancing susceptibility to Adherent-Invasive E. coli infection and intestinal inflammation. *Sci Rep* **6**, 19032 (2016).

424. D. Haenen, *et al.*, A Diet High in Resistant Starch Modulates Microbiota Composition, SCFA Concentrations, and Gene Expression in Pig Intestine. *J Nutr* **143**, 274–283 (2013).

425. A. Koh, F. De Vadder, P. Kovatcheva-Datchary, F. Bäckhed, From dietary fiber to host physiology: Short-chain fatty acids as key bacterial metabolites. *Cell* **165**, 1332–1345 (2016).

426. F. Sommer, *et al.*, Altered mucus glycosylation in core 1 O-glycan-deficient mice affects microbiota composition and intestinal architecture. *PLoS One* **9** (2014).

427. A. J. MacPherson, E. Slack, M. B. Geuking, K. D. McCoy, The mucosal firewalls against commensal intestinal microbes. *Semin Immunopathol* **31**, 145–149 (2009).

428. M. E. V. Johansson, *et al.*, The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc Natl Acad Sci U S A* **105**, 15064–15069 (2008).

429. C. K. Heazlewood, *et al.*, Aberrant mucin assembly in mice causes endoplasmic reticulum stress and spontaneous inflammation resembling ulcerative colitis. *PLoS Med* **5**, 0440–0460 (2008).

430. M. Wlodarska, *et al.*, Indoleacrylic Acid Produced by Commensal Peptostreptococcus Species Suppresses Inflammation. *Cell Host Microbe* **22**, 25-37.e6 (2017).

431. J. Reunanen, *et al.*, Akkermansia muciniphila adheres to enterocytes and strengthens the integrity of the epithelial cell layer. *Appl Environ Microbiol* **81**, 3655–3662 (2015).

432. K. Atarashi, Y. Umesaki, K. Honda, Microbiotal influence on T cell subset development. *Semin Immunol* **23**, 146–53 (2011).

433. R. J. Perry, *et al.*, Acetate mediates a microbiome-brain-β-cell axis to promote metabolic syndrome. *Nature* **534**, 213–217 (2016).

434. M. Trajkovski, C. B. Wollheim, Physiology: Microbial signals to the brain control weight. *Nature* **534**, 185–187 (2016).

435. C. M. van der Beek, *et al.*, Distal, not proximal, colonic acetate infusions promote fat oxidation and improve metabolic markers in overweight/obese men. *Clin Sci* **130**, 2073–2082 (2016).

436. R. Sender, S. Fuchs, R. Milo, Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol* **14** (2016).

437. M. B. Miller, B. L. Bassler, Quorum Sensing in Bacteria. *Annu Rev Microbiol* **55**, 165–199 (2001).

438. L. Lin, R. X. Tan, Cross-kingdom actions of phytohormones: A functional scaffold exploration. *Chem Rev* **111**, 2734–2760 (2011).

439. K. P. Rumbaugh, G. F. Kaufmann, Exploitation of host signaling pathways by microbial quorum sensing signals. *Curr Opin Microbiol* **15**, 162–168 (2012).

440. A. Bandyopadhaya, *et al.*, The quorum sensing volatile molecule 2-amino acetophenon modulates host immune responses in a manner that promotes life with unwanted guests. *PLoS Pathog* **8**, e1003024 (2012).

441. K. Papenfort, B. L. Bassler, Quorum sensing signal-response systems in Gram-negative bacteria. *Nat Rev Microbiol* **14**, 576–588 (2016).

442. R. P. Novick, E. Geisinger, Quorum Sensing in Staphylococci. *Annu Rev Genet* **42**, 541–564 (2008).

443. C. M. Waters, B. L. Bassler, QUORUM SENSING: Cell-to-Cell Communication in Bacteria. *Annu Rev Cell Dev Biol* **21**, 319–346 (2005).

444. V. V. Kravchenko, *et al.*, Modulation of gene expression via disruption of NF-κB signaling by a bacterial small molecule. *Science (1979)* **321**, 259–263 (2008).

445. T. Karlsson, F. Musse, K.-E. Magnusson, E. Vikström, N -Acylhomoserine lactones are potent neutrophil chemoattractants that act via calcium mobilization and actin remodeling . *J Leukoc Biol* **91**, 15–26 (2012).

446. K. Okamoto, *et al.*, Competence and sporulation factor derived from Bacillus subtilis improves epithelial cell injury in intestinal inflammation via immunomodulation and cytoprotection. *Int J Colorectal Dis* **27**, 1039–1046 (2012).

447. A. Zargar, *et al.*, Bacterial secretions of nonpathogenic Escherichia coli elicit inflammatory pathways: A closer investigation of interkingdom signaling. *mBio* **6** (2015).

448. J. L. Coombes, *et al.*, A functionally specialized population of mucosal CD103+ DCs induces Foxp3+ regulatory T cells via a TGF-β -and retinoic acid-dependent mechanism. *Journal of Experimental Medicine* **204**, 1757–1764 (2007).

449. J. Dai, R. J. Mumper, Plant phenolics: Extraction, analysis and their antioxidant and anticancer properties. *Molecules* **15**, 7313–7352 (2010).

450. C. A. Rice-evans, N. J. Miller, P. G. Bolwell, P. M. Bramley, J. B. Pridham, The relative antioxidant activities of plant-derived polyphenolic flavonoids. *Free Radic Res* **22**, 375–383 (1995).

451. S. Banihani, S. Swedan, Z. Alguraan, Pomegranate and type 2 diabetes. *Nutrition Research* **33**, 341–348 (2013).

452. X. Zheng, *et al.*, The footprints of gut microbial-mammalian co-metabolism. *J Proteome Res* **10**, 5512–5522 (2011).

453. C. Fernández, E. Díaz, J. L. García, Insights on the regulation of the phenylacetate degradation pathway from Escherichia coli. *Environ Microbiol Rep* **6**, 239–250 (2014).

454. R. Teufel, *et al.*, Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc Natl Acad Sci U S A* **107**, 14390–14395 (2010).

455. R. Teufel, T. Friedrich, G. Fuchs, An oxygenase that forms and deoxygenates toxic epoxide. *Nature* **483**, 359–362 (2012).

456. J. M. Navarro-Llorens, *et al.*, Phenylacetate catabolism in Rhodococcus sp. strain RHA1: A central pathway for degradation of aromatic compounds. *J Bacteriol* **187**, 4497–4504 (2005).

457. A. Kabir, G. Suresh Kumar, Binding of the Biogenic Polyamines to Deoxyribonucleic Acids of Varying Base Composition: Base Specificity and the Associated Energetics of the Interaction. *PLoS One* **8**, e70510 (2013).

458. N. Guan, L. Liu, Microbial response to acid stress: mechanisms and applications. *Appl Microbiol Biotechnol* **104**, 51–65 (2020).

459. C. W. Tabor, H. Tabor, Polyamines in microorganisms. *Microbiol Rev* **49**, 81–99 (1985).

460. A. Álvarez-Ordóñez, A. Fernández, A. Bernardo, M. López, Arginine and lysine decarboxylases and the Acid Tolerance Response of Salmonella Typhimurium. *Int J Food Microbiol* **136**, 278–282 (2010).

461. J. W. Foster, Escherichia coli acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol* **2**, 898–907 (2004).

462. C. C. Hanfrey, *et al.*, Alternative spermidine biosynthetic route is critical for growth of Campylobacter jejuni and is the dominant polyamine pathway in human gut microbiota. *Journal of Biological Chemistry* **286**, 43301–43312 (2011).

463. K. Igarashi, K. Kashiwagi, Characteristics of cellular polyamine transport in prokaryotes and eukaryotes. *Plant Physiology and Biochemistry* **48**, 506–512 (2010).

464. G. Iacomino, G. Picariello, L. D'Agostino, DNA and nuclear aggregates of polyamines. *Biochim Biophys Acta Mol Cell Res* **1823**, 1745–1755 (2012).

465. A. E. Pegg, Mammalian polyamine metabolism and function. *IUBMB Life* **61**, 880–894 (2009).

466. K. Igarashi, K. Kashiwagi, Effects of polyamines on protein synthesis and growth of Escherichia coli. *Journal of Biological Chemistry* **293**, 18702–18709 (2018).

467. K. Igarashi, K. Kashiwagi, Modulation of cellular function by polyamines. *International Journal of Biochemistry and Cell Biology* **42**, 39–51 (2010).

468. T. M. Fletcher, J. C. Hansen, The nucleosomal array: Structure/function relationships. *Crit Rev Eukaryot Gene Expr* **6**, 149–188 (1996).

469. K. J. Pollard, M. L. Samuels, K. A. Crowley, J. C. Hansen, C. L. Peterson, "Functional interaction between GCN5 and polyamines: a new role for core histone acetylation" (1999).

470. L. Liu, *et al.*, Activation of TGF-β-Smad signaling pathway following polyamine depletion in intestinal epithelial cells. *Am J Physiol Gastrointest Liver Physiol* **285** (2003).

471. M. Barbagallo, *et al.*, A new piece of the shigella pathogenicity puzzle: Spermidine accumulationby silencing of the speg gene. *PLoS One* **6** (2011).

472. K. A. Bliven, A. T. Maurelli, Antivirulence genes: Insights into pathogen evolution through gene loss. *Infect Immun* **80**, 4061–4070 (2012).

473. M. Linsalata, C. Messa, F. Russo, A. Cavallini, A. Dileo, Estrogen receptors and polyamine levels in human gastric carcinoma. *Scand J Gastroenterol* **29**, 67–70 (1994).

474. T. Thomas, T. J. Thomas, Polyamines in cell growth and cell death: Molecular mechanisms and therapeutic applications. *Cellular and Molecular Life Sciences* **58**, 244–258 (2001).

475. H. M. Wallace, R. Caslake, Polyamines and colon cancer. *Eur J Gastroenterol Hepatol* **13**, 1033–1039 (2001).

476. T. Thomas, T. J. Thomas, Polyamine metabolism and cancer. *J Cell Mol Med* **7**, 113–126 (2003).

477. J. A. Nilsson, *et al.*, Targeting ornithine decarboxylase in Myc-induced lymphomagenesis prevents tumor formation. *Cancer Cell* **7**, 433–444 (2005).

478. A. D. L. M Linsalata, A Cavallini, Polyamine Oxidase Activity and Polyamine Levels in Human Colorectal Cancer and in Normal Surrounding Mucosa. *Anticancer Res* **17**, 3757–3760 (1997).

479. M. Linsalata, *et al.*, Polyamine Biosynthesis in relation to K-ras and p-53 mutations in colorectal carcinoma. *Scand J Gastroenterol* **39**, 470–477 (2004).

480. M. Linsalata, *et al.*, Prognostic value of tissue polyamine levels in human colorectal carcinoma. *Anticancer Res* **22**, 2465–9 (2002).

481. Aitor Blanco-Miguez, *et al.*, Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4. *Preprint* (2022) https:/doi.org/10.1101/2022.08.22.504593.

482. T. Magoc, S. L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

483. J. G. Caporaso, *et al.*, QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336 (2010).

484. R. C. Edgar, B. J. Haas, J. C. Clemente, C. Quince, R. Knight, UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).

485. B. J. Haas, *et al.*, Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**, 494–504 (2011).

486. R. C. Edgar, UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**, 996–998 (2013).

487. Q. Wang, G. M. Garrity, J. M. Tiedje, J. R. Cole, Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol* **73**, 5261–5267 (2007).

488. C. Quast, *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590–D596 (2012).

489. S. T. Westreich, M. L. Treiber, D. A. Mills, I. Korf, D. G. Lemay, SAMSA2: a standalone metatranscriptome analysis pipeline. *BMC Bioinformatics* **19**, 175 (2018).

490. J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).

491. E. Kopylova, *et al.*, Open-Source Sequence Clustering Methods Improve the State Of the Art. *mSystems* **1** (2016).

492. N. A. O'Leary, *et al.*, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–D745 (2016).

493. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59–60 (2015).

494. B. Buchfink, K. Reuter, H.-G. Drost, Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368 (2021).

495. R. Overbeek, *et al.*, The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42**, D206–D214 (2014).

496. D. H. Parks, G. W. Tyson, P. Hugenholtz, R. G. Beiko, STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**, 3123–4 (2014).

497. J. R. White, N. Nagarajan, M. Pop, Statistical Methods for Detecting Differentially Abundant Features in Clinical Metagenomic Samples. *PLoS Comput Biol* **5**, e1000352 (2009).

498. E. Bolyen, *et al.*, Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852–857 (2019).

499. B. J. Callahan, *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).

500. N. A. Bokulich, *et al.*, Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).

501. V. Heidrich, P. Karlovsky, L. Beule, 'SRS' R Package and 'q2-srs' QIIME 2 Plugin: Normalization of Microbiome Data Using Scaling with Ranked Subsampling (SRS). *Applied Sciences* **11**, 11473 (2021).

502. Jari Oksanen, *et al.*, vegan: Community Ecology Package (2022).

503. C. E. Shannon, A Mathematical Theory of Communication. *Bell System Technical Journal* **27**, 379–423 (1948).

504. E. H. Simpson, Measurement of Diversity. *Nature* **163**, 688–688 (1949).

505. J. R. Bray, J. T. Curtis, An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr* **27**, 325–349 (1957).

506. G. Muyzer, E. C. de Waal, A. G. Uitterlinden, Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**, 695–700 (1993).

507. B. P. Alcock, *et al.*, CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* (2019) https:/doi.org/10.1093/nar/gkz935.

508. A. Ullmann, Complementation in beta-galactosidase: from protein structure to genetic engineering. *Bioessays* **14**, 201–205 (1992).

509. B. P. Alcock, *et al.*, CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* (2019) https:/doi.org/10.1093/nar/gkz935.

510. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

511. D. E. Madden, *et al.*, Keeping up with the pathogens: Improved antimicrobial resistance detection and prediction in Pseudomonas aeruginosa. *medRxiv*, 2022.08.11.22278689 (2022).

512. Q. Sheng, *et al.*, Multi-perspective quality control of Illumina RNA sequencing data analysis. *Brief Funct Genomics*, elw035 (2016).

513. Matthew T. F. Lamaudière, Ramesh Arasaradnam, Gareth D. Weedall, Igor Y. Morozov, The colorectal cancer gut environment regulates activity of the microbiome and promotes the multidrug resistant phenotype of ESKAPE and other pathogens . *mSphere* (2023).

514. M. Lopez-Siles, *et al.*, Cultured Representatives of Two Major Phylogroups of Human Colonic Faecalibacterium prausnitzii Can Utilize Pectin, Uronic Acids, and Host-Derived Substrates for Growth. *Appl Environ Microbiol* **78**, 420–428 (2012).

515. T. Ohkusa, *et al.*, Commensal bacteria can enter colonic epithelial cells and induce proinflammatory cytokine secretion: a possible pathogenic mechanism of ulcerative colitis. *J Med Microbiol* **58**, 535–545 (2009).

516. U. Wegmann, *et al.*, Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota ('Ruminococcus bicirculans') reveals two chromosomes and a selective capacity to utilize plant glucans. *Environ Microbiol* **16**, 2879–2890 (2014).

517. L. B. Rice, Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. *J Infect Dis* **197**, 1079–1081 (2008).

518. N. Segata, *et al.*, Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9**, 811–814 (2012).

519. M. Bergkessel, *et al.*, The dormancy-specific regulator, SutA, is intrinsically disordered and modulates transcription initiation in Pseudomonas aeruginosa. *Mol Microbiol* **112**, 992–1009 (2019).

520. J. A. Bravo, *et al.*, Communication between gastrointestinal bacteria and the nervous system. *Curr Opin Pharmacol* **12**, 667–672 (2012).

521. S. Duranti, *et al.*, Bifidobacterium adolescentis as a key member of the human gut microbiota in the production of GABA. *Sci Rep* **10**, 14112 (2020).

522. M. Urbańska, D. Gieruszczak-Białek, H. Szajewska, Systematic review with meta-analysis: Lactobacillus reuteri DSM 17938 for diarrhoeal diseases in children. *Aliment Pharmacol Ther* **43**, 1025–1034 (2016).

523. A. Pfleiderer, *et al.*, Culturomics identified 11 new bacterial species from a single anorexia nervosa stool sample. *European Journal of Clinical Microbiology & Infectious Diseases* **32**, 1471–1481 (2013).

524. S. H. Wong, J. Yu, Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat Rev Gastroenterol Hepatol* **16**, 690–704 (2019).

525. L. Petit, M. Gibert, M. R. Popoff, Clostridium perfringens: toxinotype and genotype. *Trends Microbiol* **7**, 104–110 (1999).

526. M. T. F. Lamaudière, R. Arasaradnam, G. D. Weedall, I. Y. Morozov, The colorectal cancer microbiota alter their transcriptome to adapt to the acidity, reactive oxygen species and metabolite availability of gut microenvironments. *mSphere* (2023).

527. M. J. Bibb, Understanding and manipulating antibiotic production in actinomycetes. *Biochem Soc Trans* **41**, 1355–1364 (2013).

528. M. T. Alou, *et al.*, Inediibacterium massiliense gen. nov., sp. nov., a new bacterial species isolated from the gut microbiota of a severely malnourished infant. *Antonie Van Leeuwenhoek* **110**, 737–750 (2017).

529. Q. Feng, *et al.*, Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat Commun* **6**, 6528 (2015).

530. S. Doron, D. R. Snydman, Risk and Safety of Probiotics. *Clinical Infectious Diseases* **60**, S129–S134 (2015).

531. M. M. O' Donnell, B. M. Forde, B. Neville, P. R. Ross, P. W. O' Toole, Carbohydrate catabolic flexibility in the mammalian intestinal commensal Lactobacillus ruminis revealed by fermentation studies aligned to genome annotations. *Microb Cell Fact* **10**, S12 (2011).

532. L. K. Dropulic, H. M. Lederman, Overview of Infections in the Immunocompromised Host. *Microbiol Spectr* **4** (2016).

533. M. Bassetti, M. Peghin, D. Pecori, The management of multidrug-resistant Enterobacteriaceae. *Curr Opin Infect Dis* **29**, 583–594 (2016).

534. M. S. Mulani, E. E. Kamble, S. N. Kumkar, M. S. Tawre, K. R. Pardesi, Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review. *Front Microbiol* **10** (2019).

535. S. Octavia, *et al.*, Klebsiella pneumoniae and Klebsiella quasipneumoniae define the population structure of blaKPC-2Klebsiella: a 5 year retrospective genomic study in Singapore. *Journal of Antimicrobial Chemotherapy* **74**, 3205–3210 (2019).

536. T. Irrazábal, A. Belcheva, S. E. Girardin, A. Martin, D. J. Philpott, The Multifaceted Role of the Intestinal Microbiota in Colon Cancer. *Mol Cell* **54**, 309–320 (2014).

537. M. S. Attene-Ramos, E. D. Wagner, H. R. Gaskins, M. J. Plewa, Hydrogen Sulfide Induces Direct Radical-Associated DNA Damage. *Molecular Cancer Research* **5**, 455–459 (2007).

538. J.-P. Motta, *et al.*, Hydrogen Sulfide Protects from Colitis and Restores Intestinal Microbiota Biofilm and Mucus Production. *Inflamm Bowel Dis* **21**, 1006–1017 (2015).

539. J. Xu, *et al.*, Response mechanisms to acid stress of acid-resistant bacteria and biotechnological applications in the food industry. *Crit Rev Biotechnol*, 1–17 (2022).

540. Z. Dai, *et al.*, Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome* **6**, 70 (2018).

541. I. Koliarakis, *et al.*, Oral Bacteria and Intestinal Dysbiosis in Colorectal Cancer. *Int J Mol Sci* **20**, 4146 (2019).

542. L. Wu, *et al.*, Gut microbiota as an antioxidant system in centenarians associated with high antioxidant activities of gut-resident Lactobacillus. *NPJ Biofilms Microbiomes* **8**, 102 (2022).

543. I. Munteanu, B. Mastalier, Genetics of colorectal cancer. *J Med Life* **7**, 507–11.

544. S. Brands, P. Schein, K. F. Castro-Ochoa, E. A. Galinski, Hydroxyl radical scavenging of the compatible solute ectoine generates two N-acetimides. *Arch Biochem Biophys* **674**, 108097 (2019).

545. M. Camilleri, Serotonin in the gastrointestinal tract. *Curr Opin Endocrinol Diabetes Obes* **16**, 53–59 (2009).

546. O. Atakisi, *et al.*, Effects of reduced glutathione on nitric oxide level, total antioxidant and oxidant capacity and adenosine deaminase activity. *Eur Rev Med Pharmacol Sci* **14**, 19–23 (2010).

547. Md. S. Alam, S. K. Garg, P. Agrawal, Studies on structural and functional divergence among seven WhiB proteins of Mycobacterium tuberculosis H37Rv. *FEBS Journal* **276**, 76–93 (2009).

548. J. N. Blaza, *et al.*, The mechanism of catalysis by type-II NADH:quinone oxidoreductases. *Sci Rep* **7**, 40165 (2017).

549. M. Ashoori, A. Saedisomeolia, Riboflavin (vitamin B2) and oxidative stress: a review. *British Journal of Nutrition* **111**, 1985–1991 (2014).

550. O. Almarza, D. Núñez, H. Toledo, The DNA-Binding Protein HU has a Regulatory Role in the Acid Stress Response Mechanism in Helicobacter pylori. *Helicobacter* **20**, 29–40 (2015).

551. T. Nunoshiba, T. deRojas-Walker, J. S. Wishnok, S. R. Tannenbaum, B. Demple, Activation by nitric oxide of an oxidative-stress response that defends Escherichia coli against activated macrophages. *Proceedings of the National Academy of Sciences* **90**, 9993–9997 (1993).

552. Y. Fichman, *et al.*, Selenoprotein O is a chloroplast protein involved in ROS scavenging and its absence increases dehydration tolerance in Arabidopsis thaliana. *Plant Science* **270**, 278–291 (2018).

553. K. L. Seib, *et al.*, Defenses against Oxidative Stress in Neisseria gonorrhoeae: a System Tailored for a Challenging Environment. *Microbiology and Molecular Biology Reviews* **70**, 344–361 (2006).

554. H. Ohno, *et al.*, The effects of reactive nitrogen intermediates on gene expression in Mycobacterium tuberculosis. *Cell Microbiol* **5**, 637–648 (2003).

555. S. G. NUGENT, Intestinal luminal pH in inflammatory bowel disease: possible determinants and implications for therapy with aminosalicylates and other drugs. *Gut* **48**, 571–577 (2001).

556. M. T. F. Lamaudière, R. Arasaradnam, G. D. Weedall, I. Y. Morozov, "The gut microbiome and colorectal cancer: I. Groups of pathogens gain while commensals lose activity" (2022).

557. T. A. Krulwich, G. Sachs, E. Padan, Molecular aspects of bacterial pH sensing and homeostasis. *Nat Rev Microbiol* **9**, 330–343 (2011).

558. H. Liu, *et al.*, Butyrate: A Double-Edged Sword for Health? *Advances in Nutrition* **9**, 21–29 (2018).

559. B.-K. Cho, *et al.*, The PurR regulon in Escherichia coli K-12 MG1655. *Nucleic Acids Res* **39**, 6456–6464 (2011).

560. Y. Yuan, *et al.*, Discovery of novel bacterial queuine salvage enzymes and pathways in human pathogens. *Proceedings of the National Academy of Sciences* **116**, 19126–19135 (2019).

561. S. I. Kim, Roles of YehZ, a Putative Osmoprotectant Transporter, in Tempering Growth of Salmonella enterica serovar Typhimurium. *J Microbiol Biotechnol* **23**, 1560–1568 (2013).

562. U. H. Beier, T. Görögh, Implications of galactocerebrosidase and galactosylcerebroside metabolism in cancer cells. *Int J Cancer* **115**, 6–10 (2005).

563. M. Hofreiter, D. Serre, H. N. Poinar, M. Kuch, S. Pääbo, Ancient DNA. *Nat Rev Genet* **2**, 353–359 (2001).

564. L. Hall-Stoodley, J. W. Costerton, P. Stoodley, Bacterial biofilms: from the Natural environment to infectious diseases. *Nat Rev Microbiol* **2**, 95–108 (2004).

565. R. Baghban, *et al.*, Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Communication and Signaling* **18**, 59 (2020).

566. K. M. Pruss, *et al.*, Oxidative ornithine metabolism supports non-inflammatory C. difficile colonization. *Nat Metab* **4**, 19–28 (2022).

567. M. Orlowski, A. Meister, The γ-Glutamyl Cycle: A Possible Transport System for Amino Acids. *Proceedings of the National Academy of Sciences* **67**, 1248–1255 (1970).

568. A. Boehm, *et al.*, Second messenger signalling governs Escherichia coli biofilm induction upon ribosomal stress. *Mol Microbiol* **72**, 1500–1516 (2009).

569. T. Arcondéguy, R. Jack, M. Merrick, P II Signal Transduction Proteins, Pivotal Players in Microbial Nitrogen Control. *Microbiology and Molecular Biology Reviews* **65**, 80–105 (2001).

570. T. P. N. Bui, *et al.*, Conversion of dietary inositol into propionate and acetate by commensal Anaerostipes associates with host health. *Nat Commun* **12**, 4798 (2021).

571. M.-P. Castanie-Cornet, T. A. Penfound, D. Smith, J. F. Elliott, J. W. Foster, Control of Acid Resistance in Escherichia coli. *J Bacteriol* **181**, 3525–3535 (1999).

572. Y. Liu, H. Tang, Z. Lin, P. Xu, Mechanisms of acid tolerance in bacteria and prospects in biotechnology and bioremediation. *Biotechnol Adv* **33**, 1484–1492 (2015).

573. K. O'Riordan, J. C. Lee, Staphylococcus aureus Capsular Polysaccharides. *Clin Microbiol Rev* **17**, 218–234 (2004).

574. M. Rajagopal, S. Walker, "Envelope Structures of Gram-Positive Bacteria" in (2015), pp. 1–44.

575. H. S. Courtney, *et al.*, Lipoteichoic acid and M protein: dual adhesins of group A streptococci. *Microb Pathog* **12**, 199–208 (1992).

576. J. Wang, *et al.*, The role of the type VI secretion system vgrG gene in the virulence and antimicrobial resistance of Acinetobacter baumannii ATCC 19606. *PLoS One* **13**, e0192288 (2018).

577. J. Vogel-Scheel, C. Alpert, W. Engst, G. Loh, M. Blaut, Requirement of Purine and Pyrimidine Synthesis for Colonization of the Mouse Intestine by Escherichia coli. *Appl Environ Microbiol* **76**, 5181–5187 (2010).

578. Q. Gao, *et al.*, Roles of iron acquisition systems in virulence of extraintestinal pathogenic Escherichia coli: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiol* **12**, 143 (2012).

579. H. S. Chidwick, M. A. Fascione, Mechanistic and structural studies into the biosynthesis of the bacterial sugar pseudaminic acid (Pse5Ac7Ac). *Org Biomol Chem* **18**, 799–809 (2020).

580. C. Pesavento, *et al.*, Inverse regulatory coordination of motility and curli-mediated adhesion in Escherichia coli. *Genes Dev* **22**, 2434–2446 (2008).

581. G. A. Turnbull, J. A. W. Morgan, J. M. Whipps, J. R. Saunders, The role of bacterial motility in the survival and spread of Pseudomonas fluorescens in soil and in the attachment and colonisation of wheat roots. *FEMS Microbiol Ecol* **36**, 21–31 (2001).

582. E. A. Raleigh, K. B. Low, "Conjugation" in *Brenner's Encyclopedia of Genetics*, (Elsevier, 2013), pp. 144–151.

583. F. Hille, E. Charpentier, CRISPR-Cas: biology, mechanisms and relevance. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150496 (2016).

584. N. I. Kunonga, R. J. Sobieski, S. S. Crupper, Prevalence of the multiple antibiotic resistance operon (marRAB) in the genus Salmonella. *FEMS Microbiol Lett* **187**, 155–160 (2000).

585. N. Allali, H. Afif, M. Couturier, L. van Melderen, The Highly Conserved TldD and TldE Proteins of Escherichia coli Are Involved in Microcin B17 Processing and in CcdA Degradation. *J Bacteriol* **184**, 3224–3231 (2002).

586. A. Pourbaix, *et al.*, Unexpected Activity of Oral Fosfomycin against Resistant Strains of Escherichia coli in Murine Pyelonephritis. *Antimicrob Agents Chemother* **63** (2019).

587. C. Schäffer, P. Messner, Emerging facets of prokaryotic glycosylation. *FEMS Microbiol Rev* **41**, 49–91 (2017).

588. A. Ebihara, *et al.*, Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Science* **15**, 1494–1499 (2006).

589. M. L. Faron, N. A. Ledeboer, B. W. Buchan, Resistance Mechanisms, Epidemiology, and Approaches to Screening for Vancomycin-Resistant Enterococcus in the Health Care Setting. *J Clin Microbiol* **54**, 2436–2447 (2016).

590. L. Vinué, L. M. McMurry, S. B. Levy, The 216-bp marB gene of the marRAB operon in Escherichia coli encodes a periplasmic protein which reduces the transcription rate of marA. *FEMS Microbiol Lett* **345**, 49–55 (2013).

591. J. A. N. Alexander, *et al.*, Structural analysis of avibactam-mediated activation of the bla and mec divergons in methicillin-resistant Staphylococcus aureus. *Journal of Biological Chemistry* **295**, 10870–10884 (2020).

592. S. R. Chhabra, *et al.*, "Extracellular Communication in Bacteria" in (2004), pp. 279–315.

593. J. W. Little, "Lysogeny, Prophage Induction, and Lysogenic Conversion" in *Phages*, (ASM Press, 2014), pp. 37–54.

594. N. Babbar, E. W. Gerner, "Targeting Polyamines and Inflammation for Cancer Prevention" in (2010), pp. 49–64.

595. A. Taj, *et al.*, Functional analysis and enzyme characterization of mannose-1-phosphate guanylyl transferase (ManB) from Mycobacterium tuberculosis. *Res Microbiol* **173**, 103884 (2022).

596. R. Al-Sadi, *et al.*, Interleukin-6 Modulation of Intestinal Epithelial Tight Junction Permeability Is Mediated by JNK Pathway Activation of Claudin-2 Gene. *PLoS One* **9**, e85345 (2014).

597. M. Nguyen, *et al.*, The fermentation product 2,3-butanediol alters P. aeruginosa clearance, cytokine response and the lung microbiome. *ISME J* **10**, 2978–2983 (2016).

598. W. L. Nicholson, The Bacillus subtilis ydjL (bdhA) Gene Encodes Acetoin Reductase/2,3-Butanediol Dehydrogenase. *Appl Environ Microbiol* **74**, 6832–6838 (2008).

599. H. Zafar, M. H. Saier, Gut Bacteroides species in health and disease. *Gut Microbes* **13** (2021).

600. J. Cerning, Exocellular polysaccharides produced by lactic acid bacteria. *FEMS Microbiol Lett* **87**, 113–130 (1990).

601. J. Ahn, *et al.*, Human Gut Microbiome and Risk for Colorectal Cancer. *JNCI: Journal of the National Cancer Institute* **105**, 1907–1911 (2013).

602. P. B. Eckburg, *et al.*, Diversity of the Human Intestinal Microbial Flora. *Science (1979)* **308**, 1635–1638 (2005).

603. D. B. Kell, E. Pretorius, No effects without causes: the Iron Dysregulation and Dormant Microbes hypothesis for chronic, inflammatory diseases. *Biological Reviews* **93**, 1518–1557 (2018).

604. A. M. Thomas, *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).

605. Y. Wang, *et al.*, Eubacterium rectale contributes to colorectal cancer initiation via promoting colitis. *Gut Pathog* **13**, 2 (2021).

606. M.-Z. Jin, W.-L. Jin, The updated landscape of tumor microenvironment and drug repurposing. *Signal Transduct Target Ther* **5**, 166 (2020).

607. J. Naughton, G. Duggan, B. Bourke, M. Clyne, Interaction of microbes with mucus and mucins. *Gut Microbes* **5**, 48–52 (2014).

608. Y.-I. Kim, Role of Folate in Colon Cancer Development and Progression. *J Nutr* **133**, 3731S-3739S (2003).

609. G. R. Wasson, *et al.*, Global DNA and p53 Region-Specific Hypomethylation in Human Colonic Cells Is Induced by Folate Depletion and Reversed by Folate Supplementation. *J Nutr* **136**, 2748–2753 (2006).

610. Y. Zhao, C. Wang, A. Goel, Role of gut microbiota in epigenetic regulation of colorectal Cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1875**, 188490 (2021).

611. B. W. Wolf, K. A. Garleb, D. G. Ataya, I. A. Casas, Safety and Tolerance of Lactobacillus reuteri in Healthy Adult Male Subjects. *Microb Ecol Health Dis* **8**, 41–50 (1995).

612. A. C. Ericsson, C. L. Franklin, Manipulating the Gut Microbiota: Methods and Challenges. *ILAR J* **56**, 205–217 (2015).

613. C. Zhang, *et al.*, Transfer efficiency and impact on disease phenotype of differing methods of gut microbiota transfer. *Sci Rep* **12**, 19621 (2022).

614. E. Buc, *et al.*, High Prevalence of Mucosa-Associated E. coli Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS One* **8**, e56964 (2013).

615. M. Bonnet, *et al.*, Colonization of the Human Gut by E. coli and Colorectal Cancer Risk. *Clinical Cancer Research* **20**, 859–867 (2014).

616. Z. Gao, B. Guo, R. Gao, Q. Zhu, H. Qin, Microbiota disbiosis is associated with colorectal cancer. *Front Microbiol* **6** (2015).

617. A. D. S. Cameron, R. J. Redfield, Non-canonical CRP sites control competence regulons in Escherichia coli and many other γ-proteobacteria. *Nucleic Acids Res* **34**, 6001–6014 (2006).

618. L. Spiga, *et al.*, An Oxidative Central Metabolism Enables Salmonella to Utilize Microbiota-Derived Succinate. *Cell Host Microbe* **22**, 291-301.e6 (2017).

619. M. J. A. van Hoek, R. M. H. Merks, Emergence of microbial diversity due to cross-feeding interactions in a spatial model of gut microbial metabolism. *BMC Syst Biol* **11**, 56 (2017).

620. R. Pifer, R. M. Russell, A. Kumar, M. M. Curtis, V. Sperandio, Redox, amino acid, and fatty acid metabolism intersect with bacterial virulence in the gut. *Proceedings of the National Academy of Sciences* **115** (2018).

621. N. V. Bhagavan, C.-E. Ha, "Carbohydrate Metabolism II" in *Essentials of Medical Biochemistry*, (Elsevier, 2015), pp. 205–225.

622. Z. Deng, *et al.*, Glyceraldehyde-3-Phosphate Dehydrogenase Increases the Adhesion of Lactobacillus reuteri to Host Mucin to Enhance Probiotic Effects. *Int J Mol Sci* **21**, 9756 (2020).

623. D. M. P. de Oliveira, *et al.*, Antimicrobial Resistance in ESKAPE Pathogens. *Clin Microbiol Rev* **33** (2020).

624. T. D. Lawley, W. A. Klimke, M. J. Gubbins, L. S. Frost, F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* **224**, 1–15 (2003).

625. C. Johnston, B. Martin, G. Fichant, P. Polard, J.-P. Claverys, Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* **12**, 181–196 (2014).

626. R. Meyer, Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* **62**, 57–70 (2009).

627. A. Marin-Hernandez, J. Gallardo-Perez, S. Ralph, S. Rodriguez-Enriquez, R. Moreno-Sanchez, HIF-1a Modulates Energy Metabolism in Cancer Cells by Inducing Over-Expression of Specific Glycolytic Isoforms. *Mini-Reviews in Medicinal Chemistry* **9**, 1084–1101 (2009).

628. M. de Wit, *et al.*, Cell surface proteomics identifies glucose transporter type 1 and prion protein as candidate biomarkers for colorectal adenoma-to-carcinoma progression. *Gut* **61**, 855–864 (2012).

629. Y. Litvak, M. X. Byndloss, A. J. Bäumler, Colonocyte metabolism shapes the gut microbiota. *Science (1979)* **362**, eaat9076 (2018).

630. A. el Kaoutari, F. Armougom, J. I. Gordon, D. Raoult, B. Henrissat, The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* **11**, 497–504 (2013).

631. J. H. Cummings, H. N. Englyst, Fermentation in the human large intestine and the available substrates. *Am J Clin Nutr* **45**, 1243–1255 (1987).

632. J. C. Mell, R. J. Redfield, Natural Competence and the Evolution of DNA Uptake Specificity. *J Bacteriol* **196**, 1471–1483 (2014).

633. C. B. Whitchurch, T. Tolker-Nielsen, P. C. Ragas, J. S. Mattick, Extracellular DNA Required for Bacterial Biofilm Formation. *Science (1979)* **295**, 1487–1487 (2002).

634. N. Høiby, T. Bjarnsholt, M. Givskov, S. Molin, O. Ciofu, Antibiotic resistance of bacterial biofilms. *Int J Antimicrob Agents* **35**, 322–332 (2010).

635. F. Carbonero, A. C. Benefiel, A. H. Alizadeh-Ghamsari, H. R. Gaskins, Microbial pathways in colonic sulfur metabolism and links with health and disease. *Front Physiol* **3** (2012).

636. Alhinai, Walton, Commane, The Role of the Gut Microbiota in Colorectal Cancer Causation. *Int J Mol Sci* **20**, 5295 (2019).

637. R. Tanaka, *et al.*, Feed additives with thermophile-fermented compost enhance concentrations of free amino acids in the muscle of the flatfish Paralichthys olivaceus. *J Gen Appl Microbiol* **56**, 61–65 (2010).

638. T. Satoh, *et al.*, Impact of oral administration of compost extract on gene expression in the rat gastrointestinal tract. *J Biosci Bioeng* **114**, 500–505 (2012).

639. H. Miyamoto, *et al.*, Potential probiotic thermophiles isolated from mice after compost ingestion. *J Appl Microbiol* **114**, 1147–1157 (2013).

640. J. J. Bunker, *et al.*, Innate and Adaptive Humoral Responses Coat Distinct Commensal Bacteria with Immunoglobulin A. *Immunity* **43**, 541–553 (2015).

641. G. Deinhard, P. Blanz, K. Poralla, E. Altan, Bacillus acidoterrestris sp. nov., a New Thermotolerant Acidophile Isolated from Different Soils. *Syst Appl Microbiol* **10**, 47–53 (1987).

642. P. A. Ramsland, *et al.*, Structural basis for evasion of IgA immunity by Staphylococcus aureus revealed in the complex of SSL7 with Fc of human IgA1. *Proceedings of the National Academy of Sciences* **104**, 15051–15056 (2007).

643. K. E. Huntington, *et al.*, Colorectal cancer extracellular acidosis decreases immune cell killing and is partially ameliorated by pH-modulating agents that modify tumor cell cytokine profiles. *Am J Cancer Res* **12**, 138–151 (2022).

644. Press, *et al.*, Gastrointestinal pH profiles in patients with inflammatory bowel disease. *Aliment Pharmacol Ther* **12**, 673–678 (1998).

645. J. Qin, *et al.*, A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).

646. A. P. A. Oxley, *et al.*, Halophilic archaea in the human intestinal mucosa. *Environ Microbiol* **12**, 2398–2410 (2010).

647. C. Bang, K. Weidenbach, T. Gutsmann, H. Heine, R. A. Schmitz, The Intestinal Archaea Methanosphaera stadtmanae and Methanobrevibacter smithii Activate Human Dendritic Cells. *PLoS One* **9**, e99411 (2014).

648. P. J. Turnbaugh, *et al.*, An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).

649. M. O. A. Sommer, G. Dantas, G. M. Church, Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science (1979)* **325**, 1128–1131 (2009).

650. A. A. Salyers, C. F. Amábile-Cuevas, Why are antibiotic resistance genes so resistant to elimination? *Antimicrob Agents Chemother* **41**, 2321–2325 (1997).

651. J. L. Martínez, Antibiotics and Antibiotic Resistance Genes in Natural Environments. *Science (1979)* **321**, 365–367 (2008).

652. B. Stecher, *et al.*, Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proceedings of the National Academy of Sciences* **109**, 1269–1274 (2012).

653. C. Liu, *et al.*, Clinical Practice Guidelines by the Infectious Diseases Society of America for the Treatment of Methicillin-Resistant Staphylococcus aureus Infections in Adults and Children. *Clinical Infectious Diseases* **52**, e18–e55 (2011).

654. R. Ohki, *et al.*, The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in Bacillus subtilis. *Mol Microbiol* **49**, 1135–1144 (2003).

655. M. E. Falagas, E. K. Vouloumanou, G. Samonis, K. Z. Vardakas, Fosfomycin. *Clin Microbiol Rev* **29**, 321–347 (2016).

656. V. Liévin-Le Moal, A. L. Servin, The Front Line of Enteric Host Defense against Unwelcome Intrusion of Harmful Microorganisms: Mucins, Antimicrobial Peptides, and Microbiota. *Clin Microbiol Rev* **19**, 315–337 (2006).

657. E. A. Groisman, C. Parra-Lopez, M. Salcedo, C. J. Lipps, F. Heffron, Resistance to host antimicrobial peptides is necessary for Salmonella virulence. *Proceedings of the National Academy of Sciences* **89**, 11939–11943 (1992).

658. R. K. Ernst, T. Guina, S. I. Miller, Salmonella typhimurium outer membrane remodeling: role in resistance to host innate immunity. *Microbes Infect* **3**, 1327–1334 (2001).

659. L. Guo, *et al.*, Regulation of Lipid A Modifications by Salmonella typhimurium Virulence Genes phoP-phoQ. *Science (1979)* **276**, 250–253 (1997).

660. L. Guo, *et al.*, Lipid A Acylation and Bacterial Resistance against Vertebrate Antimicrobial Peptides. *Cell* **95**, 189–198 (1998).

661. Y. Shi, M. J. Cromie, F.-F. Hsu, J. Turk, E. A. Groisman, PhoP-regulated Salmonella resistance to the antimicrobial peptides magainin 2 and polymyxin B. *Mol Microbiol* **53**, 229–241 (2004).

662. M. W. Bader, *et al.*, Regulation of Salmonella typhimurium virulence gene expression by cationic antimicrobial peptides. *Mol Microbiol* **50**, 219–230 (2003).

663. Y. Li, *et al.*, Metabolic adaptability shifts of cell membrane fatty acids of Komagataeibacter hansenii HDM1-3 improve acid stress resistance and survival in acidic environments. *J Ind Microbiol Biotechnol* **46**, 1491–1503 (2019).

664. L. A. Royce, P. Liu, M. J. Stebbins, B. C. Hanson, L. R. Jarboe, The damaging effects of short chain fatty acids on Escherichia coli membranes. *Appl Microbiol Biotechnol* **97**, 8317–8327 (2013).

665. E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, A. van Oudenaarden, Multistability in the lactose utilization network of Escherichia coli. *Nature* **427**, 737–740 (2004).

666. M. Laurent, G. Charvin, J. Guespin-Michel, Bistability and hysteresis in epigenetic regulation of the lactose operon. Since Delbrück, a long series of ignored models. *Cell Mol Biol (Noisy-le-grand)* **51**, 583–94 (2005).

667. Z. Avramova, Transcriptional 'memory' of a stress: transient chromatin and memory (epigenetic) marks at stress-response genes. *The Plant Journal* **83**, 149–159 (2015).

668. R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, U. Sauer, Multidimensional Optimality of Microbial Metabolism. *Science (1979)* **336**, 601–604 (2012).

669. E. Kussell, S. Leibler, Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. *Science (1979)* **309**, 2075–2078 (2005).

670. C. A. Brennan, W. S. Garrett, Gut Microbiota, Inflammation, and Colorectal Cancer. *Annu Rev Microbiol* **70**, 395–411 (2016).

671. M. Perše, Oxidative Stress in the Pathogenesis of Colorectal Cancer: Cause or Consequence? *Biomed Res Int* **2013**, 1–9 (2013).

672. A. K. Singh, R. Y. Hertzberger, U. G. Knaus, Hydrogen peroxide production by lactobacilli promotes epithelial restitution during colitis. *Redox Biol* **16**, 11–20 (2018).

673. R. Fattouh, *et al.*, Rac2-Deficiency Leads to Exacerbated and Protracted Colitis in Response to Citrobacter rodentium Infection. *PLoS One* **8**, e61629 (2013).

674. J.-S. Pan, M.-Z. Hong, J.-L. Ren, Reactive oxygen species: A double-edged sword in oncogenesis. *World J Gastroenterol* **15**, 1702 (2009).

675. J. W. Collins, *et al.*, Citrobacter rodentium: infection, inflammation and the microbiota. *Nat Rev Microbiol* **12**, 612–623 (2014).

676. E. E. Olsan, *et al.*, Colonization resistance: The deconvolution of a complex trait. *Journal of Biological Chemistry* **292**, 8577–8581 (2017).

677. M. J. Brookes, Modulation of iron transport proteins in human colorectal carcinogenesis. *Gut* **55**, 1449–1460 (2006).

678. L. CHARDALIAS, I. PAPACONSTANTINOU, A. GKLAVAS, M. POLITOU, T. THEODOSOPOULOS, Iron Deficiency Anemia in Colorectal Cancer Patients: Is Preoperative Intravenous Iron Infusion Indicated? A Narrative Review of the Literature. *Cancer Diagnosis & Prognosis* **3**, 163–168 (2023).

679. M. Malaguarnera, Decrease of serum carnitine levels in patients with or without gastrointestinal cancer cachexia. *World J Gastroenterol* **12**, 4541 (2006).

680. N. Longo, M. Frigeni, M. Pasquali, Carnitine transport and fatty acid oxidation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1863**, 2422–2435 (2016).

681. M. Bhat, *et al.*, Targeting the translation machinery in cancer. *Nat Rev Drug Discov* **14**, 261–278 (2015).

682. C. Valadon, O. Namy, The Importance of the Epi-Transcriptome in Translation Fidelity. *Noncoding RNA* **7**, 51 (2021).

683. C. Corbet, O. Feron, Tumour acidosis: from the passenger to the driver's seat. *Nat Rev Cancer* **17**, 577–593 (2017).

684. L. A. Liotta, E. C. Kohn, The microenvironment of the tumour–host interface. *Nature* **411**, 375–379 (2001).

685. B. W. P. Reeve, S. J. Reid, Glutamate and histidine improve both solvent yields and the acid tolerance response of Clostridium beijerinckii NCP 260. *J Appl Microbiol* **120**, 1271–1281 (2016).

686. X. Shi, G. N. Bennett, Effects of rpoA and cysB mutations on acid induction of biodegradative arginine decarboxylase in Escherichia coli. *J Bacteriol* **176**, 7017–7023 (1994).

687. J. M. Kirsch, *et al.*, Bacteriophage-Bacteria Interactions in the Gut: From Invertebrates to Mammals. *Annu Rev Virol* **8**, 95–113 (2021).

688. J.-P. Claverys, M. Prudhomme, B. Martin, Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria. *Annu Rev Microbiol* **60**, 451–475 (2006).

689. E. C. Keen, *et al.*, Novel "Superspreader" Bacteriophages Promote Horizontal Gene Transfer by Transformation. *mBio* **8** (2017).

690. N. C. Pyenson, K. Gayvert, A. Varble, O. Elemento, L. A. Marraffini, Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host Microbe* **22**, 343-353.e3 (2017).

691. E. C. Keen, *et al.*, Novel "Superspreader" Bacteriophages Promote Horizontal Gene Transfer by Transformation. *mBio* **8** (2017).

692. A. Ebihara, *et al.*, Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Science* **15**, 1494–1499 (2006).

693. B. Bozic, J. Repac, M. Djordjevic, Endogenous Gene Regulation as a Predicted Main Function of Type I-E CRISPR/Cas System in E. coli. *Molecules* **24**, 784 (2019).

694. O. Musharova, *et al.*, Systematic analysis of Type I-E Escherichia coli CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. *Mol Microbiol* **111**, 1558–1570 (2019).

695. B. Bozic, J. Repac, M. Djordjevic, Endogenous Gene Regulation as a Predicted Main Function of Type I-E CRISPR/Cas System in E. coli. *Molecules* **24**, 784 (2019).

696. N. C. Pyenson, K. Gayvert, A. Varble, O. Elemento, L. A. Marraffini, Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host Microbe* **22**, 343-353.e3 (2017).

697. A. Millman, *et al.*, Bacterial Retrons Function In Anti-Phage Defense. *Cell* **183**, 1551-1561.e12 (2020).

698. D. Ziech, R. Franco, A. Pappa, M. I. Panayiotidis, Reactive Oxygen Species (ROS)—Induced genetic and epigenetic alterations in human carcinogenesis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **711**, 167–173 (2011).

699. H. Liu, *et al.*, Redox Imbalance in the Development of Colorectal Cancer. *J Cancer* **8**, 1586–1597 (2017).

700. J. Zeng, *et al.*, Aberrant ROS Mediate Cell Cycle and Motility in Colorectal Cancer Cells Through an Oncogenic CXCL14 Signaling Pathway. *Front Pharmacol* **12** (2021).

701. C. N. Heiss, L. E. Olofsson, Gut Microbiota-Dependent Modulation of Energy Metabolism. *J Innate Immun* **10**, 163–171 (2018).

702. E. Neis, C. Dejong, S. Rensen, The Role of Microbial Amino Acid Metabolism in Host Metabolism. *Nutrients* **7**, 2930–2946 (2015).

703. D. R. Donohoe, *et al.*, The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab* **13**, 517–526 (2011).

704. J. Miro-Blanch, O. Yanes, Epigenetic Regulation at the Interplay Between Gut Microbiota and Host Metabolism. *Front Genet* **10** (2019).

705. C. C. Gillis, *et al.*, Dysbiosis-Associated Change in Host Metabolism Generates Lactate to Support Salmonella Growth. *Cell Host Microbe* **23**, 54-64.e6 (2018).

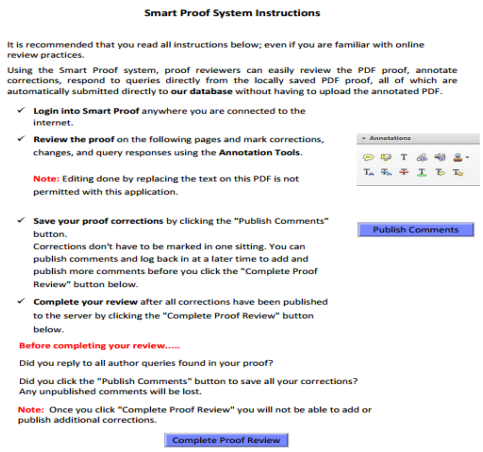
706. D. R. Donohoe, A. Wali, B. P. Brylawski, S. J. Bultman, Microbial Regulation of Glucose Metabolism and Cell-Cycle Progression in Mammalian Colonocytes. *PLoS One* **7** (2012).

10.0 Appendix

10.1 Publications

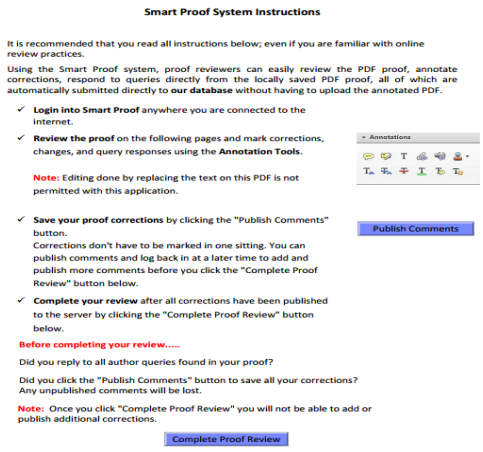
M. T. F. Lamaudière, R. Arasaradnam, G. D. Weedall, I. Y. Morozov, The colorectal cancer gut environment regulates activity of the microbiome and promotes the multidrug resistant phenotype of ESKAPE and other pathogens. *mSphere* (2023). https://doi.org/10.1128/msphere.00626-22

Link to paper, double-click image:



M. T. F. Lamaudière, R. Arasaradnam, G. D. Weedall, I. Y. Morozov, The colorectal cancer microbiota alter their transcriptome to adapt to the acidity, reactive oxygen species and metabolite availability of gut microenvironments. *mSphere* (2023). https://doi.org/10.1128/msphere.00627-22

Link to paper, double-click image:



11.0 Supplementary Data

Metatranscriptome	**Df**	**Sum Sq**	**R2**	**F**	**Pr(>F)**
Group (healthy/CRC)	1	0.23246	0.09193	1.9591	0.02760 *
Age (<73/>73)	1	0.16018	0.06334	1.35	0.16288
Sex (F/M)	1	0.08078	0.03195	0.6808	0.82972
Has smoked (no/yes)	1	0.20691	0.08182	1.7438	0.06169
BMI (<30/>30)	1	0.18727	0.07406	1.5783	0.09409
Residual	14	1.66117	0.65691	-	-
Total	19	2.52877	1	-	-
16S rRNA genes					
Group (healthy/CRC)	1	0.29558	0.1205	2.5009	0.0005 ***
Age (<73/>73)	1	0.11281	0.04599	0.9545	0.5152
Sex (F/M)	1	0.12597	0.05136	1.0659	0.3639
Has smoked (no/yes)	1	0.11105	0.04527	0.9396	0.5412
BMI (<30/>30)	1	0.15287	0.06232	1.2935	0.1523
Residual	14	1.65461	0.67455	-	-
Total	19	2.45288	1	-	-

Table S1 | PERMANOVA analysis of patient metadata influence over distribution of metatranscriptome- and 16S rRNA gene sequence profiling-based community taxonomy. * Denotes statistical significance.

Organism name	16S rDNA detected	mRNA detected
Acidovorax delafieldii	+	-
Acinetobacter junii	-	+
Actinophytocola sp.	+	-
Actinoplanes globisporus	+	-
Agrobacterium pusense	+	-
Akkermansia muciniphila	-	+
Alistipes shahii	-	+
Amycolatopsis mediterranei	+	-
Anaerococcus hydrogenalis	-	+
Anaerococcus obesiensis	-	+
Anaerococcus prevotii	-	+
Anaerococcus sp.	-	+
Anaerofustis sp.	+	-
Anaerotruncus colihominis	-	+
Angelakisella massiliensis	-	+
Asticcacaulis excentricus	-	+
Asticcacaulis sp.	-	+
Bacteroides finegoldii	-	+
Bacteroides fluxus	-	+
Bacteroides gallinaceum	+	-
Bacteroides stercorirosoris	-	+
Bdellovibrio bacteriovorus	-	+
Bdellovibrio exovorus	-	+
Bdellovibrio sp.	-	+
Bifidobacterium angulatum	-	+
Bifidobacterium subtile	-	+
Bryobacter aggregatus	-	+
Burkholderia cenocepacia	+	-
Butyricicoccus pullicaecorum	-	+
Caedimonas varicaedens	+	-
Catabacter hongkongensis	-	+
Chthoniobacter flavus	-	+
Clostridium beijerinckii	-	+
Colidextribacter massiliensis	+	-
Corynebacterium amycolatum	+	-
Corynebacterium durum	+	-
Dialister sp.	+	-
Dielma sp.	+	-

Dyella sp.	+	-
Enorma massiliensis	-	+
Enterococcus faecium	-	+
Ezakiella massiliensis	-	+
Ezakiella massiliensis	-	+
Faecalitalea sp.	+	-
Finegoldia magna	-	+
Granulicatella	-	+
Granulicatella adiacens	-	+
Granulicatella elegans	-	+
Granulicatella sp.	-	+
Haliangium ochraceum	-	+
Helicobacter canadensis	+	-
Hyphomicrobium denitrificans	-	+
Hyphomicrobium sp.	-	+
Intestinimonas massiliensis	-	+
Intestinimonas massiliensis	-	+
Kaistia granuli	-	+
Kaistia soli	-	+
Lachnoclostridium phocaeense	-	+
Lactococcus chungangensis	-	+
Lacunisphaera limnophila	-	+
Longilinea arvoryzae	-	+
Marmoricola sp.	-	+
Megasphaera cerevisiae	-	+
Megasphaera sp.	-	+
Mesorhizobium	-	+
Mesorhizobium australicum	-	+
Mesorhizobium ciceri	-	+
Mesorhizobium loti	-	+
Mesorhizobium plurifarium	+	-
Mesorhizobium sp.	-	+
Methylobacterium sp.	-	+
Methylocystis sp.	-	+
Millionella massiliensis	-	+
Mitsuaria chitosanitabida	+	-
Monoglobus pectinilyticus	+	-
Nakamurella multipartita	+	-
Nitrospira defluvii	+	-

Nitrospira japonica	-	+
Nitrospira moscoviensis	-	+
Nitrospira sp.	-	+
Nocardia vaccinii	+	-
Paeniclostridium sordellii	-	+
Pajaroellobacter abortibovis	-	+
Pantoea dispersa	+	-
Paraburkholderia sp.	+	-
Paraburkholderia tropica	+	-
Promicromonospora sp.	+	-
Propionibacterium acidifaciens	-	+
Propionibacterium sp.	-	+
Prosthecobacter debontii	-	+
Pseudopropionibacterium propionicum	-	+
Pseudoxanthomonas sp.	-	+
Ralstonia pickettii	+	-
Rhodomicrobium vannielii	-	+
Rhodoplanes sp.	-	+
Romboutsia ilealis	+	-
Schaalia odontolytica	+	-
Schwartzia succinivorans	-	+
Selenomonadales bacterium	+	-
Slackia heliotrinireducens	-	+
Sphingopyxis terrae	+	-
Streptococcus constellatus	-	+
Streptomyces griseorubiginosus	+	-
Streptomyces nanshensis	+	-
Thermobaculum terrenum	-	+
Variovorax paradoxus	-	+
Variovorax sp.	-	+
Verrucomicrobium sp.	-	+
Verrucomicrobium spinosum	-	+

Table S2 | Microbial species detected at either the DNA (16S rRNA gene sequence profiling) level or RNA (metatranscriptome) level bot not both. (+) Detected, (-) not detected.