Trends in Microbiology

CellPress

Review

Computational methods and challenges in analyzing intratumoral microbiome data

Qi Wang,[1] Zhaoqian Liu,[1] Anjun Ma,[2,3] Zihai Li,[3] Bingqiang Liu,[1,4,*] and Qin Ma[2,3,*]

The human microbiome is intimately related to cancer biology and plays a vital role in the efficacy of cancer treatments, including immunotherapy. Extraordinary evidence has revealed that several microbes influence tumor development through interaction with the host immune system, that is, immuno–oncology–microbiome (IOM). This review focuses on the intratumoral microbiome in IOM and describes the available data and computational methods for discovering biological insights of microbial profiling from host bulk, single-cell, and spatial sequencing data. Critical challenges in data analysis and integration are discussed. Specifically, the microorganisms associated with cancer and cancer treatment in the context of IOM are collected and integrated from the literature. Lastly, we provide our perspectives for future directions in IOM research.

Highlights

The gut and intratumoral microbiota significantly affect cancer development and progression by interacting with the host's immune system, that is, the immuno–oncology–microbiome (IOM).

Available microbial data in IOM studies are mined from existing host bulk sequencing and single-cell sequencing datasets to provide unprecedented opportunities for investigating IOM in various cancer types.

The development of rigorous computational methods for characterizing and elucidating IOM is urgently needed to guide researchers to unravel new biological insights and develop precision cancer therapeutics.

A reliable benchmarking system is needed to model IOM interactions, analyze IOM data, and evaluate computational predictions.

In-depth functional analyses of IOM mechanisms are necessary to develop novel therapeutic strategies targeting microbiota to improve cancer treatment outcomes.

Significance of exploring the immuno–oncology–microbiome

The human microbiome is the collection of all the microorganisms, which are non-negligible components of the human body, residing on or within human tissues and biofluids, such as the skin, oral mucosa, lung, and gastrointestinal tract [1]. Human microbes are believed to play a broad role in cancer diagnosis, pathogenesis, and treatment by interacting with the host immune system [2]. The immune-mediated interactions among immune cells, tumors, and microbes in the **tumor microenvironment** (**TME**) (see Glossary) are defined as the immuno–oncology–microbiome (IOM) [2]. In the context of the IOM, the interactions between the host microbiome and tumor mainly include two categories [2,3] (Figure 1, Key figure): (i) interactions involving gut microbes, which affect both local and distant tumor growth and survival by impacting host immune system, and (ii) interactions involving intratumoral microbes, which either reside in the TME or tumor/immune cells to influence tumor progression and antitumor immunity. As polymorphic microbes become new cancer hallmarks, research into the intricate relationship in the IOM is receiving increasing attention [4].

The gut microbes can manipulate and infiltrate the gut epithelial barrier, relocate to other tissues and organs through the blood, and influence the host immune context [2]. They exhibit broad effects on primary lymphoid organs [5,6], the TME [2,7,8], adaptive immune responses [9,10], and inflammatory responses in the intestines and other organs [11]. Existing computational analyses of gut microbes mainly investigate the associations between the gut microbial diversity/composition and cancer types/immune responses [12–14] through 16S ribosomal RNA sequencing, **metagenomic shotgun sequencing**, flow cytometry, immunohistochemistry, and cytokine assays from host stool samples. With decades of development, the analyses of gut microbes are now becoming more mature [15,16], and projects have been created for archiving gut microbiome data related to cancer, such as the Human Microbiome Project [17] and MetaHIT [18].

Conversely, intratumoral microbes have been observed in multiple cancer types, such as colorectal [19], pancreatic [20], breast [21,22], and lung cancer [23–25]. They not only reside in tumor

[1]School of Mathematics, Shandong University, Jinan, Shandong, 250100, China
[2]Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, USA
[3]Pelotonia Institute for Immuno-Oncology, The James Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA
[4]Shandong National Center for Applied Mathematics, Jinan, Shandong, 250100, China

*Correspondence:
bingqiang@sdu.edu.cn (B. Liu) and
qin.ma@osumc.edu (Q. Ma).

cells, contributing to tumor progression, metastasis, oncogenesis, and drug resistance, but also interact with immune cells in the TME to promote or inhibit immune responses and activities [2,25,26]. The role and functions of intratumoral microbes have been confirmed to be largely tumor-specific [24,27,28], yet the intratumoral microbiome's direct causal roles and underlying mechanisms remain unclear [29]. As such, attention should be focused on analyzing data directly derived from TME: (i) microbial profiling extracted from host **whole-genome sequencing** (**WGS**), **whole-exome sequencing** (**WES**), and **whole-transcriptome sequencing** (**RNA-seq**), known as 'microbial profiling derived from **bulk sequencing**', (ii) microbial profiling extracted from host **single-cell RNA sequencing** (**scRNA-seq**) data, known as 'microbial profiling derived from single-cell sequencing', and (iii) microbial profiling extracted from host **spatially resolved transcriptomics** (**SRT**) data, known as 'microbial profiling derived from spatial sequencing'. Specific challenges and limitations lead to an increasing need for computational methods development to discover novel insight into modes of action, functionalities, and causal relations of intratumor microbes to specific cancer types, as well as for the successful prediction and diagnosis of cancers. Here, we provide a timely review of computational methods for the intratumoral microbiome in IOM, alongside the challenges and future perspectives.

Analysis of intratumoral microbial profiling derived from bulk sequencing

To leverage the large databases of clinically annotated samples with bulk data, such as **The Cancer Genome Atlas** (**TCGA**), computational tools and pipelines have been developed to analyze the IOM involving intratumoral microbes [23,30–32]. The preprocessing for deriving microbial profiling from host bulk sequencing data involves four steps. (i) Identify microbial reads from WGS, WES, or RNA-seq data of tumor/normal tissues. This step aligns sequencing reads to host reference genomes using a short-read aligner Mapping and Assembly with Quality (MAQ) [33] or the ultrafast RNA-seq aligner Spliced Transcripts Alignment to a Reference (STAR) [34]. The unmapped reads are considered candidate microbial reads. (ii) Characterize microbial taxonomic profiles. Metagenomic analysis is performed for the taxonomic profiling of microbes by mapping candidate microbial reads from the previous step to all known microbial genomes using Kraken2 [35], Metagenomic Phylogenetic Analysis (MetaPhlAn2) [36], or SHallow shOtGUN profiler (SHOGUN) [37]. The remaining unmapped reads can be assembled as novel microbial genomes using MetaVelvet [38] or MEGAHIT [39]. An existing bioinformatics tool, PathSeq [40], can provide an integrative analysis of the first two steps. (iii) Remove contaminants. Due to the low microbial biomass in host sequencing samples [32], contamination in the laboratory environment and the sequencing process significantly impacts downstream analysis. The contaminants can be removed using decontam [41] or SourceTracker [42]. (iv) Normalize decontaminated microbial data. The decontaminated microbial data are renormalized, alleviating batch effects while preserving biological signals. Commonly used normalization methods include log-ratio transformation, log upper quartile, cumulative sum scaling, and variance stabilization [43,44]. Similar pipelines can also be applied for microbiome profiling from single-cell and spatial data.

Several studies have been conducted using the above methods to investigate the effect of tumor-associated microbes on host tumors. For example, Poore *et al*. analyzed microbial profiling from 4831 WGS and 13 285 RNA-seq datasets across 10 481 patients and 33 cancer types from the TCGA compendium [30]. The read counts at the genus taxonomic level of each dataset were accessible. These bulk sequencing-derived microbial profiles were used to discover tumor-type-specific microbial signatures [30]. Pan-cancer analyses showed that *Fusobacterium* was overabundant in gastrointestinal (GI) cancers compared with non-GI cancers in both primary tumor tissue and adjacent solid-tissue normal samples. In addition, Dohlman *et al*. extracted microbial reads by retrieving raw WGS and WES sequencing data of different cancer types

Glossary

Bulk sequencing: examines the sequence information of bulk samples, usually containing multiple cells.
Deep learning: a type of machine-learning algorithm that uses multiple layers to extract higher-level features from the raw input progressively.
Graph neural network (GNN): a class of neural networks for processing data best represented by graph data structures. They were popularized by their use in supervised learning on the properties of various molecules.
Metagenomic shotgun sequencing: a method in which nucleic acid from a sample is sequenced to identify and characterize microorganisms present in the sample; it is being evaluated and used with increasing frequency for clinical microbiology diagnostics.
Single-cell RNA sequencing (scRNA-seq): examines the expression profiles of individual cells with optimized next-generation sequencing technologies, providing a higher resolution of cellular diversity.
Spatially resolved transcriptomics (SRT): an overarching term for a range of methods designed for assigning cell types (identified by the mRNA readouts) to their locations in the histological sections. This method can also be used to determine the subcellular localization of mRNA molecules.
The Cancer Genome Atlas (TCGA): a landmark cancer genomics program which has molecularly characterized over 20 000 primary cancer and matched normal samples spanning 33 cancer types.
Tumor microenvironment (TME): the environment around a tumor, including the surrounding blood vessels, immune cells, fibroblasts, signaling molecules, and the extracellular matrix (ECM).
Whole-exome sequencing (WES): also known as exome sequencing, a genomic technique for sequencing all of the protein-coding regions of genes in a genome (known as the exome).
Whole-genome sequencing (WGS): also known as full genome sequencing, complete genome sequencing, or entire genome sequencing, is the process of determining the entirety, or nearly the entirety, of the DNA sequence of an organism's genome at a single time.
Whole-transcriptome sequencing (RNA-seq): a sequencing technique that uses next-generation sequencing (NGS) to reveal the presence and

from the TCGA database [32]. Combined with the matched host mRNA expression data, they investigated host genes whose transcriptional patterns highly correlated with the abundance of identified tumor-associated microbial species [32]. They found that pathways significantly and consistently enriched by these genes were related to immune system activation, such as antigen presentation and natural killer (NK) cell-mediated cytotoxicity [32]. They also showcased that microbes that are equally prevalent across cancer types and blood samples are generally contaminants [32].

The qualitative comparison of existing computational methods used to analyze microbial profiling derived from host data have been summarized in Table 1. Nevertheless, microbiome profiling and analysis from host bulk samples are still facing challenges. First, the selection of a microbial reference genome is critical, which may cause significant differences in sequence mapping as the species/strains and the version of microbial genomes included in the reference varies. Secondly, the existing pipeline is designed for microbiota mapping suitable for discovery-driven questions, for example, what are the signature microbes involved in cancer tissue? While for

quantity of RNA in a biological sample at a given moment.

Wilcoxon rank-sum test: a nonparametric test of the null hypothesis that, for randomly selected values *X* and *Y* from two populations, the probability of *X* being greater than *Y* is equal to the probability of *Y* being greater than *X*.

Key figure

Schematic overview of computational studies of immuno–oncology–microbiome (IOM)

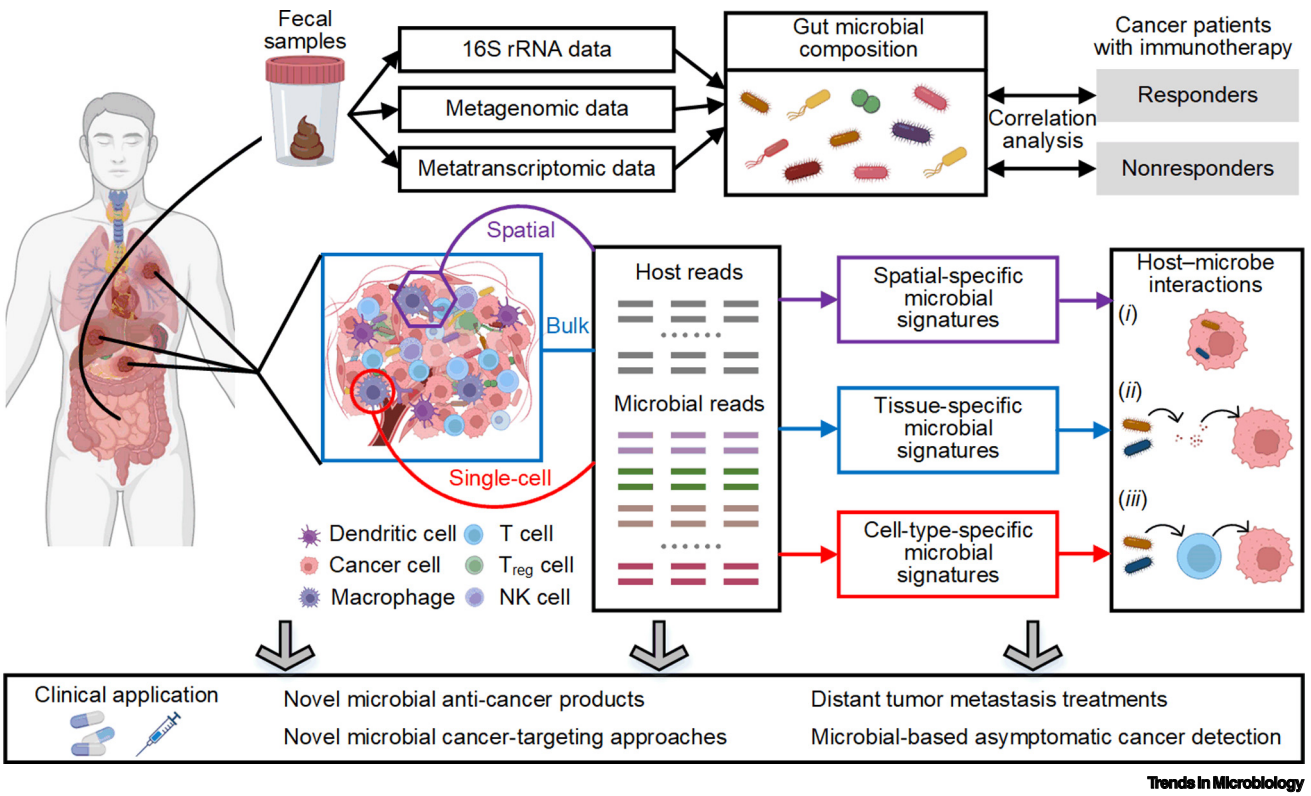

Figure 1. To analyze the effects of gut microbiota on host tumors, researchers can acquire the gut microbial composition using the 16S ribosomal RNA (rRNA) data, metagenomic data, or metatranscriptomic data from stool samples. Correlation analysis between gut microbial composition and cancer patients undergoing immunotherapy can help researchers to understand clinical response heterogeneity. For host tumor tissues, microbial profiling can be derived from existing host sequencing data, including bulk sequencing, single-cell sequencing, and spatial transcriptome data. The computational analyses of these data can enable researchers to obtain the tissue-specific, cell-type-specific, or spatial-specific microbial signatures for further IOM studies and clinical applications. Abbreviation: NK, natural killer.

Table 1. The qualitative comparison of different computational methods used to analyze microbial profiling derived from host data

Tools	Description	Strength	Limitation	Code	Refs
STAR	An RNA-seq aligner based on sequential maximum mappable seed search in suffix arrays	Fast mapping speed; high alignment precision and sensitivity; good default performance	Memory intensive	C++	[34,72,73]
MAQ	A short-read aligner based on hash table and mapping quality scores	Accurate and feature-rich short reads alignment	Unsuitable for the alignment of longer reads where indels may occur frequently; single-threaded software; slow running speed	C++	[33,74,75]
Kraken2	A taxon binning tool based on the exact alignment of *k*-mers	High accuracy at genus and species level; fast running speed	Unsuitable for running in the personal computer environment due to the RAM limitation	C++	[35,76]
MetaPhlAn2	A taxon profiling tool based on the alignment of unique taxonomic marker genes	High accuracy at genus and species level; efficient memory usage	Limited identification of microbial eukaryotes	Python	[36,76–78]
SHOGUN	A microbiome quantification framework including contaminating read filtering and relative abundance profiling	A modular, accurate, and scalable pipeline; data analysis and transformation steps can be run individually or together in an automated workflow	Computationally intensive taxonomy assignments	Python	[30,37]
MetaVelvet	A short-read *de novo* assembler for metagenomic data based on the de Bruijn graph	High sensitivity for sequence diversity	Low assembly length statistics	C++	[38,79]
MEGAHIT	An NGS *de novo* assembler for large and complex metagenomic data based on succinct de Bruijn graph	Efficient memory usage and fast running speed	Suboptimal assembly of genomes of high abundance population members on very large datasets	C++	[39,76,79,80]
decontam	Simple statistical methods to identify and remove contaminant sequences in marker-gene and metagenomic data	Easy integration with existing metagenomic sequencing workflows	Auxiliary data from DNA quantitation and negative control data are required	R	[41,81]
Source Tracker	A Bayesian approach to estimate the proportion of contaminants in a given community	Directly estimate source proportions; model the uncertainty of known and unknown source environments	Limited for discerning sources with similar bacterial communities; high running time, only applicable to datasets between small and medium in size with few sources	R	[42,82,83]
Shortread	A Bioconductor package for input, quality assessment and exploration of high-throughput sequence data	Suitable for removing low-complexity reads, low-quality reads, and PCR duplicates tagged with the same unique molecular identifier (UMI) and cellular barcode	Lack of sophisticated and flexible programming frameworks	R	[84,85]
Trimmomatic	A flexible trimmer for Illumina sequence data	A more flexible and efficient preprocessing tool, which could correctly handle paired-end data	Relatively slow and overly time-consuming	Java	[86,87]
UMI-tools	A software package modeling sequencing errors in UMIs	Easily be integrated into existing pipelines for analysis of sequencing techniques utilizing UMIs	High UMI preprocessing runtime cost	Python	[86,88]
Wilcoxon rank-sum test	A nonparametric method used to test the differences between two populations	It can be used for the comparison of a non-normally distributed, but at least ordinally scaled, parameter in two unpaired samples	The correlation obtained by statistical tests is typically a 'spurious correlation'	R/Python	[45,46,52,89,90]
Spearman correlation	A method used to test whether there is a monotonous relationship between two variables	It is preferable when variables feature heavy-tailed distributions or when outliers are present		R/Python	[46,52,90,91]

individual species/strain mapping, the pipeline needs to be optimized with a more flexible sequence alignment threshold due to the extremely short reference genome. Such adjustment may bring more false positives in quantifying the species/strain's abundance. Lastly, reliable and systematic benchmarking standards and protocols for evaluating and comparing results from different sequencing technologies and protocols are still lacking.

Analysis of intratumoral microbial profiling derived from single-cell sequencing

The primary advantage of microbial profiling analysis from single-cell sequencing compared to bulk data analyses is the ability of cell barcodes to pair microbes with corresponding somatic cells. Deriving the sequencing reads of microbial profiling from the host single-cell sequencing data can help researchers uncover cell-type-specific microbial signatures and infer crosstalk between microbes, immune cells, and tumor cells [45,46].

Currently, studies using host single-cell data to discover the heterogeneity of microbial diversities and abundances in different cancer types are limited. Robinson *et al*. developed cell-type-specific intracellular microbes to extract microbial reads from 21 human scRNA-seq datasets of cancer patients across three cancer types (i.e., Merkel cell carcinoma, colorectal carcinoma, and non-small-cell lung carcinoma) and produced a list of candidate cell-type-specific intracellular microbial taxa (from class-level to species-level) [45]. A **Wilcoxon rank-sum test** on the cell-type-specific microbial taxa abundance revealed that tumor samples from patients receiving immunotherapy exhibited more abundant bacterial taxa in the immune cells than in tumor cells [45]. Then, Ghaddar *et al*. developed a single-cell analysis of host–microbiome interactions (SAHMI) to extract microbial reads from human scRNA-seq data for two pancreatic cancer cohorts, including 41 pancreatic ductal adenocarcinomas (PDA) tumor samples and 14 normal pancreatic tissues samples [46]. They identified host-cell-associated bacteria in a subset of tumors by examining microbial reads paired with host somatic cell barcodes. In addition, by investigating differentially expressed genes (DEGs) in cells associated with bacteria, the strongest bacteria-associated DEGs are linked to PDA or microbiome-related inflammation, indicating that microbes can be involved in growth and inflammatory processes in PDA [46]. Specifically, Table 2 presents the host sequencing datasets that have been used for mining microbial profiling.

IOM analysis of microbial profiling derived from single-cell sequencing faces similar limitations and difficulties as the bulk data, such as the direct mapping of human papillomavirus (HPV) from the host scRNA-seq data [47]. Choosing the viral reference genome and annotation files wisely and optimizing tool parameters are crucial to the mapping results, considering the much lower virus read depths involved in single cells than in bulk samples. Additional inherent limitations of data at the single-cell level include the following. (i) It is difficult to determine the condition and cellular localization of microbes extracted from host sequencing data. For example, we cannot determine whether detected microbial nucleic acids come from living, lysed, intracellular, or extracellular microorganisms [30,46]. (ii) The decontamination step remains challenging in analyzing microbial profiling derived from host data. Removal of contaminants *in silico* cannot replace gold-standard wet experiments, such as sterile processing, sterile-certified reagents, and negative blanks of reagents [46]. Many technical operations in decontamination limit the analysis of individual-specific, region-specific, low-abundance, or difficult-to-detect microbes [30,46]. (iii) The low-biomass microorganisms mined from host data will also impact the subsequent analysis, especially for single-cell data. Pan-cancer analysis of intratumoral microbiome abundance estimation showed only one bacterial cell per 147 tumor cells in the TME [4]. The low microbial biomass issue will significantly impact the capture of microbial profiling for a specific species mining from host tissue samples infected by microbes, requiring additional reliability validation for data mining. (iv) The mechanisms for capturing microbial nucleic acids in scRNA-seq data

Table 2. Overview of host sequencing datasets that have been used for mining microbial profiling

Cancer type[a]	Sample counts	Cancer stage	Survival time	Sex	Source	Refs
Bulk sequencing data						
COAD	1006	Yes	Yes	Yes	TCGA (https://portal.gdc.cancer.gov/)	–
READ	372	Yes	Yes	Yes		
KIRC	1140	Yes	Yes	Yes		
THCA	880	Yes	Yes	Yes		
STAD	1091	Yes	Yes	Yes		
BRCA	1497	Yes	Yes	Yes		
HNSC	907	Yes	Yes	Yes		
LUAD	951	Yes	Yes	Yes		
Single-cell sequencing data						
Pancreatic ductal adenocarcinomas	24	Yes	NA	Yes	Genome Sequence Archive under project CRA001160	[92]
Merkel cell carcinoma	2	Yes	NA	Yes	NCBI BioProject PRJNA483959 (patient 2586-4), PRJNA484204 (patient 9245-3)	[93]
Colorectal carcinoma	6	Yes	NA	Yes	ArrayExpress EMTAB-8410	[94]
Non-small-cell lung carcinoma	13	Yes	NA	Yes	NCBI BioProject PRJNA591860	[95]

[a]Cancer type of different samples. BRCA, breast invasive carcinoma; COAD, colon adenocarcinoma; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; LUAD, lung adenocarcinoma; READ, rectum adenocarcinoma; STAD, stomach adenocarcinoma; THCA, thyroid carcinoma.

are still debatable. Several explanations have been discussed, including more polyadenylation than previously believed in prokaryotic transcripts and L-form switching in microbes [46]. More efforts are needed to investigate the origin of microbial acids in host sequencing techniques.

Analysis of intratumoral microbial profiling derived from spatial sequencing

The emerging SRT provides the spatial information that bulk and single-cell RNA-sequencing approaches cannot deliver [48]. Particularly, SRT defines the organization of tissue functional niches and crosstalk that modulate cellular function in human tumor studies [48,49]. For example, Shi *et al*. found that antibiotic treatment can disrupt the spatial networks in the gut microbiome of mice [50]. They reported an altered spatial association with the most significant fold change between *Oscillibacter* and *Veillonella*, which has been linked to altered inflammatory responses and metabolic activities in the host [50]. Another study developed a novel spatial meta-transcriptomic analysis method that captures intratumoral microbes and host transcriptomic data with spatial coordinates [51]. By examining tumor tissue samples from 12 patients with early-stage lung cancer, they found that specific species or strains are significantly enriched in tumor cells compared with other cell types, and the bacterial burden is strongly positively associated with the expression of oncogenic β-catenin [51]. Therefore, mining microbial nucleic acids from SRT data of host tissue samples is a promising research aspect for tumor–immune–microbiome crosstalk in the TME with direct evidence of locations that it can provide.

So far, mature and popularized computational tools have yet to be established for host spatial microbiome data analysis. As the SRT data analysis is still in its infancy, deriving microbiome profiling from host SRT becomes more challenging in building connections between microbes and functional spatially variable genes in the context of tissue architecture. Moreover, current

computational tools in IOM studies, including bulk, single-cell, and spatial data, mainly rely on basic statistical tests for exploring host–microbiome interactions, such as the Wilcoxon rank-sum test [45]. Yet, the correlation between microbial signatures and host tissue type or cell type obtained by statistical tests could be a 'spurious correlation' [52]. This spurious correlation between microorganisms and the host is typically not caused by the intrinsic biological association between them but is generated from the involvement of confounding factors (a coincidence or the presence of a certain third, unseen factor). As an emerging method for investigating the scRNA-seq [53], STR data [54], as well as host-microbiome data [55], **deep learning**, such as **graph neural network** (**GNN**), shows great potential in representing host gene and microbial signatures and building their relations, simultaneously. It is an ideal way to identify reliable tumor-associated immuno-microbiome interactions by examining genes or pathways associated with immunity and metabolism.

Integrated analysis of IOM data

Unlike conventional microbial analysis that investigates the different diversity and composition of gut microbes between healthy and diseased individuals, IOM studies provide a more systematic and mechanistic way to explore the intricate interactions between microbes and cancer hallmarks (Figure 2). Computational mining of microbiome from host bulk, single-cell, and SRT data of host tissue samples can help researchers investigate the identity and spatial distribution of the cell-/cancer-associated microbes, the host cell types with which they interact, and the specific host genes that can be regulated by intracellular microbes. Hence, computational analyses that leverage complementary information of different kinds of data will provide great opportunities for studying intracellular and extracellular microbe–microbe interactions, as well as microbe–host interactions. In addition, analysis of host antitumor response is a critical aspect in IOM studies, which is also not involved in conventional microbial analysis. The integrated analysis of various host and microbial data types is a trend that can bring unique features to solve IOM problems. Specifically, three categories of data integration can be expected, as follows.

(i) Integration of different microbial sequencing data. Newsome *et al*. collected fecal samples from 65 non-small-cell lung cancer (NSCLC) patients (undergoing immune checkpoint inhibitor/inhibition therapy). These fecal samples were investigated in terms of microbial composition and their transcriptome activities using 16S rRNA gene amplicon sequencing and metatranscriptomic sequencing [56]. They revealed that the genus *Ruminococcus* is the strongest associated taxa in responders, and the responders' enriched microbial pathways are carbon-fixation pathways in the prokaryotes [56].

(ii) Integration of microbiome data with host data. As the 16S rRNA gene amplicon sequencing identifies the composition of intratumoral microbes, the resulting microbial composition can guide the size of chosen microbial reference genomes when computationally mining microbial profiling from host SRT data, enabling a less time-consuming and precise alignment. In a recent study, spatial transcriptomics and scRNA-seq were modified to capture microbial rRNA with spatial coordinates in host SRT data and microbial 16S rRNA with host cell barcodes at the single-cell level from oral squamous cell carcinoma (OSCC) and colorectal cancer (CRC) patients [57]. The results showed that bacteria are enriched in the TME niches with immune and epithelial cell functions and promote tumor growth rather than being distributed randomly in the TME. In addition, they found that cell-associated bacteria affect the expression of host genes involved in inflammation, metastasis, cell dormancy, and DNA repair pathways [57].

(iii) Integration of gut and intratumoral microbiome and host data. Coanalyzing data from multiple sequencing technologies on both fecal and tumor tissue samples enables us to

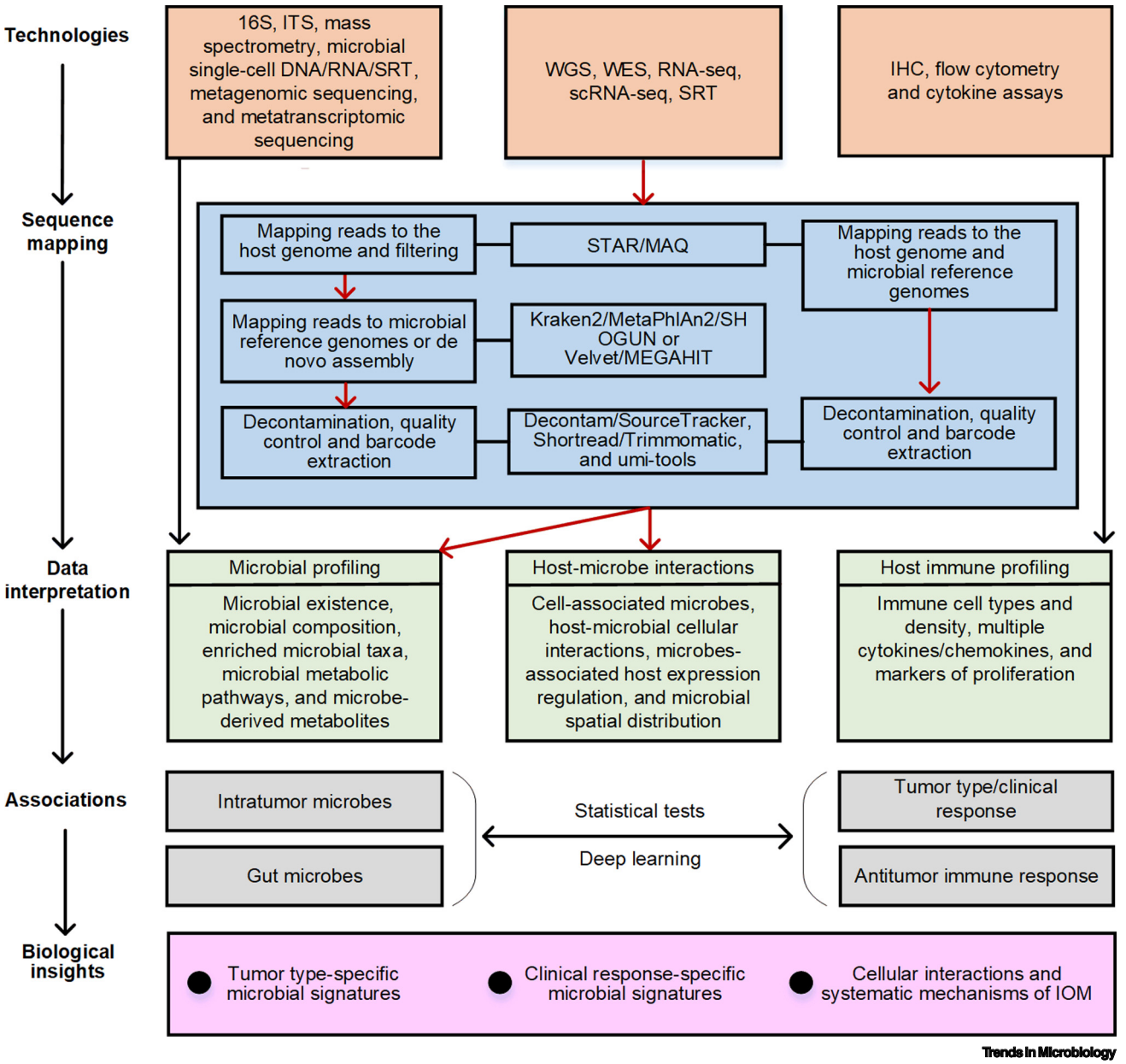

Figure 2. Roadmap for computationally studying immuno–oncology–microbiome (IOM). The tasks in IOM studies are to investigate tumor type-specific and clinical response-specific microbial signatures as well as cellular interactions and systematic mechanisms of IOM. By using various sequencing technologies on host fecal and tissue samples, researchers can obtain data interpretation of IOM, including microbial profiling, host–microbe interactions, and host immune profiling. Computational methods, such as statistical tests and deep learning, reveal the associations between microbes and tumor-type/clinical response or host antitumor immune response by using the collected data interpretation. The processes highlighted in red arrows represent the microbial profiling mining of IOM analysis involving intratumoral microbes, which fills in the missing part of current IOM studies (i.e., the effect of the intratumoral microbiota in the TME). Abbreviations: IHC, immunohistochemistry; MAQ, mapping and assembly with quality; SRT, spatially resolved transcriptomics; STAR, spliced transcripts alignment to a reference; WES, whole-exome sequencing; WGS, whole-genome sequencing.

explore the TME–intestinal transmission of microbes and metabolites and reveals the systematic interactions of the IOM [9]. Uribe-Herranz *et al*. analyzed 16S rRNA gene amplicon sequencing and mass spectrometry data from fecal and tumor tissue samples of mice with

melanoma and lung/cervical cancer [9]. They revealed that vancomycin treatment could eliminate Gram-positive bacteria and decrease short-chain fatty acid (SCFA) concentrations in the gut, enhancing the host's antitumor immune response. Additionally, by applying immunohistochemistry (IHC), flow cytometry, and cytokine assays, researchers can examine the microbe-associated immune cell types, the density of immune cells and markers of antigen processing and presentation in the TME, and the frequency of immune cells in the host systemic circulation to reveal the modulation role of microorganisms on the antitumor immune response.

Integrating different types of microbiome and host data not only inherits the challenges in bulk, single-cell, and spatial data mentioned above but more for data harmonization. The proper alignment of data types from multiple samples requires the careful removal of sample bias. The sequencing and modality bias still exist even for data profiled from the same data. More importantly, how to align microbial profiling in spatial spots and single cells is still unclear, as the microbiome obtained from scRNA-seq data is intracellular only while SRT can capture microbes residing in the TME. All these things make the challenge for integrated analysis squared.

Clinical application of IOM studies

Designing robust computational methods in IOM studies enables the identification of tumor-associated microbial signatures and reveals more undiscovered IOM. These outcomes can aid the understanding of the underlying mechanisms of tumor development and progression, as well as the heterogeneity in the clinical response of patients. For example, the microbiota can promote lung cancer development via γδ T cell [24] (Figure 3A), and the commensal microbiome, such as *Bifidobacterium longum*, *Collinsella aerofaciens*, and *Enterococcus faecium*, may have a mechanistic impact on antitumor immunity in metastatic melanoma patients [12] (Figure 3B). Moreover, the modulation of pulmonary microbiota using antibiotic treatment can promote immunosurveillance against lung metastases, shedding light on new precise treatment designs for cancer prevention [25] (Figure 3C). Studies have showcased that fecal microbiota transplantation (FMT) promotes the immunotherapy response in refractory melanoma patients by modulating the gut microbiome [58,59]. *Enterococcus gallinarum* MRx0518 treatment can cause the upregulation of genes and metagenes associated with antitumor activity in solid tumors [60], showing potent immunostimulatory activity and antitumorigenic efficacy in a clinical trial (NCT03934827). Another clinical trial (NCT03829111) suggested that CBM588, a bifidogenic live bacterial product, appears to improve the overall survival of patients with metastatic renal cell carcinoma who were receiving nivolumab plus ipilimumab [61]. Microbial modulation in immunotherapy can be further reinforced by investigating the mechanism of action, long-term efficacy, and stability of gut microbiome modulation in the cancer treatment [62].

Given that current cancer treatments cannot address refractory metastatic cancers, drug-resistant cancers, and cancers that evade immune clearance, 'bugs as drugs' (such as microbial therapies) may provide solutions to these unresolvable clinical needs [2,63]. Microbial therapies, including oncolytic viral therapy and bacterial antitumor therapy, treat cancer by exploiting tumor-specific infectious microbes [2,63]. For oncolytic viral treatment, talimogene laherparepvec (T-VEC), a modified herpesvirus, is used to treat advanced melanoma by destroying tumor cells and triggering tumor-specific immune responses [64] (Figure 3D). For bacterial antitumor therapy, attenuated recombinant *Listeria monocytogenes* bacteria have been proven to induce long-lasting tumor-specific cytolytic T lymphocyte (CTL) responses by efficiently delivering recombinantly expressed tumor antigens [63].

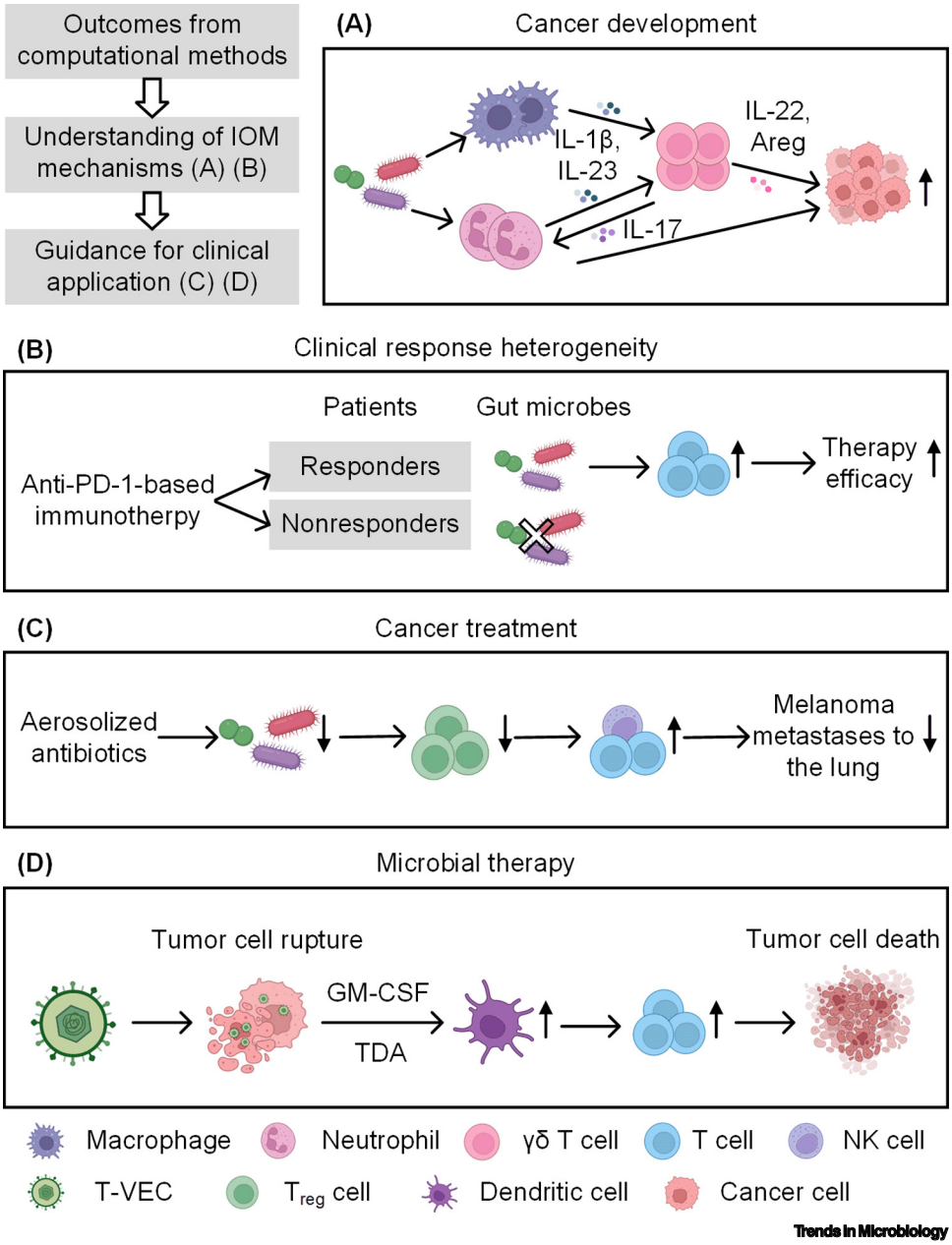

Figure 3. Clinical application of immuno–oncology–microbiome (IOM) studies. Outcomes from computational methods in IOM studies help researchers to understand the IOM and guide more precise cancer treatments. (A) IOM can help researchers understand the mechanisms of tumor development. For example, the dysregulation of local microbiota can promote lung cancer development via γδ T cells [24]. (B) Microbes may contribute to the heterogeneity in the clinical response of cancer patients receiving the same treatments. For example, the differential composition of the commensal microbiome of metastatic melanoma patients may affect the effect of the immunotherapy [12]. (C) IOM can guide cancer treatment. An example is the modulation of pulmonary microbiota by antibiotic treatment, which promotes immunosurveillance against melanoma metastases to the lung [25]. (D) Microbial therapies provide a new opportunity for treating cancers. For example, tumor cell lysis triggered by oncolytic T-VEC releases TDA, GM-CSF, and new viral particles, which can enhance the activation of dendritic cells and initiate a systemic antitumor adaptive immune response in advanced melanoma patients [64]. Abbreviations: Areg, amphiregulin; GM-CSF, granulocyte-macrophage colony-stimulating factor; NK, natural killer; T-VEC, talimogene laherparepvec; TDA, tumor-derived antigens.

Recently, researchers have engineered genetically attenuated, auxotrophic, and inducible versions of *Escherichia*, *Bifidobacterium*, *Listeria*, *Shigella*, *Clostridium*, *Lactococcus*, *Vibrio*, and *Salmonella* species, which exhibited antitumor efficacy in preclinical models [65]. For example, attenuated *Salmonella* expressing aquatic flagellin has been demonstrated to destroy tumor cells by activating the human immune system [65]. *Bifidobacterium bifidum* strains can induce an antitumor host immune response, thus improving the efficacy of checkpoint inhibitors in mice [66]. Moreover, *Clostridium* bacteria can lyse tumor cells growing in hypoxic environments, and *Clostridium novyi* spores have been used to treat patients with solid tumors [67]. These known bacterial–host interactions can be valuable resources for evaluating computational predictions from bulk and single-cell data. Thus, we summarized a list of microorganisms linked to cancer and cancer treatment in the context of the IOM according to published literature (Table 3). These gathered microorganisms contribute to overcoming the lack of valid benchmark standards for evaluating and comparing results from different sequencing technologies and computational approaches.

Concluding remarks

Microorganisms in the GI tract and TME niches contribute to tumor development and progression by interacting with the host's immune system. Although it has been deeply investigated and developed, the gut microbiome only measures microbiome cohort and indirect connections to cancers, which is insufficient to characterize IOM computationally. With the development of sequencing techniques, investigating the intratumoral microbiome from host samples is receiving more attention. Current analyses of host bulk, single-cell, and spatial intratumoral microbial profiling provide a collection of decontaminated microbial compositions of tumor tissues. Microbial profiling mined from host intratumoral samples shows two strengths: (i) it can identify cell-type-specific intracellular microbes that match with host data in the same tissue, providing new insights into the investigation of the multi-omic IOM in host tissue samples, and (ii) it is easier to obtain these data than it is to obtain clinical biopsies.

Yet, computational methods and tools are critically needed to conquer issues in intratumoral microbiome prediction. Several questions are to be answered using computational strategies in elucidating IOM (see Outstanding questions). It is also worth noting that new sequencing technologies, such as microbial single-cell DNA sequencing [68], microbial single-cell RNA sequencing [69], and microbial SRT sequencing [70], provide a different perspective on data for investigating IOM. When these technologies are applied to patient fecal samples, individual-microbe scale data provided by them help researchers to elucidate the mechanistic interactions of IOM. Similar to the single-cell multimodal omics (scMulti-omics) technologies used in the host somatic cells [71], microbial scMulti-omics technologies can measure multiple molecular types from a single microbe (e.g., genomics, transcriptomics, and proteomics), thus enabling researchers to explore the links between microbial transcriptional regulatory mechanisms, tumor, and the host immune system. Although these technologies still need to be developed for general use in microbiological analyses or IOM research, we anticipate that more experimental data will be generated for in-depth functional analysis soon.

Moreover, the promising trends for studying the cellular interactions and systematic mechanisms of IOM mainly include three aspects: (i) the analysis of microbial profiling from host single-cell and spatial sequencing data, (ii) the coanalysis of data from multiple sequencing technologies on different samples, and (iii) the utilization of deep learning methods to investigate biological associations of IOM. Overall, the computational analysis of IOM using high-throughput sequencing data is paving the way to a better understanding of host–microbiome relationships and interactions and how microbes are involved in the TME and cancer treatment.

Outstanding questions

How should computational methods for better profiling microbial communities and investigating the IOM at a higher level of taxonomic resolution (such as at the strain level) be developed?

What advanced computational tools should be generated to effectively mine microbes from the spatially resolved transcriptomics data to investigate IOM interactions?

What is the best practice for revealing undiscovered IOM mechanisms by integrating microbial data from bulk sequencing, single-cell sequencing, and spatially resolved transcriptomics data of host samples?

How should more sophisticated computational methods be designed for elucidating and interpreting the intrinsic biological associations among microorganisms, the host immune system, and transformed cells?

What is the sensitive and reliable method to apply scMulti-omics sequencing technologies to single-cell microbes to support IOM research?

Table 3. A list of microorganisms linked with cancer and cancer treatment in the context of the IOM based on published literature

Cancer type	Microbes associated with cancer			Microbes associated with cancer treatment			Refs
	Microbe	Taxa	Type	Microbe	Taxa	Type	
Stomach cancer, gastric mucosa-associated lymphoid tissue (MALT) lymphoma, and cancer of the esophagus	*Helicobacter pylori*[a]	s	Bacteria				[96,97]
Lung cancer	*Prevotella*	g	Bacteria	*Clostridia*	c	Bacteria	[9,24, 98–100]
	Veillonella	g	Bacteria	*Enterococcus hirae*[b]	s	Bacteria	
	Haemophilus	g	Bacteria	*Firmicutes*	p	Bacteria	
	Streptococcus	g	Bacteria				
	Leuconostocaceae	f	Bacteria				
	Sphingomonadaceae	f	Bacteria				
	Alphaproteobacteria	c	Bacteria				
	Herbaspirillum	g	Bacteria				
	Staphylococcus	g	Bacteria				
	Delftia	g	Bacteria				
	Burkholderiales	o	Bacteria				
	Proteobacteria	p	Bacteria				
Gastrointestinal tumor	*Fusobacterium*	g	Bacteria	–			[23,30]
	Candida	g	Fungi				
Liver hepatocellular carcinoma (LIHC)	*Orthohepadnavirus*	g	Viruses	–			[30]
	Hepatotoxic microcystis	g	Bacteria				
Pancreatic adenocarcinoma (PDAC)	*Sachharopolyspora*	g	Bacteria	*Sachharopolyspora*	g	Bacteria	[20,101]
	Pseudoxanthomonas	g	Bacteria	*Pseudoxanthomonas*	g	Bacteria	
	Streptomyces	g	Bacteria	*Streptomyces*	g	Bacteria	
	Bacillus clausii	s	Bacteria	*Bacillus clausii*	s	Bacteria	
	Malassezia	g	Fungi				
Pancreatic ductal adenocarcinoma (PDA)	*Proteobacteria*	p	Bacteria	*Malassezia*	g	Fungi	[20,26,27]
	Actinobacteria	p	Bacteria	*Gammaproteobacteria*[a]	c	Bacteria	
	Fusobacteria	p	Bacteria				
	Verrucomicrobia	p	Bacteria				
	Malassezia	g	Fungi				
Breast cancer	*Fusobacterium nucleatum*	s	Bacteria	–			[21]
Colorectal cancer (CRC)	*Fusobacterium nucleatum*	s	Bacteria	*E. coli*[a]	s	Bacteria	[8,19,30, 102–111]
	Faecalibacterium	g	Bacteria	*Comamonas*[a]	g	Bacteria	
	Fusobacterium nucleatum	s	Bacteria	*Fusobacterium nucleatum*	s	Bacteria	
	Bacteroides	g	Bacteria	*Bifidobacterium pseudolongum*	s	Bacteria	
	Enterotoxigenic *B. fragilis*	s	Bacteria	*Lactobacillus johnsonii*	s	Bacteria	
	Campylobacter jejuni	s	Bacteria	*Olsenella*	g	Bacteria	
	Genotoxic *pks*$^+$ *E. coli*	s	Bacteria	*Bacteroides vulgatus*[a]	s	Bacteria	
				Clostridium ramosum[a]	s	Bacteria	

Table 3. (continued)

Cancer type	Microbes associated with cancer			Microbes associated with cancer treatment			Refs
	Microbe	Taxa	Type	Microbe	Taxa	Type	
Melanoma	*Bifidobacteria*	g	Bacteria	*Bifidobacterium pseudolongum*	s	Bacteria	[7,8,12, 112–115]
	Bifidobacterium longum	s	Bacteria	*Bifidobacteria*	g	Bacteria	
	Enterococcus faecium	s	Bacteria	*Akkermansia muciniphila*	s	Bacteria	
	Collinsella aerofaciens	s	Bacteria	*Collinsella aerofaciens*	s	Bacteria	
	Bacteroides	g	Bacteria	*Enterococcus faecium*	s	Bacteria	
	Parabacteroides	g	Bacteria	*Olsenella*	g	Bacteria	
				Lactobacillus johnsonii	s	Bacteria	
				Bifidobacterium longum	s	Bacteria	
				Bacteroides[a]	g	Bacteria	
				Burkholderiales[a]	o	Bacteria	
				Bifidobacterium[a]	g	Bacteria	
Cancer at anogenital sites, cancer of the upper aerodigestive tract, and cancer of the skin	*Human papillomavirus*[a]	s	Viruses	–	–		[97]
Cervical cancer	*Human papillomavirus*[a]	s	Viruses	*Clostridia*	c	Bacteria	[9,97,116]
				Firmicutes	p	Bacteria	
Cervical squamous cell carcinoma (CESC)	*Alphapapillomavirus*	g	Viruses	–			[30]
Burkitt lymphoma, immunosuppression-related non-Hodgkin lymphoma, extranodal NK/T-cell lymphoma, Hodgkin lymphoma, and nasopharyngeal carcinoma	*Epstein–Barr virus*[a]	s	Viruses	–			[97]
Hepatocellular carcinoma, cholangiocarcinoma, and non-Hodgkin lymphoma	*Hepatitis B virus*[a]	s	Viruses	–			[97]
Hepatocellular carcinoma, lymphoid malignancies, leukemias, and cancer of the thyroid	*Hepatitis C virus*[a]	s	Viruses	–			[97]
Kaposi sarcoma, primary effusion lymphoma, and multiple myeloma	*Kaposi sarcoma herpesvirus*[a]	s	Viruses	–			[97]
Kaposi sarcoma, non-Hodgkin lymphoma, Hodgkin lymphoma, cervical and anogenital cancers, cancer of the skin, cancer of the conjunctiva, cancer of the lung, and cancer of the liver	*Human immunodeficiency virus I*[a]	s	Viruses	–			[97]
T-cell malignancies, cutaneous T-cell lymphoma, B- and T-cell lymphomas, and non-lymphomatous tumors	*Human T-cell leukemia virus type I*[a]	–	Viruses	–			[97]
Cancer of the urinary bladder and cancers of the female genital tract	*Schistosoma haematobium*[a]	s	Eukaryota	–			[90]
Cholangiocarcinoma and hepatocellular carcinoma	*Opisthorchis viverrini*[a]	s	Eukaryota	–			[90]
	Clonorchis sinensis[a]	s	Eukaryota				
Bladder cancer	–			*Bifidobacterium pseudolongum*	s	Bacteria	[8]
				Lactobacillus johnsonii	s	Bacteria	
				Olsenella	g	Bacteria	
Liver cancer	–			*Clostridium scindens*	s	Bacteria	[117]

(continued on next page)

Table 3. (continued)

Cancer type	Microbes associated with cancer			Microbes associated with cancer treatment			Refs
	Microbe	Taxa	Type	Microbe	Taxa	Type	
Epithelial tumor	–			*Akkermansia muciniphila*[a]	s	Bacteria	[118]
Prostate cancer	–			*Akkermansia muciniphila*	s	Bacteria	[119]
Sarcoma	–			*Enterococcus hirae*[a]	s	Bacteria	[99,100]

[a]Microorganisms labeled as human carcinogens ('oncomicrobes') by the International Association for Cancer Registries (IACR).
[b]Microorganisms associated with cancer drug treatment. p, phylum; c, class; o, order; f, family; g, genus; s, species.

Code availability

We provided a pipeline for analyzing microbial profiling mined from host single-cell sequencing datasets on GitHub at https://github.com/OSU-BMBL/IOM.

Acknowledgments

This work was supported by awards R01-GM131399 from the National Institute of General Medical Sciences of the National Institutes of Health. This work was supported by the Pelotonia Institute for Immuno-Oncology (PIIO) at the Ohio State University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the PIIO. We thank Yihua Wang, Yuhan Sun, and Yuli Zhang from Shandong University for collecting materials. Figures 1 and 3 were created with BioRender.com.

Declaration of interests

No interests are declared.

References

1. Aldars-García, L. *et al.* (2021) Systematic review: the gut microbiome and its potential clinical application in inflammatory bowel disease. *Microorganisms* 9, 977
2. Sepich-Poore, G.D. *et al.* (2021) The microbiome and human cancer. *Science* 371, eabc4552
3. Toker, J. *et al.* (2020) The microbiome in immuno-oncology. In *Immunotherapy. Advances in Experimental Medicine and Biology* (1244) (Naing, A. and Hajjar, J., eds), pp. 325–334, Springer
4. Lythgoe, M.P. *et al.* (2022) Polymorphic microbes: a new emerging hallmark of cancer. *Trends Microbiol.* 30, 1131–1134
5. Peled, J.U. *et al.* (2020) Microbiota as predictor of mortality in allogeneic hematopoietic-cell transplantation. *N. Engl. J. Med.* 382, 822–834
6. Schluter, J. *et al.* (2020) The gut microbiota is associated with immune cell dynamics in humans. *Nature* 588, 303–307
7. Li, Y. *et al.* (2019) Gut microbiota dependent anti-tumor immunity restricts melanoma growth in $Rnf5^{-/-}$ mice. *Nat. Commun.* 10, 1492
8. Mager, L.F. *et al.* (2020) Microbiome-derived inosine modulates response to checkpoint inhibitor immunotherapy. *Science* 369, 1481–1489
9. Uribe-Herranz, M. *et al.* (2020) Gut microbiota modulate dendritic cell antigen presentation and radiotherapy-induced antitumor immune response. *J. Clin. Investig.* 130, 466–479
10. Li, Y. *et al.* (2022) Effects of gut microbiota on host adaptive immunity under immune homeostasis and tumor pathology state. *Front. Immunol.* 13, 844335
11. Kim, C.H. (2021) Control of lymphocyte functions by gut microbiota-derived short-chain fatty acids. *Cell. Mol. Immunol.* 18, 1161–1171
12. Matson, V. *et al.* (2018) The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* 359, 104–108
13. Gopalakrishnan, V. *et al.* (2018) Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* 359, 97–103
14. Kartal, E. *et al.* (2022) A faecal microbiota signature with high specificity for pancreatic cancer. *Gut* 71, 1359–1372
15. Lee, K.A. *et al.* (2021) The gut microbiome: what the oncologist ought to know. *Br. J. Cancer* 125, 1197–1209
16. Cammarota, G. *et al.* (2020) Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nat. Rev. Gastroenterol. Hepatol.* 17, 635–648
17. Turnbaugh, P.J. *et al.* (2007) The human microbiome project. *Nature* 449, 804–810
18. Qin, J. *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65
19. Hamada, T. *et al.* (2018) *Fusobacterium nucleatum* in colorectal cancer relates to immune response differentially by tumor microsatellite instability status. *Cancer Immunol. Res.* 6, 1327–1336
20. Aykut, B. *et al.* (2019) The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. *Nature* 574, 264–267
21. Parhi, L. *et al.* (2020) Breast cancer colonization by *Fusobacterium nucleatum* accelerates tumor growth and metastatic progression. *Nat. Commun.* 11, 3259
22. Fu, A. *et al.* (2022) Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. *Cell* 185, 1356–1372. e26
23. Dohlman, A.B. *et al.* (2022) A pan-cancer mycobiome analysis reveals fungal involvement in gastrointestinal and lung tumors. *Cell* 185, 3807–3822. e12
24. Jin, C. *et al.* (2019) Commensal microbiota promote lung cancer development via γδ T cells. *Cell* 176, 998–1013. e16
25. Le Noci, V. *et al.* (2018) Modulation of pulmonary microbiota by antibiotic or probiotic aerosol therapy: a strategy to promote immunosurveillance against lung metastases. *Cell Rep.* 24, 3528–3538
26. Pushalkar, S. *et al.* (2018) The pancreatic cancer microbiome promotes oncogenesis by induction of innate and adaptive immune suppressionmicrobiome influences pancreatic oncogenesis. *Cancer Discov.* 8, 403–416
27. Geller, L.T. *et al.* (2017) Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science* 357, 1156–1160

28. Nejman, D. *et al.* (2020) The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* 368, 973–980
29. Liu, J. and Zhang, Y. (2022) Intratumor microbiome in cancer progression: current developments, challenges and future trends. *Biomark. Res.* 10, 37
30. Poore, G.D. *et al.* (2020) Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* 579, 567–574
31. Narunsky-Haziza, L. *et al.* (2022) Pan-cancer analyses reveal cancer-type-specific fungal ecologies and bacteriome interactions. *Cell* 185, 3789–3806. e17
32. Dohlman, A.B. *et al.* (2021) The cancer microbiome atlas: a pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants. *Cell Host Microbe* 29, 281–298. e5
33. Li, H. *et al.* (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18, 1851–1858
34. Dobin, A. *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21
35. Wood, D.E. and Salzberg, S.L. (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15, R46
36. Truong, D.T. *et al.* (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903
37. Hillmann, B. *et al.* (2018) Evaluating the information content of shallow shotgun metagenomics. *Msystems* 3 e00069–18
38. Namiki, T. *et al.* (2011) MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. In *Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, pp. 116–124
39. Li, D. *et al.* (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676
40. Kostic, A.D. *et al.* (2011) PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat. Biotechnol.* 29, 393–396
41. Davis, N.M. *et al.* (2018) Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6, 226
42. Knights, D. *et al.* (2011) Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8, 761–763
43. Paulson, J.N. *et al.* (2013) Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* 10, 1200–1202
44. Weiss, S. *et al.* (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27
45. Robinson, W. *et al.* (2021) CSI-Microbes: identifying cell-type specific intracellular microbes from single-cell RNA-seq data. *bioRxiv* Published online May 14, 2021. https://doi.org/10.1101/2020.05.14.096230
46. Ghaddar, B. *et al.* (2022) Tumor microbiome links cellular programs and immunity in pancreatic cancer. *Cancer Cell* 40, 1240–1253. e5
47. Bost, P. *et al.* (2020) Host–viral infection maps reveal signatures of severe COVID-19 patients. *Cell* 181, 1475–1488. e12
48. Hunter, M.V. *et al.* (2021) Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nat. Commun.* 12, 6278
49. Ji, A.L. *et al.* (2020) Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* 182, 497–514. e22
50. Shi, H. *et al.* (2020) Highly multiplexed spatial mapping of microbial communities. *Nature* 588, 676–681
51. Wong-Rolle, A. *et al.* (2022) Spatial meta-transcriptomics reveal associations of intratumor bacteria burden with lung cancer cells showing a distinct oncogenic signature. *J. Immunother. Cancer* 10, e004698
52. Haig, Brian D. (2003) What Is a Spurious Correlation? *Underst. Stat.* 2, 125–132
53. Wang, J. *et al.* (2021) scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nat. Commun.* 12, 1882
54. Chang, Y. *et al.* (2022) Define and visualize pathological architectures of human tissues from spatially resolved transcriptomics using deep learning. *Comput. Struct. Biotechnol. J.* 20, 4600–4617
55. Sudhakar, P. *et al.* (2021) Computational biology and machine learning approaches to understand mechanistic microbiome-host interactions. *Front. Microbiol.* 12, 618856
56. Newsome, R.C. *et al.* (2022) Interaction of bacterial genera associated with therapeutic response to immune checkpoint PD-1 blockade in a United States cohort. *Genome Med.* 14, 35
57. Niño, J.L.G. *et al.* (2022) Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature* 611, 810–817
58. Davar, D. *et al.* (2021) Fecal microbiota transplant overcomes resistance to anti-PD-1 therapy in melanoma patients. *Science* 371, 595–602
59. Baruch, E.N. *et al.* (2021) Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. *Science* 371, 602–609
60. Lythgoe, M. *et al.* (2021) 543P Neoadjuvant MRx0518 treatment is associated with significant gene and metagene signature changes in solid tumours. *Ann. Oncol.* 32, S607
61. Dizman, N. *et al.* (2022) Nivolumab plus ipilimumab with or without live bacterial supplementation in metastatic renal cell carcinoma: a randomized phase 1 trial. *Nat. Med.* 28, 704–712
62. Lythgoe, M.P. *et al.* (2022) The potential of fecal microbiota transplantation in oncology. *Trends Microbiol.* 30, 10–12
63. Forbes, N.S. *et al.* (2018) White paper on microbial anti-cancer therapy and prevention. *J. Immunother. Cancer* 6, 1–24
64. Ferrucci, P.F. *et al.* (2021) Talimogene laherparepvec (T-VEC): an intralesional cancer immunotherapy for advanced melanoma. *Cancers* 13, 1383
65. Kramer, M.G. *et al.* (2018) Bacterial therapy of cancer: promises, limitations, and insights for future directions. *Front. Microbiol.* 9, 16
66. Lee, S.-H. *et al.* (2021) *Bifidobacterium bifidum* strains synergize with immune checkpoint inhibitors to reduce tumour burden in mice. *Nat. Microbiol.* 6, 277–288
67. Janku, F. *et al.* (2021) Intratumoral injection of *Clostridium novyi*-NT spores in patients with treatment-refractory advanced solid tumors Phase I study of intratumoral *Clostridium novyi*-NT. *Clin. Cancer Res.* 27, 96–106
68. Zheng, W. *et al.* (2022) High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science* 376, eabm1483
69. Dohn, R. *et al.* (2021) mDrop-seq: massively parallel single-cell RNA-seq of *Saccharomyces cerevisiae* and *Candida albicans*. *Vaccines* 10, 30
70. Dar, D. *et al.* (2021) Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science* 373, eabi4882
71. Ma, A. *et al.* (2020) Integrative methods and practical challenges for single-cell multi-omics. *Trends Biotechnol.* 38, 1007–1022
72. Baruzzo, G. *et al.* (2017) Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat. Methods* 14, 135–139
73. Alser, M. *et al.* (2021) Technology dictates algorithms: recent developments in read alignment. *Genome Biol.* 22, 249
74. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760
75. Prasad, V.V.S. and Loshma, G. (2011) (2011) HPC-MAQ: a parallel short-read reference assembler. *CCSEA* 2011, 84
76. Meyer, F. *et al.* (2022) Critical assessment of metagenome interpretation: the second round of challenges. *Nat. Methods* 19, 429–440
77. Zeng, H. *et al.* (2021) Physiological and metagenomic strategies uncover the rhizosphere bacterial microbiome succession underlying three common environmental stresses in cassava. *J. Hazard. Mater.* 411, 125143
78. Lind, A.L. and Pollard, K.S. (2021) Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. *Microbiome* 9, 58
79. Vollmers, J. *et al.* (2017) Comparing and evaluating metagenome assembly tools from a microbiologist's perspective-not only size matters! *PLoS One* 12, e0169662
80. Quince, C. *et al.* (2017) Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844
81. Cao, Q. *et al.* (2021) Effects of rare microbiome taxa filtering on statistical analysis. *Front. Microbiol.* 11, 607325
82. Xu, Y. *et al.* (2022) Application of fast expectation-maximization microbial source tracking to discern fecal contamination in rivers exposed to low fecal inputs. *J. Microbiol.* 60, 594–601

83. Shenhav, L. *et al.* (2019) FEAST: fast expectation-maximization for microbial source tracking. *Nat. Methods* 16, 627–632
84. Morgan, M. *et al.* (2009) ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* 25, 2607–2608
85. McKenna, A. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303
86. Chen, S. *et al.* (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890
87. Bolger, A.M. *et al.* (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120
88. Smith, T. *et al.* (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* 27, 491–499
89. Wysock, B.M. *et al.* (1995) Statistical procedures for corrosion studies. *J. Am. Water Works Assoc.* 87, 99–112
90. Parab, S. and Bhalerao, S. (2010) Choosing statistical test. *Int. J. Ayurveda Res.* 1, 187
91. De Winter, J.C. *et al.* (2016) Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. *Psychol. Methods* 21, 273
92. Peng, J. *et al.* (2019) Single-cell RNA-seq highlights intratumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* 29, 725–738
93. Paulson, K. *et al.* (2018) Acquired cancer resistance to combination immunotherapy from transcriptional loss of class I HLA. *Nat. Commun.* 9, 3868
94. Lee, H.-O. *et al.* (2020) Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat. Genet.* 52, 594–603
95. Maynard, A. *et al.* (2020) Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* 182, 1232–1251. e22
96. Suzuki, M. *et al.* (2009) *Helicobacter pylori* CagA phosphorylation-independent function in epithelial proliferation and inflammation. *Cell Host Microbe* 5, 23–34
97. IARC (2012) *IARC Working Group on the Evaluation of Carcinogenic Risks to Humans: Biological Agents*. vol 100 B. *A Review of Human Carcinogens. IARC Monogr. Eval. Carcinog. Risks Hum.* 100, 1
98. Tsay, J.-C.J. *et al.* (2021) Lower airway dysbiosis affects lung cancer progression. *Cancer Discov.* 11, 293–307
99. Viaud, S. *et al.* (2013) The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science* 342, 971–976
100. Daillère, R. *et al.* (2016) *Enterococcus hirae* and *Barnesiella intestinihominis* facilitate cyclophosphamide-induced therapeutic immunomodulatory effects. *Immunity* 45, 931–943
101. Riquelme, E. *et al.* (2019) Tumor microbiome diversity and composition influence pancreatic cancer outcomes. *Cell* 178, 795–806. e12
102. Pleguezuelos-Manzano, C. *et al.* (2020) Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. *Nature* 580, 269–273
103. Dejea, C.M. *et al.* (2018) Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* 359, 592–597
104. Barrett, M. *et al.* (2020) Mutagenesis by microbe: the role of the microbiota in shaping the cancer genome. *Trends Cancer* 6, 277–287
105. Yu, T. *et al.* (2017) *Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell* 170, 548–563. e16
106. He, Z. *et al.* (2019) *Campylobacter jejuni* promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut* 68, 289–300
107. Rubinstein, M.R. *et al.* (2013) *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe* 14, 195–206
108. Nanjundappa, R.H. *et al.* (2017) A gut microbial mimic that hijacks diabetogenic autoreactivity to suppress colitis. *Cell* 171, 655–667. e17
109. Scott, T.A. *et al.* (2017) Host–microbe co-metabolism dictates cancer drug efficacy in *C. elegans*. *Cell* 169, 442–456. e18
110. García-González, A.P. *et al.* (2017) Bacterial metabolism affects the *C. elegans* response to cancer chemotherapeutics. *Cell* 169, 431–441. e8
111. Wallace, B.D. *et al.* (2010) Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* 330, 831–835
112. Snyder, A. *et al.* (2014) Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189–2199
113. Li, Y. *et al.* (2020) Prebiotic-induced anti-tumor immunity attenuates tumor growth. *Cell Rep.* 30, 1753–1766. e6
114. Vétizou, M. *et al.* (2015) Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* 350, 1079–1084
115. Sivan, A. *et al.* (2015) Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti–PD-L1 efficacy. *Science* 350, 1084–1089
116. Lowy, D.R. and Schiller, J.T. (2017) Preventing cancer and other diseases caused by human papillomavirus infection: 2017 Lasker-DeBakey Clinical Research Award. *JAMA* 318, 901–902
117. Ma, C. *et al.* (2018) Gut microbiome-mediated bile acid metabolism regulates liver cancer via NKT cells. *Science* 360, eaan5931
118. Routy, B. *et al.* (2018) Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors. *Science* 359, 91–97
119. Daisley, B.A. *et al.* (2020) Abiraterone acetate preferentially enriches for the gut commensal *Akkermansia muciniphila* in castrate-resistant prostate cancer patients. *Nat. Commun.* 11, 4822