Trends in Cancer

CellPress REVIEWS

Opinion

Mutagenesis by Microbe: the Role of the Microbiota in Shaping the Cancer Genome

Maurice Barrett,[1,2] Collette K. Hand,[3] Fergus Shanahan,[1,4] Thomas Murphy,[5] and Paul W. O'Toole[1,2,*]

Cancers arise through the process of somatic evolution fueled by the inception of somatic mutations. We lack a complete understanding of the sources of these somatic mutations.
Humans host a vast repertoire of microbes collectively known as the microbiota. The microbiota plays a role in altering the tumor microenvironment and proliferation. In addition, microbes have been shown to elicit DNA damage which provides the driver for somatic mutations. An understanding of microbiota-driven mutational mechanisms would contribute to a more complete understanding of the origins of the cancer genome. Here, we review the modes by which microbes stimulate DNA damage and the effect of these phenomena upon the cancer genomic architecture, specifically in the form of mutational spectra and mutational signatures.

Origin of the Cancer Genome and Role of the Microbiota

Oncogenesis (see Glossary) is driven by the Darwinian selection of **somatic mutations** over time [1]. Mutations arise through the formation of genetic aberrations and their subsequent interactions with the **DNA repair** machinery and cell-cycle-related pathways including DNA synthesis [2]. **Mutational mechanisms** alter the DNA in distinguishing manners resulting in genetic patterns known as **mutational signatures** (Box 1).

The origin of mutations allows them to be classified into three categories: (i) inherited genetic variants that lead to an increase in the risk of cancer development; (ii) environmental factors, exogenous factors including UV light, tobacco smoking, and diet that mutate the DNA and that are directly linked to cancer; and (iii) stochastic errors associated with DNA replication and other phenomena. These are seemingly inevitable random mutations that arise due to the intrinsic properties of DNA biology. Seminal work by Tomasetti and Vogelstein showed that about two-thirds of the mutations in the cancer genome originate from stochastic events [3,4]. Lung and cervical adenocarcinoma genomes harbor median values of 33% and 83% stochastic mutations, respectively [3]. However, epidemiological evidence indicates that a high proportion (~90%) of cases are attributable to environmental factors, namely, tobacco smoking and human papillomavirus (HPV) infection, respectively. The managing of **environmental risk factors** is thus crucial in cancer prevention, even though stochastic/replicative mechanisms are the major drivers [3]. However, a complete catalogue of environmental factors that contribute to cancer risk is lacking. A large number of known carcinogens promote oncogenesis by causing mutagenesis; for example, UV light, ethanol, tobacco smoke, and radioactive substances.

The human **microbiota** is increasingly recognized as an emerging environmental risk factor. The human microbiota is home to about $3.8\acute{}10^{13}$ bacterial cells and it is estimated that the collective metagenome of these bacteria encompasses about 100 times more genes than the human genome [5,6]. Although the majority of studies have focused on bacteria, upon which this review

Highlights

The literature describing the differences in microbiota features between individuals with cancer and matched controls has undergone dramatic recent expansion. Mechanistic models for how microbes promote cancer formation and progression are being developed and experimentally tested.

Microbes have been implicated in mutational mechanisms namely in the formation of DNA damage. These mechanisms include the production of crosslinking genotoxic colibactin by *Escherichia coli* or ectopic expression of activation-induced cytidine caused by *Helicobacter pylori* infection.

Developments in bioinformatics have allowed for the elucidation of the mutational mechanisms that act upon the cancer genome through oncogenesis, particularly by identifying mutational signatures.

Elucidation of microbe-associated mechanisms will allow for a more complete understanding of the forces behind the etiology of the cancer genome.

[1]APC Microbiome Ireland, University College Cork, National University of Ireland, Cork, Ireland
[2]School of Microbiology, University College Cork, National University of Ireland, Cork, Ireland
[3]Department of Pathology, University College Cork, Cork, Ireland
[4]Department of Medicine, University College Cork, National University of Ireland, Cork, Ireland
[5]Department of Surgery, Mercy University Hospital, Cork, Ireland

*Correspondence:
pwotoole@ucc.ie (P.W. O'Toole).

Box 1. Mutational Signatures

Specific mutational mechanisms produce characteristic patterns in the genome known as mutational signatures. Recent advances in mathematical modelling and bioinformatics have led to improvements in our ability to identify mutational signatures from cancer genomic data. There are six defined classes of base substitutions: C>A, C>G, C>T, T>A, T>C, and T>G [note: In accordance with the Catalogue of Somatic Mutations in Cancer (COSMIC) system, all substitutions are referred to by the pyrimidine of the mutated Watson–Crick base pair]. The incorporation of the 5′ and 3′ bases flanking the mutated base of the six originally defined classes gives an expanded classification system of 96 possible mutations. Utilizing this 96-class system as the framework and applying nonnegative matrix factorization and model selection, with input from genomic data from 7042 cancer samples from 31 different cancer types, 21 mutational signatures were initially identified [82]. With the inclusion of more genomes for a heterogeneity of cancers, as well as the consideration of single base insertion/deletions and double base substitutions, the number of mutational signatures has expanded [55]. Currently, the number and type of mutational signatures characterized are as follows: 49 single base substitutions, 11 doublet base substitutions, four clustered base substitutions, and 17 small insertion and deletion (indels) mutational signatures [55]. Structural variants also occur in cancer genomes and they include insertions, deletions, inversions, balanced or unbalanced translocations, amplifications, and complex rearrangements on a scale of >50 bp in size [88]. Efforts have also been made to define the signatures of these events [89]. Mutational signatures provide an insight into the mutational mechanisms that act on a cancer genome over time. Mutational signatures are typically displayed as histogram with the frequency of base substations (or indels or doublet base substitutions) with respect to the genomic context. SBS signature 1 is characterized by C>T transversions at methylated CpG sites within an NpCpG trinucleotide context. The putative mechanisms behind SBS signature 1 is spontaneous or enzymatic deamination of 5-methylcytosine to thymine. This newly formed thymine may be base-paired with adenine during replication, provided DNA repair is not executed. Many mutational signatures described do not have a known etiology.

Glossary

Base substitutions: a type of mutation in which one base is replaced by another in DNA.
Chromosomal instability: a phenomenon which leads to alterations in chromosome number and/or structure.
DNA adduct: formed by the addition of a chemical moiety to a DNA base.
Alkylation: in the context of DNA is the addition of an alkyl group (C_nH2_{n+1}) to a DNA base.
DNA crosslinking: formation of covalent bonds between two nucleotides. This bond can be formed between nucleotides on the same DNA stand (intrastrand crosslinks) or different strands (interstrand crosslinks).
Deamination: in the context of DNA is the removal of an amino group from a DNA base.
DNA repair: a diverse collection of pathways with the purpose of addressing DNA damage and maintaining genome stability.
Double-strand breaks (DSBs): This is where both strands of DNA which are juxtaposed to each other contain a break in their phosphate backbone.
Environmental risk factor: a thing or process that is not inherited that increases the risk for a particular disease.
Microbes: microorganisms including bacteria, fungi, protists and virus. Usually exist as a single cell organism.
Microbiome: the combined genetic material of the microorganisms in a particular niche.
Microbiota: the collection of organisms in a niche.
Mutational mechanism: biological phenomenon that leads to the generation of mutations. Usually involving DNA damage, DNA repair and DNA replication.
Mutational signature: the characteristic DNA pattern of mutations produced by a mutational mechanism.
Oncogenesis: the transformation of a normal cell into a cancer cell.
Oxidative base lesions: DNA Bases that occur due to a reaction with reactive oxygen species
Somatic mutation: mutation which occurs in a somatic cell and is thus not heritable.

is focused, the human microbiota includes members from all five kingdoms of life as well as viruses. A large number of studies demonstrate that microbiota features are involved in the development and progression of a range of cancers. The term oncobiome has been coined to describe the relationship between the microbiota and cancer [7]. However, oncobiome research has identified relationships that are primarily correlative rather than causative in nature. With regard to the putative mechanistic role that the microbiota has in cancer development, immunomodulation in the form of inflammation caused by the microbiota is an intense area of research [8]. Efforts have also been made in defining the role of the microbiota in cell proliferation [9].

The microbiota is known to be involved in a diverse assortment of mutational mechanisms (Table 1, Key Table). Known variation in cancer risk due to unknown environmental factors could be explained in part by variations in the ability of the microbiota of individual subjects to induce DNA damage and thus somatic mutations. Here, we describe the current state of knowledge on **microbes** and their ability to compromise the stability of the human genome ultimately leading to cancer. We describe the microbiota influences on genome integrity through; (i) direct DNA damage, (ii) immune-cell-induced DNA damage, (iii) dietary interaction, and (iv) disruption to the DNA damage response.

Direct DNA Damage

Members of the microbiota can produce proteins, molecules, and secondary metabolites that can directly cause DNA damage. These products can interact directly with the host DNA thereby mutating it.

Colibactin

Escherichia coli is classified into four phylogenetic groups, A, B1, B2, and D. About 30–50% of *E. coli* strains identified in stool microbiota of individuals from high-income nations belong to group B2. Within the B2 group, 35% of isolates possess genomic islands known as *pks* islands [10]. The 54-kb *pks* island is a biosynthetic gene cluster encoding for a nonribosomal peptide synthetase (NRPS)–polyketide synthase (PKS) hybrid gene cluster, which encodes colibactin [11]. Colibactin can cause **double-strand breaks (DSBs)** in mammalian DNA, thereby promoting genome instability and an increase in mutation rate [12,13]. It is not currently known how colibactin

Key Table

Table 1. Microbe-Associated Mechanisms and Genomic Consequences

Source	Involvement of microbiota features	Key role in a mutational mechanism	Postulated effect on cancer genomic landscape	Refs
AID	*Helicobacter pylori* infection causes ectopic expression of AID	Cytosine deamination at specific motifs	Mutational signatures SBS84 and SBS85	[53,55]
Acetaldehyde	Various inhabitants produce ethanol and metabolically act on it to produce acetaldehyde	N2-ethylidenedeoxyguanosine, guanine–guanine intrastrand crosslinks	G:G-to-T:T base substitution. Mutational signature DBS2	[73]
Colibactin	Expressed by *Escherichia coli* containing a *pks* island	Adenine–adenine intrastrand crosslinks, DSBs	DSBs at an AAWWTT pentanucleotide motif. Mutational signatures SBS28 and SBS41	[22]
CDT	Produced by various Gram-negative bacteria including enteropathogenic *E. coli*, *Campylobacter* species, *Shigella* species and *Haemophilus ducreyi*	SSBs and DSBs	Infidelity of DNA repair can lead to structural variants such as indels	[55]
Disruption of DNA mismatch repair	*H. pylori* and Enteropathogenic *E. coli* can disrupt MMR	Deletion of MMR proteins	Microsatellite instability, mutational signature SBS6, ID1, and ID2	[79,80,82]
N_2O_3	Metabolic activities of the microbiota can produce precursors to N_20_3, e.g., denitrifying bacteria	Nitrosative deamination	Various base substitutions; e.g., adenine nitrosative deamination to hypoxanthine can lead to T>C substitution	[39,42]
HOBO	Eosinophils produce HOBO. The microbiota can influence eosinophil biology	8-bromoguanine	G>T primarily but also G>C, G>A, and delG	[50]
HOCl	HOCl is produced by neutrophils. The microbiota can influence neutrophil inflammatory status	Formation of 5ClC, formation of MDA	C>T, G >A, G>T substitutions	[45,46]
NOCs	Microbes play a role in the production of nitrosating agents and produce biogenic amine	Alkylated DNA bases	Various base substitutions e.g., O6-methylguanine (O6-MeG) can cause a G(C)>A(T) transition	[69]
ROS	Various metabolic activities	Oxidative base lesions	G to T transversion, SBS mutational signatures 18 and 36	[95]
4-HNE	*Enterococcus faecalis* induces the bystander effect via polarizing macrophages which then produce 4-HNE	Exocyclic HNE–DNA adducts	Chromosomal instability	[60]

is transported from the cell exterior all the way into the nucleus. The *pks*$^+$ *E. coli* strains are over-represented in the gut of individuals with colorectal cancer (CRC); being detected at a rate of 20% in the mucosa of healthy individuals but 55–67% in patients with CRC [14,15]. Furthermore, *pks*+ *E. coli* is disproportionally frequently identified in subjects with familial adenomatous polyposis compared with healthy controls [16]. Monocolonization of azoxymethane (AOM)-treated *IL10*$^{-/-}$ mice with *pks*+ *E. coli* promotes tumorigenesis, while challenge with strains lacking *pks* reduces the frequency of tumorigenesis [14].

Colibactin cross links directly with DNA through an electrophilic cyclopropane moiety 'warhead' [17]. Liquid chromatography–mass spectrometry-based methodologies have identified that colibactin **alkylation** of DNA via the cyclopropane warhead results in adenine–colibactin adducts [18,19]. This phenomenon was identified in both HeLa cells and in mouse models [19]. Colibactin can also induce DNA interstrand crosslinks and activation of the DNA damage response including Fanconi anemia DNA repair [20]. Recent structural analysis revealed that colibactin contains two conjoined warheads enabling it to cause **DNA crosslinks** [21]. DSBs are not believed to be a direct consequence of colibactin activity but rather occur due to replication stress caused by DNA crosslinks [20]. Recent sequencing analysis of colibactin-induced DSB sites revealed that

these DSBs occurred at AT-rich regions and in particularly at the pentanucleotides motif containing AAWWTT [22]. Single nucleotide variants at the AAWWTT are enriched in a number of cancers, including CRC and stomach cancer, compared with a WWWWW motif. Two mutational signatures have been linked with the AAWWTT colibactin motifs, SBS28 and SBS41 [22]. Mutational signature SBS28 has been associated with POLE mutation while mutational signature SBS41 has no known etiology.

Cytolethal Distending Toxin (CDT)

CDT is produced by an array of Gram-negative bacteria within the gamma and epsilon classes of the phylum Proteobacteria [23]. It is a heat-labile exotoxin whose properties lead it to be classified as a both a cyclomodulin and a genotoxin. The Proteobacteria that can produce CDT are subdominant members of the human gut microbiota.

CDT is a heteromultimeric protein comprising three subunits, CdtA, CdtB, and CdtC, which are encoded within a bacterial single operon [24,25]. Subunits CdtA and CdtC function to allow delivery and internalization of CDT into target cells [25]. CdtB shares sequence, structural, and functional homology with DNase I and is highly conserved among bacteria [26,27]. Furthermore, nuclear localization signals have been identified in CdtB proteins [28]. Studies with ApcMin/+ mice that are genetically susceptible to small bowel cancer found that a *Campylobacter jejuni* strain harboring the CDT operon promoted colorectal tumorigenesis compared with treatment with non-CDT bacterial controls, while mutation of the CdtB subunit attenuated this phenomenon [29]. CdtB has been shown to promote DSB *in vitro* and *in vivo* [26,30,31]. However, the current model of CdtB activity holds that CdtB acts in a dose-dependent manner and tends not to induce DSBs directly [32]. At low to moderate doses, CdtB causes single-strand breaks (SSBs), which are addressed by single-strand break repair (SSBR) [33]. If CDT-induced SSBs are not addressed before replication or occur during replication, they may cause a stalled replication fork [32,33]. At high doses, CDT can induce DSBs directly by two cuts to the DNA backbone that are juxtaposed to each other [32].

Reactive Oxygen Species (ROS)

ROS are a chemically reactive family of molecules containing oxygen, which includes the highly reactive hydroxyl radical (OH^-), superoxide radical (O_2^-), and nonradical H_2O_2. Reactions of ROS with DNA generates oxidative DNA base lesions. To date, more than 30 **oxidative DNA base lesions** have been identified (Box 2) [34].

Microbiota activity is known to elicit ROS through varied means. For example, primary bile acids, cholic acid (CA) and chenodeoxycholic acid (CDCA), are synthesized by the liver and secreted into the small intestine from the gall bladder. A small proportion of these bile salts is transformed into secondary bile salts by the gut microbiota. These secondary bile salts are thought to be involved in the production of ROS [35]. The production of secondary bile in the colon where the bacterial metabolic repertoire exists may be one of the reasons that CRC is more prevalent than small intestine cancer, although differences in stem cell turnover is likely a more important factor [3].

H_2S is produced by the metabolic activity of colonic bacteria including taurine desulfonation by *Bilophila wadsworthia*, cysteine degradation by *Fusobacterium nucleatum*, and sulfonate degradation by sulfate-reducing bacterium such as *Desulfovibrio desulfuricans*. Increased relative abundance of such bacteria has been linked to CRC development [36,37]. Evidence suggests that H_2S production leads to DNA damage partly due to ROS generation [37,38].

Box 2. Oxidative DNA Base Lesions

Guanine has the lowest redox potential of the native bases and is thus the most readily oxidized. Two common **oxidative base lesions** that are generated by the oxidation of guanine include 8-oxo-7,8-dihydro-2′-deoxyguanosine and 2,6-diamino-4-oxo-5-formamidopyrimidine (FapyG), which occur at an estimated rate of 1000–2000 and 1500–2500 per cell/per day in normal tissues, respectively [90]. Furthermore, the occurrence and the mutagenicity of these oxidative DNA base lesions vary considerable. For example, 7,8-dihydro-8-oxo-guanine is about four times as mutagenic and four times more frequent in its occurrence than 7,8-dihydro-8-oxo-adenine [90,91]. Replication of DNA containing FapyG is shown to induce G:C to T:A (C >A) and G:C to T:A (C >A) [92].

The nucleobases within the cellular nucleotide pool may also undergo oxidation. Misincorporation of these nucleoside triphosphates can induce mutations. The two major products of nucleotide pool oxidation are 8-hydroxy-2′-deoxyguanosine 5′-triphosphate (8-OH-dGTP) and 2-hydroxydeoxy-ATP. 8-OH-dGTP has been demonstrated to induce A:T to C:G transversions when introduced into COS-7 mammalian cells [93]. *In vitro* analysis using HeLa cell extract showed that 2-OH-dATP within the nucleotide pool can led to G:C to A:T (C>T) transitions and G:C to T:A(C>A) [94].

Mutational signatures 18 and 36 have been suggested to be caused by ROS. Mutational signature 36 has been specifically attributed to ROS in the context of MUTYH-associated polyposis (MAP) syndrome [95]. MAP syndrome is defined by biallelic germline mutation of *MUTYH* gene and is a colorectal polyposis that predisposes individuals to CRC. MUTYH DNA glycosylase is coded by the *MUTYH* gene and functions to prevent 8-oxoguanine-related mutagenesis by scanning the newly-synthesized daughter strand in order locate and remove incorporated adenine paired with 8-oxoguanine [92].

Dinitrogen trioxide (N_2O_3) and Nitrosative Deamination

Nitrosative **deamination** is deamination mediated by N_2O_3 (nitrous anhydride). In this phenomenon, N_2O_3 can react with nucleotides and induce deamination by nucleophilic aromatic substitution. These events are mutagenic because the resulting deaminated bases may be read incorrectly if not repaired [39].

N_2O_3 can be generated from the auto-oxidation of nitric oxide ($NO^{\cdot}$) or the condensation of nitrous acid (HNO_2) [40]. Gut microbes can produce endogenous NO and/or HNO_2 by four mechanisms. (i) Haem thiolate monooxygenase, NO synthase (NOS), oxidizes L-arginine (Arg) to produce NO [41]. (ii) Denitrification of nitrate (NO_3^-) to N_2, which is an important part of the nitrogen cycle and is carried out by denitrifying bacteria and plants. During denitrification, NO is produced by one-electron reduction of nitrite (NO_2^-) by heme or Cu-containing nitrite reductases [42]. (iii) Respiratory NO_2^- ammonification (also referred to as dissimilatory NO_3^- reduction to ammonium) [42]. (iv) Acidic nonenzymatic reduction of NO_2^- to NO, which is driven by lactic acid bacteria such as lactobacilli and bifidobacteria [43].

Immune-Cell-Induced DNA Damage

The microbiota and immune system closely interact from the early stages of human development. In this section we review mechanisms by which the microbiota can influence immune cells to behave in a genotoxic manner.

Hypochlorous Acid (HOCl) Production

Neutrophils, which are a type of polymorphonuclear leukocyte, accumulate at sites of injury with the primary function of promoting inflammation. Neutrophils produce a potent antimicrobial known as HOCl, which is produced by myeloperoxidase using as substrates the chloride ions and H_2O_2 produced by NADPH oxidase [44]. HOCl is highly reactive and readily interacts with DNA. HOCl has been shown to cause a cytosine to 5-chlorocytosine (5ClC) conversion [45]. This is in turn can cause a C to T transition during replication.

In addition, HOCl can induce the peroxidation of lipids, leading to the formation of malondialdehyde (MDA). Studies in both cellular and animal models have found that such production of MDA can lead to a significant increase in the formation of 3-(2-deoxy-β-D-erythro-

pentofuranosyl)pyrimido[1,2-α]purin-10(3H)-one (M1dG), a damaged guanine [46]. M1dG adducts are mutagenic causing G>T and G >A substitutions [47].

The microbiota is now known to be a modulator of neutrophil biology [48]. A recent study in a mouse model demonstrated that neutrophil proinflammatory activity correlates positively with neutrophil ageing while in circulation [49]. Furthermore, the study found that the microbiota regulates neutrophil ageing by Toll-like receptor and myeloid differentiation factor 88-mediated signaling pathways [49]. Depletion of the microbiota was mirrored in the number of aged neutrophils and an improvement in inflammatory disease.

Hypobromous Acid (HOBO) Production

Eosinophils are granular leukocytes with a multifunctional role in immune biology. Eosinophils secrete eosinophil peroxidase that catalyzes the formation of HOBO from H_2O_2 and halide ions (Br^-) in solution. HOBO can also be produced by reaction of HOCl with Br^- ions. Like HOCl, HOBO is an oxidant and functions to oxidize the cellular components of invading pathogens; however, excess production of HOBO can also lead to host damage, including DNA damage, namely the formation of 8-bromo-2′-deoxyguanosine and 5-bromo-2′-deoxycytidine. A SupF forward mutation assay in human cells found that the prominent mutation induced was G>T mutation but HOBO also induces G>C, G>A, and delG [50].

Activation-Induced Cytidine Deaminase (AID)

AID is a member of the cytidine deaminase family of enzymes with a role in somatic hypermutation. Immunohistochemistry identified the ectopic overexpression of AID in inflamed tissue derived from patients with Crohn's disease and ulcerative colitis, as well as colitis-associated CRC [51]. The expression of AID in colonic epithelial cell lines induced an increase in the mutation rates in these cells [51]. Knockout of AID in *IL10* null mice attenuates the mutation rate in their colonic cells and also inhibits CRC development [52]. Inflammation seems to be key to this aberrant activity. *Helicobacter pylori* infection, which is known to induce inflammation, promotes ectopic expression of AID in nontumorous epithelial tissues [53].

Whole genome analyses in chronic lymphocytic leukemia have revealed that the activity of AID may produce two types of substitution pattern: (i) a canonical AID signature characterized by C to T/G substitutions at WRCY motifs near active transcriptional start sites; and (ii) a noncanonical AID signature characterized by A to C mutations at WA (W=A or T) motifs occurring genome-wide in a nonclustered fashion [54]. These mutational processes have been assigned to mutational signatures SBS84 and SBS85 [55].

Bystander Effect and *Enterococcus faecalis*

E. faecalis is known to promote CRC oncogenesis in interleukin $10^{-/-}$ mice [56]. *E. faecalis* can promote the bystander effect that leads to DNA DSBs, tetraploidy, and **chromosomal instability**. In this model, *E. faecalis* production of extracellular superoxide induces polarization of macrophages to an M1 phenotype [57–59]. In turn, macrophages produce 4-hydroxy-2-nonenal (4-HNE), a diffusible breakdown product of ω-6 polyunsaturated fatty acids whose expression in this context is dependent on cyclooxygenase-2 [60,61]. Primary murine colon epithelial cells exposed to polarized macrophages or purified 4-HNE undergo transformation [62].

Dietary Interaction

The diets of the host and the gut microbiota are inextricably linked. Gut bacteria depend almost exclusively on the host diet for their nutritional substrates (a restricted number of taxa can metabolize mucins and glycoproteins), and indeed, the composition of the **microbiome** is correlated strongly

with diet. Diet is a key modulator of cancer risk. In the cases described below, microbiota–diet interactions lead to the formation of genotoxic compounds capable of mutating the host genome.

N-Nitroso Compounds (NOCs)

NOCs, such as nitrosamines and nitrosamide, are known to be potent carcinogens. NOCs are formed by the nitrosation of secondary amines and amides via nitrosating agents, such as N_2O_3 and N_2O_4 [63]. NOCs can be found in foods such as processed meats, smoked/cured fish, and German beer [64]. Additional compounds such as NO_3^- and NO_2^-, which are precursors to nitrosating agents, can be found in food, including vegetables, which may account for 50–70% of an individual's intake of NO_3^- and NO_2^- [65]. Endogenous NOCs are also formed, and in many cases, this is because of the activities of microbes. Bacteria produce nitrosating agents [see 'Dinitrogen Trioxide (N_2O_3) and Nitrosative Deamination']. Further amines and amides are produced by bacterial decarboxylation of amino acids [65]. Heme has been suggested to catalyze the formation of NOCs [66]. Inhibitors of nitrosation are ingested as part of a diet and include vitamin C, vitamin E, and polyphenols [67]. The activated form of NOCs induce a number of methylated **DNA adducts** (of which over 12 are known) by SN1-nucleophilic substitution [68]. These alkylated DNA bases can be mutagenic if not repaired before replication [69]. SBS mutational signature 11 has been linked to the mutagenic activity of alkylating agents [70].

Acetaldehyde

Alcohol is classified as a group 1 human carcinogen. Worldwide, 3.6% of all cancer deaths and 3.5% of all cancer cases are attributable to alcohol consumption [71]. Ethanol, the psychoactive ingredient in alcoholic beverages, is believed to be the major causative compound of cancer in alcoholic beverages.

Ethanol is introduced into a catabolic pathway where it is broken down and the metabolites expelled via the urinary system. Ethanol is first metabolized by alcohol dehydrogenase (ADH), cytochrome P4502E1 (CYP2E1), and catalase, thereby forming acetaldehyde (ethanal). Acetaldehyde is further oxidized by aldehyde dehydrogenase to produce acetate. Aldehydes cause DNA damage in the form of DSBs and the Fanconi anemia pathway is responsible for the repair of this damage [72]. Aldehydes have been demonstrated to cause intrastrand crosslinking between adjacent guanine bases [73]. This can lead to the mutagenic event of GG>TT double **base substitution** which is a characteristic of the DBS2 mutational signature [55,73].

Bacteria can not only produce ethanol but also break it down into acetaldehyde. Oral taxa are known to be able to produce acetaldehyde from ethanol or glucose [74]. In addition, gut microbes can also produce acetaldehyde from sugars [75]. Indeed there have been reports of bacterial autobrewery syndrome (intoxication by ethanol formed by fermentation by microbes in the gut) in which a strain of *Klebsiella pneumoniae* was implicated [76]. This strain was also strongly associated with nonalcoholic fatty liver disease and fatty liver disease symptoms in a mouse model. Mutational signature 16 has been linked to alcohol consumption [77].

Disruption of the DNA Damage Response

Human DNA experiences repeated events of DNA damage throughout the cell cycle. The cell has a complex network of systems whose purpose is to ensure the fidelity of DNA. Known as the DNA damage response, this cellular system is responsible for detecting DNA damage, signaling its presence, and promoting a DNA repair cell cycle checkpoint and/or apoptosis.

The mismatch repair (MMR) mechanism is responsible for addressing base–base mismatches and insertion/deletion mispairs generated during DNA replication and recombination [78].

Enteropathogenic *E. coli* promotes the depletion of MSH2 and MLH1 proteins, which are crucially important for MMR, in cell models [79]. This phenomenon is dependent on the bacterial type-3 secretion effector EspF [79]. Furthermore, mitochondrial targeting of EspF is necessary for this activity. Colonic epithelial cells infected with enteropathogenic *E. coli* display an increased mutation rate particularly in microsatellite DNA sequences.

The human gastric pathogen *H. pylori* also inhibits the expression of MMR genes, in part through the modulation of miRNAs [80,81].

Mutational signature 6 is characterized by C>T transitions in a NpCpG trinucleotide context [82]. This mutational signature is associated with small indels (usually 1–3 bp) at nucleotide repeats. This indel pattern is equivalent to a phenomenon known as microsatellite instability. Microsatellite instability is caused by aberrations in the DNA MMR machinery. The origin of MMR deficiencies is genetic and/or epigenetic alterations in MMR genes. Microsatellite instability occurs in 15% of CRC genomes; 3% are associated with Lynch syndrome and 12% are associated with sporadic CRC [83].

Mutational Signatures as a Tool to Study the Effect of Microbes on the Human Genome

Multiomic experimental designs are well placed to delineate the relationship between the microbiota and the architecture of the cancer genome. Population studies in which both cancer genomic and microbiome data are assessed can provide information on the interaction between the cancer genetic architecture and the microbiota. However, there is a fundamental caveat with this type of experimental design. Cancer can take many years to form, and mutational mechanisms act at different time points of the natural tumor history. Furthermore, composition of the microbiota at most body sites is usually dynamic. Thus, a single snap shot of the microbiota may not be wholly related to the mutational signatures then identified. A prospective study where individual's microbiota are determined at pre- and post-transformation stages would allow for more informative comparisons between the microbiota and pretransformation mutational mechanisms. Additionally, individuals with precancer lesions such as Barrett's esophagus may be prime candidates to study due to their increased propensity to develop cancer. Studying cancer heterogeneity and evolutionary dynamics could allow for the identification of the timing of mutational mechanisms. Furthermore, recent advancements have allowed for mutational signature extraction from noncancerous tissue, thus allowing elucidation of microbe-associated mechanisms prior to transformation [84]. Experiments in which a microbe or a community of microbes are grown in the context of a model such as a cell line or organoids would help to eliminate confounders and make more direct correlations. Indeed, cultured cell lines were exposed to colibactin to identify its DNA sequence targets [22]. This target sequence was then cross-referenced with mutational signatures derived in population cancer genomic data to find clinically associated mutational signatures.

Concluding Remarks

Cancer prevention is under-researched when compared with therapeutic development, with only 2–9% of funding directed towards this area [85]. A high proportion of cancer cases and deaths could be avoided through modification of environmental risk factors. About 42% of cancer incidence in the USA is estimated as being attributable to modifiable risk factors – this figure is also reflected in the UK population [86]. Evidence is building in favor of the microbiota as an environmental modulator of cancer risk. We have outlined the multitude of ways that the metabolic activities of members of the human microbiota can lead to mutations.

Our ability to modulate the microbiota is improving steadily through the use of diet, antibiotics, phage therapy, fecal microbiota transplantation, prebiotics, probiotics, and live biotherapeutics

Outstanding Questions

What is the complete repertoire of modes by which the microbiota promotes DNA damage or compromises DNA integrity?

What is the exact mutational mechanisms by which microbes elicit mutations?

What are the mutational signatures that result from a particular microbiota-associated mutational mechanism?

How does the mutagenic potential of the microbiota vary within the population? This would need to take into consideration epidemiological factors such as age, diet, genetics, and other modifiers/risk factors.

How does this variation in the mutagenic capacity of the microbiota contribute to cancer risk?

What proportion of cancer genomes have microbial influence in their formation? Further, in cancer genomes with microbial influences, what is the quantitative impact it has (frequency per Mbp/ overall abundance)?

How might the microbiota protect genome stability and prevent cancer?

What are the necessary interventions that would be required in order to address these microbiota-associated mutational mechanisms?

[87]. Thus, one could plausibly develop strategies to alter the structure of an individual's microbiota in order to reduce its mutagenic potential (see Outstanding Questions). In order to make informed decisions on therapeutic interventions, a complete catalogue of microbe-associated mutational mechanisms is required. Furthermore, the relative impact of each mutational mechanism on the cancer genome needs to be delineated. Microbe-associated mutational mechanisms that have been found in a wide range of cancers as well as contributing to many mutations will take priority when deciding which mechanisms need to be addressed first.

We propose to leverage advancements in cancer genomics, namely in the form of mutational signatures, to associate microbes to mutational mechanisms. These can provide qualitative and quantitative information on the mutagenic effect that microbes undoubtedly have. It is possible that certain aspects of the microbiota activity protect against mutagenesis and cancer. These potential mechanisms need to be elucidated to enable harnessing the microbiota as prophylactic agents.

Acknowledgements

This work was supported by The Health Research Board of Ireland under Grant ILP-POR-2017-034, and by Science Foundation Ireland (APC/SFI/12/RC/2273) in the form of a research center, the APC Microbiome Ireland.

References

1. Williams, M.J. *et al.* (2019) Measuring clonal evolution in cancer with genomics. *Annu. Rev. Genomics Hum. Genet.* 20, 309–329
2. Helleday, T. *et al.* (2014) Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* 15, 585–598
3. Tomasetti, C. *et al.* (2017) Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* 355, 1330–1334
4. Tomasetti, C. and Vogelstein, B. (2015) Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347, 78–81
5. Sender, R. *et al.* (2016) Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* 14, e1002533
6. Gilbert, J.A. *et al.* (2018) Current understanding of the human microbiome. *Nat. Med.* 24, 392–400
7. Thomas, R.M. and Jobin, C. (2015) The microbiome and cancer: is the 'oncobiome' mirage real? *Trends Cancer* 1, 24–35
8. Belkaid, Y. and Hand, T.W. (2014) Role of the microbiota in immunity and inflammation. *Cell* 157, 121–141
9. Thomas, J.P. *et al.* (2018) The gut microbiota influences intestinal epithelial proliferative potential. *Gut* 67, A204
10. Tenaillon, O. *et al.* (2010) The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* 8, 207–217
11. Putze, J. *et al.* (2009) Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infect. Immun.* 77, 4696–4703
12. Cuevas-Ramos, G. *et al.* (2010) *Escherichia coli* induces DNA damage *in vivo* and triggers genomic instability in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 107, 11537–11542
13. Nougayrede, J.P. *et al.* (2006) *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* 313, 848–851
14. Arthur, J.C. *et al.* (2012) Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 338, 120–123
15. Buc, E. *et al.* (2013) High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. *PLoS One* 8, e56964
16. Dejea, C.M. *et al.* (2018) Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* 359, 592–597
17. Vizcaino, M.I. and Crawford, J.M. (2015) The colibactin warhead crosslinks DNA. *Nat. Chem.* 7, 411–417
18. Xue, M. *et al.* (2018) Characterization of natural colibactin-nucleobase adducts by tandem mass spectrometry and isotopic labeling. support for DNA alkylation by cyclopropane ring opening. *Biochemistry* 57, 6391–6394
19. Wilson, M.R. *et al.* (2019) The human gut bacterial genotoxin colibactin alkylates DNA. *Science* 363
20. Bossuet-Greif, N. *et al.* (2018) The colibactin genotoxin generates DNA interstrand cross-links in infected cells. *mBio* 9, e02393-17
21. Xue, M. *et al.* (2019) Structure elucidation of colibactin and its DNA cross-links. *Science* 365
22. Dziubańska-Kusibab, P.J. *et al.* (2019) Colibactin DNA damage signature indicates causative role in colorectal cancer. *bioRxiv* 819854
23. Jinadasa, R.N. *et al.* (2011) Cytolethal distending toxin: a conserved bacterial genotoxin that blocks cell cycle progression, leading to apoptosis of a broad range of mammalian cell lineages. *Microbiology-Sgm* 157, 1851–1875
24. Scott, D.A. and Kaper, J.B. (1994) Cloning and sequencing of the genes encoding Escherichia coli cytolethal distending toxin. *Infect. Immun.* 62, 244–251
25. Lara-Tejero, M. and Galan, J.E. (2001) CdtA, CdtB, and CdtC form a tripartite complex that is required for cytolethal distending toxin activity. *Infect. Immun.* 69, 4358–4365
26. Lara-Tejero, M. and Galan, J.E. (2000) A bacterial toxin that controls cell cycle progression as a deoxyribonuclease I-like protein. *Science* 290, 354–357
27. Nesic, D. *et al.* (2004) Assembly and function of a bacterial genotoxin. *Nature* 429, 429–433
28. Boesze-Battaglia, K. *et al.* (2016) A journey of cytolethal distending toxins through cell membranes. *Front. Cell. Infect. Microbiol.* 6, 81
29. He, Z. *et al.* (2019) *Campylobacter jejuni* promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut* 68, 289–300
30. Elwell, C.A. and Dreyfus, L.A. (2000) DNase I homologous residues in CdtB are critical for cytolethal distending toxin-mediated cell cycle arrest. *Mol. Microbiol.* 37, 952–963
31. Frisan, T. *et al.* (2003) The *Haemophilus ducreyi* cytolethal distending toxin induces DNA double-strand breaks and promotes ATM-dependent activation of RhoA. *Cell. Microbiol.* 5, 695–707
32. Fedor, Y. *et al.* (2013) From single-strand breaks to double-strand breaks during S-phase: a new mode of action of the *Escherichia coli* cytolethal distending toxin. *Cell. Microbiol.* 15, 1–15
33. Bezine, E. *et al.* (2016) Cell resistance to the cytolethal distending toxin involves an association of DNA repair mechanisms. *Sci. Rep.* 6, 36022
34. Dizdaroglu, M. (2015) Oxidatively induced DNA damage and its repair in cancer. *Mutat. Res. Rev. Mutat. Res.* 763, 212–245
35. Bernstein, H. *et al.* (2009) Bile acids as endogenous etiologic agents in gastrointestinal cancer. *World J. Gastroenterol.* 15, 3329–3340
36. Yazici, C. *et al.* (2017) Race-dependent association of sulfidogenic bacteria with colorectal cancer. *Gut* 66, 1983–1994

37. Ridlon, J.M. *et al.* (2016) Taurocholic acid metabolism by gut microbes and colon cancer. *Gut Microbes* 7, 201–215
38. Attene-Ramos, M.S. *et al.* (2010) DNA damage and toxicogenomic analyses of hydrogen sulfide in human intestinal epithelial FHs 74 Int cells. *Environ. Mol. Mutagen.* 51, 304–314
39. Cao, W. (2013) Endonuclease V: an unusual enzyme for repair of DNA deamination. *Cell. Mol. Life Sci.* 70, 3145–3156
40. Weiss, B. (2006) Evidence for mutagenesis by nitric oxide during nitrate metabolism in *Escherichia coli*. *J. Bacteriol.* 188, 829–833
41. Crane, B.R. *et al.* (2010) Bacterial nitric oxide synthases. *Annu. Rev. Biochem.* 79, 445–470
42. Kuypers, M.M.M. *et al.* (2018) The microbial nitrogen-cycling network. *Nat. Rev. Microbiol.* 16, 263–276
43. Tiso, M. and Schechter, A.N. (2015) Nitrate reduction to nitrite, nitric oxide and ammonia by gut bacteria under physiological conditions. *PLoS One* 10, e0119712
44. Klebanoff, S.J. *et al.* (2013) Myeloperoxidase: a front-line defender against phagocytosed microorganisms. *J. Leukoc. Biol.* 93, 185–198
45. Fedeles, B.I. *et al.* (2015) Intrinsic mutagenic properties of 5-chlorocytosine: a mechanistic connection between chronic inflammation and cancer. *Proc. Natl. Acad. Sci. U. S. A.* 112, E4571–E4580
46. Gungor, N. *et al.* (2010) Genotoxic effects of neutrophils and hypochlorous acid. *Mutagenesis* 25, 149–154
47. VanderVeen, L.A. *et al.* (2003) Induction of frameshift and base pair substitution mutations by the major DNA adduct of the endogenous carcinogen malondialdehyde. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14247–14252
48. Zhang, D. and Frenette, P.S. (2019) Cross talk between neutrophils and the microbiota. *Blood* 133, 2168–2177
49. Zhang, D.C. *et al.* (2015) Neutrophil ageing is regulated by the microbiome. *Nature* 525, 528 -+
50. Shinmura, K. *et al.* (2017) Mutation spectrum induced by 8-bromoguanine, a base damaged by reactive brominating species, in human cells. *Oxidative Med. Cell. Longev.* 2017, 7308501
51. Endo, Y. *et al.* (2008) Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. *Gastroenterology* 135, 889–898
52. Takai, A. *et al.* (2012) Targeting activation-induced cytidine deaminase prevents colon cancer development despite persistent colonic inflammation. *Oncogene* 31, 1733–1742
53. Matsumoto, Y. *et al.* (2007) *Helicobacter pylori* infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat. Med.* 13, 470–476
54. Kasar, S. *et al.* (2015) Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat. Commun.* 6, 8866
55. Alexandrov, L.B. *et al.* (2019) The repertoire of mutational signatures in human cancer. *bioRxiv* 322859
56. Kim, S.C. *et al.* (2005) Variable phenotypes of enterocolitis in interleukin 10-deficient mice monoassociated with two different commensal bacteria. *Gastroenterology* 128, 891–906
57. Wang, X. and Huycke, M.M. (2007) Extracellular superoxide production by *Enterococcus faecalis* promotes chromosomal instability in mammalian cells. *Gastroenterology* 132, 551–561
58. Wang, X.M. *et al.* (2008) *Enterococcus faecalis* induces aneuploidy and tetraploidy in colonic epithelial cells through a bystander effect. *Cancer Res.* 68, 9909–9917
59. Yang, Y.H. *et al.* (2013) Colon macrophages polarized by commensal bacteria cause colitis and cancer through the bystander effect. *Transl. Oncol.* 6, 596–606
60. Wang, X. *et al.* (2012) 4-hydroxy-2-nonenal mediates genotoxicity and bystander effects caused by *Enterococcus faecalis*-infected macrophages. *Gastroenterology* 142, 543–551 e7
61. Wang, X. *et al.* (2013) Cyclooxygenase-2 generates the endogenous mutagen trans-4-hydroxy-2-nonenal in *Enterococcus faecalis*-infected macrophages. *Cancer Prev. Res. (Phila.)* 6, 206–216
62. Wang, X. *et al.* (2015) Commensal bacteria drive endogenous transformation and tumour stem cell marker expression through a bystander effect. *Gut* 64, 459–468
63. Lundberg, J.O. *et al.* (2004) Nitrate, bacteria and human health. *Nat. Rev. Microbiol.* 2, 593–602
64. Lijinsky, W. (1999) N-Nitroso compounds in the diet. *Mutat. Res.* 443, 129–138
65. Habermeyer, M. *et al.* (2015) Nitrate and nitrite in the diet: how to assess their benefit and risk for human health. *Mol. Nutr. Food Res.* 59, 106–128
66. Bastide, N.M. *et al.* (2011) Heme iron from meat and risk of colorectal cancer: a meta-analysis and a review of the mechanisms involved. *Cancer Prev. Res. (Phila.)* 4, 177–184
67. Bartsch, H. *et al.* (1993) Inhibition of nitrosation. *Basic Life Sci.* 61, 27–44
68. Fahrer, J. and Kaina, B. (2013) O6-methylguanine-DNA methyltransferase in the defense against N-nitroso compounds and colorectal cancer. *Carcinogenesis* 34, 2435–2442
69. Shrivastav, N. *et al.* (2010) Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis* 31, 59–70
70. Kucab, J.E. *et al.* (2019) A compendium of mutational signatures of environmental agents. *Cell* 177, 821–836 e16
71. Boffetta, P. *et al.* (2006) The burden of cancer attributable to alcohol drinking. *Int. J. Cancer* 119, 884–887
72. Garaycoechea, J.I. *et al.* (2018) Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nature* 553, 171–177
73. Sonohara, Y. *et al.* (2019) Acetaldehyde forms covalent GG intrastrand crosslinks in DNA. *Sci. Rep.* 9, 660
74. Moritani, K. *et al.* (2015) Acetaldehyde production by major oral microbes. *Oral Dis.* 21, 748–754
75. Elshaghabee, F.M. *et al.* (2016) Ethanol production by selected intestinal microorganisms and lactic acid bacteria growing under different nutritional conditions. *Front. Microbiol.* 7, 47
76. Yuan, J. *et al.* (2019) Fatty liver disease caused by high-alcohol-producing *Klebsiella pneumoniae*. *Cell Metab.* 30, 675–688 e7
77. Letouze, E. *et al.* (2017) Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* 8, 1315
78. Li, G.M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res.* 18, 85–98
79. Maddocks, O.D. *et al.* (2013) An *Escherichia coli* effector protein promotes host mutation via depletion of DNA mismatch repair proteins. *MBio* 4, e00152-13
80. Kim, J.J. *et al.* (2002) *Helicobacter pylori* impairs DNA mismatch repair in gastric epithelial cells. *Gastroenterology* 123, 542–553
81. Santos, J.C. *et al.* (2017) *Helicobacter pylori* infection modulates the expression of miRNAs associated with DNA mismatch repair pathway. *Mol. Carcinog.* 56, 1372–1379
82. Alexandrov, L.B. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500, 415–421
83. Moller, P. *et al.* (2017) Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. *Gut* 66, 464–472
84. Lee-Six, H. *et al.* (2019) The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* 574, 532–537
85. Eckhouse, S. *et al.* (2008) Trends in the global funding and activity of cancer research. *Mol. Oncol.* 2, 20–32
86. Islami, F. *et al.* (2018) Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA Cancer J. Clin.* 68, 31–54
87. O'Toole, P.W. *et al.* (2017) Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. *Nat. Microbiol.* 2, 17057
88. Baker, M. (2012) Structural variation: the genome's hidden architecture. *Nat. Methods* 9, 133–137
89. Nik-Zainal, S. *et al.* (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54
90. van Loon, B. *et al.* (2010) Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. *DNA Repair* 9, 604–616
91. Tan, X. *et al.* (1999) Comparison of the mutagenic properties of 8-oxo-7,8-dihydro-2′-deoxyadenosine and 8-oxo-7,8-dihydro-2′-deoxyguanosine DNA lesions in mammalian cells. *Carcinogenesis* 20, 2287–2292

92. Whitaker, A.M. *et al.* (2017) Base excision repair of oxidative DNA damage: from mechanism to disease. *Front. Biosci. (Landmark Ed.)* 22, 1493–1522
93. Satou, K. *et al.* (2007) Mutagenic effects of 8-hydroxy-dGTP in live mammalian cells. *Free Radic. Biol. Med.* 42, 1552–1560
94. Satou, K. *et al.* (2003) Mutagenic effects of 2-hydroxy-dATP on replication in a HeLa extract: induction of substitution and deletion mutations. *Nucleic Acids Res.* 31, 2570–2575
95. Viel, A. *et al.* (2017) A specific mutational signature associated with DNA 8-oxoguanine persistence in MUTYH-defective colorectal cancer. *EBioMedicine* 20, 39–49