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]*

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

* Corresponding author. E-mail: pwotoole@ucc.ie

Keywords

microbiota; microbiome; DNA damage; mutational mechanism; mutational signatures

Abstract

Cancers arise through the process of somatic evolution fuelled 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 tumour microenvironment and proliferation. In addition, microbes have been shown to elicit DNA damage which provides the substrate for somatic mutations. An understanding of microbiota-driven mutational mechanism 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 the role of the microbiota

Oncogenesis is driven by the Darwinian selection of somatic mutations (see Glossary) 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, which are (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. (ii) Stochastic errors associated with DNA replication and other phenomena. These are seemingly inevitable random mutations which 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 harbour median values of 33% and 83% stochastic mutations respectively [3]. However, epidemiologic evidence indicates that a high proportion (~90%) of cases are attributable to environmental factors, i.e. tobacco smoking and HPV infection, respectively. The managing of environmental risk factors is thus crucial in cancer prevention, even though stochastic/replicative mechanisms are the major drivers (See ref 3 for a more detailed discussion). However a complete catalogue of environmental factors that contribute cancer risk is lacking. A large number of known carcinogens promote oncogenesis by causing mutagenesis e.g. ultraviolet 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 3.8×10^{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 focus on bacteria, upon which this review is focussed, the human microbiota includes members from all 5 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 cancers[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, immune modulation in the form of inflammation caused by the microbiota is an intense area of research [8]. Effort has 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). 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.

In this review 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 4 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* (for polyketide synthase) islands[10]. The 54-kb *pks* island is a biosynthetic gene cluster encoding for a non-ribosomal peptide synthetase (NRPS)–polyketide synthase (PKS) hybrid gene cluster, which encodes for colibactin [11]. Colibactin can cause Double-strand breaks (DSB) in mammalian DNA thereby promoting genome instability and an increase in mutation

rate [12, 13]. It is not currently known how colibactin 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, being detected at a rate of 20% in the mucosa of healthy individuals but 55%-67% in patients with colorectal cancer (CRC) [14, 15]. Furthermore, pks+ *E. coli* was disproportionally frequently identified in subjects with familial adenomatous polyposis (FAP) compared to healthy controls [16]. Monocolonization of azoxymethane (AOM)–treated IL10−/− mice with pks+ *E. coli* promoted 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 inter-strand cross-links and activation of the DNA damage response including Fanconi anemia DNA repair [20]. Recent structural analysis revealed that colibactin contains two conjoined warheads enabling its ability to cause DNA crosslinks [21]. Double strands breaks are not believed to be a direct consequence of colibactin activity but rather occur due to replication stress caused by DNA cross-links [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 were found to be enriched in a number of cancers including CRC and stomach cancer compared with a WWWWW motif. Two mutational signatures were found to be linked with the AAWWTT colibactin motif, SBS28 and SBS41[22]. Mutational signature SBS28 has been associated with POLE mutation while Mutational signature SBS41 has no known aetiology.

Cytolethal distending toxin (CDT)

The cytolethal distending toxin (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 sub-dominant members of the human gut microbiota.

CDT is a heteromultimeric protein comprised of 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 harbouring the CDT operon promoted colorectal tumorigenesis compared to 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 double strand breaks directly [32]. At low to moderate doses, CdtB causes single strand breaks (SSB) 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 DSB directly by two cuts to the DNA backbone that are juxtaposed to each other [32] .

Reactive oxygen species

Reactive oxygen species (ROS) are a chemically reactive family of molecules containing oxygen which include the highly reactive hydroxyl radical (OH−), superoxide radical (O2−), and non-radical hydrogen peroxide (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 produce reactive oxygen species through varied means. For example, primary bile acids, cholic acid (CA) and chenodeoxycholic acid; (CDCA) are synthesised by the liver and are secreted into the small intestine from the gall bladder. A small proportion of these bile salts are 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 exist maybe 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].

Hydrogen sulphide (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].

Dinitrogen trioxide and nitrosative deamination

Nitrosative deamination is deamination mediated by dinitrogen trioxide (N_2O_3, nitrous anhydride). In this phenomenon, dinitrogen trioxide 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].

Dinitrogen trioxide can be generated from the autooxidation of nitric oxide (NO-) or the condensation of nitrous acid (HNO_2)[40]. GIT microbes can produce endogenous nitric oxide and/or nitrous acid by four mechanisms: (i) The hemethiolate monooxygenase, nitric oxide synthase (NOS), oxidises L-arginine (Arg) to produce nitric oxide [41] (ii) Denitrification of nitrate (NO_3^-) to nitrogen (N_2), which is an important part of the nitrogen cycle and is carried out by denitrifying bacteria and plants. During denitrification, nitric oxide is produced by one-electron reduction of nitrite (NO_2^-) by heme or Cu-containing nitrite reductases[42]. (iii) Respiratory nitrite ammonification (also referred to as dissimilatory nitrate reduction to ammonium)[42]. (iv) Acidic non-enzymatic reduction of nitrite 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 hypochlorous acid (HOCl) which is produced by myeloperoxidase using as substrates the chloride ions and hydrogen peroxide (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 found that such a 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 neutrophilic biology[48]. A recent study in a mouse model demonstrated that neutrophil pro-inflammatory 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 signalling pathways[49]. A depletion of the microbiota was mirrored in the number of aged neutrophils and an improvement in inflammatory disease.

Hypobromous acid production

Eosinophils are granular leukocytes with a multifunctional role in immune biology. Eosinophils secrete eosinophil peroxidase which catalyzes the formation of hypobromous acid (HOBO) from hydrogen peroxide 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

Activation-induced cytidine deaminase (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 colorectal cancers [51]. The expression of AID in colonic epithelial cell lines induced an increase in the mutation rates in these cells [51]. Knock-out of AID in IL10 null mice attenuated the mutation rate in their colonic cells and also inhibits CRC development[52]. Inflammation seems to be key to this aberrant activity. *H. pylori* infection, which is known to induce inflammation, promotes ectopic expression of AID in non-tumorous epithelial tissues [53]

Whole genome analyses in chronic lymphocytic leukaemia revealed that the activity of AID may produces two types of substitution pattern (i) a 'canonical AID signature' characterised by C to T/G substitutions at WRCY motifs near active transcriptional start sites and (ii) a 'non-canonical AID signature' characterised by A to C mutations at WA (W=A or T) motifs occurring genome-wide in a non-clustered fashion [54]. These mutational processes have been assigned to mutational signatures SBS84 and SBS85[55].

By-stander effect and Enterococcus faecalis

Enterococcus faecalis is known to promote CRC oncogenesis in interleukin 10 -/- mice [56]. *E. faecalis* can promote the bystander effect which leads to double-stranded DNA breaks, 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 diet of the host and the gut microbiota are inextricably linked. GIT 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 nitrate and nitrite 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 nitrate and nitrite [65]. Endogenous NOCs are also formed and in many cases, this is because of the activities of microbes. Firstly, bacteria produce nitrosating agents (See Dinitrogen trioxide and nitrosative deamination). Further amines and amides are produced by bacterial decarboxylation of amino acids [65]. Heme has been suggested to catalyse 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 carcinogen (carcinogenic to humans). Worldwide, 3.6% of all cancer deaths and 3.5% of all cancer cases are attributable to alcohol consumption[71]. Ethanol (C_2H_5OH), 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 oxidised by aldehyde dehydrogenase to produce acetate. Aldehydes cause DNA damage in the form of double strand breaks and the Fanconi anaemia pathway is responsible for the repair of this damage [72]. Aldehydes has been demonstrated to cause intrastrand crosslink between adjacent guanine bases[73]. This can lead to the mutagenic event of GG>TT double base substitution which is a characteristics of Mutational signature DBS2 [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 non-alcoholic fatty liver disease and fatty liver disease symptoms in a mouse model. Mutational signature 16 has been link to alcohol consumption [77].

Disruption to 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, signalling its presence, promoting DNA repair cell cycle checkpoint and/or apoptosis.

The mismatch repair mechanism is responsible for addressing base-base mismatches and insertion/deletion mis-pairs generated during DNA replication and recombination[78].

Enteropathogenic *Escherichia coli* was found to promote the depletion of MSH2 and MLH1 proteins, which are crucially important for mismatch repair in cell models[79]. This phenomenon was found to be dependent on the bacterial type-3 secretion effector EspF[79]. Furthermore, mitochondrial targeting of EspF was 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 *Helicobacter pylori* also inhibits the expression of MMR gene expression, in part through the modulation of miRNAs [80, 81].

Mutational signature 6 is characterised by C>T transitions at an NpCpG trinucleotide context [82]. This mutational signature is associated with small indels (usually 1-3bp) at nucleotide repeats. This indel pattern is equivalent to phenomena known as microsatellite instability. Microsatellite instability is caused by aberrations in the DNA mismatch repair (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 while 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 supremely placed to delineate the relationship between the microbiota and the architecture of the cancer genome. Population studies in which both cancer genomic and microbiome 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 tumour 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 pre-transformation mutational mechanisms. Additionally, individuals with pre-cancer lesions such as Barrett's oesophagus 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. Futhermore, recent advancements have allowed for mutational signature extraction from non-cancerous tissue thus allowing elucidation of microbial 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. Dziubańska-Kusibab and colleagues used cultured cell lines exposed to colibactin to identify its DNA sequence targets. Furthermore this target sequence was then cross-referenced with mutational signatures derived in population cancer genomic data to find clinically associated mutational signatures (See colibactin section).

Concluding Remarks

Cancer prevention is relatively under-researched when compared to therapeutic development, with only 2 to 9% of funding put towards this area [85]. A high proportion of cancer cases and cancer deaths could be avoided through modification of environmental risk factors. About 42% of cancer incidences in the US are estimated as being attributable to modifiable risk factors - this figure is also reflected in the UK population [86]. Evidence is building in favour of the microbiota as an environmental modulator of cancer risk. We 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, featuring diet, antibiotics, phage therapy, faecal microbiota transplantation (FMT), prebiotics, probiotics and Live Biotherapeutics[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 microbial-associated mutational mechanisms is required. Furthermore, the relative impact of each mutational mechanisms on the cancer genome need to be delineated. Microbial-associated mutational mechanisms which have both been found in a wide range of cancers as well as contributing to a great number of mutations will take priority when deciding what 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 mechanism need to be elucidated to enable the harnessing the microbiota as prophylactic agents.

Acknowledgments

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 centre, the APC Microbiome Institute.

Glossary

Base substitutions: A type of mutation in which one base is replaced by another in DNA.

Chromosomal instability: A phenomena which leads to alterations in chromosome number and/or structure.

DNA adduct: Formed by the addition of a chemical moiety to a DNA base

DNA alkylation: 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)

DNA deamination: 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: This is where both strands of DNA which are juxtaposed to each other

Environmental risk factor: A thing or process which 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 phenomena which lead 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: A mutation which occurs in a somatic cell and is thus not heritable.

Box1 | 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 great 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 non-negative 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 characterised are as follows: 49 single base substitutions, 11 doublet base substitutions, four clustered base substitutions (DBS), 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 characterised 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 maybe base-paired with adenine during replication, provided DNA repair is not executed. Many mutational signatures described do not have a known aetiology.

Table 1. Microbial associated mechanisms and genomic consequences

Source	Involvement of microbiota features	Key role in a mutational mechanism	Postulated effected on cancer genomic landscape	Reference
Activation-induced cytidine deaminase (AID)	*Helicobacter pylori* infection cause ectopic expression of AID	Cytosine deamination at specific motifs	Mutational signatures SBS84 and SBS85	[53, 55]
Acetaldehyde	Various inhabitants of produce ethanol and are capable metabolic act on it to produces acetaldehyde	N2-ethylidenedeoxyguanosine, Guanine- guanine intrastrand crosslinks	GG-to-TT base substitution. Mutational signature DBS2	[73]
Colibactin	Expressed by *Escherichia coli* containing a *pks* island	Adenine – adenine intra-strand crosslinks, Double strand breaks,	DSBs at an AAWWTT pentanucleotides motif. Mutational signatures SBS28 and SBS41	[22]
Cytolethal distending toxin (CDT)	Produced by various Gram-negative bacteria including enteropathogenic *Escherichia coli*, *Campylobacter* species, *Shigella* species and *Haemophilus ducreyi*	Single strand breaks and Double-strand breaks	Infidelity of DNA repair can lead to structural variants such as indels	[55]
Disruption of DNA mismatch repair	*Helicobacter pylori* and Enteropathogenic	Deletion of MMR proteins	Microsatellite instability, Mutational	[79, 80, 82]

	Escherichia coli can disrupt mismatch repair		signature SBS6, ID1 and ID2	
Dinitrogen trioxide	Metabolic activities of the microbiota can produces precursors to N203 e.g. denitrifying bacteria	Nitrosative deamination	Various base substitutions e.g. Adenine nitrosative deamination to Hypoxanthine can lead to T>A substitution	[39, 42]
Hypobromous acid	Eosinophil's produce Hypobromous acid. The microbiota can influence eosinophic biology	8-bromoguanine	G > T primarily but also G > C, G > A, and delG	[50]
Hypochlorous acid	HOCL is produce by Neutrophils. The microbiota can influence neutrophil inflammatory status	Formation of 5-chlorocytosine (5ClC), formation of malondialdehyde	C>T, G >A, G>T substitutions	[45, 46]
N-nitroso compounds (NOCs)	Microbes play a role in the production of nitrosating agents and produces biogenic amine	Alkylated DNA base	Various base substitutions e.g O6-methylguanine (O6-MeG) can cause a G(C)>A(T) transition	[69]
Reactive oxygen species	Various metabolic activities	Oxidative Base Lesions	G to T transversion, SBS Mutational signatures 18 and 36	[90]
4-hydroxy-2-nonenal	*Enterococcus faecalis* induces the bystander effect via polarising	Exocyclic HNE-DNA adducts	Chromosomal instability	[60]

	marcophages. Polarised marcophages produces 4-hydroxy-2-nonenal			

Box 2 | Oxidative DNA Base Lesions

Guanine has the lowest redox potential of the native bases and is thus the most readily oxidised. Two common oxidative base lesions which 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[91]. 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[91, 92]. Replication of DNA containing 8-oxo-7,8-dihydro-2'-deoxyguanosine and 2,6-diamino-4-oxo-5-formamidopyrimidine (FapyG) are shown to induce G:C to T:A (C >A) and G:C to T:A (C >A) respectively[93].

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-hydroxydeoxyadenosine 5'-triphosphate (2-OH-dATP). 8-OH-dGTP has been demonstrated to induce A:T to C:G transversions when introduced into COS-7 mammalian cells[94]. *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)[95].

Mutational signatures 18 and 36 have been suggested to be attributed to reactive oxygen species. Mutational signature 36 has been specifically attributed to ROS in the context of MUTYH-Associated Polyposis (MAP) syndrome [90]. MAP syndrome is defined by biallelic germline mutation of MUTYH gene and is a colorectal polyposis which 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[93].

References

1. Williams, M.J. et al. (2019) Measuring Clonal Evolution in Cancer with Genomics. Annual Review of Genomics and Human Genetics, Vol 20, 2019 20, 309-329.
2. Helleday, T. et al. (2014) Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet 15 (9), 585-98.
3. Tomasetti, C. et al. (2017) Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science 355 (6331), 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 (6217), 78-81.
5. Sender, R. et al. (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol 14 (8), e1002533.
6. Gilbert, J.A. et al. (2018) Current understanding of the human microbiome. Nat Med 24 (4), 392-400.
7. Thomas, R.M. and Jobin, C. (2015) The Microbiome and Cancer: Is the 'Oncobiome' Mirage Real? Trends Cancer 1 (1), 24-35.
8. Belkaid, Y. and Hand, T.W. (2014) Role of the microbiota in immunity and inflammation. Cell 157 (1), 121-41.
9. Thomas, J.P. et al. (2018) The Gut Microbiota Influences Intestinal Epithelial Proliferative Potential. Gut 67, A204-A204.
10. Tenaillon, O. et al. (2010) The population genetics of commensal Escherichia coli. Nat Rev Microbiol 8 (3), 207-17.
11. Putze, J. et al. (2009) Genetic Structure and Distribution of the Colibactin Genomic Island among Members of the Family Enterobacteriaceae. Infection and Immunity 77 (11), 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 (25), 11537-42.
13. Nougayrede, J.P. et al. (2006) Escherichia coli induces DNA double-strand breaks in eukaryotic cells. Science 313 (5788), 848-51.
14. Arthur, J.C. et al. (2012) Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 338 (6103), 120-3.
15. Buc, E. et al. (2013) High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. PLoS One 8 (2), e56964.
16. Dejea, C.M. et al. (2018) Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science 359 (6375), 592-597.
17. Vizcaino, M.I. and Crawford, J.M. (2015) The colibactin warhead crosslinks DNA. Nat Chem 7 (5), 411-7.
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 (45), 6391-6394.
19. Wilson, M.R. et al. (2019) The human gut bacterial genotoxin colibactin alkylates DNA. Science 363 (6428).
20. Bossuet-Greif, N. et al. (2018) The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells. Mbio 9 (2).
21. Xue, M. et al. (2019) Structure elucidation of colibactin and its DNA cross-links. Science 365 (6457).
22. Dziubańska-Kusibab, P.J. et al. (2019) Colibactin DNA damage signature indicates causative role in colorectal cancer. bioRxiv, 819854.

509 23. Jinadasa, R.N. et al. (2011) Cytolethal distending toxin: a conserved bacterial genotoxin that
510 blocks cell cycle progression, leading to apoptosis of a broad range of mammalian cell lineages.
511 Microbiology-Sgm 157, 1851-1875.
512 24. Scott, D.A. and Kaper, J.B. (1994) Cloning and sequencing of the genes encoding Escherichia coli
513 cytolethal distending toxin. Infect Immun 62 (1), 244-51.
514 25. Lara-Tejero, M. and Galan, J.E. (2001) CdtA, CdtB, and CdtC form a tripartite complex that is
515 required for cytolethal distending toxin activity. Infection and Immunity 69 (7), 4358-4365.
516 26. Lara-Tejero, M. and Galan, J.E. (2000) A bacterial toxin that controls cell cycle progression as a
517 deoxyribonuclease I-like protein. Science 290 (5490), 354-357.
518 27. Nesic, D. et al. (2004) Assembly and function of a bacterial genotoxin. Nature 429 (6990), 429-33.
519 28. Boesze-Battaglia, K. et al. (2016) A Journey of Cytolethal Distending Toxins through Cell
520 Membranes. Front Cell Infect Microbiol 6, 81.
521 29. He, Z. et al. (2019) Campylobacter jejuni promotes colorectal tumorigenesis through the action
522 of cytolethal distending toxin. Gut 68 (2), 289-300.
523 30. Elwell, C.A. and Dreyfus, L.A. (2000) DNase I homologous residues in CdtB are critical for
524 cytolethal distending toxin-mediated cell cycle arrest. Mol Microbiol 37 (4), 952-63.
525 31. Frisan, T. et al. (2003) The Haemophilus ducreyi cytolethal distending toxin induces DNA double-
526 strand breaks and promotes ATM-dependent activation of RhoA. Cell Microbiol 5 (10), 695-707.
527 32. Fedor, Y. et al. (2013) From single-strand breaks to double-strand breaks during S-phase: a new
528 mode of action of the Escherichia coli Cytolethal Distending Toxin. Cellular Microbiology 15 (1), 1-15.
529 33. Bezine, E. et al. (2016) Cell resistance to the Cytolethal Distending Toxin involves an association
530 of DNA repair mechanisms. Sci Rep 6, 36022.
531 34. Dizdaroglu, M. (2015) Oxidatively induced DNA damage and its repair in cancer. Mutat Res Rev
532 Mutat Res 763, 212-45.
533 35. Bernstein, H. et al. (2009) Bile acids as endogenous etiologic agents in gastrointestinal cancer.
534 World Journal of Gastroenterology 15 (27), 3329-3340.
535 36. Yazici, C. et al. (2017) Race-dependent association of sulfidogenic bacteria with colorectal cancer.
536 Gut 66 (11), 1983-1994.
537 37. Ridlon, J.M. et al. (2016) Taurocholic acid metabolism by gut microbes and colon cancer. Gut
538 Microbes 7 (3), 201-15.
539 38. Attene-Ramos, M.S. et al. (2010) DNA damage and toxicogenomic analyses of hydrogen sulfide in
540 human intestinal epithelial FHs 74 Int cells. Environ Mol Mutagen 51 (4), 304-14.
541 39. Cao, W. (2013) Endonuclease V: an unusual enzyme for repair of DNA deamination. Cell Mol Life
542 Sci 70 (17), 3145-56.
543 40. Weiss, B. (2006) Evidence for mutagenesis by nitric oxide during nitrate metabolism in
544 Escherichia coli. J Bacteriol 188 (3), 829-33.
545 41. Crane, B.R. et al. (2010) Bacterial nitric oxide synthases. Annu Rev Biochem 79, 445-70.
546 42. Kuypers, M.M.M. et al. (2018) The microbial nitrogen-cycling network. Nat Rev Microbiol 16 (5),
547 263-276.
548 43. Tiso, M. and Schechter, A.N. (2015) Nitrate reduction to nitrite, nitric oxide and ammonia by gut
549 bacteria under physiological conditions. PLoS One 10 (3), e0119712.
550 44. Klebanoff, S.J. et al. (2013) Myeloperoxidase: a front-line defender against phagocytosed
551 microorganisms. Journal of Leukocyte Biology 93 (2), 185-198.
552 45. Fedeles, B.I. et al. (2015) Intrinsic mutagenic properties of 5-chlorocytosine: A mechanistic
553 connection between chronic inflammation and cancer. Proceedings of the National Academy of
554 Sciences of the United States of America 112 (33), E4571-E4580.
555 46. Gungor, N. et al. (2010) Genotoxic effects of neutrophils and hypochlorous acid. Mutagenesis 25
556 (2), 149-54.
557 47. VanderVeen, L.A. et al. (2003) Induction of frameshift and base pair substitution mutations by
558 the major DNA adduct of the endogenous carcinogen malondialdehyde. Proceedings of the National
559 Academy of Sciences of the United States of America 100 (24), 14247-14252.

48. Zhang, D. and Frenette, P.S. (2019) Cross talk between neutrophils and the microbiota. Blood 133 (20), 2168-2177.
49. Zhang, D.C. et al. (2015) Neutrophil ageing is regulated by the microbiome. Nature 525 (7570), 528-+.
50. Shinmura, K. et al. (2017) Mutation Spectrum Induced by 8-Bromoguanine, a Base Damaged by Reactive Brominating Species, in Human Cells. Oxid 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 (3), 889-898.
52. Takai, A. et al. (2012) Targeting activation-induced cytidine deaminase prevents colon cancer development despite persistent colonic inflammation. Oncogene 31 (13), 1733-42.
53. Matsumoto, Y. et al. (2007) Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med 13 (4), 470-6.
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 (4), 891-906.
57. Wang, X. and Huycke, M.M. (2007) Extracellular superoxide production by Enterococcus faecalis promotes chromosomal instability in mammalian cells. Gastroenterology 132 (2), 551-61.
58. Wang, X.M. et al. (2008) Enterococcus faecalis Induces Aneuploidy and Tetraploidy in Colonic Epithelial Cells through a Bystander Effect. Cancer Research 68 (23), 9909-9917.
59. Yang, Y.H. et al. (2013) Colon Macrophages Polarized by Commensal Bacteria Cause Colitis and Cancer through the Bystander Effect. Translational Oncology 6 (5), 596-+.
60. Wang, X. et al. (2012) 4-hydroxy-2-nonenal mediates genotoxicity and bystander effects caused by Enterococcus faecalis-infected macrophages. Gastroenterology 142 (3), 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 (3), 206-16.
62. Wang, X. et al. (2015) Commensal bacteria drive endogenous transformation and tumour stem cell marker expression through a bystander effect. Gut 64 (3), 459-68.
63. Lundberg, J.O. et al. (2004) Nitrate, bacteria and human health. Nat Rev Microbiol 2 (7), 593-602.
64. Lijinsky, W. (1999) N-Nitroso compounds in the diet. Mutat Res 443 (1-2), 129-38.
65. Habermeyer, M. et al. (2015) Nitrate and nitrite in the diet: How to assess their benefit and risk for human health. Molecular Nutrition & Food Research 59 (1), 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 (2), 177-84.
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 (11), 2435-42.
69. Shrivastav, N. et al. (2010) Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. Carcinogenesis 31 (1), 59-70.
70. Kucab, J.E. et al. (2019) A Compendium of Mutational Signatures of Environmental Agents. Cell 177 (4), 821-836 e16.
71. Boffetta, P. et al. (2006) The burden of cancer attributable to alcohol drinking. Int J Cancer 119 (4), 884-7.
72. Garaycoechea, J.I. et al. (2018) Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. Nature 553 (7687), 171-177.
73. Sonohara, Y. et al. (2019) Acetaldehyde forms covalent GG intrastrand crosslinks in DNA. Sci Rep 9 (1), 660.
74. Moritani, K. et al. (2015) Acetaldehyde production by major oral microbes. Oral Dis 21 (6), 748-54.

611 75. Elshaghabee, F.M. et al. (2016) Ethanol Production by Selected Intestinal Microorganisms and
612 Lactic Acid Bacteria Growing under Different Nutritional Conditions. Front Microbiol 7, 47.
613 76. Yuan, J. et al. (2019) Fatty Liver Disease Caused by High-Alcohol-Producing Klebsiella
614 pneumoniae. Cell Metab 30 (4), 675-688 e7.
615 77. Letouze, E. et al. (2017) Mutational signatures reveal the dynamic interplay of risk factors and
616 cellular processes during liver tumorigenesis. Nat Commun 8 (1), 1315.
617 78. Li, G.M. (2008) Mechanisms and functions of DNA mismatch repair. Cell Res 18 (1), 85-98.
618 79. Maddocks, O.D. et al. (2013) An Escherichia coli effector protein promotes host mutation via
619 depletion of DNA mismatch repair proteins. MBio 4 (3), e00152-13.
620 80. Kim, J.J. et al. (2002) Helicobacter pylori impairs DNA mismatch repair in gastric epithelial cells.
621 Gastroenterology 123 (2), 542-53.
622 81. Santos, J.C. et al. (2017) Helicobacter pylori infection modulates the expression of miRNAs
623 associated with DNA mismatch repair pathway. Mol Carcinog 56 (4), 1372-1379.
624 82. Alexandrov, L.B. et al. (2013) Signatures of mutational processes in human cancer. Nature 500
625 (7463), 415-21.
626 83. Moller, P. et al. (2017) Cancer incidence and survival in Lynch syndrome patients receiving
627 colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome
628 database. Gut 66 (3), 464-472.
629 84. Lee-Six, H. et al. (2019) The landscape of somatic mutation in normal colorectal epithelial cells.
630 Nature 574 (7779), 532-537.
631 85. Eckhouse, S. et al. (2008) Trends in the global funding and activity of cancer research. Mol Oncol
632 2 (1), 20-32.
633 86. Islami, F. et al. (2018) Proportion and number of cancer cases and deaths attributable to
634 potentially modifiable risk factors in the United States. CA Cancer J Clin 68 (1), 31-54.
635 87. O'Toole, P.W. et al. (2017) Next-generation probiotics: the spectrum from probiotics to live
636 biotherapeutics. Nat Microbiol 2, 17057.
637 88. Baker, M. (2012) Structural variation: the genome's hidden architecture. Nat Methods 9 (2), 133-
638 7.
639 89. Nik-Zainal, S. et al. (2016) Landscape of somatic mutations in 560 breast cancer whole-genome
640 sequences. Nature 534 (7605), 47-54.
641 90. Viel, A. et al. (2017) A Specific Mutational Signature Associated with DNA 8-Oxoguanine
642 Persistence in MUTYH-defective Colorectal Cancer. EBioMedicine 20, 39-49.
643 91. van Loon, B. et al. (2010) Oxygen as a friend and enemy: How to combat the mutational potential
644 of 8-oxo-guanine. DNA Repair 9 (6), 604-616.
645 92. Tan, X. et al. (1999) Comparison of the mutagenic properties of 8-oxo-7,8-dihydro-2'-
646 deoxyadenosine and 8-oxo-7,8-dihydro-2'-deoxyguanosine DNA lesions in mammalian cells.
647 Carcinogenesis 20 (12), 2287-92.
648 93. Whitaker, A.M. et al. (2017) Base excision repair of oxidative DNA damage: from mechanism to
649 disease. Front Biosci (Landmark Ed) 22, 1493-1522.
650 94. Satou, K. et al. (2007) Mutagenic effects of 8-hydroxy-dGTP in live mammalian cells. Free Radic
651 Biol Med 42 (10), 1552-60.
652 95. Satou, K. et al. (2003) Mutagenic effects of 2-hydroxy-dATP on replication in a HeLa extract:
653 induction of substitution and deletion mutations. Nucleic Acids Res 31 (10), 2570-5.

654