Journal of Microbiology (2021) Vol. 59, No. 3, pp. 292–297
DOI 10.1007/s12275-021-1004-0
eISSN 1976-3794
pISSN 1225-8873

REVIEW

Omics in gut microbiome analysis

Tae Woong Whon[1†], Na-Ri Shin[2†], Joon Yong Kim[1], and Seong Woon Roh[1*]

[1]*Microbiology and Functionality Research Group, World Institute of Kimchi, Gwangju 61755, Republic of Korea*
[2]*Biological Resource Center, Korea Research Institute of Bioscience and Biotechnology, Jeongeup-si, Jeollabuk-do 56212, Republic of Korea*

(Received Jan 4, 2021 / Revised Feb 9, 2021 / Accepted Feb 10, 2021)

Our understanding of the interactions between microbial communities and their niche in the host gut has improved owing to recent advances in environmental microbial genomics. Integration of metagenomic and metataxonomic sequencing data with other omics data to study the gut microbiome has become increasingly common, but downstream analysis after data integration and interpretation of complex omics data remain challenging. Here, we review studies that have explored the gut microbiome signature using omics approaches, including metagenomics, metataxonomics, metatranscriptomics, and metabolomics. We further discuss recent analytics programs to analyze and integrate multi-omics datasets and further utilization of omics data with other advanced techniques, such as adaptive immune receptor repertoire sequencing, microbial culturomics, and machine learning, to evaluate important microbiome characteristics in the gut.

Keywords: omics, gut microbiome, metagenome, metataxonome, metatranscriptome, metabolome

Introduction

The human gut harbors various microbial entities, including bacteria, archaea, unicellular eukaryotes, and viruses (Cani, 2018). These biological entities are essential components, attributing the inherent features of the gastrointestinal tract. Over the last decades, the human gut has attracted increasing attention, with studies revealing the genetic and functional traits of gut bacteria (The Human Microbiome Project Consortium, 2012a). After the introduction of next-generation sequencing technologies in the field of gut microbiology, environmental microbial genomics first sheds light on "who they are": metataxonomics (i.e., amplification and sequencing of marker genes, such as bacterial 16S rRNA genes) and metagenomics (i.e., shotgun sequencing of DNA extracted from samples) have provided valuable insights regarding the taxonomic diversity of the microbial community in a defined environment, called the microbiota. These omics-based gut microbial studies conducted in normal healthy populations (Zhernakova *et al.*, 2016; Deschasaux *et al.*, 2018) as well as in subjects with illness (Duvallet *et al.*, 2017) have well described the characteristics of the gut microbiota in both eubiotic and dysbiotic conditions. Accumulating evidence regarding the natural members of the bacterial microbiota has positioned gut microbial studies in the next research step, identifying "what they do".

The gut microbiome can be defined as a collection of information regarding the biotic (i.e., microbes and the surrounding host gut environment), genomic (i.e., the collection of genes and genomes of members of the microbiota), and abiotic factors (i.e., clinical and environmental metadata). In this context, a combination of the environmental microbial genomics with other omics approaches (such as metatranscriptomics, metabolomics, and metaproteomics) is a promising approach to understand genomic, transcriptomic, chemical/metabolic, and proteomic interactions between microbes and/or microbial communities and their niche in the gut. This review discusses gut microbiome studies based on several recent omics approaches and attempts to describe how to analyze and integrate multi-omics datasets to interrogate microbial signatures in the gut.

Omics in gut microbiome analyses

In humans, studies have been largely weighted toward taxonomic and functional profiling of the microbiome in collected fecal samples. Although this non-invasive protocol ensures the safety of fecal donors, feces only mirror the microbiome in the luminal content of large bowels, with a small part of host cells being shed from the gut epithelium. However, studies based on experimental animals have enabled us to explore the compartmentally different microbiome from proximal to the distal gut with dissection of the luminal and mucosal parts of the intestines. Below, we discuss several omics approaches frequently used in the field of the gut microbiome (Fig. 1).

Metagenomics of gut microbiota

Metagenomics is a tool used to analyze the collection of genomes and genes obtained via shotgun sequencing of ex-

†These authors contributed equally to this work.
*For correspondence. E-mail: swroh@wikim.re.kr

tracted DNA from samples. It usually comprises an annotation step after assembly or mapping of sequences to a reference database (Marchesi and Ravel, 2015) (Fig. 1). For the bacterial microbiome, metagenomics can be used to obtain genomic information of unculturable microbes in the gut. A recent study by Almeida *et al.* (2020) has reported over 200,000 non-redundant metagenome-assembled genomes (MAGs) from the human gut microbiome by binning de novo-assembled contigs into putative genomes. Although the human gut microbiota consists of several hundreds of individual bacterial taxa, it exerts its effects as a whole on the human host. Metagenomics enabled profiling of the whole gene repertoire of a study group, referred to as the pan-genome. Pan-genomic analyses have advantages not only for a precise determination of the whole genomic contents within a sample but also for species definition from closely related taxa and analyzing pathogenic microbes (Rouli *et al.*, 2015). Additionally, by aligning the metagenomic reads to protein databases, researchers can assess the enzymatic functions of the entire gut microbiome (Tanes *et al.*, 2021).

Metagenomics has become more useful when applied to the field of viral ecology because there are no marker genes to amplify, such as 16S rRNA genes in bacteria, in the viral genome. In the human gut, viruses infecting bacteria (bacteriophages or phages) are central members of the gut microbiota and play essential roles in the relationship among viruses, bacteria, and gut epithelial cells (Mirzaei and Maurice, 2017). Viral metagenomics has shed light on viral ecology in the gut. Kim *et al.* (2011) investigated viral communities in fecal samples from five healthy Korean subjects based on viral metagenomics and reported diverse single-stranded DNA bacteriophages in the human gut. By comparing the prophage genomes with the metagenome of free bacteriophages, they further reported that bacteriophage generalists contribute to the prevalence of lysogeny in the mammalian gut ecosystem (Kim and Bae, 2018). Considering these viral metagenome studies, our views on gut bacteriophages now converge on Kill-the-Winner dynamics, i.e., the bacterium-phage infection networks of lytic models (Weitz *et al.*, 2013), coupled with the Piggyback-the-Winner model, i.e., the high-density/high-growth lysogeny (Knowles *et al.*, 2016).

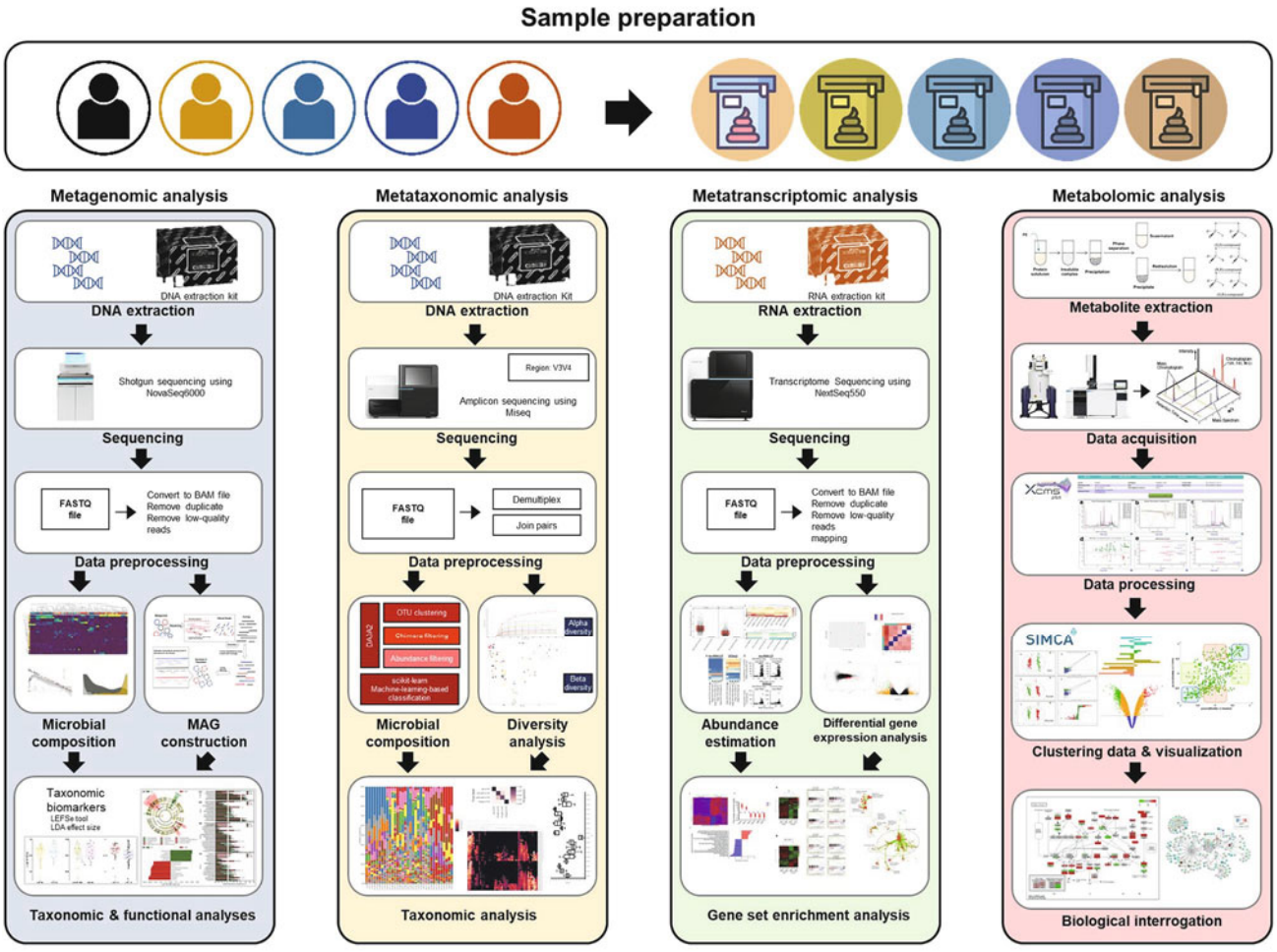

Fig. 1. A schematic model for omics approaches to study the gut microbiome. General procedures for metagenomics, metataxonomics, metatranscriptomics, and metabolomics used to study the gut microbiome are shown.

Metataxonomics of gut microbiota

Metataxonomics can be defined as a high-throughput process for characterization of the entire microbiota (Marchesi and Ravel, 2015). Metataxonomics is based mostly on the amplification and sequencing of marker genes (i.e., bacterial 16S rRNA genes), where the conserved and variable regions exist (Hilton *et al.*, 2016) (Fig. 1). Metataxonomic tree-based hierarchical clustering analyses enable community-wide taxonomic classifications and demonstrate the phylogenetic relationships between all sequences obtained. Therefore, metataxonomics has shed light on the structure and composition of gut bacterial communities in humans (Kim *et al.*, 2013; Lee *et al.*, 2020a), mice (Shin *et al.*, 2014; Stacy *et al.*, 2021), and insects (Yun *et al.*, 2014; Whon *et al.*, 2017). Along with identification of the bacterial microbiota, metataxonomics can be applied to archaeal community analysis in the gut. Recently, Kim *et al.* (2020) surveyed archaeal communities of fecal samples collected from 897 East Asian subjects living in South Korea. The archaeal 16S rRNA gene-targeted amplicon sequencing identified extensive colonization of the Korean gut by halophilic archaea. In terms of the fungal component in the gut, called the gut mycobiota, the nuclear ribosomal internal transcribed spacer (ITS) region and fungal 18S rRNA genes are mainly subjected to amplification and sequencing for metataxonomic analysis (Richard and Sokol, 2019). Importantly, neither metataxonomics is the same as metagenomics, nor is metataxonomics included in the category of metagenomics; therefore, the two words must not be used interchangeably.

Metatranscriptomics of gut microbiota

Metatranscriptomics is designed to profile the regulation and expression of RNA in complex microbes within natural environments. In general, synthesized clonal DNA (cDNA), followed by RNA extraction from samples, are subjected to high-throughput sequencing (Fig. 1). While the transcriptome mirrors a comprehensive set of RNA encoded by the genome of an organism, the metatranscriptome encompasses all transcripts encoded by genes of a group of microbial communities. Given that the majority of the extracted RNA is ribosomal RNA (rRNA represents almost 95% of total RNA), removal of this unnecessary part using commercial kits was a key procedure to obtain as many messenger RNAs (mRNAs) as possible. However, in the last decade, the above patterns have changed quickly because the cost of high-throughput sequencing is decreasing and the sequencing output is increasing. At present, the method that sequences much and excludes much is preferred. Specifically, researchers prefer to conduct RNA-Seq on intact RNA and remove *in silico* rRNA prior to sequencing using programs such as SortMeRNA (Kopylova *et al.*, 2012).

Extracted gut RNA (e.g., from intestinal specimens, intestinal luminal contents, or fecal samples) inevitably possesses the host RNA of gut epithelial cells (Williams *et al.*, 2015; Stauber *et al.*, 2016). In the case of gut (or fecal) samples collected from the host with antibiotic treatment, extracted gut RNA probably mainly consisted of host RNA with a very small proportion of microbial RNA. A high host RNA background can simply be depleted using commercial kits targeting the polyA tail of eukaryotic RNA (Marsh *et al.*, 2017). Interestingly, increasing sequencing output now makes host RNA removal unnecessary and facilitates the use of another option, such as dual transcriptomics, to profile transcriptomic changes in the host and microbes simultaneously. Recently, we conducted dual transcriptomics on rectal luminal samples of diarrheic calves to understand the multifactorial nature of calf diarrhea (Whon *et al.*, 2021). The inter-transcriptomic relationship between the bovine host and gut bacteria indicated that the diarrheic gut constitutes a distinct environmental niche, as exemplified by elevated sulfur metabolism, immune responses, and gut motility. These conditions favor the growth of aerobes and/or facultative anaerobes, such as those belonging to the genus *Escherichia*.

Other omics approaches to study the gut microbiome

Gut metabolites can be regarded as the products of gut microbial activities followed by transcriptomic and proteomic regulations. The omics approach to identify the metabolite profiles in any given strain or single tissue, called metabolomics, is also frequently used in gut microbiome studies (Fig. 1). Metabolomics is the measurement of the amounts (or concentrations) and locations of all the metabolites in cells or tissues. Metabolites are the small molecules transformed in the process of metabolism (in most cases they are substrates and products of enzymes). The term "metabolomics" must not be interchangeably used with "metabonomics", the approach to generate metabolite profiles from complex systems (Marchesi and Ravel, 2015). However, the terms metabonomics or metametabolomics are not frequently used when describing the analysis of metabolites derived from environmental samples because, in most cases, we are not able to differentiate the metabolites depending on the microorganism. Nuclear magnetic resonance (NMR) spectroscopy and liquid chromatography-mass spectroscopy (LC-MS) separation systems are the most favored platforms for metabolome characterization. In addition, metaproteomic data can represent microbial activity more directly in human gut environments than any other omics approaches. For instance, Lobel *et al.* (2020) have recently reported kidney protection by dietary modification of gut microbial metabolism. They could not observe changes in the composition of the gut microbiota from the metataxonomic datasets. However, by using metaproteomics, they could successfully demonstrate that a diet with high levels of sulfur-containing amino acids is able to modify a microbial enzyme in the gut bacteria that can modulate renal function.

Such omics approaches become more powerful when applied together. Given that metataxonomics, the most frequently used approach in the field of gut microbial ecology, only identifies microbial abundance with phylogenetic identity relative to the total microbial abundance and phylogeny within a sample, this approach could not define whether the sample contains a responsible microbial community or mostly the microbial DNA from dead cells. In this case, a combination of metataxonomics with other omics, including metatranscriptomics and/or metabolomics, can provide information regarding the transcriptionally/metabolically active microbes. Gut microbiome studies based on a combination of omics approaches have been increasingly reported. Metataxono-

mics combined with metatranscriptomics was used to identify microbial compositional dysbiosis and functional dysbiosis in the gut of diarrheic calves (Whon *et al.*, 2021), and metataxonomics combined with metabolomics to identify castration-induced gut microbial alteration and further metabolic phenotypes in male cattle (Whon *et al.*, 2020), as well as long-term effects of fecal microbial transplantation on calf diarrhea and further growth performance (Kim *et al.*, 2021). The abovementioned bovine gut microbiome studies collectively imply that understanding crosstalk between microbes and/or host and microbes by using multi-omics approaches is key to produce high-quality animal products. In addition, the combined work of metataxonomics and metabolomics has identified the role of the gut microbiome and metabolites involved in liver fibrosis pathogenesis in humans (Lee *et al.*, 2020b) and radio-protective gut microbes and metabolites in mice that received a high dose of total body radiation (Guo *et al.*, 2020). Metagenomics combined with metabolomics and genome sequencing identified several biomarkers of *Clostridioides difficile* infection in pediatric inflammatory bowel disease patients (Bushman *et al.*, 2020).

Immune repertoire sequencing

Given that a large amount of microbial biomass in the gut microbiota comprises non-self antigens, it is reasonable to assume that the gut microbiota is one of the biggest challenges to the host immune system (Savage, 1977). The mammalian intestinal tract is equipped with a variety of effective and efficient defense/immune mechanisms to maintain a precise balance between immunity and tolerance (Martens *et al.*, 2018). The interaction between the gut microbiota and the host immune system is a principal requirement for their symbiotic homeostasis; thus, it is necessary to gain an integrated insight into the bidirectional response of microbes and their host.

Adaptive immunity, mainly mediated by B- and T-cells, is responsible for determining the mode of the immune response in an antigen-specific manner. Because B- and T-cells express antigen-specific receptors, these cells undergo somatic rearrangement of the complementarity-determining region 3 (CDR3), comprising variable (V), joining (J), and/or diversity (D) gene segments, in T cell receptor (TCR) or B cell receptor (BCR)/immunoglobulin (Ig)-encoding loci to diversify the range of recognizable antigens (Tonegawa, 1983; Patten *et al.*, 1984). The integration of multi-omics approaches is being leveraged to decipher the whole landscape of adaptive immune receptor repertoires, estimated from a few thousand to more than billions, at a sequence level based on the high-throughput nature of next-generation sequencing. The introduction of this advanced methodology to immune profiling, i.e., high-throughput sequencing of CDR3 amplicons, enables us to unveil the mechanism or principle of host immune response against the gut microbiota, which was previously unrecognized by targeted investigation using qPCR for mRNA expression, fluorescence activated cell sorter or microarray. For example, recent studies have revealed how the BCR repertoire is shaped by the microbiota (Chen *et al.*, 2018; Li *et al.*, 2020) and the contribution of the gut microbiota to liver fibrosis in terms of the TCR repertoire (Liang *et al.*, 2020).

In future studies, analyses combining the developing multi-omics techniques, such as cellular indexing of transcriptomes and epitopes using sequencing (CITE-seq), single-cell assay for transposase-accessible chromatin using sequencing (ATAC-seq), or high-throughput liquid chromatography tandem mass spectroscopy (LC-MS/MS), will provide more clarified insights into how the immune system works at the single-cell and amino acid-sequence levels.

Omics data analysis

QIIME2 (Bolyen *et al.*, 2019) is currently the most commonly used bioinformatics platform to analyze metataxonomic sequencing data. QIIME2 is a free, open-source, and community-developed program. One outstanding feature different from those in QIIME1 is the view function in the QIIME2 that allows users to securely share and interact with analytical results without installing QIIME2. The source code is available at https://github.com/qiime2, and help for QIIME2 is provided at https://forum.qiime2.org. In terms of the metagenomic sequencing data, recovery of microbial genomes from the human gut metagenomes can be achieved by using IDBA-UD (Peng *et al.*, 2012) and DAS Tool (Sieber *et al.*, 2018). MetaPhlAn provides reference to genome-based analytical methods (Segata *et al.*, 2012). Nubeam provides a clustering analysis of metagenomic sequencing data with a reference-free approach (Dai and Guan, 2020).

There exist several workflow pipelines to comprehensively analyze the omics data. QIIME2 plugins currently provide initial support to analyze metabolomic and metagenomic data (Bolyen *et al.*, 2019). The HUMAnN2 program employs metagenomic or metatranscriptomic sequencing data to profile the presence/absence and abundance of microbial pathways in a community (Franzosa *et al.*, 2018). The original version of this program was developed during The Human Microbiome Project Consortium (2012b); thus, the program is advantageous for analyzing human samples, including the gut metagenome and/or metatranscriptome. The source code is available at https://huttenhower.sph.harvard.edu/humann2.

Lastly, we recommend that researchers studying the gut microbiome pay more attention to the machine learning (ML) approach for analyzing multi-omics datasets. ML focuses on how computers learn and improve from available data. The models created by learning algorithms are able to make decisions and predict results without performing experimental tasks. Deep learning is a subfield of ML algorithms that builds on large multi-layer neural networks inspired by the brain structure and function (Cammarota *et al.*, 2020; Oh and Zhang, 2020). Moreover, ML algorithms are frequently used to predict disease state based on multi-omics datasets including random forest, multi-layer perceptron, and support vector machine. Metagenomic, metataxonomic, metatranscriptomic, and metabolomic data can all be sources of ML training sets, and trained ML models enable researchers to analyze new data and identify important microbiome characteristics. ML has increasingly been applied to gut microbiome studies for the diagnosis and prediction of a variety of diseases (Aryal *et*

al., 2020; Cammarota *et al.*, 2020). Moreover, ML combined with rapidly expanding public biological databases can bridge the scattered open reference data to explore specific hypotheses without conducting new experiments (Cammarota *et al.*, 2020).

Conclusion

With recent advances in sequencing and analytical technologies, the integration of metagenomics/metataxonomic sequencing data with other omics data to study the gut microbiome is becoming common. However, downstream analysis after data integration and interpretation of complex omics data still remain challenging because i) most analytics programs introduced in this review are script-based and too computer-demanding, ii) the step-by-step approach to analyzing large multi-omics datasets by individual researchers lacks standardized protocols, and iii) most importantly, multi-omics data analyses are largely dependent on correlation-based analysis. These problems are not exclusive of gut microbiome study, but of all microbiology studies, including environmental microbiology. Along with the correlation results on multi-omics datasets, evaluation of the causative role of gut microbes in the metabolic/disease/clinical phenotypes of the host will become increasingly important.

Given that most of the microorganisms present in the environment cannot be cultured using the existing culture methods (Amann *et al.*, 1995; Nichols *et al.*, 2010), microbial culturomics (i.e., a high-throughput culture approach) to obtain key microbial taxa from the gut samples continues to be an essential technique. Such a culture-dependent approach with the combined use of a gnotobiotic animal model, i.e., animals harboring defined microbial communities in their gut, will reciprocally elucidate the causative roles of microbes. Taken as a whole, a combinatory use of multi-omics datasets, microbial culturomics, and ML can be an accurate and cost-effective approach that reduces the use of experimental animals to study the signature of the gut microbiome.

Acknowledgments

This study was supported by the Main Research Program of the World Institute of Kimchi (KE2101-2 to T.W.W., J.Y.K., and S.W.R.) and the KRIBB Research Initiative Program (KGM5232113 to N.-R.S.), Republic of Korea.

Conflict of Interest

We have no conflicts of interest to report.

References

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., Pollard, K.S., Sakharova, E., Parks, D.H., Hugenholtz, P., *et al.* 2020. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114.

Amann, R.I., Ludwig, W., and Schleifer, K.H. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169.

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.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., *et al.* 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857.

Bushman, F.D., Conrad, M., Ren, Y., Zhao, C.Y., Gu, C., Petucci, C., Kim, M.S., Abbas, A., Downes, K.J., Devas, N., *et al.* 2020. Multiomic analysis of the interaction between *Clostridioides difficile* infection and pediatric inflammatory bowel disease. *Cell Host Microbe* **28**, 422–433.

Cammarota, G., Ianiro, G., Ahern, A., Carbone, C., Temko, A., Claesson, M.J., Gasbarrini, A., and Tortora, G. 2020. Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 635–648.

Cani, P.D. 2018. Human gut microbiome: hopes, threats and promises. *Gut* **67**, 1716–1725.

Chen, Y., Chaudhary, N., Yang, N., Granato, A., Turner, J.A., Howard, S.L., Devereaux, C., Zuo, T., Shrestha, A., Goel, R.R., *et al.* 2018. Microbial symbionts regulate the primary Ig repertoire. *J. Exp. Med.* **215**, 1397–1415.

Dai, H. and Guan, Y. 2020. The Nubeam reference-free approach to analyze metagenomic sequencing reads. *Genome Res.* **30**, 1364–1375.

Deschasaux, M., Bouter, K.E., Prodan, A., Levin, E., Groen, A.K., Herrema, H., Tremaroli, V., Bakker, G.J., Attaye, I., Pinto-Sietsma, S.J., *et al.* 2018. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nat. Med.* **24**, 1526–1531.

Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J. 2017. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* **8**, 1784.

Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., *et al.* 2018. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968.

Guo, H., Chou, W.C., Lai, Y., Liang, K., Tam, J.W., Brickey, W.J., Chen, L., Montgomery, N.D., Li, X., Bohannon, L.M., *et al.* 2020. Multi-omics analyses of radiation survivors identify radioprotective microbes and metabolites. *Science* **370**, eaay9097.

Hilton, S.K., Castro-Nallar, E., Perez-Losada, M., Toma, I., McCaffrey, T.A., Hoffman, E.P., Siegel, M.O., Simon, G.L., Johnson, W.E., and Crandall, K.A. 2016. Metataxonomic and metagenomic approaches vs. culture-based techniques for clinical pathology. *Front. Microbiol.* **7**, 484.

Kim, M.S. and Bae, J.W. 2018. Lysogeny is prevalent and widely distributed in the murine gut microbiota. *ISME J.* **12**, 1127–1141.

Kim, M.S., Hwang, S.S., Park, E.J., and Bae, J.W. 2013. Strict vegetarian diet improves the risk factors associated with metabolic diseases by modulating gut microbiota and reducing intestinal inflammation. *Environ. Microbiol. Rep.* **5**, 765–775.

Kim, M.S., Park, E.J., Roh, S.W., and Bae, J.W. 2011. Diversity and abundance of single-stranded DNA viruses in human feces. *Appl. Environ. Microbiol.* **77**, 8062–8070.

Kim, J.Y., Whon, T.W., Lim, M.Y., Kim, Y.B., Kim, N., Kwon, M.S., Kim, J., Lee, S.H., Choi, H.J., Nam, I.H., *et al.* 2020. The human gut archaeome: identification of diverse haloarchaea in Korean subjects. *Microbiome* **8**, 114.

Kim, H.S., Whon, T.W., Sung, H., Jeong, Y.S., Jung, E.S., Shin, N.R., Hyun, D.W., Kim, P.S., Lee, J.Y., Lee, C.H., *et al.* 2021. Longitudinal evaluation of fecal microbiota transplantation for ameliorating calf diarrhea and improving growth performance. *Nat. Commun.* **12**, 161.

Knowles, B., Silveira, C.B., Bailey, B.A., Barott, K., Cantu, V.A., Cobian-Guemes, A.G., Coutinho, F.H., Dinsdale, E.A., Felts, B., Furby, K.A., *et al.* 2016. Lytic to temperate switching of viral communities. *Nature* **531**, 466–470.

Kopylova, E., Noe, L., and Touzet, H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217.

Lee, J.Y., Cevallos, S.A., Byndloss, M.X., Tiffany, C.R., Olsan, E.E., Butler, B.P., Young, B.M., Rogers, A.W.L., Nguyen, H., Kim, K., *et al.* 2020a. High-fat diet and antibiotics cooperatively impair mitochondrial bioenergetics to trigger dysbiosis that exacerbates pre-inflammatory bowel disease. *Cell Host Microbe* **28**, 273–284.e6.

Lee, G., You, H.J., Bajaj, J.S., Joo, S.K., Yu, J., Park, S., Kang, H., Park, J.H., Kim, J.H., Lee, D.H., *et al.* 2020b. Distinct signatures of gut microbiome and metabolites associated with significant fibrosis in non-obese NAFLD. *Nat. Commun.* **11**, 4982.

Li, H., Limenitakis, J.P., Greiff, V., Yilmaz, B., Scharen, O., Urbaniak, C., Zund, M., Lawson, M.A.E., Young, I.D., Rupp, S., *et al.* 2020. Mucosal or systemic microbiota exposures shape the B cell repertoire. *Nature* **584**, 274–278.

Liang, Q., Zhang, M., Hu, Y., Zhang, W., Zhu, P., Chen, Y., Xue, P., Li, Q., and Wang, K. 2020. Gut microbiome contributes to liver fibrosis impact on T cell receptor immune repertoire. *Front. Microbiol.* **11**, 571847.

Lobel, L., Cao, Y.G., Fenn, K., Glickman, J.N., and Garrett, W.S. 2020. Diet posttranslationally modifies the mouse gut microbial proteome to modulate renal function. *Science* **369**, 1518–1524.

Marchesi, J.R. and Ravel, J. 2015. The vocabulary of microbiome research: a proposal. *Microbiome* **3**, 31.

Marsh, J.W., Humphrys, M.S., and Myers, G.S.A. 2017. A laboratory methodology for dual RNA-sequencing of bacteria and their host cells *in vitro*. *Front. Microbiol.* **8**, 1830.

Martens, E.C., Neumann, M., and Desai, M.S. 2018. Interactions of commensal and pathogenic microorganisms with the intestinal mucosal barrier. *Nat. Rev. Microbiol.* **16**, 457–470.

Mirzaei, M.K. and Maurice, C.F. 2017. Menage a trois in the human gut: interactions between host, bacteria and phages. *Nat. Rev. Microbiol.* **15**, 397–408.

Nichols, D., Cahoon, N., Trakhtenberg, E., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K., and Epstein, S. 2010. Use of ichip for high-throughput *in situ* cultivation of "uncultivable" microbial species. *Appl. Environ. Microbiol.* **76**, 2445–2450.

Oh, M. and Zhang, L. 2020. DeepMicro: deep representation learning for disease prediction based on microbiome data. *Sci. Rep.* **10**, 6026.

Patten, P., Yokota, T., Rothbard, J., Chien, Y., Arai, K., and Davis, M.M. 1984. Structure, expression and divergence of T-cell receptor beta-chain variable regions. *Nature* **312**, 40–46.

Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428.

Richard, M.L. and Sokol, H. 2019. The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 331–345.

Rouli, L., Merhej, V., Fournier, P.E., and Raoult, D. 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* **7**, 72–85.

Savage, D.C. 1977. Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.* **31**, 107–133.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814.

Shin, N.R., Lee, J.C., Lee, H.Y., Kim, M.S., Whon, T.W., Lee, M.S., and Bae, J.W. 2014. An increase in the *Akkermansia* spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice. *Gut* **63**, 727–735.

Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843.

Stacy, A., Andrade-Oliveira, V., McCulloch, J.A., Hild, B., Oh, J.H., Perez-Chaparro, P.J., Sim, C.K., Lim, A.I., Link, V.M., Enamorado, M., *et al.* 2021. Infection trains the host for microbiota-enhanced resistance to pathogens. *Cell* **184**, 615–627.

Stauber, J., Shaikh, N., Ordiz, M.I., Tarr, P.I., and Manary, M.J. 2016. Droplet digital PCR quantifies host inflammatory transcripts in feces reliably and reproducibly. *Cell. Immunol.* **303**, 43–49.

Tanes, C., Bittinger, K., Gao, Y., Friedman, E.S., Nessel, L., Paladhi, U.R., Chau, L., Panfen, E., Fischbach, M.A., Braun, J., *et al.* 2021. Role of dietary fiber in the recovery of the human gut microbiome and its metabolome. *Cell Host Microbe* doi: 10.1016/j.chom.2020.12.012.

The Human Microbiome Project Consortium. 2012a. A framework for human microbiome research. *Nature* **486**, 215–221.

The Human Microbiome Project Consortium. 2012b. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214.

Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature* **302**, 575–581.

Weitz, J.S., Poisot, T., Meyer, J.R., Flores, C.O., Valverde, S., Sullivan, M.B., and Hochberg, M.E. 2013. Phage-bacteria infection networks. *Trends Microbiol.* **21**, 82–91.

Whon, T.W., Kim, H.S., Shin, N.R., Jung, E.S., Tak, E.J., Sung, H., Jung, M.J., Jeong, Y.S., Hyun, D.W., Kim, P.S., *et al.* 2020. Male castration increases adiposity via small intestinal microbial alterations. *EMBO Rep.* **22**, e50663.

Whon, T.W., Kim, H.S., Shin, N., Sung, H., Kim, M., Kim, J.Y., Kang, W., Kim, P.S., Hyun, D., Seong, H.J., *et al.* 2021. Calf diarrhea caused by prolonged expansion of autochthonous gut *Enterobacteriaceae* and their lytic bacteriophages. *mSystems*. Doi: 10.1128/mSystems.00816-20

Whon, T.W., Shin, N.R., Jung, M.J., Hyun, D.W., Kim, H.S., Kim, P.S., and Bae, J.W. 2017. Conditionally pathogenic gut microbes promote larval growth by increasing redox-dependent fat storage in high-sugar diet-fed drosophila. *Antioxid. Redox Signal.* **27**, 1361–1380.

Williams, J.M., Duckworth, C.A., Burkitt, M.D., Watson, A.J., Campbell, B.J., and Pritchard, D.M. 2015. Epithelial cell shedding and barrier function: a matter of life and death at the small intestinal villus tip. *Vet. Pathol.* **52**, 445–455.

Yun, J.H., Roh, S.W., Whon, T.W., Jung, M.J., Kim, M.S., Park, D.S., Yoon, C., Nam, Y.D., Kim, Y.J., Choi, J.H., *et al.* 2014. Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Appl. Environ. Microbiol.* **80**, 5254–5264.

Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., *et al.* 2016. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569.