Immunology & Cell Biology 2024; 1–12
www.wileyonlinelibrary.com/journal/icb

SPECIAL FEATURE REVIEW

OPEN

Integrating functional metagenomics to decipher microbiome–immune interactions

Puspendu Sardar[1,2], Alexandre Almeida[3] & Virginia A Pedicord[1,2]

1 Cambridge Institute of Therapeutic Immunology and Infectious Disease, Jeffrey Cheah Biomedical Centre, Cambridge, UK
2 Department of Medicine, University of Cambridge School of Clinical Medicine, Cambridge, UK
3 Department of Veterinary Medicine, University of Cambridge School of Biological Sciences, Cambridge, UK

Keywords
Bioinformatics, Functional metagenomics, Immunology, Microbiome

Correspondence
Virginia A Pedicord, Cambridge Institute of Therapeutic Immunology and Infectious Disease, Jeffrey Cheah Biomedical Centre, Cambridge, UK.
E-mail: vap33@cam.ac.uk

Received 5 April 2024;
Revised 4 June 2024;
Accepted 13 June 2024

doi: 10.1111/imcb.12798

Immunology & Cell Biology 2024; 1–12

Abstract

Microbial metabolites can be viewed as the cytokines of the microbiome, transmitting information about the microbial and metabolic environment of the gut to orchestrate and modulate local and systemic immune responses. Still, many immunology studies focus solely on the taxonomy and community structure of the gut microbiota rather than its functions. Early sequencing-based microbiota profiling approaches relied on PCR amplification of small regions of bacterial and fungal genomes to facilitate identification of the microbes present. However, recent microbiome analysis methods, particularly shotgun metagenomic sequencing, now enable culture-independent profiling of microbiome functions and metabolites in addition to taxonomic characterization. In this review, we showcase recent advances in functional metagenomics methods and applications and discuss the current limitations and potential avenues for future development. Importantly, we highlight a few examples of key areas of opportunity in immunology research where integrating functional metagenomic analyses of the microbiome can substantially enhance a mechanistic understanding of microbiome–immune interactions and their contributions to health and disease states.

INTRODUCTION

The human gastrointestinal tract houses a vast community of resident microbes, collectively known as the microbiota, that supports host metabolism, immune development and infection resistance.[1] Gut microbiota are composed of several kingdoms of microorganisms, including bacteria, archaea, fungi and viruses, with bacteria being the most abundant and harboring over 99% of the microbial genes in the human gut.[2,3] Along with their genes and metabolites, these microbes form the gut microbiome, a dynamic ecosystem that changes spatially (in different regions of the gut), temporally (within the diurnal cycle and over the host lifetime) and in response to environmental factors such as diet and antibiotic treatment.[4,5] The microbiome extensively contributes to host metabolism and nutrient absorption by converting and fermenting dietary substrates that are not directly usable by the host and producing essential vitamins and metabolites that the host can absorb and utilize.[6] Resident gut microbes also contribute to colonization resistance against enteric pathogens both by occupying similar microbial niches and by triggering microbe–microbe and host–microbe defenses.[7]

The high degree of host-specificity of gut microbiomes suggests an intricate co-evolution with the host, based on a number of distinguishing factors affecting the gut microenvironment and consequent microbial community assembly.[8] These factors likely include, but are not limited to, diet, mucus production and glycosylation, host-restricted pathogenic threats and host epithelial and immune antimicrobial pressures. Pioneering studies in humans and 59 other species of mammals revealed that the interplay between diet and host phylogeny largely dictates the bacterial diversity of the microbiota,[9] and several detailed studies have gone on to characterize such associations.[10,11] For example, Desai *et al.* showed that a low-fiber diet leads to a depletion of fiber-fermenting

bacteria and an expansion of mucus-degrading bacteria in mice that results in increased susceptibility to pathogenic infection with *Citrobacter rodentium*,[12] demonstrating how host differences in dietary fiber abundance and mucus can select for different microbial populations. The restoration of a "healthy" microbiome through fecal microbiota transplant (FMT) has been trialed with varying success to treat recurrent infection, inflammatory bowel disease and graft-versus-host disease,[13] and its efficacy in the resolution of *Clostridioides difficile* infection has been shown to depend on host adaptive immune cells.[14]

The host, in turn, has likely evolved both to facilitate occupation by beneficial microbes and to respond to cues from the resident microbial community regarding the metabolic and ecological status of the gut. As home to the largest population of lymphocytes in the human body,[15] the intestinal cellular environment, including intestinal epithelial cells, is perfectly poised to receive and transmit information from the microbiota. The gut therefore constitutes a major communication hub between the microbiota and the immune system, with nearly endless possible combinations of direct and indirect interactions. Truly understanding the impacts of the gut microbiome on host immunity therefore requires functionally characterizing these interactions from both microbe and immune cell perspectives. Early studies relied on the isolation and culture of gut bacteria and the phenotyping of immune cell responses, largely *in vitro*,[16,17] limiting analyses to targeted immune read-outs with the few commensal bacteria species that were culturable. With the advent of improved and higher-throughput techniques for anaerobic bacteria culture, more recent studies have been able to conduct larger *in vivo* screens, monocolonizing germ-free mice with commensal bacteria to assess the effects on specific immune cell subsets such as colonic regulatory T cells.[18] As multi-omic approaches such as single-cell RNA sequencing (scRNA-seq) and proteomics continue to improve our appreciation of immune cell plasticity and heterogeneity, full utilization of the wealth of microbial information encoded in metagenomes serves as an important next step in uncovering the mechanisms of microbiome contributions to immune function.

ADVANCES IN CULTURE-INDEPENDENT MICROBIOME ANALYSES

The advent of high-throughput DNA sequencing has revolutionized the ability to characterize the composition of the human microbiome and improved our understanding of its relationship with human health. Historically, the first major studies analyzing the human microbiome through DNA-based methods relied on a technology pioneered by Carl Woese, known as 16S (ribosomal RNA) rRNA genotyping.[19] The 16S rRNA gene encodes the RNA component of the small subunit of a prokaryotic ribosome. As it is a highly conserved gene present in all bacteria and archaea species, it was originally utilized to obtain the first insights into the stability and variability of the human microbiome in large population cohorts such as the Human Microbiome Project.[20] Since then, it has been commonly used to link the taxonomic composition of a microbiome sample to health and disease states and to profile microbiomes across various habitats and geographical locations.[21–24] However, 16S rRNA genotyping can only reliably classify samples up to a genus-level resolution,[25] and does not provide meaningful functional insights, which limits any biological interpretation. In addition, disentangling correlation from causation is a major challenge in the field, as many microbiome taxonomic associations may simply reflect a shared microbial response to a change in the host state rather than a causative role.[26]

Further developments in sequencing and computational analyses have led to an increasing adoption of whole-genome shotgun metagenomics.[27] Using this technique, the genetic repertoire of a microbiome sample is sequenced without targeted amplification. Therefore, metagenomics not only provides a greater resolution to characterize the microbiome diversity at the species and subspecies levels, but also offers additional insights into the functional capacity of the microbiota community.[28] One major advantage of metagenomics over 16S rRNA genotyping is the ability to reconstruct new genes and genomes missing from reference databases, which has led to the rise of what is known as genome-resolved metagenomics (Figure 1). In essence, genome-resolved metagenomics involves assembling and binning DNA fragments into metagenome-assembled genomes (MAGs). The process of genome binning consists of grouping individual DNA fragments (contigs) based on features such as sequence composition (e.g. tetranucleotide frequencies) and sequencing depths estimated within a single sample or across multiple metagenomic samples. Pioneering work from Tyson *et al.* in 2004[29] was among the first to successfully recover near-complete microbial genomes from metagenomic data.

Within the past decade, genome-resolved metagenomics has since been applied to generate new catalogs of MAGs from various habitats,[30] including cow rumen,[31] pig[32] and mouse.[33] With regards to the human gut microbiome, three major studies[34–36] used genome-resolved metagenomics to reconstruct tens of thousands of MAGs revealing the existence of over 3000 uncultured species.[37] Although the quality of MAGs is

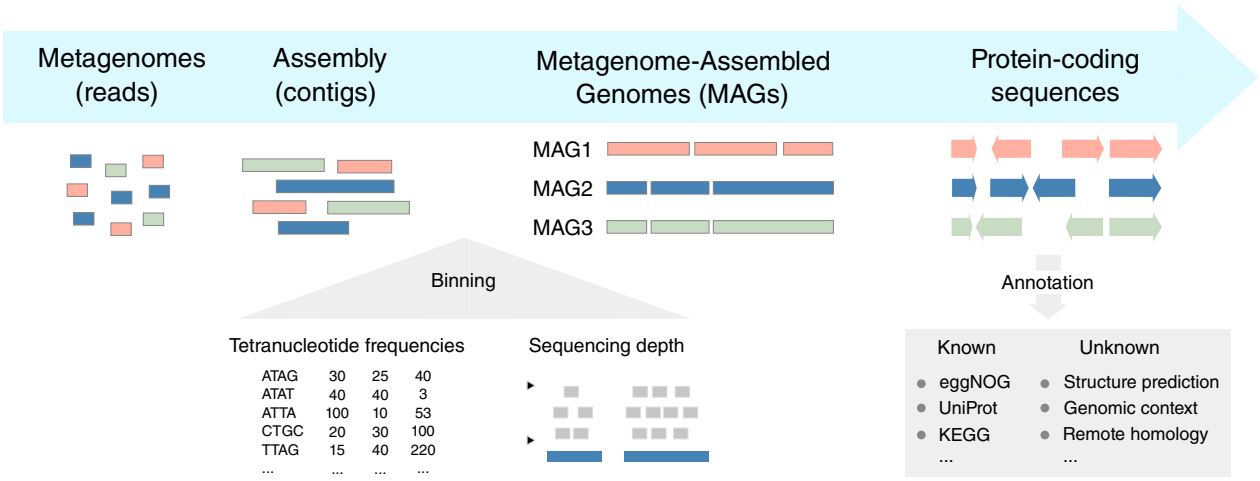


Figure 1. Overview of a typical genome-resolved metagenomics workflow. Generating metagenome-assembled genomes (MAGs) begins with the assembly of short sequence fragments (reads) into longer contiguous sequences (contigs). Thereafter, genome binning is performed by grouping contigs with similar tetranucleotide frequencies and sequencing depths. Protein-coding sequences can be extracted from MAGs, which offer additional information on the genomic context and taxonomic affiliation of the predicted genes. Diverse sequence databases are available to annotate the extracted sequences, while additional approaches such as protein structure prediction and remote homology analyses can be used to further investigate the unknown diversity.

inherently lower than that of pure isolate genomes, they enable characterization of the sequence diversity of unknown bacterial species that currently lack any cultured representatives. In addition, a wider implementation of long-read metagenomic sequencing (e.g. Oxford Nanopore) will help to improve the quality of generated MAGs, with some studies even reporting the successful assembly of complete, closed bacterial genomes from metagenomic data.[38] Although long-read sequencing still has generally lower per-read accuracy than short-read sequencing, combining both technologies can lead to the highest quality reference genomes. Ultimately, the ability to create custom, habitat-specific reference databases can substantially improve the coverage and accuracy of metagenomic analyses. Therefore, having access to the genomes of these uncharacterized species opens new opportunities to understand the biology of the human microbiome through more focused approaches such as functional metagenomics.

BIOINFORMATICS TOOLS AND APPROACHES FOR FUNCTIONAL METAGENOMICS

Functional characterization of metagenomes can be performed through either reference-based mapping methods or annotation of *de novo* assembled genes (Figure 2), with each approach presenting its own advantages and limitations. Mapping-based techniques (e.g. HUMAnN3)[39] are more computationally efficient and enable the detection, as well as quantification, of low-abundance, rare genes. However, analyses of *de novo* assembled genomes (e.g. MAGs) generally provide greater accuracy in gene annotation, as these approaches are able to capture full-length genes and to analyze new genes and functions missing from reference databases. Annotation of metagenome assemblies is achieved through a combination of gene and functional predictions. The gene-finding algorithm created by Hyatt *et al.* in 2010 (known as Prodigal[40]) remains the most popular approach for extracting protein-coding sequences in prokaryotic genomes. This method utilizes a combination of Hidden Markov Models (HMMs) and dynamic programming to identify protein-coding genes through three main steps: (1) training an HMM on known protein-coding sequences, (2) scanning the input DNA sequences to identify potential coding regions using the trained model, and (3) refining the predictions based on features such as start and stop codons, and coding potential scores. More recently, new tools, such as Bakta,[41] have also been developed to improve the standardization of gene annotations and the recovery of small (< 50 amino acids) protein-coding sequences.

Once a set of genes is identified, a wide range of methods and databases are available to functionally annotate the predicted genes. Alignment-based methods (e.g. BLAST[42] or HMMER[43]), can be used to match individual sequences to large databases such as eggNOG,[44]

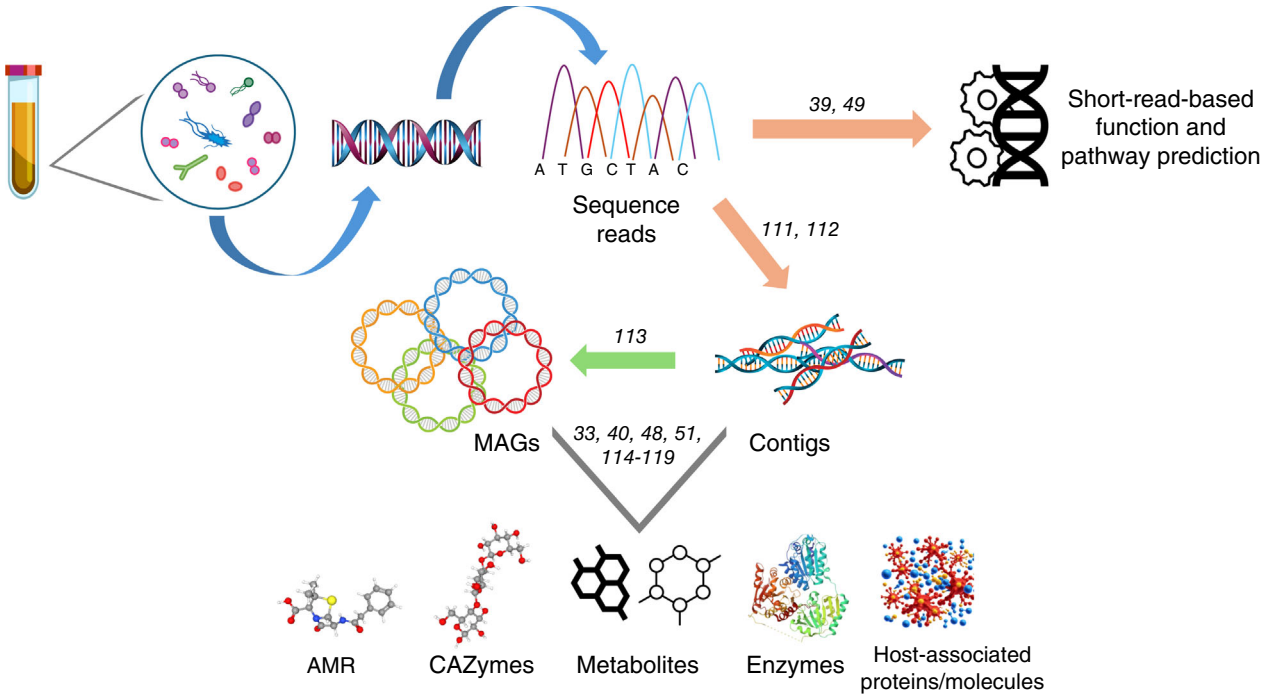


Figure 2. Schematic diagram of the major steps and bioinformatics tools used for functional metagenomics. The major bioinformatics tools and/or approaches used in different steps are referenced in *italics*. Workflows such as HUMAnN and MetaBGC can leverage quality-controlled short sequencing reads directly by implementing mapping of reads to reference databases or profile hidden Markov models (pHMMs), respectively, for functional annotation and biological pathway prediction. On the other hand, a *de novo* assembly-based approach employs multiple steps, starting with contig assembly, but provides more flexibility in choosing different tools and databases. In the assembly-based approach, functional annotation of a metagenome can be performed either using the assembled contigs or after producing MAGs. Specialized tools and biological databases offer more comprehensive functional profiling of the metagenomic samples. Some of these databases provide annotation of antimicrobial resistance (AMR) genes, carbohydrate-active enzymes (CAZymes), primary metabolites and metabolic pathways.[111–119]

InterPro,[45] UniProt[46] and KEGG,[47] which collectively encompass millions of protein families recovered from organisms belonging to all domains of life. Beyond these broad protein collections, there are dedicated approaches and tools for profiling specific functional attributes such as the primary metabolism of gut-associated species (e.g. gutSMASH[48]), secondary metabolites (e.g. MetaBGC[49] and antiSMASH[50]), antimicrobial resistance genes (AMRFinderPlus[51] and CARD[52]) and carbohydrate utilization enzymes (dbCAN[53]). Furthermore, we have recently published a dedicated tool that analyzes the functional aspects of mouse metagenomes and allows users to identify the functional equivalents of mouse gut microbes in the human microbiota.[33] Altogether, these tools provide an integrated view of the functional metagenomic diversity within a microbiome, which can be further linked to specific phenotypic characteristics. To achieve this, the number of annotated genes and functions must be subsequently pre-processed before performing differential abundance analyses. This may involve filtering out very rare features (< 1%) and performing additional normalization approaches such as total sum scaling and centered-log ratio transformation. However, careful consideration should be made to the choice of statistical methods when performing differential abundance analyses, as it has been shown previously that different tools can identify varying sets of significant features.[54] Nevertheless, differential abundance methods can still provide useful insights into the association between gut microbiome function with host phenotype. For instance, a meta-analysis of ~2000 human gut metagenomes from case–control cohorts revealed that certain microbial pathways, such as lipopolysaccharide biosynthesis and iron transport, robustly distinguish diseased individuals from healthy controls.[55] These pathways were found to be more prevalent in individuals from multiple disease populations, including colorectal cancer, liver cirrhosis, Crohn's disease, type 2 diabetes and obesity. However, as mentioned previously, correlation does not equal causation. Therefore, additional experimentation is still needed to validate new disease-associated mechanisms and to gain insights into key host–microbe interactions.

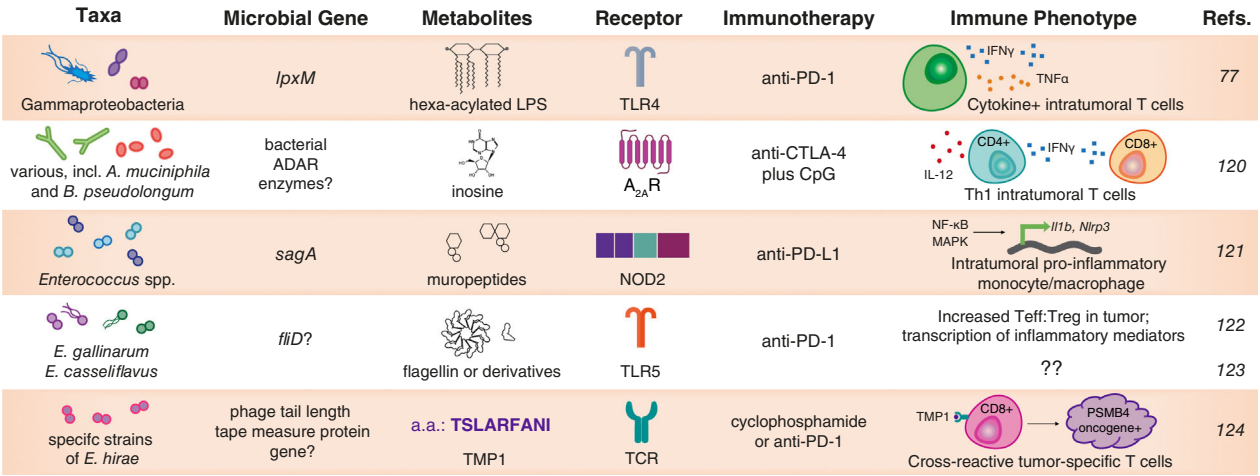
As metagenomic sequencing and functional annotation improve and become more broadly adopted, opportunities to identify mechanistic links between microbiota metabolites and functions and effects on host physiology will continue to expand. Rather than attempt to cover the wealth of recent studies investigating the role of the microbiome in various aspects of gut and systemic immune phenotypes, we will highlight a few key areas of microbiome–immune research that have been intensely investigated and have the potential to quickly benefit from the integration of functional metagenomics approaches.

OPPORTUNITIES FOR UNDERSTANDING MICROBIOME MODULATION OF IMMUNITY IN DISEASE

Perhaps the first area in which the gut microbiota was recognized to contribute to immunity was in the context of enteric infections. Soon after their introduction for the treatment of bacterial infections, antibiotics were observed to be associated with diarrhea and an increased incidence of pseudomembranous enterocolitis.[56] As early as the 1950s, antibiotic administration was demonstrated to increase susceptibility to intestinal pathogens such as *Vibrio cholerae*[57] and *Salmonella*.[58] While antibiotic-mediated depletion and germ-free animal models continue to demonstrate the importance of a replete microbiota in protection against intestinal infections, only a few of the likely many mechanisms have been explored in terms of specific microbiota metabolites and the immune pathways they modulate. Although antibiotic treatment has long been a known risk factor for *Clostridioides difficile* infection (CDI), only in the past 10 years was a specific microbial gene and function, production of secondary bile acids via the *baiCD*-encoded 7α-hydroxysteroid dehydrogenase enzyme, identified as a key mechanism of microbiota-mediated protection against CDI.[59] Likewise, the activity of the peptidoglycan hydrolase SagA from commensal *Enterococcus* has been characterized recently for its ability to generate unique muropeptides that stimulate host anti-microbial activity against *Salmonella* infection and CDI via the innate immune receptor nucleotide binding oligomerization domain containing 2 (NOD2).[60,61] In both cases, the identification of a bacterial gene and enzyme, along with their immunomodulatory products, enables broad, taxonomy-independent analyses of metagenomic datasets. Specifically, it allows quantitation of the abundance of the bacterial gene of interest, and its functional equivalents, in patient microbiomes and subsequent association with clinical phenotypes. Measurement of corresponding metabolites in biofluids or peripheral tissues could then be explored for further mechanistic insights.

Similarly, inflammatory bowel disease (IBD), an umbrella term that encompasses the chronic inflammatory diseases ulcerative colitis and Crohn's disease, has been extensively investigated for associations with patient microbiota composition. Numerous studies have shown differences between IBD patients and healthy controls in overall microbiota composition and diversity or the abundance of specific bacterial taxa.[62–65] For example, an early association was drawn between a decreased abundance of *Faecalibacterium prausnitzii* and an increased risk of recurrence in patients with Crohn's disease, and a secreted factor from *F. prausnitzii* was demonstrated to be anti-inflammatory.[66] *F. prausnitzii* is an abundant producer of the short-chain fatty acid (SCFA) butyrate, and subsequent work has gone on to show that both *F. prausnitzii* and other phylogenetically-diverse, butyrate-producing species are also reduced in patients with ulcerative colitis.[67,68] Microbiome SCFAs, including butyrate, have been demonstrated to cue intestinal epithelial cell production of TGF-β and differentiation of anti-inflammatory colonic $Foxp3^+$ regulatory $CD4^+$ T cells (Tregs) that are protective in mouse models of colitis.[69] However, recent work revealing high levels of inter- and intra-species variation in bacterial butyrate biosynthesis capacity[70] reinforces the need to evaluate microbiome SCFA production beyond the abundance of individual taxa such as *F. prausnitzii*. Although further multi-omic analyses have revealed a general reduction in SCFAs, especially butyrate, in IBD patients,[64] taxonomy-independent, systematic assessment of microbiome SCFA biosynthesis in IBD patients has been limited, primarily utilizing mass spectrometry-based metabolomics methods in small cohort studies. As metagenomic predictions of SCFA biosynthesis have recently been validated to correspond to SCFA abundance in patient stool samples, as measured by mass spectrometry,[71] functional metagenomics approaches may offer a more accessible proxy for evaluating SCFA levels and matching these to inflammatory versus anti-inflammatory immune phenotypes in larger prospective cohort studies.

In the context of cancer, several landmark studies revealed a negative impact of antibiotic usage on the efficacy of immune-dependent therapies[72,73] and identified specific alterations in the gut microbiota to be associated with the clinical response to immune checkpoint inhibitors (ICI).[74] Since then, numerous different taxa have been correlated with response or non-response to ICI therapy, but no true consensus has surfaced regarding specific beneficial species.[75] Several of studies have gone a step further to identify and validate microbial functions and metabolites that may contribute to immune activation and ICI efficacy



$A_{2A}R$, adenosine A2A receptor; ADAR, adenosine deaminase RNA-specific; CpG, CpG oligodeoxynucleotides; LPS, lipopolysaccharide; NOD2, nucleotide-binding oligomerization domain-containing 2; PSMB4, proteasome subunit beta type-4; TCR, T cell receptor; Teff, effector T cell; Treg, regulatory T cell; Th1, type 1 helper T cell; TLR4 or 5, toll-like receptor 4 or 5; TMP1, tape measure protein 1

Figure 3. A graphic table of microbiome metabolites that have been shown to impact cancer immunotherapy. In the studies referenced in *italics*, specific functions or metabolites from the gut microbiota have been demonstrated to impact anti-tumor immune responses in the context of immune checkpoint inhibitors or immunomodulatory chemotherapies.[77,120–124]

(Figure 3 and recently reviewed by Yang Q *et al.*[76]), opening up possibilities to leverage functional metagenomics approaches to globally assess these metabolites across patient cohorts in a taxonomy-independent manner. Functional metagenomics can also be utilized to uncover metabolites linked to responses in patients that can be further investigated in "reverse translation" studies using *in vitro* and *in vivo* models to gain a more thorough understanding of the immune mechanisms involved. For example, using functional metagenomic analyses across multiple patient cohorts and datasets, we recently uncovered a correlation between a structural variant of the bacterial cell wall component lipopolysaccharide (LPS) and the response to anti-PD-1 ICI therapy in melanoma patients.[77] A causal relationship was then established *in vitro* and in mouse tumor models, with enhanced toll-like receptor 4 (TLR4)-dependent stimulation of the NF-kB pathway in monocytes and macrophages and increases in intra-tumoral cytokine-secreting T cells, respectively.[77] Future functional metagenomic analyses may reveal additional biomarkers of responsiveness and potential therapeutic targets.

OPPORTUNITIES FOR UNDERSTANDING MICROBIOME MODULATION OF IMMUNE HOMEOSTASIS

Beyond their impact on the outcomes of perturbed or diseased states, gut-resident microbes are also increasingly appreciated to shape immune development[78] and to function under homeostatic conditions.[79] Although a vast majority of studies are correlative, utilizing broad-spectrum antibiotic treatment or germ-free mice to eliminate the microbiota altogether and/or to colonize with a singular or cohort of bacterial species, a number of microbial functions and metabolites have been mechanistically linked to the differentiation of specialized $CD4^+$ T cell subsets in the gut. For example, maturation of IL-17-secreting Th17 cells was shown to depend on intestinal epithelial cell (IEC) production of serum amyloid A (SAA) in response to sensing of adherent commensal bacterial and fungal species.[80,81] Subsequently, Clostridia and Proteobacteria were demonstrated to have opposing effects on IEC retinoic acid production,[82] and retinoic acid sensing in IECs was shown to regulate SAA production.[83] Microbiome-modulated IEC production of SAA has recently been further linked to pathogenic pro-inflammatory Th17 function.[84] Similarly, differentiation of colonic Tregs has also been shown to be driven by multiple cues from the gut microbiota. In 2010, polysaccharide A from commensal *Bacteroides fragilis* was shown to induce Treg differentiation in a toll-like receptor (TLR) 2-dependent manner,[85] and in 2013, three different studies demonstrated roles for SCFA-producing commensal bacteria in the differentiation and anti-inflammatory function of colonic Tregs.[69,86,87]

Both of these examples illustrate how dynamic and plastic immune cell phenotypes can be in the gut, with

multiple region-specific microbial functions and metabolites, combined with host cytokine cues, likely fine-tuning immune responses *in situ* to suit different contexts. Recent work has begun to explore how the specific microbial environment of the less-studied small intestine may also play a critical role in balancing inflammatory and anti-inflammatory responses at steady state and in the context of disease. While most studies again focus on presence or absence of specific taxa or the gut microbiome in its entirety, a few have examined microbial metabolites of interest and their mechanisms of interaction with host immune cells. In the case of small intestine intraepithelial lymphocytes (IELs), stimulation of the aryl hydrocarbon receptor (AhR) by dietary and microbial ligands has long been shown to be required for the maintenance of IEL numbers,[88] but a recent study highlights the need to regulate AhR activity for optimal IEL survival and function.[89] Isomerization of linoleic acid by numerous small intestine-inhabiting bacterial species has likewise recently been demonstrated to contribute to IEL homeostasis.[90] By leveraging functional approaches alongside increased metagenomic sequencing and immune phenotyping of different regions of the gut, future studies will likely reveal additional site-specific mechanisms of microbiome immune modulation.

In addition, the immune receptors and mechanisms for host sensing of different microbial products, in the form of pattern recognition receptor (PRR) sensing of microbe-associated molecular patterns (MAMPs), have largely been characterized in the context of microbial pathogens, likely leaving many pathways that contribute to immune homeostasis yet to be discovered. For example, NOD2, a cytosolic PRR, was initially described for its structural similarity to NOD1 and plant R gene products,[91] which play a role in cellular responses to pathogens.[92] The ligand for NOD2 was subsequently identified to be muramyl dipeptide (MDP), a component released from peptidoglycan in bacterial cell walls by enzymes such as lysozyme.[93,94] NOD2 has since been demonstrated to influence the composition of the intestinal microbiota,[95,96] and recent structural and biochemical approaches have revealed that smaller muropeptides produced by commensal bacteria also stimulate NOD2.[61] As multi-omic and high-throughput methods improve, some recent studies have employed more systematic approaches to identifying the host factors required for the detection and distinction of different MAMPs such as LPS,[97] and host responses to different variants of LPS from the microbiome have been shown to influence autoimmune diabetes.[98] Combined with functional metagenomics methods, these approaches could identify novel microbial metabolite–host receptor pairings and provide powerful mechanistic insights into how changes in commensal microbial communities modulate host immune development and homeostasis.

CURRENT CHALLENGES AND LIMITATIONS TO FUNCTIONAL METAGENOMIC ANALYSES

The main challenge in functional metagenomic analyses is understanding how to translate functional predictions into mechanistic hypotheses and accurately link genotype to phenotype. For instance, issues with misannotations, which are particularly problematic when analyzing short sequence fragments, can severely affect any resulting biological interpretations. In addition, the identification of functional genes through metagenomics does not necessarily equate to their expression or activity within the microbial community. Genes may also be present in the genome but remain dormant or inactive under specific environmental conditions. Therefore, functional annotation of metagenomic data may not always reflect the active metabolic pathways or functions occurring within the community and need to be complemented with additional multi-omics approaches, such as metaproteomics, metatranscriptomics and metabolomics. However, the analysis of multi-omics datasets in the context of the human gut microbiome is still an active area of development, as there is currently a lack of tools and standardized bioinformatics approaches to effectively integrate different data types.

Another major limitation in the field is that functional metagenomic data generally require predicting gene functions based on sequence homology to known genes. However, many microbial genes have no known homologs or functional annotations in existing databases, which can introduce potential biases towards well-characterized functions. For instance, a common notion inferred from metagenomic studies is that there is a large degree of functional redundancy within the human microbiome, as the functional diversity is found to be more consistent between individuals than their respective taxonomic compositions.[99] However, this has been argued to represent to some extent a technical artefact,[100] as functional metagenomic studies generally discard all unknown/uncharacterized genes. This is particularly problematic for the human microbiome field, as 40% of genes detected in gut metagenomes are estimated to lack any meaningful annotations.[37] To circumvent this issue, new computational approaches leveraging protein-structure predictions (e.g. AlphaFold2[101]), genomic context and remote homology are being actively developed to obtain further insights into this unknown, hidden functional diversity. Recently, two large-scale studies[102,103] generated hundreds of thousands of newly

discovered protein families from a variety of habitats. However, only 15% and 7% of the unknown protein families could be annotated using structural similarity and genomic-context information, respectively. Therefore, despite computational advancements, there is a pressing need to further improve the biological insights that can be derived from existing sequence data, which may ultimately only be resolved using targeted, hypothesis-driven experimental approaches.

FUTURE PERSPECTIVES

As high-throughput and computational methods begin to more efficiently yield accurate predictions of host factors and mechanisms in patient and experimental datasets, functional interrogation of corresponding microbiomes constitutes an exceptional opportunity to uncover key links between microbial metabolites and effects on host physiology. Compared with more challenging and costly methods, such as scRNA-seq of patient samples, microbiome sample collection and preservation for shotgun metagenomic sequencing have become progressively more streamlined and affordable. As such, opportunities to build microbiome biobanks and to conduct longitudinal sampling from large patient and population cohorts are increasingly being explored.[104,105] For instance, a large prospective cohort with ~18 years of follow-up clinical records recently showed that integrating metagenomic information about the gut microbiome improved risk prediction for diseases such as type 2 diabetes and prostate cancer.[106] Such efforts will in turn provide resources from which biological mechanisms can be explored and computational applications can be refined to improve functional microbiome profiling.

Moving forward, expanding the isolation and culture of commensal microbes represents another important potential avenue for future development. A substantial percentage of both mouse and human gut microbiotas remains uncultured,[107] limiting the ability to experimentally measure specific microbial functions and their effects on the host. Using cultured microbial isolates for whole-genome sequencing, comparative genomics, metabolomics and other phenotypic assays would enable improved functional annotation of microbial genes. In addition, while multi-omic approaches can be exceptional for hypothesis generation and validation, experimental testing is required to move beyond correlative analyses to causal relationships. More comprehensive culture collections of commensal gut microbes will enable both functional characterization of novel microbial species and mechanistic experimental investigations of associated immune phenotypes. Importantly, these culture collections must be sampled from diverse human populations, geographies and diets to capture a wider degree of microbial diversity and functions.

Of note, although this review focuses primarily on the larger-biomass, bacterial component of the microbiota, the viral and fungal contributions to microbiome functions are also becoming more accessible via shotgun metagenomic sequencing tools and datasets.[108–110] Likewise, while T cell-centered studies have dominated much of the mechanistic investigation of microbiome–immune interactions, functional metagenomics approaches hold a great deal of promise for defining the effects of the commensal microbiota on a multitude of immune and non-immune cell types and functions. Integrated multi-omic approaches and computational analyses of both hosts and microbiomes will provide unprecedented granularity for deciphering the mechanisms of microbiome–immune interactions in both disease and homeostasis and ultimately uncover potential strategies for modulating and optimizing these interactions.

ACKNOWLEDGMENTS

PS is supported in part by a project grant from the Rosetrees Trust [PGS22/100116], AA is supported by a Career Development Award by the Medical Research Council [MR/W016184/1] and VAP was supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society [206245/Z/17/Z] and is currently funded by the Wellcome Trust [302351/Z/23/Z]. We thank B Beresford-Jones for helpful discussions in the initial stages of manuscript preparation.

AUTHOR CONTRIBUTIONS

Puspendu Sardar: Conceptualization; visualization; writing – original draft; writing – review and editing. **Alexandre Almeida:** Conceptualization; funding acquisition; visualization; writing – original draft; writing – review and editing. **Virginia A Pedicord:** Conceptualization; funding acquisition; supervision; visualization; writing – original draft; writing – review and editing.

CONFLICT OF INTEREST

The authors confirm that there are no conflicts of interest to disclose.

REFERENCES

1. de Vos WM, Tilg H, Van Hul M, Cani PD. Gut microbiome and health: mechanistic insights. *Gut* 2022; **71**: 1020–1032.
2. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature* 2007; **449**: 804–810.

3. Qin J, Li R, Raes J, *et al*. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010; **464**: 59–65.
4. Gutierrez Lopez DE, Lashinger LM, Weinstock GM, Bray MS. Circadian rhythms and the gut microbiome synchronize the host's metabolic response to diet. *Cell Metab* 2021; **33**: 873–887.
5. Rothschild D, Weissbrod O, Barkan E, *et al*. Environment dominates over host genetics in shaping human gut microbiota. *Nature* 2018; **555**: 210–215.
6. Rowland I, Gibson G, Heinken A, *et al*. Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr* 2018; **57**: 1–24.
7. Khan I, Bai Y, Zha L, *et al*. Mechanism of the gut microbiota colonization resistance and enteric pathogen infection. *Front Cell Infect Microbiol* 2021; **11**: 716299.
8. Mallott EK, Amato KR. Host specificity of the gut microbiome. *Nat Rev Microbiol* 2021; **19**: 639–653.
9. Ley RE, Hamady M, Lozupone C, *et al*. Evolution of mammals and their gut microbes. *Science* 2008; **320**: 1647–1651.
10. Youngblut ND, Reischer GH, Walters W, *et al*. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nat Commun* 2019; **10**: 2200.
11. Groussin M, Mazel F, Sanders JG, *et al*. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat Commun* 2017; **8**: 14319.
12. Desai MS, Seekatz AM, Koropatkin NM, *et al*. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* 2016; **167**: 1339–1353.e1321.
13. Porcari S, Benech N, Valles-Colomer M, *et al*. Key determinants of success in fecal microbiota transplantation: from microbiome to clinic. *Cell Host Microbe* 2023; **31**: 712–733.
14. Littmann ER, Lee JJ, Denny JE, *et al*. Host immunity modulates the efficacy of microbiota transplantation for treatment of *Clostridioides difficile* infection. *Nat Commun* 2021; **12**: 755.
15. Lockhart A, Mucida D, Bilate AM. Intraepithelial lymphocytes of the intestine. *Annu Rev Immunol* 2024; **42**: 289–316.
16. Klaenhammer TR. Bacteriocins of lactic acid bacteria. *Biochimie* 1988; **70**: 337–349.
17. Brown JP. Role of gut bacterial flora in nutrition and health: a review of recent advances in bacteriological techniques, metabolism, and factors affecting flora composition. *CRC Crit Rev Food Sci Nutr* 1977; **8**: 229–336.
18. Atarashi K, Tanoue T, Shima T, *et al*. Induction of colonic regulatory T cells by indigenous clostridium species. *Science* 2011; **331**: 337–341.
19. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 1977; **74**: 5088–5090.
20. Human Microbiome Project C. Structure. Function and diversity of the healthy human microbiome. *Nature* 2012; **486**: 207–214.
21. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006; **444**: 1027–1031.
22. Wang T, Cai G, Qiu Y, *et al*. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J* 2012; **6**: 320–329.
23. Scheperjans F, Aho V, Pereira PA, *et al*. Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov Disord* 2015; **30**: 350–358.
24. Thompson LR, Sanders JG, McDonald D, *et al*. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 2017; **551**: 457–463.
25. Golob JL, Margolis E, Hoffman NG, Fredricks DN. Evaluating the accuracy of amplicon-based microbiome computational pipelines on simulated human gut microbial communities. *BMC Bioinformatics* 2017; **18**: 283.
26. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* 2017; **8**: 1784.
27. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 2017; **35**: 833–844.
28. Perez-Cobas AE, Gomez-Valero L, Buchrieser C. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microb Genom* 2020; **6**: mgen000409.
29. Tyson GW, Chapman J, Hugenholtz P, *et al*. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 2004; **428**: 37–43.
30. Parks DH, Rinke C, Chuvochina M, *et al*. Recovery of nearly 8000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2017; **2**: 1533–1542.
31. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. Compendium of 4941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol* 2019; **37**: 953–961.
32. Chen C, Zhou Y, Fu H, *et al*. Expanded catalog of microbial genes and metagenome-assembled genomes from the pig gut microbiome. *Nat Commun* 2021; **12**: 1106.
33. Beresford-Jones BS, Forster SC, Stares MD, *et al*. The mouse gastrointestinal bacteria catalogue enables translation between the mouse and human gut microbiotas via functional mapping. *Cell Host Microbe* 2022; **30**: 124–138.e128.
34. Pasolli E, Asnicar F, Manara S, *et al*. Extensive unexplored human microbiome diversity revealed by over 150 000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 2019; **176**: 649–662.e620.
35. Almeida A, Mitchell AL, Boland M, *et al*. A new genomic blueprint of the human gut microbiota. *Nature* 2019; **568**: 499–504.

36. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. *Nature* 2019; **568**: 505–510.
37. Almeida A, Nayfach S, Boland M, *et al*. A unified catalog of 204 938 reference genomes from the human gut microbiome. *Nat Biotechnol* 2021; **39**: 105–114.
38. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol* 2020; **38**: 701–707.
39. Beghini F, McIver LJ, Blanco-Miguez A, *et al*. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* 2021; **10**: e65088.
40. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 119.
41. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom* 2021; **7**: 685.
42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.
43. Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998; **14**: 755–763.
44. Huerta-Cepas J, Szklarczyk D, Heller D, *et al*. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019; **47**: D309–D314.
45. Jones P, Binns D, Chang HY, *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014; **30**: 1236–1240.
46. The UniProt C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2017; **45**: D158–D169.
47. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017; **45**: D353–D361.
48. Pascal Andreu V, Augustijn HE, Chen L, *et al*. gutSMASH predicts specialized primary metabolic pathways from the human gut microbiota. *Nat Biotechnol* 2023; **41**: 1416–1423.
49. Sugimoto Y, Camacho FR, Wang S, *et al*. A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. *Science* 2019; **366**: eaax9176.
50. Blin K, Shaw S, Steinke K, *et al*. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* 2019; **47**: W81–W87.
51. Feldgarden M, Brover V, Gonzalez-Escalona N, *et al*. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* 2021; **11**: 12728.
52. Alcock BP, Raphenya AR, Lau TTY, *et al*. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020; **48**: D517–D525.
53. Zhang H, Yohe T, Huang L, *et al*. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 2018; **46**: W95–W101.
54. Nearing JT, Douglas GM, Hayes MG, *et al*. Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun* 2022; **13**: 342.
55. Armour CR, Nayfach S, Pollard KS, Sharpton TJ. A metagenomic meta-analysis reveals functional signatures of health and disease in the human gut microbiome. *mSystems* 2019; **4**: e00332-18.
56. Fekety FR Jr. Gastrointestinal complications of antibiotic therapy. *JAMA* 1968; **203**: 210–212.
57. Freter R. The fatal enteric cholera infection in the Guinea pig, achieved by inhibition of normal enteric flora. *J Infect Dis* 1955; **97**: 57–65.
58. Bohnhoff M, Drake BL, Miller CP. Effect of streptomycin on susceptibility of intestinal tract to experimental salmonella infection. *Proc Soc Exp Biol Med* 1954; **86**: 132–137.
59. Buffie CG, Bucci V, Stein RR, *et al*. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* 2015; **517**: 205–208.
60. Pedicord VA, Lockhart AAK, Rangan KJ, *et al*. Exploiting a host-commensal interaction to promote intestinal barrier function and enteric pathogen tolerance. *Sci Immunol* 2016; **1**: eaai7732.
61. Kim B, Wang YC, Hespen CW, *et al*. *Enterococcus faecium* secreted antigen a generates muropeptides to enhance host immunity and limit bacterial pathogenesis. *Elife* 2019; **8**: e45343.
62. Lee M, Chang EB. Inflammatory bowel diseases (IBD) and the microbiome-searching the crime scene for clues. *Gastroenterology* 2021; **160**: 524–537.
63. Ning L, Zhou YL, Sun H, *et al*. Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts. *Nat Commun* 2023; **14**: 7135.
64. Lloyd-Price J, Arze C, Ananthakrishnan AN, *et al*. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 2019; **569**: 655–662.
65. Gevers D, Kugathasan S, Denson LA, *et al*. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014; **15**: 382–392.
66. Sokol H, Pigneur B, Watterlot L, *et al*. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci USA* 2008; **105**: 16731–16736.
67. Sokol H, Seksik P, Furet JP, *et al*. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm Bowel Dis* 2009; **15**: 1183–1189.
68. Machiels K, Joossens M, Sabino J, *et al*. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* 2014; **63**: 1275–1283.
69. Atarashi K, Tanoue T, Oshima K, *et al*. Treg induction by a rationally selected mixture of clostridia strains from the human microbiota. *Nature* 2013; **500**: 232–236.

70. Sorbara MT, Littmann ER, Fontana E, *et al.* Functional and genomic variation between human-derived isolates of Lachnospiraceae reveals inter- and intra-species diversity. *Cell Host Microbe* 2020; **28**: 134–146.e134.
71. Nogal A, Asnicar F, Vijay A, *et al.* Genetic and gut microbiome determinants of SCFA circulating and fecal levels, postprandial responses and links to chronic and acute inflammation. *Gut Microbes* 2023; **15**: 2240050.
72. Iida N, Dzutsev A, Stewart CA, *et al.* Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* 2013; **342**: 967–970.
73. Viaud S, Saccheri F, Mignot G, *et al.* The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science* 2013; **342**: 971–976.
74. Sivan A, Corrales L, Hubert N, *et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* 2015; **350**: 1084–1089.
75. Oh B, Boyle F, Pavlakis N, *et al.* The gut microbiome and cancer immunotherapy: can we use the gut microbiome as a predictive biomarker for clinical response in cancer immunotherapy? *Cancers (Basel)* 2021; **13**: 4824.
76. Yang Q, Wang B, Zheng Q, *et al.* A review of gut microbiota-derived metabolites in tumor progression and cancer therapy. *Adv Sci (Weinh)* 2023; **10**: e2207366.
77. Beresford-Jones BS, Sardar P, Xia W, *et al.* Hexa-acylated lipopolysaccharides from the gut microbiota enhance cancer immunotherapy responses. *bioRxiv* 2024; 2024.2006.2024.600291.
78. Donald K, Finlay BB. Early-life interactions between the microbiota and immune system: impact on immune system development and atopic disease. *Nat Rev Immunol* 2023; **23**: 735–748.
79. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. *Cell Res* 2020; **30**: 492–506.
80. Atarashi K, Tanoue T, Ando M, *et al.* Th17 cell induction by adhesion of microbes to intestinal epithelial cells. *Cell* 2015; **163**: 367–380.
81. Sano T, Huang W, Hall JA, *et al.* An IL-23R/IL-22 circuit regulates epithelial serum amyloid a to promote local effector Th17 responses. *Cell* 2015; **163**: 381–393.
82. Grizotte-Lake M, Zhong G, Duncan K, *et al.* Commensals suppress intestinal epithelial cell retinoic acid synthesis to regulate Interleukin-22 activity and prevent microbial dysbiosis. *Immunity* 2018; **49**: 1103–1115.e1106.
83. Gattu S, Bang YJ, Pendse M, *et al.* Epithelial retinoic acid receptor beta regulates serum amyloid a expression and vitamin A-dependent intestinal immunity. *Proc Natl Acad Sci USA* 2019; **116**: 10911–10916.
84. Lee JY, Hall JA, Kroehling L, *et al.* Serum amyloid a proteins induce pathogenic Th17 cells and promote inflammatory disease. *Cell* 2020; **180**: 79–91.e16.
85. Round JL, Mazmanian SK. Inducible Foxp3^{+} regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci USA* 2010; **107**: 12204–12209.
86. Smith PM, Howitt MR, Panikov N, *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* 2013; **341**: 569–573.
87. Arpaia N, Campbell C, Fan X, *et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* 2013; **504**: 451–455.
88. Li Y, Innocentin S, Withers DR, *et al.* Exogenous stimuli maintain intraepithelial lymphocytes via aryl hydrocarbon receptor activation. *Cell* 2011; **147**: 629–640.
89. Panda SK, Peng V, Sudan R, *et al.* Repression of the aryl-hydrocarbon receptor prevents oxidative stress and ferroptosis of intestinal intraepithelial lymphocytes. *Immunity* 2023; **56**: 797–812.e794.
90. Song X, Zhang H, Zhang Y, *et al.* Gut microbial fatty acid isomerization modulates intraepithelial T cells. *Nature* 2023; **619**: 837–843.
91. Ogura Y, Inohara N, Benito A, Chen FF, Yamaoka S, Nunez G. Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. *J Biol Chem* 2001; **276**: 4812–4818.
92. DeYoung BJ, Innes RW. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat Immunol* 2006; **7**: 1243–1249.
93. Inohara N, Ogura Y, Fontalba A, *et al.* Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J Biol Chem* 2003; **278**: 5509–5512.
94. Girardin SE, Boneca IG, Viala J, *et al.* Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem* 2003; **278**: 8869–8872.
95. Petnicki-Ocwieja T, Hrncir T, Liu YJ, *et al.* Nod2 is required for the regulation of commensal microbiota in the intestine. *Proc Natl Acad Sci USA* 2009; **106**: 15813–15818.
96. Ramanan D, Tang MS, Bowcutt R, Loke P, Cadwell K. Bacterial sensor Nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal *Bacteroides vulgatus*. *Immunity* 2014; **41**: 311–324.
97. Jost M, Jacobson AN, Hussmann JA, Cirolia G, Fischbach MA, Weissman JS. CRISPR-based functional genomics in human dendritic cells. *Elife* 2021; **10**: e65856.
98. Vatanen T, Kostic AD, d'Hennezel E, *et al.* Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* 2016; **165**: 842–853.
99. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature* 2012; **489**: 220–230.
100. Walker AW, Hoyles L. Human microbiome myths and misconceptions. *Nat Microbiol* 2023; **8**: 1392–1396.
101. Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021; **596**: 583–589.
102. Pavlopoulos GA, Baltoumas FA, Liu S, *et al.* Unraveling the functional dark matter through global metagenomics. *Nature* 2023; **622**: 594–602.
103. Rodriguez Del Rio A, Giner-Lamia J, Cantalapiedra CP, *et al.* Functional and evolutionary significance of unknown genes from uncultivated taxa. *Nature* 2024; **626**: 377–384.

104. Ryan MJ, Schloter M, Berg G, *et al.* Development of microbiome biobanks - challenges and opportunities. *Trends Microbiol* 2021; **29**: 89–92.
105. Tigchelaar EF, Zhernakova A, Dekens JA, *et al.* Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* 2015; **5**: e006772.
106. Liu Y, Ritchie SC, Teo SM, *et al.* Integration of polygenic and gut metagenomic risk prediction for common diseases. *Nat Aging* 2024; **4**: 584–594.
107. Bellali S, Lagier JC, Million M, *et al.* Running after ghosts: are dead bacteria the dark matter of the human gut microbiota? *Gut Microbes* 2021; **13**: 1–12.
108. Saary P, Mitchell AL, Finn RD. Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. *Genome Biol* 2020; **21**: 244.
109. Nayfach S, Paez-Espino D, Call L, *et al.* Metagenomic compendium of 189 680 DNA viruses from the human gut microbiome. *Nat Microbiol* 2021; **6**: 960–970.
110. Zeng S, Almeida A, Li S, *et al.* A metagenomic catalog of the early-life human gut virome. *Nat Commun* 2024; **15**: 1864.
111. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015; **31**: 1674–1676.
112. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017; **27**: 824–834.
113. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018; **6**: 158.
114. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res* 2023; **51**: W115–W121.
115. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 2021; **18**: 366–368.
116. Tamames J, Puente-Sanchez F. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front Microbiol* 2018; **9**: 3349.
117. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011; **7**: e1002195.
118. Caspi R, Altman T, Dreher K, *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 2012; **40**: D742–D753.
119. Darzi Y, Falony G, Vieira-Silva S, Raes J. Towards biome-specific analysis of meta-omics data. *ISME J* 2016; **10**: 1025–1028.
120. Mager LF, Burkhard R, Pett N, *et al.* Microbiome-derived inosine modulates response to checkpoint inhibitor immunotherapy. *Science* 2020; **369**: 1481–1489.
121. Griffin ME, Espinosa J, Becker JL, *et al.* *Enterococcus* peptidoglycan remodeling promotes checkpoint inhibitor cancer immunotherapy. *Science* 2021; **373**: 1040–1046.
122. Stevenson A, Panzica A, Holt A, *et al.* Host–microbe interactions mediating antitumorigenic effects of MRX0518, a gut microbiota-derived bacterial strain, in breast, renal and lung carcinoma. *J Clin Oncol* 2018; **36**: e15006.
123. Okumura R, Mitsunobu H, Ishii C, Miyazaki S, Tamiya T, Bernier F. Base-edited live microbiome therapeutics with target-aid -live biotherapeutic products (LBPS) with synergistic effects of anti-PD-1 antibodies in a syngeneic model of mice. *J Clin Oncol* 2023; **41**: e15104. doi:10.1200/JCO.2023.41.16_suppl.e15104
124. Fluckiger A, Daillere R, Sassi M, *et al.* Cross-reactivity between tumor MHC class I-restricted antigens and an enterococcal bacteriophage. *Science* 2020; **369**: 936–942.