The implementation of omics technologies in cancer microbiome research

Benjamin H Mullish[1], Laura S Osborne[2], Julian R Marchesi[1,2] and Julie AK McDonald[1]

[1]Division of Integrative Systems Medicine and Digestive Disease, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London SW7 2AZ, UK
[2]Microbiomes, Microbes and Informatics Research Department, School of Biosciences, Cardiff University, Cardiff CF10 3AT, UK

Correspondence to: Julie AK McDonald. Email: julie.mcdonald@imperial.ac.uk

Abstract

Whilst the interplay between host genetics and the environment plays a pivotal role in the aetiopathogenesis of cancer, there are other key contributors of importance as well. One such factor of central and growing interest is the contribution of the microbiota to cancer. Even though the field is only a few years old, investigation of the 'cancer microbiome' has already led to major advances in knowledge of the basic biology of cancer risk and progression, opened novel avenues for biomarkers and diagnostics, and given a better understanding of mechanisms underlying response to therapy. Recent developments in microbial DNA sequencing techniques (and the bioinformatics required for analysis of these datasets) have allowed much more in-depth profiling of the structure of microbial communities than was previously possible. However, for more complete assessment of the functional implications of microbial changes, there is a growing recognition of the importance of the integration of microbial profiling with other omics modalities, with metabonomics (metabolite profiling) and proteomics (protein profiling) both gaining particular recent attention. In this review, we give an overview of some of the key scientific techniques being used to unravel the role of the cancer microbiome. We have aimed to highlight practical aspects related to sample collection and preparation, choice of the modality of analysis, and examples of where different omics technologies have been complementary to each other to highlight the significance of the cancer microbiome.

Keywords: *microbiome, metataxonomics, metagenomics, metabonomics, metaproteomics*

Review

Introduction

The past few years have been associated with a surge of interest in the potential contribution of the microbiota to a number of human diseases, including a wide range of cancers [1]. The streamlining of protocols for the optimal collection and processing of samples from mucosal sites—coupled with the rapid refinement of culture-independent techniques for profiling the microbiota from these samples—has meant that microbial profiling has become an area of focus for clinicians and scientists engaged in cancer research. One central recurring question in cancer microbiome research is regarding the mechanisms of interaction between the microbiota and the host, and the role of specific metabolites and proteins appears to be key [1]. As such, in addition to microbial profiling, the use of other omics technologies—in particular, metabonomics and proteomics—contributes highly to understanding the functional implications of microbiome changes. In this review, we will give a summary of some popular omics technologies available to researchers as they apply to cancer microbiome research (metataxonomics, metagenomics, metabonomics and metaproteomics).

Study design and sample types

There are a variety of variables that have been shown to influence the composition and/or functionality of the gut microbiota, including (but not limited to) diet, medications (especially antibiotics), age, surgery, stress, chemotherapeutic drugs, probiotics, body mass index, pregnancy and microbial infections [2–4]. Detailed clinical data must be recorded for each study participant to identify these confounding variables, and these variables must be considered when matching participants in case–control studies. Due to the large impact that antibiotics have on the gut microbiota, many studies exclude participants which have used antibiotics within the previous 3–6 months. Moreover, the microbiota present in both faecal and biopsy samples is affected by bowel preparation procedures performed prior to colonoscopies. Shobar *et al* [5] found short-term changes in the composition of the microbiota following bowel preparation; however, changes affected diversity metrics differently in mucosal and faecal samples, and in healthy and inflammatory bowel disease (IBD) samples.

The most common sample type used in human gut microbiome studies is faecal samples. Faecal samples are noninvasive and can be collected by participants at home using commercial kits. Although there are significant differences in the composition of faecal samples and mucosal biopsy samples [6], faecal samples are widely accepted as a more practical alternative to biopsy samples for diagnostic purposes. Indeed, studies have shown differences in the composition of the faecal microbiota of colorectal cancer patients and healthy participants, with studies consistently reporting enrichment of *Peptostreptococcus stomatis*, *Parvimonas micra*, *Porphyromonas* spp. and *Fusobacterium nucleatum* in the faecal microbiota of colorectal cancer patients [6–10].

The gut mucosal microbiota are thought to be particularly important to host health due to the proximity of microbial cells to the host epithelium. Therefore, biopsy samples are thought to provide more meaningful mechanistic data to studies investigating host–microbe interactions compared to faecal samples. Mucosal microbiota samples are often taken through pinch biopsy during colonoscopy or during bowel surgery, and therefore characterisation of the mucosal microbiota could also be useful as a diagnostic tool [11]. Comparison of biopsies from colorectal cancer patients and healthy controls can be difficult, as it is not ethical to ask healthy volunteers to undergo a colonoscopy to collect healthy mucosal biopsy samples. However, some studies recruit healthy volunteers undergoing colonoscopies for colon cancer screening [12]. Other options for sampling the mucosal microbiota include luminal brushing (superficial sampling of tissues) [13, 14], submucosal sampling (deeper sampling of tissues) [15] or colonic lavage samples [12]. Previous studies have explored changes in the mucosal microbiota in patients with colorectal cancer. Nakatsu *et al* [16] found that *Fusobacterium*, *Parvimonas*, *Gemella* and *Leptotrichia* were most significantly enriched in early-stage colorectal cancer. In another study, Kinross *et al* [17] demonstrated that *Fusobacteria* and ε-*Proteobacteria* were enriched on tumour tissue compared to adjacent normal mucosal tissue, and the relative abundance of *Fusobacteria* and β-*Proteobacteria* increased with advancing cancer stage.

DNA extraction

We recommend extracting DNA from samples within 3 months of storage at −80°C to avoid potential changes in the composition of the stool sample over time [18]. It is important to include a mechanical lysis step (bead beating) in the DNA extraction protocol to ensure proper cell lysis of Gram-positive bacteria and archaea [19, 20]. Bead beating is performed by mixing faecal samples with 0.1-mm glass or silica beads and buffer and homogenising with a bead beating instrument. There are many options for the DNA extraction protocol used; however,

researchers often choose to use commercial kits. Following both the mechanical and chemical lysis using these kits, the DNA is purified by mixing the cell lysate with a series of buffers, binding of the DNA to an immobilised matrix (e.g. a column, magnetic beads), washing the bound DNA, and eluting the purified DNA.

Unfortunately, the DNA extraction protocol used for gut microbiota studies is not standardised, and there are a wide variety of DNA extraction kits used in published studies [21]. It is important to use the same DNA extraction protocol for all samples to be compared for DNA sequencing studies, as differences in the DNA extraction kits results in differences in the DNA yield and profile of the microbial community [22, 23]. However, regardless of the DNA extraction protocol chosen for a study, the microbial DNA extracted will still be subject to some degree of bias, e.g. some bacterial cells are easier to lyse than others.

It can be challenging to obtain adequate DNA concentrations from small biopsy samples. In metataxonomic studies, sequencing of low biomass samples results in sequencing data containing contaminant bacterial DNA sequences originating from DNA extraction kit reagents [24, 25]. We advise researchers sequencing low biomass samples to also sequence several DNA kit controls where the DNA extraction protocol is carried out in the absence of a sample. In metagenomic studies, low bacterial biomass biopsy samples result in the undesired sequencing of human DNA. Therefore, we do not advise researchers to perform metagenomic analysis on these samples as it is very difficult to obtain adequate sequencing coverage of bacterial DNA, unless the researcher is prepared to pay for the sequencing data.

Metataxonomic sequencing (16S rRNA gene sequencing)

Metataxonomic sequencing is a technique where researchers determine the bacterial composition of a sample by polymerase chain reaction (PCR) amplifying and sequencing one or more variable regions of the 16S rRNA gene present in bacterial genomes from the extracted DNA sample. The 16S rRNA gene is present in all prokaryotes, and consists of nine hypervariable regions of differing sequences and length, interspaced by highly conserved regions [26, 27]. Primers are designed to bind to conserved regions of the gene that flank one or more of the hypervariable regions [28]. Because the DNA sequences of these hypervariable regions are phylogenetically distinct for a given species, sequencing of these regions allows researchers to determine the bacterial composition of each sample. Metataxonomic sequencing can reliably classify bacterial taxa down to the genus level in human samples (and in some taxonomic cases, species level), and down to the family level in murine samples [29].

Following amplification and purification of one or more variable regions of 16S rRNA gene, the PCR product for each sample is used as a template in a second PCR reaction that adds a unique combination of barcoded indices to each sample [30]. These barcoded indices allow many samples to be pooled and sequenced in parallel on a single flow cell (up to 384 samples for one sequencing run). The amplicons are purified, mixed together at equimolar concentrations, denatured and sequenced. Illumina and ThermoFisher produce several sequencing instruments, with the Illumina MiSeq and the ThermoFisher Ion Personal Genome Machine being the most popular for metataxonomic sequencing. These instruments are capable of sequencing 300–400 bp paired-end reads [31, 32].

A disadvantage of metataxonomic sequencing is that different bacteria contain different copy numbers of the 16S rRNA gene (2–15 copies per bacterial genome) [33], which can bias the composition of the sample by making some bacteria look more abundant than they actually are in the sample. The composition of the sample also depends on how well the universal primers can amplify a bacterial species, and therefore this method is susceptible to PCR bias [34]. Moreover, the choice of hypervariable regions, primer sequences and PCR conditions are not standardised, and there can be significant differences in the sequencing results obtained from the same sample [34, 35]. Therefore, researchers are advised to use the same primers and PCR conditions for all samples from the same study and are cautioned when comparing data between studies where samples were prepared following different protocols.

Shotgun metagenomic sequencing

Shotgun metagenomic sequencing is a technique where researchers fragment and randomly sequence DNA from the collection of genomes and genes present in the extracted DNA sample. In addition to gaining taxonomic data from a sample, characterisation of the metagenome allows researchers to gain additional information including the functional potential of the microbiota. This is because shotgun metagenomics sequences most of the genes present in the sample, in contrast to metataxonomic sequencing which only targets part of the 16S rRNA gene. Moreover, shotgun metagenomic sequencing can more reliably classify bacteria down to species/strain level compared to metataxonomic sequencing [36].

Shotgun metagenomic sequencing also requires a larger number of sequencing reads to provide adequate sequencing coverage of all the genes in the sample. Therefore, researchers often use the Illumina HiSeq to perform these sequencing runs. Following sequencing, the DNA sequences of these fragments are assembled or mapped to a reference database and annotated, or a database-independent approach may be used [37].

However, there are several challenges associated with shotgun metagenomic sequencing. Annotation of genes from data sets is dependent on the accuracy and completion of reference databases and genomes. Therefore, metagenomic data sets can have a large proportion of reads corresponding to genes with unknown functions, and in some cases, this can be as high as 60% of the total reads [38]. In contrast to metataxonomic sequencing (which PCR amplifies a region of the 16S rRNA gene), shotgun metagenomic sequencing shears the genomic DNA in the sample, and therefore is not subject to PCR bias at this step. However, shotgun metagenomic sequencing is still subject to biases due to the DNA extraction step, index PCR step, sequencing errors and so on. Moreover, while shotgun metagenomic sequencing can provide users with more information than metataxonomic sequencing, the sequencing process itself is significantly more expensive (shotgun metagenomic sequencing is generally 20–30 times more expensive than metataxonomic sequencing) [35]. It is also challenging to process, store and analyse the large data sets resulting from shotgun metagenomic sequencing.

Although it is unarguable that metataxonomic and shotgun metagenomic sequencing have helped to achieve significant advances in the field of cancer regarding diagnosis, prognosis and treatment, there are intrinsic limitations associated with these techniques that do not allow us to perceive the complete picture [39]. Shotgun metagenomic sequencing helps us to identify genes that might be involved in cancer, but it is not clear if these genes are being actively expressed. Another method to assess whether a given microbial gene is being actively expressed is to use metatranscriptomics, where complementary DNA (cDNA) is synthesised from transcribed microbial genes (mRNA) and sequenced. However, this technique is more challenging, complex and expensive, especially for intestinal biopsy samples where microbial mRNA is present at much lower levels compared to mRNA levels from host tissue, and with respect to the significant abundance of cDNA derived from rRNA, which can be the major component of an analysis if it is not removed. To understand gut microbiota function and interactions in their entirety in a given pathological state other techniques are required, such as metabonomics and metaproteomics.

Metabonomics

Metabonomics may be defined as 'the quantitative measurement of the dynamic multi-parametric metabolic response of multicellular systems to pathophysiological stimuli' [40]. More specifically, it enables the detection, identification and quantification of metabolites responsible for mediating the phenotypic expression of altered metabolism resulting from a biological challenge. Metabolites that are identified may represent products of host metabolism, microbial metabolism or co-metabolism between microbiota and host. Metabonomics, therefore, provides a valuable tool in elucidating potential mechanistic links between alterations in microbial composition and changes to host physiology, including the development of the disease. Metabonomics is conducted by analysing biofluids using powerful spectroscopic techniques—in particular, nuclear magnetic resonance (NMR) spectroscopy and/or mass spectrometry (MS)—with the subsequent use of advanced multivariate statistical tools to interrogate the spectral data produced.

The metabolite profile obtained from biofluids is highly sensitive to factors that may vary in sample collection, including the length of time between the collection of sample and preparation for spectroscopy, or the temperature at which the sample has been collected and stored. As such, standardised protocols for sample collection and preparation for metabolite analysis have been designed and optimised for a range of biofluids, including plasma, serum, urine, tissue extracts [41], stool/faecal supernatant [42] and breath [43]. It is even possible to metabolically analyse intact tissue using a technique called 'magic angle spinning' [44]. A wide range of different forms of spectroscopic analysis are available (targeted versus untargeted assays, negative versus positive mode and so on; Table 1), with the most suitable assays dependent upon the metabolites of interest, biofluids that are available, and cost amongst other factors. In many studies comparing biofluids from those with cancer to those of matched healthy participants, one strategy taken has been to perform a global metabolite screen with a technique such as NMR first, identify possible metabolite groups of interests that differentiate samples, and then to perform a targeted technique (often MS) to explore these metabolites in greater detail [45]. The raw spectral data that is generated is complex and multi-stage analysis under expert supervision is required, including the separation of signal from noise, alignment of peaks, normalisation of data and the use of both unsupervised and supervised statistical techniques to allow full interpretation [46]. Other complexities lie in understanding how the metabonomic data obtained link to the biology of the system under investigation—for example, do metabolites that change after an intervention reflect modifications in the host, microbial modifications or alterations in host-microbial crosstalk?

Table 1. Summary of common spectroscopic assays that may be of use to cancer microbiome research.

Technique	Specific form	Notes
NMR		• Most of use for broad profiling of a range of metabolites • Higher reproducibility than MS.
MS	Gas chromatography	• This is particularly useful for the identification and quantification of volatile metabolites such as short-chain fatty acids

	Liquid chromatography	• Can be used in 'profiling' (i.e. detection, identification and relative quantification of a particular group of metabolites, e.g. bile acids, lipids, etc), or as a 'targeted' technique (i.e. specific identification and accurate quantification of each member of a group of metabolites, e.g. individual amino acids)
		• The instrument can be used in positive or negative ionisation mode, depending upon the specific metabolites of interest
		• May be performed with chromatography step performed in a number of different ways, including
		  – Hydrophilic interaction chromatography: Useful for identifying small polar molecules
		  – Reverse phase chromatography: Detects a broad range of metabolites, ranging from moderately polar to moderately apolar molecules

	Desorption electrospray ionisation (DESI)	• A form of ambient MS, where intact samples, in their native environment, are directly ionised. The instrument used to perform ionisation is directly linked to a mass spectrometer
	Rapid evaporative ionisation MS (REIMS)	• A further form of ambulatory MS. A radiofrequency electrical current is applied to a sample via a stainless steel monopolar probe. The vapour generated contains gas phase ions of metabolites and structural lipids. The vapour is channelled to a mass spectrometer attached to the instrument using an incorporated vacuum system

Review

There have been a variety of applications of metabonomics as a means to better understand the contribution of the microbiome to disease pathogenesis or outcomes in cancer. The gradual emergence of studies co-reporting microbiome and metabolite profiles on the same sample set is the first step towards progress in understanding this relationship [17]; further studies applying this same approach—but using longitudinal sampling, and/or samples across a range of tumour progression (from pre-malignant through to metastatic malignant states)—promise further elucidation of relevant mechanisms.

One future avenue of interest is likely to be an expansion of 'pharmacomicrobiomics' [47], i.e. using metabonomics as a contributory tool to better understand the interactions between the metabolic and immune functions that the microbiota exert on a host with or at risk of cancer, and the drugs used to treat it. If successful, this could allow a new paradigm for all levels of cancer care, where manipulation of microbiota metabolic functionality could be applied to slow progression of cancer, optimise selection of or response to systemic treatment and so on [47].

Exoproteomics and oncoproteomics

Exoproteomics is a technique that studies the protein content found in the extracellular proximity of a given biological system, e.g. the luminal content of the gut. The emerging use of exoproteomics in a clinical cancer setting (or 'oncoproteomics') enables the identification of specific proteins, abundances, structures, interactions and post-translational modifications in a given biological state. There is a growing body of evidence demonstrating the utility of proteomics in cancer research, including the identification of biomarkers for early diagnosis. One key example includes the first Food and Drug Administration approved protein biomarker, OVA1, discovered following proteome profiling from sera samples, and now used for the detection of early-stage ovarian cancer [48].

Oncoproteomic methodologies used to discover such biomarkers are becoming increasingly powerful and robust, including the bioinformatics pipelines required for data interpretation. The major methods involved include gel-based and gel-free analysis, MS and microarray-based methods [49]. While detailed descriptions of the methodologies are beyond the scope of this article, they have been described in detail elsewhere [49–53]. The general workflow of a proteomics pipeline involves several steps, one of the most important of which is sample preparation. Sample preparation involves the extraction and separation of protein usually from tissue biopsies, serum and/or plasma. Sample collection, storage and processing should be conducted with consistency and under stringent standard operating

procedures [49]. Separation of proteins can be achieved via gel-based analysis, namely, 2D gel electrophoresis. Gel-free methods coupled with MS—such as high-performance liquid chromatography, isotope-coded affinity tags and stable isotope labelling by amino acids in cell culture (SILAC) [39]—may also be used. Protein identification techniques using MS have become the 'gold standard' analytical methodologies in oncoproteomic pipelines. Following MS, the entire proteome of a particular sample can be assayed via computational analysis to give a library of peptides with protein type, abundance, functions and structure. Matrix-assisted laser desorption/ionisation time-of-flight MS (described in detail by Singhal *et al* [54]) has emerged as an efficient and accurate means for routine proteomics. The technique has assisted in the analysis and validation of several biomarkers, including those related to the identification and diagnosis of infectious diseases, as well as those related to cancer [55].

However, it is important to note that the application of proteomics to cancer microbiome research is certainly less advanced than that of DNA-based analyses. One of the main reasons for this is the complexity of the proteome. The host proteome alone consists of an estimated one to several million proteins encoded by around 20,000 genes, with differing biochemical and physical properties, often interacting as a network rather than on their own [56], and these estimations are even higher when extending the proteome to include the microbiome. Projects such as the Human Proteome project (https://hupo.org/human-proteome-project) have aimed to define the human proteome, and are making progress via a concerted effort to coordinate different research laboratories regarding sharing and exchanging the generated data and methodologies used. However, discrepancies in proteome profiles generated using varied methodologies in different research laboratories is still a major problem [57]. Furthermore, MS techniques may also introduce significant bias to a complex biological sample due to a lack of sensitivity for detecting low abundance proteins. In addition, at present, the process is also significantly lower throughput than DNA sequencing platforms, and the technology too complex for routine use in a clinical setting.

High-throughput microarray-based technologies as a complement to MS represent a means by which cancer microbiome proteomics may advance. The most developed version of these, reverse phase protein microarrays, immobilise a number of different cellular lysates or biological specimens onto the microarray and screen using a range of antibodies for the expression of proteins of interest, all within a single chip (for a detailed description of the methodologies, see Sutandy *et al* [58]).

One of the key ways that proteomic technologies can aid in cancer microbiome research is by studying the effects of the gut microbiota on the pharmacodynamics of cancer therapy drugs. The gut microbiota have been shown to be central to the bioavailability of certain anticancer drugs and the efficacy of certain treatments, particularly immunotherapy [59, 60]. In a study by Stringer *et al* [61], a combination of traditional proteomics and metatransciptomics was utilised to characterise dysbiosis of the gut microbiota during chemotherapy. They found that chemotherapy-induced diarrhoea could be attributed to alterations in the gut microbiota, resulting in increased levels of several proteins which could act as toxicity biomarkers for patients at risk of developing chemotherapy-induced diarrhoea, including faecal calprotectin, circulating matrix metalloproteases, NF-κβ, IL-1β and TNF.

Proteomic techniques can also provide valuable information for cancer research *in vitro*. Kim *et al* [62] implemented MS proteome profiling of *Escherichia coli* and *Staphylococcus aureus* culture supernatants to screen for outer membrane vesicles (OMVs). OMVs are extracellular vesicles often produced by Gram-negative bacteria with lipid-bilayers containing numerous immunostimulatory factors. They demonstrated that trypsin-sensitive surface proteins of these extracellular vesicles were capable of inducing anticancer cytokines (interferon-γ and CXCL10) in a mouse model of colon adenocarcinoma, resulting in a significant reduction in tumour burden [63].

Bacterial oncoproteomics holds the potential to not only accelerate diagnosis and increase treatment efficacy but also to identify mechanisms of cancer prevention by disrupting certain key microbial pathways [64]. For oncoproteomics to reach its full potential in a clinical setting, however, there is a global need for the standardisation of sample collection, preparation and bioinformatic processing.

Example of the integrated application of omic technologies to cancer microbiome research

A relevant application of omics technologies to cancer research is the study of the microbial contribution to the toxicity of cancer chemotherapeutics. For example, Lin *et al* [65] explored the mechanisms by which CPT-11 (also known as irinotecan, a chemotherapeutic used to treat colorectal malignancy) has a common side effect of chronic diarrhoea. Initial studies showed CPT-11 increased the relative abundances of *Clostridium* cluster XI and *Enterobacteriaceae* in the caecal content of tumour-bearing rats. Studies also showed that chronic

diarrhoea can be minimised after co-administration of dietary fibre with CPT-11 [66]. However, the reduction in CPT-11 toxicity with dietary fibre did not correlate with stimulation of specific bacterial taxa, although there was a close correlation with caecal concentrations of the short chain fatty acid butyrate, and the bacterial butyryl CoA gene [66]. Proteomic studies have also shown that β-glucuronidases produced by the gut microbiota are responsible for reactivating CPT-11, which plays a role in diarrhoea and can prevent dose intensification and effi-cacy [67]. Such toxic side effects can be ameliorated with the use of inhibitors targeting these bacterial β-glucuronidases. These and other similar studies have emphasised that rather than just the presence or absence of specific bacterial taxa as a predictor of chemotherapeutic toxicity, the metabolic functionality of the microbiota may be more important.

Conclusion

Whilst there have been several interesting studies investigating the role of the gut microbiota in cancer pathogenesis, there are still many questions that need to be answered. For example, while the application of omic technologies can show differences in the mucosal micro-biota of colorectal cancer patients, it is not clear whether these bacteria are the cause of cancer initiation, or whether a developing tumour creates an environmental niche that favours the growth of opportunistic bacteria. Omic technologies are a powerful tool to identify bacteria, metabolites or proteins that correlate with cancer presence or cancer stage progression. However, carefully controlled animal studies or longitudinal studies in humans are required to confirm the role of these identified targets in cancer pathogenesis.

Acknowledgments

Authors' contributions

BHM, LSO and JAKM wrote the initial draft of the manuscript and received editorial input from JRM. All authors read and edited the final version of the manuscript.

Conflicts of interest

Nil for all authors.

References

1. Garrett WS (2015) **Cancer and the microbiota** *Science* **348** 80–86 https://doi.org/10.1126/science.aaa4972 PMID: 25838377 PMCID: 5535753

2. DuPont AW and DuPont HL (2011) **The intestinal microbiota and chronic disorders of the gut**. *Nat Rev Gastroenterol Hepatol* **8** 523–531 https://doi.org/10.1038/nrgastro.2011.133 PMID: 21844910

3. Costello EK, Stagaman K, and Dethlefsen L, *et al* (2012) **The application of ecological theory toward an understanding of the human microbiome** *Science* **336** 1255–1262 https://doi.org/10.1126/science.1224203 PMID: 22674335 PMCID: 4208626

4. Dave M, Higgins PD, and Middha S, *et al* (2012) **The human gut microbiome: current knowledge, challenges, and future direc-tions** *Transl Res* **160** 246–257 https://doi.org/10.1016/j.trsl.2012.05.003 PMID: 22683238

5. Shobar RM, Velineni S, and Keshavarzian A, *et al* (2016) **The effects of bowel preparation on microbiota-related metrics differ in health and in inflammatory bowel disease and for the mucosal and luminal microbiota compartments** *Clin Transl Gastroenterol* **7** e143 https://doi.org/10.1038/ctg.2015.54 PMID: 26866392 PMCID: 4817412

6. Flemer B, Lynch DB, and Brown JM, *et al* (2017) **Tumour-associated and non-tumour-associated microbiota in colorectal cancer** *Gut* **66** 633–643 https://doi.org/10.1136/gutjnl-2015-309595 PMCID: 5529966

7. Zeller G, Tap J, and Voigt AY, *et al* (2014) **Potential of fecal microbiota for early-stage detection of colorectal cancer** *Mol Syst Biol* **10** 766 https://doi.org/10.15252/msb.20145645 PMID: 25432777 PMCID: 4299606

8. Yu J, Feng Q, and Wong SH, *et al* (2017) **Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer** *Gut* **66** 70–78 https://doi.org/10.1136/gutjnl-2015-309800

9. Baxter NT, Ruffin MT, and Rogers MAM, *et al* (2016) **Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions** *Genome Med* **8** 37 https://doi.org/10.1186/s13073-016-0290-3 PMID: 27056827 PMCID: 4823848

10. Zackular JP, Rogers MAM, and Ruffin MT, *et al* (2014) **The human gut microbiome as a screening tool for colorectal cancer** *Cancer Prev Res (Phila)* **7** 1112–1121 https://doi.org/10.1158/1940-6207.CAPR-14-0129

11. Lu Y, Chen J, and Zheng J, *et al* (2016) **Mucosal adherent bacterial dysbiosis in patients with colorectal adenomas** *Sci Rep* **6** 26337 https://doi.org/10.1038/srep26337 PMID: 27194068 PMCID: 4872055

12. Watt E, Gemmell MR, and Berry S, *et al* (2016) **Extending colonic mucosal microbiome analysis-assessment of colonic lavage as a proxy for endoscopic colonic biopsies** *Microbiome* **4** 61 https://doi.org/10.1186/s40168-016-0207-9 PMID: 27884202 PMCID: 5123352

13. Lavelle A, Lennon G, and O'Sullivan O, *et al* (2015) **Spatial variation of the colonic microbiota in patients with ulcerative colitis and control volunteers** *Gut* **64** 1553–1561 https://doi.org/10.1136/gutjnl-2014-307873 PMID: 25596182 PMCID: 4602252

14. Huse SM, Young VB, and Morrison HG, *et al* (2014) **Comparison of brush and biopsy sampling methods of the ileal pouch for assessment of mucosa-associated microbiota of human subjects** *Microbiome* **2** 5 https://doi.org/10.1186/2049-2618-2-5 PMID: 24529162 PMCID: 3931571

15. Chiodini RJ, Dowd SE, and Chamberlin W, *et al* (2015) **Microbial population differentials between mucosal and submucosal intestinal tissues in advanced Crohn's disease of the ileum** *PLoS One* **10** e0134382 https://doi.org/10.1371/journal.pone.0134382

16. Nakatsu G, Li X, and Zhou H, *et al* (2015) **Gut mucosal microbiome across stages of colorectal carcinogenesis** *Nat Commun* **6** 8727 https://doi.org/10.1038/ncomms9727 PMID: 26515465 PMCID: 4640069

17. Kinross J, Mirnezami R, and Alexander J, *et al* (2017) **A prospective analysis of mucosal microbiome-metabonome interactions in colorectal cancer using a combined MAS 1HNMR and metataxonomic strategy** *Sci Rep* **7** 8979 https://doi.org/10.1038/s41598-017-08150-3 PMID: 28827587 PMCID: 5566496

18. Bahl MI, Bergström A, and Licht TR (2012) **Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis** *FEMS Microbiol Lett* **329** 193–197 https://doi.org/10.1111/j.1574-6968.2012.02523.x PMID: 22325006

19. Lim MY, Song E-J, and Kim SH, *et al* (2018) **Comparison of DNA extraction methods for human gut microbial community profiling** *Syst Appl Microbiol* **41**(2) 151–157 [doi:10.1016/j.syapm.2017.11.008] https://doi.org/10.1016/j.syapm.2017.11.008 PMID: 29305057

20. Salonen A, Nikkilä J, and Jalanka-Tuovinen J, *et al* (2010) **Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis** *J Microbiol Methods* **81** 127–134 https://doi.org/10.1016/j.mimet.2010.02.007 PMID: 20171997

Review

21. Costea PI, Zeller G, and Sunagawa S, *et al* (2017) **Towards standards for human fecal sample processing in metagenomic studies** *Nat Biotechnol* **35** 1069–1076 PMID: 28967887

22. Wesolowska-Andersen A, Bahl MI, and Carvalho V, *et al* (2014) **Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis** *Microbiome* **2** 19 https://doi.org/10.1186/2049-2618-2-19 PMID: 24949196 PMCID: 4063427

23. Kennedy NA, Walker AW, and Berry SH, *et al* (2014) **The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing** *PLoS One* **9** e88982 https://doi.org/10.1371/journal.pone.0088982 PMID: 24586470 PMCID: 3933346

24. Salter SJ, Cox MJ, and Turek EM, *et al* (2014) **Reagent and laboratory contamination can critically impact sequence-based microbiome analyses** *BMC Biol* **12** 87 https://doi.org/10.1186/s12915-014-0087-z PMID: 25387460 PMCID: 4228153

25. Lauder AP, Roche AM, and Sherrill-Mix S, *et al* (2016) **Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota** *Microbiome* **4** 29 https://doi.org/10.1186/s40168-016-0172-3 PMID: 27338728 PMCID: 4917942

26. Baker GC, Smith JJ, and Cowan DA (2003) **Review and re-analysis of domain-specific 16S primers** *J Microbiol Methods* **55** 541–555 https://doi.org/10.1016/j.mimet.2003.08.009 PMID: 14607398

27. Wang Y and Qian P-Y (2009) **Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies** *PLoS One* **4** e7401 https://doi.org/10.1371/journal.pone.0007401 PMID: 19816594 PMCID: 2754607

28. Van de Peer Y, Chapelle S, and De Wachter R (1996) **A quantitative map of nucleotide substitution rates in bacterial rRNA** *Nucleic Acids Res* **24** 3381–3391 https://doi.org/10.1093/nar/24.17.3381 PMID: 8811093 PMCID: 146102

29. Janda JM and Abbott SL (2007) **16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls** *J Clin Microbiol* **45** 2761–2764 https://doi.org/10.1128/JCM.01228-07 PMID: 17626177 PMCID: 2045242

30. Illumina **16S Metagenomic sequencing library preparation** [https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html] Date accessed: 27/11/17

31. Illumina **MiSeq specifications** [https://emea.illumina.com/systems/sequencing-platforms/miseq/specifications.html] Date accessed: 1/5/18

32. ThermoFisher Scientific **Ion Personal Genome Machine (PGM) system—thermo fisher scientific**. [https://www.thermofisher.com/order/catalog/product/4462921] Date accessed: 1/5/18

33. Pei AY, Oberdorf WE, and Nossa CW, *et al* (2010) **Diversity of 16S rRNA genes within individual prokaryotic genomes** *Appl Environ Microbiol* **76** 3886–3897 https://doi.org/10.1128/AEM.02953-09 PMID: 20418441 PMCID: 2893482

34. Claesson MJ, Wang Q, and O'Sullivan O *et al* (2010) **Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions** *Nucleic Acids Res* **38** e200 https://doi.org/10.1093/nar/gkq873 PMID: 20880993 PMCID: 3001100

35. Clooney AG, Fouhy F, and Sleator RD, *et al* (2016) **Comparing apples and oranges?: next generation sequencing and its impact on microbiome analysis** *PLoS One* **11** e0148028 https://doi.org/10.1371/journal.pone.0148028 PMID: 26849217 PMCID: 4746063

36. Tu Q, He Z, and Zhou J (2014) **Strain/species identification in metagenomes using genome-specific markers** *Nucleic Acids Res* **42** e67 https://doi.org/10.1093/nar/gku138 PMID: 24523352 PMCID: 4005670

37. Claesson MJ, Clooney AG, and O'Toole PW (2017) **A clinician's guide to microbiome analysis** *Nat Rev Gastroenterol Hepatol* **14** 585–595 PMID: 28790452

Review

38. Prakash T and Taylor TD (2012) **Functional assignment of metagenomic data: challenges and applications** *Brief Bioinform* **13** 711–727 https://doi.org/10.1093/bib/bbs033 PMID: 22772835 PMCID: 3504928

39. Joshi S, Tiwari AK, and Mondal B, *et al* (2011) **Oncoproteomics** *Clin Chim Acta* **412** 217–226 https://doi.org/10.1016/j.cca.2010.10.002

40. Nicholson JK, Lindon JC, and Holmes E (1999) **'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data** *Xenobiotica* **29** 1181–1189 https://doi.org/10.1080/004982599238047 PMID: 10598751

41. Beckonert O, Keun HC, and Ebbels TM, *et al* (2007) **Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts** *Nat Protoc* **2** 2692–2703 https://doi.org/10.1038/nprot.2007.376 PMID: 18007604

42. Gratton J, Phetcharaburanin J, and Mullish BH, *et al* (2016) **Optimized sample handling strategy for metabolic profiling of human feces** *Anal Chem* **88**(9) 4661–4668 https://doi.org/10.1021/acs.analchem.5b04159 PMID: 27065191

43. Kumar S, Huang J, and Abbassi-Ghadi N, *et al* (2013) **Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esophago-gastric cancer** *Anal Chem* **85** 6121–6128 https://doi.org/10.1021/ac4010309 PMID: 23659180

44. Beckonert O, Coen M, and Keun HC, *et al* (2010) **High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues** *Nat Protoc* **5** 1019–1032 https://doi.org/10.1038/nprot.2010.45 PMID: 20539278

45. Jacob M, Lopata AL, and Dasouki M, *et al* (2017) **Metabolomics toward personalized medicine** *Mass Spectrom Rev* https://doi.org/10.1002/mas.21548 PMID: 29073341

46. Nicholson JK, Connelly J, and Lindon JC, *et al* (2002) **Metabonomics: a platform for studying drug toxicity and gene function** *Nat Rev Drug Discov* **1** 153–161 https://doi.org/10.1038/nrd728 PMID: 12120097

47. Alexander JL, Wilson ID, and Teare J, *et al* (2017) **Gut microbiota modulation of chemotherapy efficacy and toxicity** *Nat Rev Gastroenterol Hepatol* **14** 356–365 https://doi.org/10.1038/nrgastro.2017.20 PMID: 28270698

48. Zhang Z, Bast RC Jr, and Yu Y, *et al* (2004) **Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer** *Cancer Res* **64** 5882–5890 https://doi.org/10.1158/0008-5472.CAN-04-0746 PMID: 15313933

49. Jin P, Wang K, and Huang C, *et al* (2017) **Mining the fecal proteome: from biomarkers to personalised medicine** *Expert Rev Proteomics* **14** 445–459 https://doi.org/10.1080/14789450.2017.1314786 PMID: 28361558

50. Cho WCS and Cheng CHK (2007) **Oncoproteomics: current trends and future perspectives** *Expert Rev Proteomics* **4** 401–410 https://doi.org/10.1586/14789450.4.3.401 PMID: 17552924

51. Guo S, Zou J, and Wang G (2013) **Advances in the proteomic discovery of novel therapeutic targets in cancer** *Drug Des Devel Ther* **7** 1259 https://doi.org/10.2147/DDDT.S52216 PMID: 24187485 PMCID: 3810204

52. Bensimon A, Heck AJR, and Aebersold R (2012) **Mass spectrometry–based proteomics and network biology** *Annu Rev Biochem* **81** 379–405 https://doi.org/10.1146/annurev-biochem-072909-100424

53. Ahrens CH, Brunner E, and Qeli E, *et al* (2010) **Generating and navigating proteome maps using mass spectrometry** *Nat Rev Mol Cell Biol* **11** 789–801 https://doi.org/10.1038/nrm2973 PMID: 20944666

54. Singhal N, Kumar M, and Kanaujia PK, *et al* (2015) **MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis** *Front Microbiol* **6** 791 https://doi.org/10.3389/fmicb.2015.00791 PMID: 26300860 PMCID: 4525378

55. Frantzi M, Bhat A, and Latosinska A (2014) **Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development** *Clin Transl Med* **3** 7 https://doi.org/10.1186/2001-1326-3-7

Review

56. Betzen C, Alhamdani MS, and Lueong S, *et al* (2015) **Clinical proteomics: promises, challenges and limitations of affinity arrays** *PROTEOMICS Clin Appl* **9** 342–347 https://doi.org/10.1002/prca.201400156 PMID: 25594918 PMCID: 5024047

57. Kim MS, Pinto SM, and Getnet D, *et al* (2014) **A draft map of the human proteome** *Nature* **509** 575–581 https://doi.org/10.1038/nature13302 PMID: 24870542 PMCID: 4403737

58. Sutandy FXR, Qian J, and Chen CS, *et al* (2013) **Overview of protein microarrays** *Curr Protoc Protein Sci* **Chapter 27** Unit 27.1 PMID: 23546620 PMCID: 3680110

59. Sivan A, Corrales L, and Hubert N, *et al* (2015) **Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy** *Science* **350** 1084–1089 https://doi.org/10.1126/science.aac4255 PMID: 26541606 PMCID: 4873287

60. Vetizou M, Pitt JM, and Daillère R, *et al* (2015) **Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota** *Science* **350** 1079–1084 https://doi.org/10.1126/science.aad1329 PMID: 26541610 PMCID: 4721659

61. Stringer AM, Al-Dasooqi N, and Bowen JM, *et al* (2013) **Biomarkers of chemotherapy-induced diarrhoea: a clinical study of intestinal microbiome alterations, inflammation and circulating matrix metalloproteinases** *Support Care Cancer* **21** 1843–1852 https://doi.org/10.1007/s00520-013-1741-7 PMID: 23397098

62. Kim JH, Lee J, and Park J *et al* (2015) **Gram-negative and Gram-positive bacterial extracellular vesicles** *Semin Cell Dev Biol* **40** 97–104 https://doi.org/10.1016/j.semcdb.2015.02.006 PMID: 25704309

63. Kim OY, Park HT, and Dinh NTH, *et al* (2017) **Bacterial outer membrane vesicles suppress tumor by interferon-γ-mediated antitumor response** *Nat Commun* **8** 626 https://doi.org/10.1038/s41467-017-00729-8

64. Vogtmann E and Goedert JJ (2016) **Epidemiologic studies of the human microbiome and cancer** *Br J Cancer* **114** 237–242 https://doi.org/10.1038/bjc.2015.465 PMID: 26730578 PMCID: 4742587

65. Lin XB, Dieleman LA, and Ketabi A, *et al* (2012) **Irinotecan (CPT-11) chemotherapy alters intestinal microbiota in tumour bearing rats** *PLoS One* **7** e39764 https://doi.org/10.1371/journal.pone.0039764 PMID: 22844397 PMCID: 3406026

66. Lin XB, Farhangfar A, and Valcheva R, *et al* (2014) **The role of intestinal microbiota in development of irinotecan toxicity and in toxicity reduction through dietary fibres in rats** *PLoS One* **9** e83644 https://doi.org/10.1371/journal.pone.0083644 PMID: 24454707 PMCID: 3891650

67. Wallace BD, Wang H, and Lane KT, *et al* (2010) **Alleviating cancer drug toxicity by inhibiting a bacterial enzyme** *Science* **330** 831–835 https://doi.org/10.1126/science.1191175 PMID: 21051639 PMCID: 3110694

Review