



OPEN ACCESS

EDITED BY

Sona Vasudevan,
Georgetown University, United States

REVIEWED BY

Fabio Gervasi,
Council for Agricultural and Economics
Research (CREA), Italy

*CORRESPONDENCE

Jingyue Wu,
✉ jingyue.wu@gwu.edu
Raja Mazumder,
✉ mazumder@gwu.edu

RECEIVED 15 November 2023

ACCEPTED 27 December 2023

PUBLISHED 19 January 2024

CITATION

Wu J, Singleton SS, Bhuiyan U, Krammer L and
Mazumder R (2024), Multi-omics approaches to
studying gastrointestinal microbiome in the
context of precision medicine and
machine learning.

Front. Mol. Biosci. 10:1337373.

doi: 10.3389/fmolb.2023.1337373

COPYRIGHT

© 2024 Wu, Singleton, Bhuiyan, Krammer and
Mazumder. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Multi-omics approaches to studying gastrointestinal microbiome in the context of precision medicine and machine learning

Jingyue Wu^{1*}, Stephanie S. Singleton¹, Urnisha Bhuiyan¹,
Lori Krammer^{1,2} and Raja Mazumder^{1,3*}

¹Department of Biochemistry and Molecular Medicine, School of Medicine and Health Sciences, The George Washington University, Washington, DC, United States, ²Milken Institute School of Public Health, The George Washington University, Washington, DC, United States, ³The McCormick Genomic and Proteomic Center, The George Washington University, Washington, DC, United States

The human gastrointestinal (gut) microbiome plays a critical role in maintaining host health and has been increasingly recognized as an important factor in precision medicine. High-throughput sequencing technologies have revolutionized -omics data generation, facilitating the characterization of the human gut microbiome with exceptional resolution. The analysis of various -omics data, including metatranscriptomics, metagenomics, glycomics, and metabolomics, holds potential for personalized therapies by revealing information about functional genes, microbial composition, glycans, and metabolites. This multi-omics approach has not only provided insights into the role of the gut microbiome in various diseases but has also facilitated the identification of microbial biomarkers for diagnosis, prognosis, and treatment. Machine learning algorithms have emerged as powerful tools for extracting meaningful insights from complex datasets, and more recently have been applied to metagenomics data via efficiently identifying microbial signatures, predicting disease states, and determining potential therapeutic targets. Despite these rapid advancements, several challenges remain, such as key knowledge gaps, algorithm selection, and bioinformatics software parametrization. In this mini-review, our primary focus is metagenomics, while recognizing that other -omics can enhance our understanding of the functional diversity of organisms and how they interact with the host. We aim to explore the current intersection of multi-omics, precision medicine, and machine learning in advancing our understanding of the gut microbiome. A multidisciplinary approach holds promise for improving patient outcomes in the era of precision medicine, as we unravel the intricate interactions between the microbiome and human health.

KEYWORDS

precision medicine, machine learning, gut microbiome, metagenomics, multi-omics, sequencing, biomarkers

Introduction

The human body hosts diverse communities of microbes and encompasses various glycomes, fostering diverse forms of communication with our organs and facilitating metabolic functions and molecular signals to maintain proper health (Kavanaugh et al., 2015; de Vos et al., 2022). The advent of high-throughput sequencing (HTS) along with research initiatives, such as the Human Microbiome Projects (iHMP Research Network Consortium, 2019), have paved a new path for microbial community characterization. The implementation of HTS strategies has equipped microbial taxonomic composition profiling at nearly any given body site, propelling the study of microbial networks, microbiome-disease associations, and host-microbiota interactions (Clooney et al., 2016). In parallel, the field of precision medicine has gained momentum, aiming to provide personalized healthcare solutions tailored to an individual's unique genetic makeup, lifestyle, and environmental exposures (Schork, 2015). As we stand at the intersection of human microbiome research and healthcare innovation, there is a growing recognition that the gut microbiome and its exploration using -omics technologies hold immense potential as a key player in achieving the goals of precision medicine.

Despite the prospect of the gut microbiome and 'omics' data used in support of precision medicine, the sheer complexity and large influx of these datasets pose formidable challenges to data interpretation and analysis. Hence, researchers have expanded their focus into the realms of bioinformatics and machine learning (ML) to address these challenges. This is done by utilizing the capacity of the aforementioned disciplines to integrate and process extensive data through different algorithms, enabling the development of models that can aid in diagnostic, prognostic, and therapeutic interventions. Harnessing these techniques enables the comprehensive analysis of intricate layers of biological information ranging from metagenomics to metabolomics, and the integration of patient record data, shedding light on the role of the gut microbiome in different aspects of precision medicine. This holistic approach ultimately improves the health trajectories of

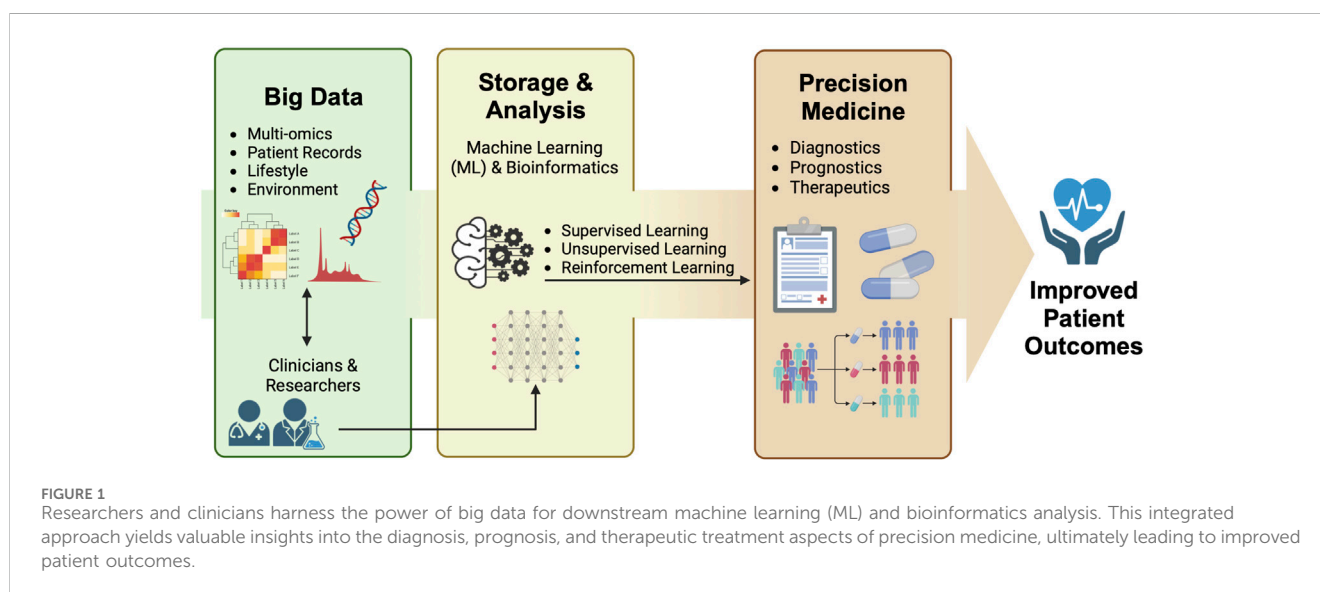
patients (Figure 1). This mini-review discusses the current status and interface between ML and bioinformatics methods of analyzing multi-omics data in advancing the understanding of the gut microbiome in relation to precision medicine.

Methods for microbiome data analysis

The continuous evolution of genomic sequencing through multiple generations of sequencing technologies has resulted in our ability to determine the abundance of microorganisms at a granular level. To date, scientific research has primarily focused on characterizing bacterial species. However, clinical research has expanded to include other microorganisms in the gut, such as viruses, fungi, and helminths (Caporaso et al., 2012; Mukhopadhyaya et al., 2019; Berg et al., 2020; Rubel et al., 2020; Zhang et al., 2022). Microbiome investigation methods have expanded from 16S rRNA gene sequencing, whole-genome shotgun metagenomic sequencing, and RNA (metatranscriptomics) sequencing (Reuter et al., 2015). Here, we outline common sequencing platforms, methods, and databases used to investigate the gut microbiome.

Sequencing technologies

Common second-generation sequencing (SGS) platforms and methods include Illumina, Ion Torrent, 454 Pyrosequencing, and SOLiD Sequencing (Liu et al., 2012; Meslier et al., 2022), where Illumina's sequencing platform became more widely used for microbiome research due to their high-throughput processing, quality and consistency, cost-effectiveness, and relevant capabilities for microbiome research (Malla et al., 2018; Caporaso et al., 2012). SGS platforms generate short-read data, typically ranging from ~50 to 300 base pairs in length and varying in sequencing depth (Ranjan et al., 2016; Johnson et al., 2019). Common third-generation sequencing (TGS) platforms include Helicos Single Molecule Sequencing (SMS), Pacific Biosciences (PacBio), and Oxford Nanopore Technologies (ONT) (Schadt et al., 2010; Koren and Phillippy, 2015; Athanasopoulou et al.,



2021). SMS enables single-molecule detection without the need for amplification and was one of the early TGS technologies that had potential applications in studying complex microbial communities (Pushkarev et al., 2009). PacBio sequencing is a single-molecule, real-time (SMRT) sequencing technology that offers long-read sequence technology, where long-read data can range from 1,000–20,000 bases or more (Udaondo et al., 2021). This feature is beneficial for resolving complex microbial communities and detecting novel microorganisms. PacBio's Sequel and Sequel II systems have been utilized in various metagenomic studies. ONT sequencing is based on nanopore technology, where DNA molecules pass through a nanopore, generating electrical signals that correspond to the nucleotide sequence. The MinION and PromethION devices from ONT have been used for real-time sequencing of microbial DNA, providing long-read lengths, portability, and potential for direct RNA sequencing. ONT sequencing also produces long-read data, where the read lengths can vary from several thousand base pairs to over a hundred thousand base pairs (Jain et al., 2016; Schmidt et al., 2017).

Current paradigm: hybrid sequencing and bioinformatics

Combining short-read and long-read data has the capacity to characterize more complete and accurate microbial genomes present in the gut microbiome. Hence, a hybrid sequencing approach leveraging the strengths of both second and third-generation technologies has been increasingly used in gut microbiome research (Bharti and Grimm, 2021; Chen et al., 2022; Jin et al., 2022). To begin with, SGS is used to generate short-read sequences from gut microbiome samples. Whole genome short-read data is particularly useful for identifying the taxonomic composition of the microbiome, as short reads can accurately distinguish among microbial species. Long-read sequences from the same samples can then be generated, as they are crucial for resolving complex genomic regions, such as repetitive elements and structural variations, which are often missed or misassembled with short-read data (Mantere et al., 2019). Synergistically, short-read and long-read data can also be combined during the genome assembly process where long reads help anchor and scaffold the assembly, ensuring that the contigs represent more accurate and contiguous genome fragments. The short reads provide accurate base-level information to correct errors in the long reads and improve the overall accuracy of the assembly (Amarasinghe et al., 2020).

Study design plays a crucial role not only in selecting sequencing platforms or sequencing types but also in analyzing the vast amounts of data generated by various sequencing technologies in microbiome research. It is important to consider capturing extensive dietary, medical history, and patient lifestyle data in tandem with microbiome sampling and sequencing. This will allow a more comprehensive view of a disease state and potential patient outcomes, particularly when applying machine learning tools in a precise use case—as discussed later in the review. Relevant analyses include taxonomic profiling, phylogenetic analysis, metagenome or genome assembly, gene prediction, functional annotation, comparative metagenomics, pathway and differential abundance analysis, metatranscriptomics analysis, metabolomics, and network

analysis (Prakash and Taylor, 2012; Maranga et al., 2023). Taxonomic profiling is one of the most important starting points for gut microbiome research; this is achieved using tools such as QIIME/QIIME2 (Quantitative Insights Into Microbial Ecology), mothur, Kraken/Kraken2 (Schloss et al., 2009; Caporaso et al., 2010; Wood and Salzberg, 2014; Bolyen et al., 2019; Wood et al., 2019; Lu and Salzberg, 2020; Schloss, 2020).

Databases and algorithms

Numerous databases are essential for identifying organisms found within the gut microbiome and determining their taxonomic composition via diverse sequencing techniques. Commonly used repositories include NCBI's non-redundant nucleotide database (NT), filtered NT, and Greengenes (McDonald et al., 2012; Shamsaddini et al., 2014). The NT is the most comprehensive sequence database and contains sequences from GenBank, RefSeq, PDB, and more, while the filtered NT is an expansion of RefSeq which contains organisms from NT whose phylogenetic lineage is clearly defined thus removing spurious and artificial sequences (Pruitt et al., 2005). While the former two databases consolidate information regarding an organism's genes and genome, Greengenes leverages full-length 16S rRNA or metatranscriptomics data from public databases that fulfill several filtering qualifications (McDonald et al., 2012). The utilization of 16S rRNA presents a cost-effective approach, as it focuses on the identification of organisms through key marker genes originating from conserved genetic regions (Balvociute and Huson, 2017). Occasionally, this technique exhibits amplification bias to some extent and has limitations in the identification of archaea and viruses, as well as overall organism resolution, in contrast to metagenomics techniques such as random shotgun sequencing and WGS (Scholz et al., 2016).

Sequence aligners, especially established ones such as BLAST and Bowtie/Bowtie2, have been extensively reviewed (Langmead and Salzberg, 2012; Alser et al., 2021). Typically, BLAST is used for a handful of reads which is subsampled to search for similarities in a large database such as NT (Shamsaddini et al., 2014), while Bowtie and other similar rapid aligners are typically used to map reads to a set of reference genomes (Langmead, 2010). BLAST utilizes local alignment techniques to identify regions of similarity that have high-scoring pairs. This is exemplified by BLAST tools such as BLASTP, BLASTN, and BLASTX which function to identify similar pairs based on protein, nucleotide, and translated nucleotide sequences, respectively (Yang et al., 2014; Altschul et al., 1990; Chen et al., 2015; Li and Lu, 2019). This differs from Bowtie, which aligns millions of short reads from a sample of interest to limited number of references quickly. In light of this information, it is imperative for researchers to fully grasp the bioinformatics task at hand when deciding whether to employ BLAST or a short-read aligner.

ML in gut microbiome for precision medicine: current research

Unlike traditional computer programming, machine learning (ML) algorithms are not explicitly programmed with rules and instructions. Instead, they autonomously learn patterns and

relationships from data, allowing for generalization and predictions on new, unseen data (Bi et al., 2019). ML techniques are used in various applications to study the gut microbiome, such as microbiome composition, analysis, and therapeutic target identification (Fukui et al., 2020). There are three principal categories of ML: supervised, unsupervised, and reinforcement learning. Supervised learning, including linear regression, decision trees, and support vector machines, uses labeled datasets to train algorithms to classify or predict unknown outcomes. Unsupervised learning, including clustering, dimensionality reduction, and anomaly detection, cluster unlabeled datasets to discover hidden patterns or data groupings (Lopez et al., 2018; Cammarota et al., 2020). Reinforcement learning is a type of artificial intelligence (AI) that achieves a goal in an uncertain and potentially complex environment to build an ML model for decision-making by maximizing a reward function (McCoubrey et al., 2021). Several deep reinforcement learning models have been utilized for biomarker discovery as well as overall microbiome characterization (Mahmud et al., 2018; Liu et al., 2022b; Pan et al., 2022). The deployment of ML has the potential to catalyze novel advancements in patient risk assessment, the discovery of pivotal diagnostic biomarkers, and the prediction of treatment response outcomes.

Personalized disease risk assessment

The unique variation in individuals' gut microbiome profiles, much like a fingerprint, can be leveraged by recommended approaches, such as a patient-centered gut microbiome report, to aid clinicians in making personalized treatment decisions (King et al., 2019). ML can be leveraged in these instances to detect specific patterns in individual patients that can help identify early disease development. Early detection of cardiovascular disease, liver disease, endometriosis, and Type 1 and 2 diabetes mellitus are some of several diseases in which ML is being implemented to detect early disease onset (Aryal et al., 2020; Fernández-Edreira et al., 2021; Huang et al., 2021; Liu et al., 2022a; Ge et al., 2022). By recognizing the initial patterns of disease development, the possibility of reducing someone's risk for a particular condition can be enhanced. This, in turn, allows for timely interventions and focused preventive measures that may impede or decelerate the disease's advancement. This proactive strategy not only enhances the individual's health outcomes but also alleviates the strain on healthcare systems and promotes overall public health.

Diagnostic biomarker identification

Timely disease detection is vital, but not all individuals have routine access to health screenings for identifying symptoms at an initial stage. After the onset of a disease, the subsequent action may involve analyzing biomarkers to detect the presence of the condition. In these scenarios, ML can reveal essential biomarkers, enhancing the precision and effectiveness of disease detection and management (Uddin et al., 2019). Previous studies have implemented ML-based analysis to identify biomarkers associated with Graves' disease, inflammatory bowel disease (IBD), and cancers such as colorectal

cancer (CRC) (Zhou et al., 2018; Maurya et al., 2021; Zhu et al., 2021).

Treatment response prediction

ML analysis of the human gut microbiome shows great potential in predicting response outcomes to interventions and medications (Ortega et al., 2014; de Jong et al., 2021). ML leverages metagenomic data to identify microbial patterns linked to treatment outcomes. These algorithms can then forecast how individual patients will react to interventions, such as dietary adjustments, probiotics, or medications, based on their unique gut microbiome profiles (Dahlin et al., 2022). This personalized approach enables healthcare providers to customize interventions for each patient, optimizing treatment effectiveness and minimizing adverse effects. Furthermore, ML can identify microbial biomarkers that indicate treatment success or failure, facilitating the development of more precise and efficient therapeutic strategies (Yi et al., 2021). As our understanding of the complex interactions between the gut microbiome and human health deepens, ML analysis becomes an invaluable asset in advancing precision medicine and enhancing patient-targeted outcomes.

Clinical applications of gut microbiome data

Diagnosis and prognosis

A variety of multi-omics approaches, including microbial metabolic modeling and phenome-wide associations, are used to identify metabolites and other biomarkers associated with distinct irritable bowel syndrome (IBS) subtypes, IBD, necrotizing enterocolitis, and late-onset sepsis (Stewart et al., 2016; Stewart et al., 2017; Grasberger et al., 2021; Jacobs et al., 2023). These findings contribute to the translation of research discoveries into clinical applications, bridging the gap between laboratory research and improved patient care. One ML approach, trans-omic network analysis, has successfully identified patterns in blood parameters, gut microbiome, and urine metabolome data to identify biomarkers associated with carotid atherosclerosis (Li et al., 2021). Additionally, gene-microbiome association methods employed multi-omics techniques of transcriptomic and metagenomic profiling to expand clinical understanding of the pathophysiology of CRC, IBD, and IBS (Priya et al., 2022).

The gut microbiome's impact on the body as a whole system is evident in its ability to influence various disease stages, with neurological and cancer-related conditions being particularly susceptible to its effects (Aho et al., 2021; Dizman et al., 2022; Tian et al., 2022). Recent advances in gut microbiome and multi-omics data allow for an augmented understanding of disease and symptom severity, disease progression, and predicted responses to therapeutic treatments. Longitudinal multi-omics was used to associate the severity of IBS symptoms with changes in bacterial relative abundance, and specific bacterial species were found to be associated with "flares" in patient symptoms (Mars et al., 2020). Favorable and unfavorable responses to IBS therapeutic classes were

identified using high-throughput -omic profiling and the predictive accuracy improved significantly with the incorporation of proteomics, metabolomics, and metagenomics data (Lee et al., 2021). Microbiota compositions of patients with ulcerative colitis were used to identify proteases associated with disease severity (Mills et al., 2022). In hematopoietic cell transplant patients, microbial diversity was used to predict critical outcomes (Adhi et al., 2019).

Therapeutic treatment

Common gut microbiome-based interventions are additive and modulatory therapies. Additive therapies, as the name suggests, involve introducing microorganisms to a patient's gut microbiome. Fecal Microbiota Transplantation (FMT) and probiotics are effective additive therapies. FMT, also known as fecal bacteriotherapy, involves the introduction of fecal matter into a patient's gut microbiome and has been found to improve patient outcomes related to *Clostridium difficile* infection (CDI), hepatic encephalopathy, and blood disorders with antibiotic-resistant bacteria (Petrof et al., 2013; Bilinski et al., 2017; Zuo et al., 2018; Bajaj et al., 2019). Studies involving probiotic therapies have shown improvements regarding obesity-related disorders and cirrhosis (Dhiman et al., 2014; Depommier et al., 2019). Live biotherapeutic properties have recently been approved by the Federal Drug Administration (U.S. Food and Drug Administration, 2023) and have been an effective treatment for IBS and recurrent CDI (Khanna et al., 2021; Khanna et al., 2022; Quigley et al., 2023). Another well-known type of therapeutic treatments are modulatory ones: diet and exercise. The introduction or restriction of certain nutrients is known to affect gut microbiome composition and improve nonalcoholic fatty liver disease and cardiovascular disease outcomes (Levitan et al., 2009; Lopez-Garcia et al., 2014; Mardinoglu et al., 2018).

Concluding remarks

The development of ML, sequencing technologies, and bioinformatics pipelines have enabled the use of the gut microbiome knowledge to improve the health outcomes of patients. However, it is important to acknowledge the current limitations in this field. A myriad of sequencing and bioinformatics combinations can be selected for the application and translation of microbiome analysis in precision medicine that can lead to widely varying results. Consequently, a principled approach should be applied based on the study design and each step of the process should be determined to ensure high-quality data collection and thoughtful algorithm selection while aiming to document all steps using technologies such as BioCompute Objects (Simonyan et al., 2017). This would allow better interpretation of results by clinicians and other researchers.

The prospect of personalized healthcare is becoming more and more tangible as our understanding of this field deepens. While there are software packages and toolkits available for multi-omics

research devoted to the clinical understanding of disease, the output from the software often lacks user-friendly reports for clinicians. Running these tools typically demands a high level of technical expertise, which is essential to maintain the validity of the results. Future endeavors in multi-omics and machine learning would be best served with a multidisciplinary approach, to develop reporting mechanisms of the results that allow evidence based clinical decision-making. These efforts have the potential to harness the full capabilities of multi-omics approaches in elucidating the gut microbiome and further advancing precision medicine. Addressing these limitations will be crucial to translate this vision into reality and benefit an extensive community.

Author contributions

JW: Writing—original draft, Writing—review and editing. SS: Writing—review and editing. UB: Writing—review and editing. LK: Writing—review and editing. RM: Funding acquisition, Project administration, Supervision, Validation, Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is partially supported by McCormick Genomic and Proteomic Center at The George Washington University.

Acknowledgments

The authors would like to thank Dr. Atin Basuchoudhary (Department of Economics and Business at The Virginia Military Institute) and Sean Kim (Department of Biochemistry and Molecular Medicine at The George Washington University) for their helpful comments on ML and AI while preparing the paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Adhi, F. I., Littmann, E. R., Taur, Y., Maloy, M. A., Markey, K. A., Fontana, E., et al. (2019). Pre-transplant fecal microbial diversity independently predicts critical illness after hematopoietic cell transplantation. *blood* 134, 3264. doi:10.1182/blood-2019-124902
- Aho, V. T. E., Houser, M. C., Pereira, P. A. B., Chang, J., Rudi, K., Paulin, L., et al. (2021). Relationships of gut microbiota, short-chain fatty acids, inflammation, and the gut barrier in Parkinson's disease. *Mol Neurodegener.* 16, 6. doi:10.1186/s13024-021-00427-6
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol.* 215, 403–410.
- Alser, M., Rotman, J., Deshpande, D., Taraszka, K., Shi, H., Baykal, P. I., et al. (2021). Technology dictates algorithms: recent developments in read alignment. *Genome Biol.* 22, 249. doi:10.1186/s13059-021-02443-7
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21, 30. doi:10.1186/s13059-020-1935-5
- Aryal, S., Alimadadi, A., Manandhar, I., Joe, B., and Cheng, X. (2020). Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease. *Hypertension* 76, 1555–1562. doi:10.1161/HYPERTENSIONAHA.120.15885
- Athanasopoulou, K., Boti, M. A., Adamopoulos, P. G., Skourou, P. C., and Scorilas, A. (2021). Third-generation sequencing: the spearhead towards the radical transformation of modern genomics. *Life (Basel)* 12, 30. doi:10.3390/life12010030
- Bajaj, J. S., Salzman, N. H., Acharya, C., Sterling, R. K., White, M. B., Gavis, E. A., et al. (2019). Fecal microbial transplant capsules are safe in hepatic encephalopathy: a phase 1, randomized, placebo-controlled trial. *Hepatology* 70, 1690–1703. doi:10.1002/hep.30690
- Balvocicute, M., and Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics* 18, 114. doi:10.1186/s12864-017-3501-4
- Berg, G., Rybakova, D., Fischer, D., Cernava, T., Verges, M. C., Charles, T., et al. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome* 8, 103. doi:10.1186/s40168-020-00875-0
- Bharti, R., and Grimm, D. G. (2021). Current challenges and best-practice protocols for microbiome analysis. *Brief. Bioinform* 22, 178–193. doi:10.1093/bib/bbz155
- Bi, Q., Goodman, K. E., Kaminsky, J., and Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *Am. J. Epidemiol.* 188, 2222–2239. doi:10.1093/aje/kwz189
- Bilinski, J., Grzesiowski, P., Sorensen, N., Madry, K., Muszynski, J., Robak, K., et al. (2017). Fecal microbiota transplantation in patients with blood disorders inhibits gut colonization with antibiotic-resistant bacteria: results of a prospective, single-center study. *Clin. Infect. Dis.* 65, 364–370. doi:10.1093/cid/cix252
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi:10.1038/s41587-019-0209-9
- Cammarota, G., Ianiro, G., Ahern, A., Carbone, C., Temko, A., Claesson, M. J., et al. (2020). Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nat. Rev. Gastroenterol. Hepatol.* 17, 635–648. doi:10.1038/s41575-020-0327-3
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi:10.1038/nmeth.f.303
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi:10.1038/ismej.2012.8
- Chen, L., Zhao, N., Cao, J., Liu, X., Xu, J., Ma, Y., et al. (2022). Short- and long-read metagenomics expand individualized structural variations in gut microbiomes. *Nat. Commun.* 13, 3175. doi:10.1038/s41467-022-30857-9
- Chen, Y., Ye, W., Zhang, Y., and Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res.* 43, 7762–7768. doi:10.1093/nar/gkv784
- Clooney, A. G., Fouhy, F., Sleator, R. D., Stanton, C., Cotter, P. D., Claesson, M. J., et al. (2016). Comparing apples and oranges? next generation sequencing and its impact on microbiome analysis. *PLoS One* 11, e0148028. doi:10.1371/journal.pone.0148028
- Dahlin, M., Singleton, S. S., David, J. A., Basuchodhary, A., Wickstrom, R., Mazumder, R., et al. (2022). Higher levels of Bifidobacteria and tumor necrosis factor in children with drug-resistant epilepsy are associated with anti-seizure response to the ketogenic diet. *EBioMedicine* 80, 104061. doi:10.1016/j.ebiom.2022.104061
- De Jong, J., Cutcutache, I., Page, M., Elmoufti, S., Dille, C., Frohlich, H., et al. (2021). Towards realizing the vision of precision medicine: AI based prediction of clinical drug response. *Brain* 144, 1738–1750. doi:10.1093/brain/awab108
- Depommier, C., Everard, A., Druart, C., Plovier, H., Van Hul, M., Vieira-Silva, S., et al. (2019). Supplementation with *Akkermansia muciniphila* in overweight and obese human volunteers: a proof-of-concept exploratory study. *Nat. Med.* 25, 1096–1103. doi:10.1038/s41591-019-0495-2
- De Vos, W. M., Tilg, H., Van Hul, M., and Cani, P. D. (2022). Gut microbiome and health: mechanistic insights. *Gut* 71, 1020–1032. doi:10.1136/gutjnl-2021-326789
- Dhiman, R. K., Rana, B., Agrawal, S., Garg, A., Chopra, M., Thumburu, K. K., et al. (2014). Probiotic VSL#3 reduces liver disease severity and hospitalization in patients with cirrhosis: a randomized, controlled trial. *Gastroenterology* 147, 1327–1337.e3. doi:10.1053/j.gastro.2014.08.031
- Dizman, N., Meza, L., Bergerot, P., Alcantara, M., Dorff, T., Lyou, Y., 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. doi:10.1038/s41591-022-01694-6
- Fernández-Edreira, D., Linares-Blanco, J., and Fernandez-Lozano, C. (2021). Machine Learning analysis of the human infant gut microbiome identifies influential species in type 1 diabetes. *Expert Syst. Appl.* 185, 115648. doi:10.1016/j.eswa.2021.115648
- Fukui, H., Nishida, A., Matsuda, S., Kira, F., Watanabe, S., Kuriyama, M., et al. (2020). Usefulness of machine learning-based gut microbiome analysis for identifying patients with irritable bowels syndrome. *J. Clin. Med.* 9, 2403. doi:10.3390/jcm9082403
- Ge, X., Zhang, A., Li, L., Sun, Q., He, J., Wu, Y., et al. (2022). Application of machine learning tools: potential and useful approach for the prediction of type 2 diabetes mellitus based on the gut microbiome profile. *Exp. Ther. Med.* 23, 305. doi:10.3892/etm.2022.11234
- Grasberger, H., Magis, A. T., Sheng, E., Conomos, M. P., Zhang, M., Garzotto, L. S., et al. (2021). DUOX2 variants associate with preclinical disturbances in microbiota-immune homeostasis and increased inflammatory bowel disease risk. *J. Clin. Invest.* 131, e141676. doi:10.1172/JCI141676
- Huang, L., Liu, B., Liu, Z., Feng, W., Liu, M., Wang, Y., et al. (2021). Gut microbiota exceeds cervical microbiota for early diagnosis of endometriosis. *Front. Cell Infect. Microbiol.* 11, 788836. doi:10.3389/fcimb.2021.788836
- Ihmp Research Network Consortium (2019). The integrative human microbiome Project. *Nature* 569, 641–648. doi:10.1038/s41586-019-1238-8
- Jacobs, J. P., Lagishetty, V., Hauer, M. C., Labus, J. S., Dong, T. S., Toma, R., et al. (2023). Multi-omics profiles of the intestinal microbiome in irritable bowel syndrome and its bowel habit subtypes. *Microbiome* 11, 5. doi:10.1186/s40168-022-01450-5
- Jain, M., Olsen, H. E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17, 239. doi:10.1186/s13059-016-1103-0
- Jin, H., You, L., Zhao, F., Li, S., Ma, T., Kwok, L. Y., et al. (2022). Hybrid, ultra-deep metagenomic sequencing enables genomic and functional characterization of low-abundance species in the human gut microbiome. *Gut Microbes* 14, 2021790. doi:10.1080/19490976.2021.2021790
- Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* 10, 5029. doi:10.1038/s41467-019-13036-1
- Kavanaugh, D., O'callaghan, J., Kilcoyne, M., Kane, M., Joshi, L., and Hickey, R. M. (2015). The intestinal glycane and its modulation by diet and nutrition. *Nutr. Rev.* 73, 359–375. doi:10.1093/nutrit/nuu019
- Khanna, S., Assi, M., Lee, C., Yoho, D., Louie, T., Knappe, W., et al. (2022). Efficacy and safety of RBX2660 in PUNCH CD3, a phase III, randomized, double-blind, placebo-controlled trial with a bayesian primary analysis for the prevention of recurrent clostridioides difficile infection. *Drugs* 82, 1527–1538. doi:10.1007/s40265-022-01797-x
- Khanna, S., Pardi, D. S., Jones, C., Shannon, W. D., Gonzalez, C., and Blount, K. (2021). RBX7455, a non-frozen, orally administered investigational live biotherapeutic, is safe, effective, and shifts patients' microbiomes in a phase 1 study for recurrent clostridioides difficile infections. *Clin. Infect. Dis.* 73, e1613–e1620. doi:10.1093/cid/ciaa1430
- King, C. H., Desai, H., Sylvestry, A. C., Lotempio, J., Ayanyan, S., Carrie, J., et al. (2019). Baseline human gut microbiota profile in healthy people and standard reporting template. *PLoS One* 14, e0206484. doi:10.1371/journal.pone.0206484
- Koren, S., and Phillippy, A. M. (2015). One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr. Opin. Microbiol.* 23, 110–120. doi:10.1016/j.mib.2014.11.014
- Langmead, B. (2010). Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinforma.* Chapter 11, Unit 11 7. doi:10.1002/0471250953.bi1107s32
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923

- Lee, J. W. J., Plichta, D., Hogstrom, L., Borren, N. Z., Lau, H., Gregory, S. M., et al. (2021). Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. *Cell Host Microbe* 29, 1294–1304 e4. doi:10.1016/j.chom.2021.06.019
- Levitant, E. B., Wolk, A., and Mittleman, M. A. (2009). Consistency with the DASH diet and incidence of heart failure. *Arch. Intern Med.* 169, 851–857. doi:10.1001/archinternmed.2009.56
- Li, R. J., Jie, Z. Y., Feng, Q., Fang, R. L., Li, F., Gao, Y., et al. (2021). Network of interactions between gut microbiome, host biomarkers, and urine metabolome in carotid atherosclerosis. *Front. Cell Infect. Microbiol.* 11, 708088. doi:10.3389/fcimb.2021.708088
- Li, Y. C., and Lu, Y. C. (2019). BLASTP-ACC: parallel architecture and hardware accelerator design for BLAST-based protein sequence alignment. *IEEE Trans. Biomed. Circuits Syst.* 13, 1771–1782. doi:10.1109/TBCAS.2019.2943539
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012). Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012, 251364. doi:10.1155/2012/251364
- Liu, Y., Meric, G., Havulinna, A. S., Teo, S. M., Aberg, F., Ruuskanen, M., et al. (2022a). Early prediction of incident liver disease using conventional risk factors and gut-microbiome-augmented gradient boosting. *Cell Metab.* 34, 719–730 e4. doi:10.1016/j.cmet.2022.03.002
- Liu, Y., Zhu, J., Wang, H., Lu, W., Lee, Y. K., Zhao, J., et al. (2022b). Machine learning framework for gut microbiome biomarkers discovery and modulation analysis in large-scale obese population. *BMC Genomics* 23, 850. doi:10.1186/s12864-022-09087-2
- Lopez, C., Tucker, S., Salameh, T., and Tucker, C. (2018). An unsupervised machine learning method for discovering patient clusters based on genetic signatures. *J. Biomed. Inf.* 85, 30–39. doi:10.1016/j.jbi.2018.07.004
- Lopez-Garcia, E., Rodriguez-Artalejo, F., Li, T. Y., Fung, T. T., Li, S., Willett, W. C., et al. (2014). The Mediterranean-style dietary pattern and mortality among men and women with cardiovascular disease. *Am. J. Clin. Nutr.* 99, 172–180. doi:10.3945/ajcn.113.068106
- Lu, J., and Salzberg, S. L. (2020). Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome* 8, 124. doi:10.1186/s40168-020-00900-2
- Mahmud, M., Kaiser, M. S., Hussain, A., and Vassanelli, S. (2018). Applications of deep learning and reinforcement learning to biological data. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 2063–2079. doi:10.1109/TNNLS.2018.2790388
- Malla, M. A., Dubey, A., Kumar, A., Yadav, S., Hashem, A., and Abd Allah, E. F. (2018). Exploring the human microbiome: the potential future role of next-generation sequencing in disease diagnosis and treatment. *Front Immunol.* 9, 2868.
- Mantere, T., Kersten, S., and Hoischen, A. (2019). Long-read sequencing emerging in medical genetics. *Front. Genet.* 10, 426. doi:10.3389/fgene.2019.00426
- Maranga, M., Szczerbiak, P., Bezshapkin, V., Gligorijevic, V., Chandler, C., Bonneau, R., et al. (2023). Comprehensive functional annotation of metagenomes and microbial genomes using a deep learning-based method. *mSystems* 8, e0117822. doi:10.1128/mSystems.01178-22
- Mardinoglu, A., Wu, H., Bjornson, E., Zhang, C., Hakkarainen, A., Rasanen, S. M., et al. (2018). An integrated understanding of the rapid metabolic benefits of a carbohydrate-restricted diet on hepatic steatosis in humans. *Cell Metab.* 27, 559–571. doi:10.1016/j.cmet.2018.01.005
- Mars, R. a. T., Yang, Y., Ward, T., Houlti, M., Priya, S., Lekatz, H. R., et al. (2020). Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell* 182, 1460–1473. doi:10.1016/j.cell.2020.08.007
- Maurya, N. S., Kushwaha, S., Chawade, A., and Mani, A. (2021). Transcriptome profiling by combined machine learning and statistical R analysis identifies TMEM236 as a potential novel diagnostic biomarker for colorectal cancer. *Sci. Rep.* 11, 14304. doi:10.1038/s41598-021-92692-0
- Mccoubrey, L. E., Elbadawi, M., Orlu, M., Gaisford, S., and Basit, A. W. (2021). Harnessing machine learning for development of microbiome therapeutics. *Gut Microbes* 13, 1–20. doi:10.1080/19490976.2021.1872323
- Mcdonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., Desantis, T. Z., Probst, A., et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618. doi:10.1038/ismej.2011.139
- Meslier, V., Quinquis, B., Da Silva, K., Plaza Onate, F., Pons, N., Roume, H., et al. (2022). Benchmarking second and third-generation sequencing platforms for microbial metagenomics. *Sci Data.* 9, 694.
- Mills, R. H., Dulai, P. S., Vazquez-Baeza, Y., Saucedo, C., Daniel, N., Germer, R. R., et al. (2022). Multi-omics analyses of the ulcerative colitis gut microbiome link *Bacteroides vulgatus* proteases with disease severity. *Nat. Microbiol.* 7, 262–276. doi:10.1038/s41564-021-01050-3
- Mukhopadhyay, I., Segal, J. P., Carding, S. R., Hart, A. L., and Hold, G. L. (2019). The gut virome: the ‘missing link’ between gut bacteria and host immunity? *Ther. Adv. Gastroenterol.* 12, 1756284819836620. doi:10.1177/1756284819836620
- Ortega, H., Li, H., Suruki, R., Albers, F., Gordon, D., and Yancey, S. (2014). Cluster analysis and characterization of response to mepolizumab. A step closer to personalized medicine for patients with severe asthma. *Ann. Am. Thorac. Soc.* 11, 1011–1017. doi:10.1513/AnnalsATS.201312-454OC
- Pan, S., Zhu, C., Zhao, X. M., and Coelho, L. P. (2022). A deep siamese neural network improves metagenome-assembled genomes in microbiome datasets across different environments. *Nat. Commun.* 13, 2326. doi:10.1038/s41467-022-29843-y
- Petrof, E. O., Gloor, G. B., Vanner, S. J., Weese, S. J., Carter, D., Daigneault, M. C., et al. (2013). Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: ‘RePOOPulating’ the gut. *Microbiome* 1, 3. doi:10.1186/2049-2618-1-3
- Prakash, T., and Taylor, T. D. (2012). Functional assignment of metagenomic data: challenges and applications. *Brief. Bioinform.* 13, 711–727. doi:10.1093/bib/bbs033
- Priya, S., Burns, M. B., Ward, T., Mars, R. a. T., Adamowicz, B., Lock, E. F., et al. (2022). Identification of shared and disease-specific host gene-microbiome associations across human diseases using multi-omic integration. *Nat. Microbiol.* 7, 780–795. doi:10.1038/s41564-022-01121-z
- Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 33, D501–D504. doi:10.1093/nar/gki025
- Pushkarev, D., Neff, N. F., and Quake, S. R. (2009). Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* 27, 847–850. doi:10.1038/nbt.1561
- Quigley, E. M. M., Markinson, L., Stevenson, A., Treasure, F. P., and Lacy, B. E. (2023). Randomised clinical trial: efficacy and safety of the live biotherapeutic product MRx1234 in patients with irritable bowel syndrome. *Aliment. Pharmacol. Ther.* 57, 81–93. doi:10.1111/apt.17310
- Ranjan, R., Rani, A., Metwally, A., Mcgee, H. S., and Perkins, D. L. (2016). Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* 469, 967–977. doi:10.1016/j.bbrc.2015.12.083
- Reuter, J. A., Spacek, D. V., and Snyder, M. P. (2015). High-throughput sequencing technologies. *Mol Cell.* 58, 586–597.
- Rubel, M. A., Abbas, A., Taylor, L. J., Connell, A., Tanes, C., Bittinger, K., et al. (2020). Lifestyle and the presence of helminths is associated with gut microbiome composition in Cameroonians. *Genome Biol.* 21, 122. doi:10.1186/s13059-020-02020-4
- Schadt, E. E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. *Hum. Mol. Genet.* 19, R227–R240. doi:10.1093/hmg/ddq416
- Schloss, P. D. (2020). Reintroducing mother: 10 Years later. *Appl. Environ. Microbiol.* 86, e02343-19. doi:10.1128/AEM.02343-19
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi:10.1128/AEM.01541-09
- Schmidt, K., Mwaigwisa, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., et al. (2017). Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* 72, 104–114. doi:10.1093/jac/dkx397
- Scholz, M., Ward, D. V., Pasolli, E., Tolio, T., Zolfo, M., Asnicar, F., et al. (2016). Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat. Methods* 13, 435–438. doi:10.1038/nmeth.3802
- Schork, N. J. (2015). Personalized medicine: time for one-person trials. *Nature* 520, 609–611. doi:10.1038/520609a
- Shamsaddini, A., Pan, Y., Johnson, W. E., Krampis, K., Shcheglovitova, M., Simonyan, V., et al. (2014). Census-based rapid and accurate metagenome taxonomic profiling. *BMC Genomics* 15, 918. doi:10.1186/1471-2164-15-918
- Simonyan, V., Goecks, J., and Mazumder, R. (2017). Biocompute objects-A step towards evaluation and validation of biomedical scientific computations. *PDA J. Pharm. Sci. Technol.* 71, 136–146. doi:10.5731/pdajpst.2016.006734
- Stewart, C. J., Embleton, N. D., Marrs, E. C., Smith, D. P., Nelson, A., Abdulkadir, B., et al. (2016). Temporal bacterial and metabolic development of the preterm gut reveals specific signatures in health and disease. *Microbiome* 4, 67. doi:10.1186/s40168-016-0216-8
- Stewart, C. J., Embleton, N. D., Marrs, E. C. L., Smith, D. P., Fofanova, T., Nelson, A., et al. (2017). Longitudinal development of the gut microbiome and metabolome in preterm neonates with late onset sepsis and healthy controls. *Microbiome* 5, 75. doi:10.1186/s40168-017-0295-1
- Tian, P., Chen, Y., Zhu, H., Wang, L., Qian, X., Zou, R., et al. (2022). Bifidobacterium breve CCFM1025 attenuates major depression disorder via regulating gut microbiome and tryptophan metabolism: a randomized clinical trial. *Brain Behav. Immun.* 100, 233–241. doi:10.1016/j.bbi.2021.11.023
- Udaondo, Z., Sittikankaew, K., Uengwetwanit, T., Wongsurawat, T., Sonthirod, C., Jenjaroenpun, P., et al. (2021). Comparative analysis of PacBio and Oxford nanopore sequencing technologies for transcriptomic landscape identification of *Penaeus monodon*. *Life (Basel)* 11, 862. doi:10.3390/life11080862
- Uddin, S., Khan, A., Hossain, M. E., and Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Med. Inf. Decis. Mak.* 19, 281. doi:10.1186/s12911-019-1004-8
- U.S. Food and Drug Administration (2023). *Fda news release FDA approves first orally administered fecal microbiota product for the prevention of recurrence of clostridioides difficile infection*. Available at: <https://www.fda.gov/news-events/press-announcements/>

fda-approves-first-orally-administered-fecal-microbiota-product-prevention-recurrence-clostridioides (Accessed April 26, 2023).

- Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257. doi:10.1186/s13059-019-1891-0
- Wood, D. E., and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15, R46. doi:10.1186/gb-2014-15-3-r46
- Yang, Y., Jiang, X. T., and Zhang, T. (2014). Evaluation of a hybrid approach using UBLAST and BLASTX for metagenomic sequences annotation of specific functional genes. *PLoS One* 9, e110947. doi:10.1371/journal.pone.0110947
- Yi, Y., Shen, L., Shi, W., Xia, F., Zhang, H., Wang, Y., et al. (2021). Gut Microbiome components predict response to neoadjuvant chemoradiotherapy in patients with locally advanced rectal cancer: a prospective, longitudinal study. *Clin Cancer Res.* 27, 1329–1340.
- Zhang, F., Aschenbrenner, D., Yoo, J. Y., and Zuo, T. (2022). The gut mycobiome in health, disease, and clinical applications in association with the gut bacterial microbiome assembly. *Lancet Microbe* 3, e969–e983. doi:10.1016/S2666-5247(22)00203-8
- Zhou, Y., Xu, Z. Z., He, Y., Yang, Y., Liu, L., Lin, Q., et al. (2018). Gut microbiota offers universal biomarkers across ethnicity in inflammatory bowel disease diagnosis and infliximab response prediction. *mSystems* 3, e00188-17. doi:10.1128/mSystems.00188-17
- Zhu, Q., Hou, Q., Huang, S., Ou, Q., Huo, D., Vazquez-Baeza, Y., et al. (2021). Compositional and genetic alterations in Graves' disease gut microbiome reveal specific diagnostic biomarkers. *ISME J.* 15, 3399–3411. doi:10.1038/s41396-021-01016-7
- Zuo, T., Wong, S. H., Lam, K., Lui, R., Cheung, K., Tang, W., et al. (2018). Bacteriophage transfer during faecal microbiota transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Gut* 67, 634–643. doi:10.1136/gutjnl-2017-313952