COLORECTAL CANCER

Microbiota and colorectal cancer: colibactin makes its mark

Janelle C. Arthur

Colibactin-producing bacteria are abundant in the gut microbiota of colorectal cancer tumours and promote colon tumorigenesis in mouse models. Now, a new study demonstrates a direct link between exposure of human intestinal epithelial cells to colibactin and two unique mutational fingerprints found in human colorectal tumours.

Refers to Pleguezuelos-Manzano, C. et al. Mutational signature in colorectal cancer caused by genotoxic *pks*$^+$ *E. coli*. *Nature* https://doi.org/10.1038/s41586-020-2080-8 (2020).

Members of the gut microbiota interact with their mammalian hosts, influencing health and disease over a lifetime. Although this diverse microbial community is essential for physiological development and immunological homeostasis, alterations to this ecosystem have been linked to chronic diseases including metabolic syndrome, inflammatory bowel diseases and cancer. Many studies suggest that colorectal cancer (CRC) in particular is influenced by the bloom or mislocalization of potentially harmful members of the gut microbiota, most notably *pks*$^+$ *Escherichia coli*, which produces the genotoxin colibactin[1]. In a new study, Pleguezuelos-Manzano et al. identify a unique mutational signature caused by exposure to colibactin-producing *E. coli* that is enriched in human CRC tumours and metastases[2]. This molecular fingerprint provides a direct link between colibactin exposure and the DNA damage patterns that drive CRC development.

First described in 2006, the genotoxic effects of colibactin were attributed to a 54 kb genomic island that encodes a multi-enzyme assembly line of non-ribosomal peptide and polyketide synthases[3]. Commensal and pathogenic *E. coli* that harboured this '*pks*' island induced mammalian cell cycle arrest and activation of DNA repair machinery in cultured cells[3]. These results suggested that the products of *pks* were microbially derived genotoxins. The pro-tumorigenic effect of *pks*$^+$ *E. coli* was revealed 6 years later in a study showing that *pks*$^+$ *E. coli* is prevalent in colon tissue of patients with CRC and also that tumour multiplicity and invasion is enhanced in mouse models[1]. This putative genotoxin continued to interest researchers, with additional papers validating the pro-tumorigenic effects of *pks* in mouse models and others defining the function of various *pks* genes required for the genotoxic effects of colibactin[4].

Although the precise chemical identity of bioactive colibactin remains elusive, analyses of inactive precursors and stable colibactin–DNA lesions in mammalian cells have suggested colibactin-induced DNA damage drives tumorigenesis[1,2,5,6]. Current data support a model in which inactive precursors termed precolibactins are deacetylated by the peptidase ClbP in the bacterial periplasm, liberating bioactive colibactin. Colibactin alkylates DNA with a 'double warhead' comprised of a cyclopropane ring conjugated to an α,β-unsaturated imine, creating adenine–colibactin adducts and DNA crosslinks[5,6]. It has been predicted that these DNA lesions lead to mutations that promote CRC in oncogenes or tumour suppressors. These lesions might also serve as biomarkers for colibactin exposure (FIG. 1).

In the new study, the researchers define two unique mutational signatures caused by colibactin exposure[2]. Human organoid cultures ('mini-guts' grown from healthy human colonic crypt stem cells) were repeatedly exposed to a human CRC-derived *pks*$^+$ *E. coli* strain and passaged for 5 months. Control cultures were exposed to a *clbQ*-deficient *E. coli* strain that was unable to produce colibactin. Whole-genome sequencing (WGS) of the organoids revealed increased single-base substitutions (SBS) in colibactin-exposed cultures compared with the control organoids. SBS were predominantly T>N substitutions, including ATA, ATT and TTT (with the middle base mutated), with a preferred presence of an adenine 3 bp upstream. This signature is termed SBS-*pks* and it is consistent with the previously observed adenine–colibactin adducts and DNA crosslinks[5,6]. A preprint manuscript describes similar AT-rich pentameric and hexameric sequence motifs in Caco-2 cells (a human colorectal adenocarcinoma cell line) exposed to *pks*$^+$ *E. coli*[7]. In the new study, WGS revealed an additional *pks* signature that contained single T deletions at T homopolymers, termed ID-*pks* to indicate insertions or deletion. As with SBS-*pks*, the ID-*pks* signature showed an enrichment of adenines immediately upstream of the indel site. SBS-*pks* exhibited transcriptional strand bias, which the authors propose might result from transcription-coupled nucleotide excision repair elicited to correct alkylated adenosines[2].

After having discovered the SBS-*pks* and ID-*pks* signatures in vitro, the researchers examined WGS datasets from the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium[2]. Using an unbiased approach to mine WGS data from 3,668 human solid cancer metastases (496 from CRC), the researchers detected a significant enrichment in SBS-*pks* and ID-*pks* signatures in CRC metastases compared with other cancer types. In fact, SBS-*pks* and ID-*pks* signatures were positively correlated with each other, with their dual presence suggesting that they have a common origin. Then, the SBS-*pks* and ID-*pks* signatures were validated in an independent cohort of the Genomics England 100,000 Genomes Project, revealing the presence of SBS-*pks* and ID-*pks* signatures in 5.0% and 4.4% of 2,208 human CRC tumours, respectively. To link *pks* mutational signatures with oncogenic mutations, the researchers examined seven cohorts of patients with CRC and found *pks* signatures in 2.4% of 4,712 CRC driver mutations. *APC*, the most commonly

mutated gene in CRC, harboured the highest amount of mutations (5.3%) containing the SBS-*pks* or ID-*pks* mutational signatures[2]. This finding clearly links the mutational signature of colibactin exposure to known CRC driver mutations.

Although virus-associated cancers are often marked by integration of viral oncogenes or disruption of tumour suppressor genes — prominent genomic features that link viral exposure to subsequent cancer development — genomic markers of bacterial genotoxin exposure are relatively unknown. This new study reveals a unique signature characterized by SBS and deletions caused by colibactin exposure and localized to CRC driver mutations. These findings are a substantial leap forward in defining mutagenic signatures from bacterially derived genotoxins of the gut microbiota. Another study examined multiple cancer types and identified 13 previously unknown SBS, many with unknown aetiologies[8]. This finding raises the possibility that other unknown genotoxins from the native microbiota can promote cancers at microbiota-rich niches, including skin, oral cavity and genital tract. Indeed, the new study also found SBS-*pks* and ID-*pks* signatures in some urinary tract-derived tumours[2]. Uropathogenic *E. coli* commonly harbour *pks*; thus, it is intriguing to speculate whether recurrent bladder infection caused by *pks*$^+$ *E. coli* might contribute to urinary cancer risk. Observations from another paper suggest that bacterial genotoxins such as colibactin might impart greater effects early in life when the gut microbiota is immature and less diverse[9]. Mutational signatures that match SBS-*pks* and ID-*pks* were detected in non-neoplastic colonic crypts from 29 of 42 healthy individuals, and data modelling revealed that these signatures were acquired before 10 years of age. Thus, it is possible that individuals who harbour a prominent *pks* mutational signature early in life might be at greater risk of developing CRC.

In conclusion, the *pks* mutational signature identified by Pleguezuelos-Manzano et al. directly links colibactin exposure to two unique mutational fingerprints whose presence might indicate greater susceptibility to developing CRC. Because *pks* mutational signatures can be detected in non-neoplastic tissue, it is likely that additional genomic insult and/or DNA repair machinery failure are required to turn a healthy colibactin-exposed cell neoplastic. Although more investigation is needed to discern the predictive power of these *pks* mutational signatures, together, these findings highlight the pervasive role of the microbiota in genetic changes that might promote neoplastic events.

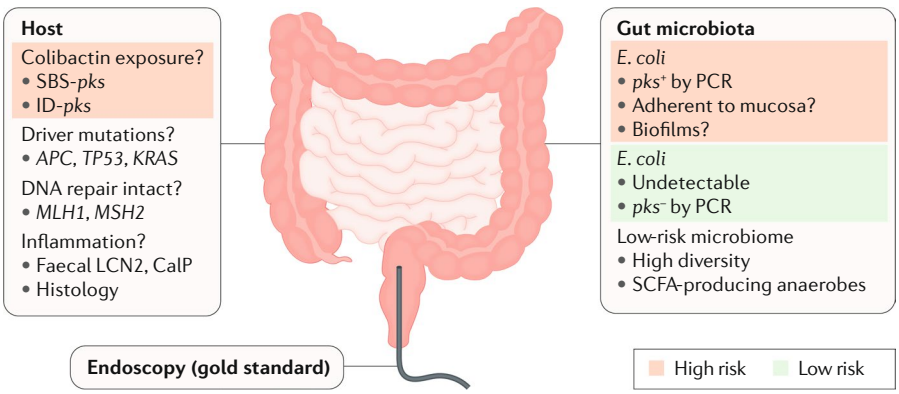

Fig. 1 | **Mutational signatures might contribute to colorectal cancer prognosis.** The figure depicts how new knowledge of *pks*$^+$ *Escherichia coli* might influence future colorectal cancer (CRC) screening. The *E. coli*-derived genotoxin colibactin has been linked to the microbiota of CRC tumours and promotes CRC in mouse models. Its mutational spectra have recently been defined in elegant in vitro models and ex vivo analyses of whole genomes of tumours. *pks* is the 'polyketide synthase' island, a biosynthetic gene cluster that is harboured by some *E. coli* strains and is responsible for production of the genotoxin colibactin. SBS-*pks* is the newly identified single-base substitution signature induced by *pks*$^+$ *E. coli*. ID-*pks* is the newly identified insertion/deletion signature induced by *pks*$^+$ *E. coli*. Pleguezuelos-Manzano et al.[2] found a significant enrichment of these mutational signatures in metastases from CRC versus other cancers, linking them to the microbiota. CalP, calprotectin; LCN2, lipocalin-2; SCFA, short-chain fatty acid.

Janelle C. Arthur [1,2,3]

[1]*Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.*

[2]*Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.*

[3]*Center for Gastrointestinal Biology and Disease, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.*

e-mail: janelle_arthur@med.unc.edu

https://doi.org/10.1038/s41575-020-0303-y

1. Arthur, J. C. et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120–123 (2012).
2. Pleguezuelos-Manzano, C. et al. Mutational signature in colorectal cancer caused by genotoxic *pks*$^+$ *E. coli*. *Nature* https://doi.org/10.1038/s41586-020-2080-8 (2020).
3. Nougayrede, J. P. et al. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848–851 (2006).
4. Healy, A. R. & Herzon, S. B. Molecular basis of gut microbiome-associated colorectal cancer: a synthetic perspective. *J. Am. Chem. Soc.* **139**, 14817–14824 (2017).
5. Wilson, M. R. et al. The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**, eaar7785 (2019).
6. Xue, M. et al. Structure elucidation of colibactin and its DNA cross-links. *Science* **365**, eaax2685 (2019).
7. Dziubańska-Kusibab, P. J. et al. Colibactin DNA damage signature indicates causative role in colorectal cancer. *bioRxiv* **1**, 819854 (2019).
8. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
9. Lee-Six, H. et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 (2019).

Competing interests
The author declares no competing interests.