Host and microbiome multi-omics integration: applications and methodologies

Qing Wang^{1,2,3} · Kaicen Wang^{1,2} · Wenrui Wu^{1,2} · Eleni Giannoulatou^{3,4} · Joshua W. K. Ho^{3,4,5} · Lanjuan Li^{1,2}

Received: 4 November 2018 / Accepted: 6 December 2018 / Published online: 9 January 2019

© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The study of the microbial community—the microbiome—associated with a human host is a maturing research field. It is increasingly clear that the composition of the human's microbiome is associated with various diseases such as gastrointestinal diseases, liver diseases and metabolic diseases. Using high-throughput technologies such as next-generation sequencing and mass spectrometry-based metabolomics, we are able to comprehensively sequence the microbiome—the metagenome—and associate these data with the genomic, epigenomics, transcriptomic and metabolic profile of the host. Our review summarises the application of integrating host omics with microbiome as well as the analytical methods and related tools applied in these studies. In addition, potential future directions are discussed.

Keywords Microbiome · Genome · Epigenome · Transcriptome · Metabolome · Network analysis · Big data

Introduction

The microbiota is a complex community of bacteria, viruses, fungi and protozoa that inhabits the host body (e.g. mucosal surface of the gut and skin) (Gilbert et al. 2018). According to a recent report, it is estimated that there are thousands of bacterial species, and each strain has thousands of genes (Gilbert et al. 2018). Thus, a microbiome carries an estimated 150-fold more genes than the human genome. The collective genome

of the microbiome, termed as metagenome, is also referred to as the second genome of the human body (Turnbaugh et al. 2007).

Studies have investigated microbiome composition in various mucosal and skin environments, among which the most studied one is the gut microbiome (Fig. 1). It is increasingly recognised that the microbiome influences the host health state and disease progression ranging from gastrointestinal diseases such as inflammatory bowel disease (IBD) and colorectal cancer to diseases related to other systems such as liver diseases, psychological disorders, metabolic diseases, autoimmune diseases, and infectious diseases (Cho and Blaser 2012). In addition, recent epidemiological studies showed the potential role of gut microbiota in determining the effectiveness of immunotherapies in cancer treatment (Kroemer and Zitvogel 2018).

Understanding the factors that influence the composition of microbiome will help us decipher the biological mechanisms of the microbial modification to disease progression and elucidate the nature of host-microbe interactions, which may ultimately lead to novel therapies. Several population-based cross-sectional studies have identified external and intrinsic influences such as diet, environment, medication and host immune and metabolic status that in total can merely account for 10–20% of microbial diversity (Zhernakova et al. 2016; Kurilshikov et al. 2017). The vast majority of microbial diversity between individuals is unexplained. Hence,

This article is part of a Special Issue on 'Big Data' edited by Joshua WK Ho and Eleni Giannoulatou.

✉ Lanjuan Li
ljli@zju.edu.cn

- ¹ State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, School of Medicine, Zhejiang University, No.79 Qingchun Road, Hangzhou 310003, Zhejiang, China
- ² Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou, China
- ³ Victor Chang Cardiac Research Institute, Darlinghurst, NSW 2010, Australia
- ⁴ St Vincent's Clinical School, University of New South Wales, Sydney, NSW 2010, Australia
- ⁵ School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong, China

Fig. 1 Overview of host-gut microbiome interaction. This figure presents an example of the possible interactions between host and microbiome in the gut. External factors such as diet and medication internal factor such as host genome can influence the microbial composition. The alterations in the microbiome later affect the metabolites that they generate, some of which are essential elements for epigenetic regulation such as DNA methylation and histone modification. Changes in the epigenome will subsequently lead to altered gene expression and modified functions. The lumen environment where the microbiome inhabits may also change correspondingly. Overall, this figure emphasises the necessity to integrate multiple omics data in microbiome studies

recent studies started to focus on the interactions between the host genome, transcriptome, epigenome and metabolome, and the microbiome.

The Integrative Human Microbiome Project (iHMP) reported three longitudinal studies of genomic, epigenomic, transcriptomic and metabolomic data sets for host and microbes (iHMP Research Network Consortium 2014). However, mere collection of such large data sets is not enough because system biology, personalised medicine and different computational analysis methods need to be adopted for multi-omics data integration between host and microbiome. In this review, we summarise the application of different omic data sets in microbiota research, and also critically evaluate the analytical methods used in current studies. In addition, suggestions will also be offered for future host-microbiome multi-omics studies.

Omics data and their application to host-microbe studies

Metagenomics

Metagenomics studies the collective genome from all the microbial species in a given sample. Since the establishment of the Human Microbiome Project in 2008, next-generation sequencing approaches enable the comprehensive analysis of all microbes without the need of cultivation (Turnbaugh et al. 2007). The most widely applied methods are 16S ribosome

RNA (rRNA) gene sequencing (16Sseq) and shotgun metagenomics sequencing (MGS). 16Sseq uses marker genes such as 16S rRNA, which is a conserved region with hypervariable sequence to distinguish between different species and to estimate the composition of microbial taxa (Barko et al. 2018). More precise and also more expensive than 16Sseq, MGS is the untargeted sequencing of the whole genome in all the microorganisms in the sample.

For 16Sseq, the sequences are then clustered into operational taxonomic units (OTU) based on similarity and annotated at the taxonomic levels (phylum, class, order, family, genus and species) based on a representative sequence within each OTU (Nguyen et al. 2016). For MGS, the sequences are binned to taxonomic assignment directly based on the reference database such as BLAST and SEED (Breitwieser et al. 2017). Then the relative abundance can be calculated as the proportion of sequencing reads of a specific taxon to overall number of reads (Hua et al. 2015). Apart from relative abundance of taxa, microbiome diversity can also be characterised by alpha diversity which measures species richness (number of species) and evenness (distribution of species) within a community or beta diversity which represents differences in microbial composition among communities (Lozupone and Knight 2008). Based on whether the diversity measures are qualitative or quantitative, alpha diversity can be calculated using Shannon diversity, Chao index and Simpson index. Beta diversity can be calculated using Bray-Curtis dissimilarity, weighted UniFrac distance and unweighted UniFrac distance (Lozupone and Knight 2008). Overall, beta diversity indices

are less vulnerable to temporal changes and reflect a more complete structure of the microbial community. Hence, they are more widely used in microbiome studies (Abdul-Aziz et al. 2016).

Host genomics

The host genome is the complete set of DNA of the human host in which a microbial sample is taken from. A genome comprises nucleic acid sequences from cell nuclei and mitochondria, and includes protein-coding and noncoding DNA (1000 Genomes Project 2012). Recent studies have highlighted the influence of host genetics (mainly coding sequence) on the inhabited microbial composition, relating to complex diseases, immune system, dietary response and metabolism. For instance, significant heritability of both overall microbial composition and individual bacteria was confirmed in a large TwinsUK cohort (Goodrich et al. 2016).

Genome-wide association study (GWAS) is the study of association between millions of single nucleotide polymorphisms (SNPs) and diseases or traits of interest using microarray or chip-based approaches. Microbiome genome-wide association study (mGWAS), which is the combination of host genetic information (genotyping or whole-genome sequencing) with microbiome data (16S rRNA or metagenome sequencing), is an emerging field (Hall et al. 2017). In the GWAS setting, similar to expression quantitative trait loci, the microbial abundance-associated genetic variants are termed as microbiota quantitative trait loci (microbiotaQTLs or mbQTL) (Imhann et al. 2018). In one of the first mGWAS, genetic variants in the nucleotide-binding oligomerization domain-containing protein 2 gene (NOD2) were significantly associated with the abundance of *Enterobacteriaceae* in IBD (Knights et al. 2014). Gene-diet-microbiome interaction was also reported with the association between the lactase gene (LCT) and *Bifidobacterium* abundance in three independent Dutch cohorts (Bonder et al. 2016). Moreover, a study from two German cohorts indicated that the variants at VDR gene which encodes the vitamin D receptor influence bacterial taxa and thus the bile acid metabolism (Wang et al. 2016). Apart from human studies, the gut microbiota was also found to differ between diverse inbred strains of mice with multiple genetic backgrounds (Org et al. 2015).

Taken together, it is strongly supported that the host genetics plays an important role in defining the microbiome composition. Thus, improving our understanding of this interaction will help us unravel the pathogenesis of disease and solve the causality of the host-microbe interaction.

Host epigenomics

Epigenomics refers to heritable and adaptive regulation of gene expression without a change in DNA sequence

(Rakyan et al. 2011). The most studied epigenomic modification is DNA methylation, which happens at the 5' cytosine of the cytosine–guanine dinucleotides (CpG sites) and is catalysed by DNA methyltransferase (DNMT) (Moore et al. 2013). DNA methylation is involved in gene expression regulation by inhibiting transcription factor binding (Moore et al. 2013). There are two main DNA methylation profiling technologies: sequencing-based techniques such as the whole-genome bisulfite sequencing (WGBS) and reduced representation bisulphite sequencing (RRBS); and array-based techniques such as the HumanMethylation450 Beadchip (450K). The genome-wide methylation level can be measured by its beta value which is a continuous variable ranging from 0 to 1 and represents the ratio of intensities between methylated reads and the overall number of reads.

Due to the fact that epigenome can be altered by external factors, many studies have shown the potential role of microbiota for epigenomic change. The host microbiota influences DNA methylation by altering the enzymes' activity and by affecting the substrate level needed for the biological reaction (Qin and Wade 2018). Thus, the epigenome may act as a mediator to link external environmental exposures and internal microbial changes to disease-related genetic and expression alterations (Kellermayer 2017).

Host transcriptomics

The host transcriptome consists of all the RNA transcripts that are transcribed from the host genome. The two common techniques for transcriptomic profiling are microarrays and RNA sequencing (RNA-seq) (Lowe et al. 2017). Integrative transcriptome-metagenome analysis is a burgeoning research field and has a significant potential to elucidate the molecular mechanism and the impact of microbial variation on gene expression, which can act as a functional validation of gene-microbe associations (Zhao and Johnson 2018).

One recent study demonstrated a framework to combine the host transcriptome and microbial sequencing data for the study of IBD aetiology (Morgan et al. 2015). It is found that inflammation-related host transcripts are inversely associated with the abundance of *Akkermansia* and *Bifidobacteria* which are beneficial and anti-inflammatory microbes. Positive association was observed between oral commensal bacteria such as *Megasphaera* and *Veillonella* and cancer-related ERK/MAPK, PI3K/AKT pathways in lung cancer patients (Tsay et al. 2018). Host immune response to opportunistic infections in asthmatic children was also reflected by differential gene expression and high *Moraxella catarrhalis* abundance in the disease group (Castro-Nallar et al. 2015). Hence, the microbiome and transcriptome studies confirm the functional modification role of microbes and also point out the necessity of including host transcriptome datasets for the study of host-microbe interaction.

Host metabolomics

The metabolome is the unbiased measurement of all the existing metabolites in a complex biological system using analytical platforms such as gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS) and ^1H Nuclear Magnetic Resonance (NMR) spectroscopy (Tian et al. 2018). The sample type varies from tissue, urine, serum to faeces depending on the aim of the study. After chromatographic peak alignment together with retention time score matching, the metabolite abundance in each sample can be quantified (Nugent et al. 2014).

As the final physiological response to internal or external changes, the host metabolome has been extensively studied in the crosstalk with microbiome in human and animal models. For instance, insulin resistance can be impacted by gut microbiome *Prevotella copri* and *Bacteroides vulgatus*, which increased the branched-chain amino acids (BCAA) concentration (Pedersen et al. 2016). Strong association between microbial genera *Bacteroides intestinalis* and *Ruminococcus* with free fatty acids and glycerol in the stool was revealed in colorectal cancer patients (Weir et al. 2013). In animal models, the *Streptococcaceae* and *Desulfovibrionaceae* families showed strong positive correlations with the level of fasting glucose in plasma in mice fed with high-fat and high-sucrose diet (Kreznar et al. 2017). Reduced secondary bile acid and increased *Ruminococcus* were shown in a longitudinal cohort of infants with genetic predisposition to type 1 diabetes (T1D) (Kostic et al. 2015). Thus, the functional influence of the microbiome on metabolic pathways offers possible explanation for the significant association between microbiome composition and disease.

Analytical methodologies

Integrating the metagenome and host genome

Prior to the integration of the metagenome and host genome, the first step is to perform a heritability analysis on the metagenomic data. Heritability is the proportion of the total phenotypic variation that can be explained by genetic variation among individuals in a population (Ge et al. 2017). This is necessary to ensure that taxa are heritable and reduce the number of microbial traits for further analysis. The heritability of the microbial taxa can be estimated by the classic ACE model, which partitions the variance components into genetic variance (A), common environmental variance (C) and unique environmental variance (E) (Goodrich et al. 2014). The heritability is defined as the ratio of genetic variance to the total variance. The intraclass correlation coefficients (ICCs) can also be compared between monozygotic (MZ) and dizygotic (DZ) twins groups to quantify the genetic impact on gut

microbiome (Goodrich et al. 2014; Xie et al. 2016). The heritability analysis has been implemented in different software (Table 1).

Following the heritability analysis, the host gene-microbe association can be tested at both candidate gene and genome-wide level. The candidate gene strategies include testing gene variants, which are related with known diseases, bacterial interaction and/or possible mechanism such as immune regulation. For instance, in one study, the tests were limited to 163 IBD risk-associated SNPs (Turpin et al. 2016). Though relatively straightforward, such a method fails to take advantage of the unbiased genome-wide nature of the data, and might neglect other potentially relevant genes (Bonder et al. 2016). Another targeted approach is to focus on functional genic regions, neglecting the effect of other intergenic regions, some of which also have regulatory effect of genes nearby. In one study, 33,814 protein-coding SNPs were selected after annotation using ANNOVAR (Igartua et al. 2017). Similar to other genome-wide scaled studies, common SNPs with minor allele frequency (MAF) larger than 1% or 5% were used for the analysis.

For the genome-wide approach, dimensionality reduction is performed to reduce the number of tests performed when studying the association between host genome (i.e. SNPs or kinship coefficients to measure genetic similarity) and the microbial traits (i.e. relative abundance, alpha diversity or beta diversity) (Igartua et al. 2017). Although it mitigates the burden of multiple testing, researchers are unable to pinpoint the impact of host genetics on specific microbes. Dimensionality reduction methods such as principal component analysis (PCA) and principal co-ordinate analysis (PCoA) (which is also known as multi-dimensional scaling, MDS) were used in several studies (Table 1, Fig. 2a) (Igartua et al. 2017). Both PCA and PCoA capture the variance within a high-dimensional data set with a linear combination of uncorrelated variables using singular value decomposition (SVD) (Jolliffe and Cadima 2016). However, PCoA, which is a SVD on the distance matrix, is more frequently used in microbiome data (Meng et al. 2016). As a common issue in dimensionality reduction, the biological interpretation of the results can be difficult. Some computational feature selection methods have also been applied in some studies to reduce the number of variables in subsequent analyses. A pipeline named HOMINID has been developed to detect host genetic variation that is correlated with microbial composition based on a penalised Lasso regression model (Table 1) (Lynch et al. 2017).

Next, the association between host genome and microbiome composition can be tested by simple correlation-based analyses such as Spearman's correlation and Pearson correlation for SNP-OTU pairs, PCs-OTU pairs or SNP-diversity pairs (Table 1, Fig. 2b) (Bonder et al. 2016; Igartua et al. 2017). In addition, the generalised linear mixed

Table 1 Summary of analytical methods and software for host-microbiome integration

	Method ^a	Software
Dimensionality reduction	PCA	R(prcomp), QIIME (Caporaso et al. 2010), EIGENSTRAT (Price et al. 2006)
	PCoA	ade4 (Dray and Dufour 2017)
	FA	ASSIGN (Shen et al. 2015)
Correlation-based approach	Spearman, Pearson	Lme4 (Bates et al. 2015)
	CCA,CIA,PA	VEGAN (Dixon 2003) QIIME (Caporaso et al. 2010) ade4 (Dray and Dufour 2017)
Regression-based approach	GLMM	PLINK (Purcell et al. 2007), microbiomeGWAS (Hua et al. 2015), VEGAN, GEMMA (Zhou and Stephens 2012), FaST-LMM (Lippert et al. 2011), MaAsLin, Matrix eQTL (Shabalin 2012), limma (Ritchie et al. 2015)
	PLS	compPLS (Ramanan et al. 2016), O2-PLS (Trygg and Wold 2003)
	LASSO	HOMINID (Lynch et al. 2017)
	Kernal association test	MiRKAT (Zhao et al. 2015)
Network-based approach	Transkingdom network	TransNetDemo (Rodrigues et al. 2018)
	WGCNA	VEGAN,WGCNA (Langfelder and Horvath 2008)
	GEM	igraph (Csardi and Nepusz), GeneNet (Schaefer et al. 2015), MIMOSA (Noecker et al. 2016), ModelSEED (Devoid et al. 2013)

^a PCA principle component analysis, PCoA principle co-ordinate analysis, FA factor analysis, CCA canonical correlation analysis, CIA co-inertia analysis, PA procrustes analysis, GLMM generalised linear mixed model, PLS partial least square, LASSO least absolute shrinkage and selection operator, WGCNA the weighted gene co-expression network analysis, GEM genome-scale metabolic models

model (GLMM), which is the standard method for conducting GWAS, can be used to identify significant association between host genetic variations and microbial traits, using commonly used analytical tools such as PLINK (Table 1) (Purcell et al. 2007; Blekhman et al. 2015; Davenport et al. 2015; Goodrich et al. 2016; Wang et al. 2016; Igartua et al. 2017). The benefit of GLMM is that it can take into account not only the relatedness between individuals, but also other non-genetic confounders. The diversity matrix, PCs and/or the normalised OTU counts (relative taxa abundance), SNP abundance or genetic risk alleles can be used as dependent or independent variables in the model (Fig. 2c). Apart from the commonly used GLMM and removing all the zero counts, various improvements have also been made to the model to deal with zero counts in low-level taxon. The summary of the tools that can be applied is shown in Table 1. GLMM was applied with a negative binomial distribution to better characterise the right-skewed distribution of the zero-inflated metagenomic data (Wang et al. 2016). Additionally, mixed two-part log-normal and logistic model was applied in some studies. When the percentage of zero counts in one bacterial taxon is more than 15% among the samples, a logistic model was used (Knights et al. 2014; Turpin et al. 2016). In the logistic model, the rare taxa were counted as binary traits (present/absent). When the zero count percentage is lower than 5%, the GLMM is used. When the zero proportion falls between 5% and 15%, the model will be selected using quasi-likelihood statistics (QIC) (Pan 2001). To adjust for the skewness and kurtosis of score statistics, a tool called

microbiomeGWAS was designed to find association between microbiome beta diversity and host genetic variants (Hua et al. 2015; Goodrich et al. 2016). MicrobiomeGWAS is computationally efficient and generates results with high confidence (Hua et al. 2015). A more sophisticated model called factored spectrally transformed linear mixed model (Fast-LMM) has been proposed for GWAS and also has been applied in mGWAS (Org et al. 2015; Rothschild et al. 2018). This method conducts a single spectral decomposition for all SNPs, leading to reduced run time and memory footprint (Lippert et al. 2011). Apart from linear regression methods, a software tool named MiRKAT based on kernel-based association test was also performed in one study. Though relatively computationally inefficient, MiRKAT is able to decipher the complex relationship between microbial diversity and host SNPs on genome-wide scale (Zhao et al. 2015; Abdul-Aziz et al. 2016).

Integrating the metagenome and host epigenome

Due to the lack of omics level integration studies in this category, the methods are limited to unsupervised clustering analysis based on correlation matrix between predefined groups. For instance, Pearson's correlation analysis was performed between age-associated differentially methylated regions (DMRs) and age-associated microbial genera during postnatal paediatric development in human (Harris et al. 2016). In a pilot study, unsupervised hierarchical clustering analysis of DNA methylation profile was conducted between two groups with different dominant microbial taxa (Kumar

Fig. 2 Key analytical methods for host-microbiome integration. This figure illustrates the graphical summary of the key analytical methods that have been adopted in the papers mentioned in this review. **a** Demonstrates the main dimensionality reduction approach which captures the variance/covariance of the high-dimensional data with low dimensional data set. **b** Explains the univariate and multivariate correlation analysis that has been used. CCA extracts linear combination of each data set separately, which will maximise their correlation. CIA identifies maximum covariance between data sets. PA is mostly used for visualisation of the multi-omics data integration and complement CCA or CIA. **c** Shows the equation of regression analysis with dependent and independent variables. The diversity matrix, PCs and/or the normalised OUT counts (relative taxa abundance), SNP abundance or genetic risk alleles can be used as dependent or independent variables in the model (X, y in the equation). Z is a random effect that incorporates the relatedness between individuals and also other non-genetic confounders. ε is the error term. $f(Z)$ is the

kernel function which is able to model OTU-OTU interaction, non-linear association or incorporate phylogenetic relationship. **d** Shows the network analysis. The nodes in the network can represent the microbial taxa or microbial genes and host transcripts or metabolites, and the edge shows the relationships (correlations, physical interactions, causal relationships, etc.) between nodes. WGCNA identifies the co-clustered gene sets (modules) for subsequent analysis based on weighted correlations between OTUs, transcripts or metabolites. TransNet network and unsupervised correlation network reconstruct the network based on Gene-OUT Spearman correlation and OTU-OTU pairwise Spearman correlations with/without the incorporation of existing public gene-gene networks. GEM predicts the metabolic potential of multiple microbial species from metabolic reactions and/or enzyme functions annotated from high-throughput metagenomic information based on literature and public databases

et al. 2014). In another study, the methylation level of selected colorectal cancer and age-related genes was compared using an unsupervised hierarchical clustering approach between the *Fusobacterium*-high group and the *Fusobacterium*-low group in ulcerative colitis patients (Tahara et al. 2017). In the same study, genome-wide differential methylation analysis was also conducted in a subset of the original two groups using Student's *t* test which generates inference with lower power and less stability especially in such small sample size (Li et al. 2015; Tahara et al. 2017). Though both the candidate and genome-scale approaches show the differential methylation status between groups with different microbial abundance, the results require further investigation with larger sample size and more sophisticated analytical method. The drawback of these studies is their relatively small sample size and the microbial species are not investigated at the individual level.

Integrating the metagenome and host transcriptome

To study host transcript-microbe relationships, dimensionality reduction can be performed in both the metagenomic and host transcriptomic data. Similar to the methods used in mGWAS, unsupervised PCA has been used in transcriptome and metagenome integration (Table 1). In another study, Bayesian factor analysis (FA) was conducted on *Moraxella catarrhalis* infection-associated gene expression signatures between asthma cases and control samples (Castro-Nallar et al. 2015). The difference between PCA and FA is that FA also captures the covariance between variables with latent factors (Fig. 2a) (Conti et al. 2014). Moreover, Bayesian FA allows the incorporation of prior knowledge in the model.

Next, correlation-based approaches and regression models have been adopted to investigate host transcript-microbe

associations (Table 1). Other than the univariate correlation analysis mentioned above, some studies employed multivariate correlation methods such as canonical correlation analysis (CCA) (Schwartz et al. 2012). CCA extracts linear combinations of each data set separately, which will maximise their correlation (Schwartz et al. 2012). The advantage of this method is that it takes into consideration the underlying multivariate relationship in a data set. However, the underlying assumptions of this method, linear independence and a small number of variables, are often not met in practise. Regression approaches, such as GLMM, have also been applied to microbial taxa and transcripts with the adjustment to possible confounding factors (Morgan et al. 2015). One study uses a multivariate linear model to fit the microbiome and host transcriptome data after unsupervised dimensionality reduction in a stepwise way (de Steenhuijsen Piters et al. 2016). Additionally, sparse and compositionally robust partial least squares regression (CompPLS) has been also used for simultaneous variable selection and multi-omics data integration (Table 1) (Ramanan et al. 2016; Zhang et al. 2018). By adding sparsity into the loading vectors of the projection of the high-dimensional data, CompPLS manages to produce easily interpretable results in a computationally efficient way (Chun and Keleş 2010).

Despite the computational intensity, network-based analysis is a valuable tool because it is an effective way to summarise, analyse and visualise the dependency among variables, such as relationships between microbial compositions and gene expressions. Nodes in a network can represent the microbial taxa, microbial genes and host transcripts, and the edge shows the relationships (correlations, physical interactions, causal relationships, etc.) between nodes (Fig. 2d). In general, a network can be constructed in a data-driven fashion using a variety of data or knowledge-driven approaches. The Weighted Gene Co-expression Network Analysis (WGCNA) is a data-driven method, which discovers the co-clustered gene sets (modules) based on weighted correlations between OTUs or transcripts (Table 1, Fig. 2d). It has been widely applied in transcriptional data sets to identify gene modules and intra-modular hub genes by pairwise correlation and hierarchical clustering (Fig. 2d) (Langfelder and Horvath 2008). The eigengenes of the modules can be subsequently used to study the crosstalk between functional transcripts and microbial traits (Molyneaux et al. 2017; Huang et al. 2017). Hence, WGCNA not only alleviates the multiple testing issue but also detects the responsible transcripts in the informative modules (Langfelder and Horvath 2008). Another newly developed method called transkingdom network (TransNet) analysis has been used to model host-microbiome interaction. As a knowledge-driven approach, it involves network reconstruction based on the existing public gene-gene network with the integration of Gene-OTU Spearman correlation and OTU-OTU pairwise Spearman correlations (Table 1, Fig. 2d)

(Dong et al. 2015; Morgun et al. 2015; Lam et al. 2018). This method is able to identify specific microbial taxa or even microbial genes affecting host gene expression.

Integrating the metagenome and host metabolome

As a popular approach in host-microbe interaction research, the first step is also dimensionality reduction for both the metagenome and host metabolome. Similar approaches as mentioned above can also be applied in this setting (Nugent et al. 2014; Chong and Xia 2017).

Univariate correlation, multivariate correlation and regression-based approaches have been applied to the reduced microbial and metabolite features to study the microbe-metabolite interactions (Chierico et al. 2017; Zierer et al. 2018). Apart from the aforementioned multivariate correlation analysis CCA, co-inertia analysis (CIA) and procrustes analysis (PA) have also been adopted in some studies. CIA identifies maximum covariance between data sets from the PCs (McHardy et al. 2013; Hill et al. 2017). PA is mostly used for visualisation of the multi-omics data integration and complements CCA or CIA (Gower 1975; McHardy et al. 2013). Even though the multivariate approaches are capable of analysing the interaction between and within data sets, the results are likely to be difficult to interpret.

Additionally, network-based methods have been widely applied to study microbe-metabolite interaction, among which the correlation-based network (WGCNA) and genome-scale metabolic models (GEMs) have been used (McHardy et al. 2013; Pedersen et al. 2016). In brief, GEMs rely on the metabolic reactions and/or enzymes functions annotated from high-throughput metagenomic information based on literature and database such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) and act as a scaffold in the integration of metabolomics and other omics data (Fig. 2d) (Shoae et al. 2013). Thus, a community-scale GEM explores the metabolic potential of multiple microbial species within a community (Larsen et al. 2011; Noecker et al. 2016). After the construction of GEMs, the competition, mutualism core and connectivity pattern can be calculated based on the topological structure of the GEM (Steinway et al. 2015). Moreover, a community-wide metabolic potential (CMP) score can be estimated to measure the relative metabolic creation or depletion capacity of a microbial community as well as identify key microbial species in the network (Noecker et al. 2016). The host-microbe interaction can also be estimated by comparing the predicted and measured metabolomics profiles (Noecker et al. 2016). Even though these network-based frameworks can incorporate prior biological knowledge and can generate more intuitive results for functional interpretation, they are unable to discover *de novo* relationships (Chong and Xia 2017).

Summary and future perspective

In summary, future microbiome analyses that integrate metagenomic data with data generated from the host genome, epigenome, transcriptome and metabolome will be important to uncover the microbiota's role in human health (Nakatsu et al. 2015). The methods that have been used to integrate host and microbiome data include dimensionality reduction methods, correlation-based approaches, regression-based approaches and network-based approaches. From our review, we note that the methods used to integrate microbiome and different host omics vary depending on the data types. However, except for the metabolic models, the methods used are not exclusive to specific omics data. For instance, WGCNA can also be applied to epigenome-microbiome integration. Similarly, PLS and LASSO can be applied in other host and microbiome data integration. Most of the methods are standard statistical methods and have been developed and used in host omics studies. Hence, more fit-for-purpose analytical methods to combine the high-dimensional data sets are required for future studies.

Moreover, efforts should be made to combine the results from multiple relevant studies to increase the power of inference which might be weak due to a small sample size. Recently, the MiBioGen consortium initiative has been announced to assemble 18 population-based microbiome GWAS (Wang et al. 2018). A standardised analytical pipeline is also proposed to process both microbiome data set and genotypes in an identical approach. However, such meta-analysis consortium for other host-microbe data integration between different studies is still lacking.

Although dimensionality reduction approaches are used as a preliminary step for downstream analysis and help correct multiple testing due to the high dimension, these methods hinder the interpretability of the results and fail to take advantage of the unbiased and untargeted omics approaches. Thus, machine learning methods such as support vector machines (SVM) can be applied to help the feature selection step to reduce the false positive rate.

Currently, most studies use linear mixed models and correlation-based approaches, which only capture the linear relationship between variables. Future exploration with kernel-based approaches will help us unveil the complex non-linear relationship in host-microbe interaction.

Among all the current methods, network-based analysis is the most biologically interpretable one. A Bayesian network model is another promising approach to model the probabilistic relationship between multiple omics as well as revealing the most probabilistic network from this graphical model (Wang et al. 2013; McGeachie et al. 2016). A constraint-based model such as flux balance analysis can also be performed to simulate metabolite flow and predict microbial

growth but its current utility is still limited to small microbial communities (Chong and Xia 2017).

The microbiome is a complex component that inhabits the host. Previous studies investigating host-microbiome interaction with multi-omics data are promising and most are health-related. Future acquisition of host omics information and metagenome data with larger sample size, together with the development and incorporation of more sophisticated analytical methods, will open opportunities to examine the mechanism of disease progression and facilitate the adoption of personalised medicine.

Compliance with ethical standards

Funding Funding was provided by the National Natural Science Foundation of China (grant nos. 81330011, 81330014, 81790631, 81790633, 81570512 and 81121002), the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (grant no. 81721091), and the National Basic Research Program of China (973 program) (grant no. 2013CB531401). This work was also supported in part by funds from the New South Wales Ministry of Health, a National Health and Medical Research Council Career Development Fellowship (1105271 to JWKH) and the National Heart Foundation (Future Leader Fellowship 100848 to JWKH and 101204 to EG).

Conflict of interest Qing Wang declares that she has no conflict of interest. Kaicen Wang declares that she has no conflict of interest. Wenrui Wu declares that she has no conflict of interest. Eleni Giannoulatou declares that she has no conflict of interest. Joshua W. K. Ho declares that he has no conflict of interest. Lanjuan Li declares that he/she has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Abdul-Aziz MA, Cooper A, Weyrich LS (2016) Exploring relationships between host genome and microbiome: new insights from genome-wide association studies. *Front Microbiol* 7:1611. <https://doi.org/10.3389/fmicb.2016.01611>
- Barko PC, McMichael MA, Swanson KS, Williams DA (2018) The gastrointestinal microbiome: a review. *J Vet Intern Med* 32:9–25. <https://doi.org/10.1111/jvim.14875>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw*. <https://doi.org/10.18637/jss.v067.i01>
- Blekhman R, Goodrich JK, Huang K et al (2015) Host genetic variation impacts microbiome composition across human body sites. *Genome Biol* 16:191. <https://doi.org/10.1186/s13059-015-0759-1>
- Bonder MJ, Kurilshikov A, Tigchelaar EF et al (2016) The effect of host genetics on the gut microbiome. *Nat Genet* 48:1407–1412. <https://doi.org/10.1038/ng.3663>
- Breitwieser FP, Lu J, Salzberg SL (2017) A review of methods and databases for metagenomic classification and assembly. *Brief Bioinform*. <https://doi.org/10.1093/bib/bbx120>

- Caporaso JG, Kuczynski J, Stombaugh J et al (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>
- Castro-Nallar E, Shen Y, Freishtat RJ et al (2015) Integrating microbial and host transcriptomics to characterize asthma-associated microbial communities. *BMC Med Genet* 8:50. <https://doi.org/10.1186/s12920-015-0121-1>
- Chierico FD, Nobili V, Vernocchi P et al (2017) Gut microbiota profiling of pediatric nonalcoholic fatty liver disease and obese patients unveiled by an integrated meta-omics-based approach. *Hepatology* 65: 451–464. <https://doi.org/10.1002/hep.28572>
- Cho I, Blaser MJ (2012) The human microbiome: at the interface of health and disease. *Nat Rev Genet* 13:260–270. <https://doi.org/10.1038/nrg3182>
- Chong J, Xia J (2017) Computational approaches for integrative analysis of the metabolome and microbiome. *Metabolites* 7(4):62. <https://doi.org/10.3390/metabo7040062>
- Chun H, Keleş S (2010) Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J R Stat Soc Ser B Stat Methodol* 72:3–25. <https://doi.org/10.1111/j.1467-9868.2009.00723.x>
- Conti G, Frühwirth-Schnatter S, Heckman JJ, Piatek R (2014) Bayesian exploratory factor analysis. *J Econom* 183:31–57. <https://doi.org/10.1016/j.jeconom.2014.06.008>
- Davenport ER, Cusanovich DA, Michelini K et al (2015) Genome-wide association studies of the human gut microbiota. *PLoS One* 10: e0140301. <https://doi.org/10.1371/journal.pone.0140301>
- Devoid S, Overbeek R, DeJongh M et al (2013) Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. *Methods Mol Biol* 985:17–45. https://doi.org/10.1007/978-1-62703-299-5_2
- Dixon P (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* 14:927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>
- Dray S, Dufour A-B (2017) The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw.* <https://doi.org/10.18637/jss.v022.i04>
- de Steenhuijsen P, WAA, Heinonen S, Hasrat R et al (2016) Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. *Am J Respir Crit Care Med* 194:1104–1115. <https://doi.org/10.1164/rccm.201602-0220OC>
- Dong X, Yambartsev A, Ramsey SA et al (2015) Reverse enGENEering of regulatory networks from big data: a roadmap for biologists. *Bioinforma Biol Insights* 9:61–74. <https://doi.org/10.4137/BBI.S12467>
- Ge T, Chen C-Y, Neale BM et al (2017) Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet* 13:e1006711. <https://doi.org/10.1371/journal.pgen.1006711>
- Genomes Project (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65. <https://doi.org/10.1038/nature11632>
- Gilbert JA, Blaser MJ, Caporaso JG et al (2018) Current understanding of the human microbiome. *Nat Med* 24:392–400. <https://doi.org/10.1038/nm.4517>
- Goodrich JK, Waters JL, Poole AC et al (2014) Human genetics shape the gut microbiome. *Cell* 159:789–799. <https://doi.org/10.1016/j.cell.2014.09.053>
- Goodrich JK, Davenport ER, Beaumont M et al (2016) Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* 19: 731–743. <https://doi.org/10.1016/j.chom.2016.04.017>
- Gower JC (1975) Generalized procrustes analysis. *Psychometrika* 40:33–51. <https://doi.org/10.1007/BF02291478>
- Hall AB, Tolonen AC, Xavier RJ (2017) Human genetic variation and the gut microbiome in disease. *Nat Rev Genet* 18:690–699. <https://doi.org/10.1038/nrg.2017.63>
- Harris RA, Shah R, Hollister EB et al (2016) Colonic mucosal epigenome and microbiome development in children and adolescents. *J Immunol Res* 2016:9170162. <https://doi.org/10.1155/2016/9170162>
- Hill CJ, Lynch DB, Murphy K et al (2017) Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET cohort. *Microbiome* 5:4. <https://doi.org/10.1186/s40168-016-0213-y>
- Hua X, Song L, Yu G, et al (2015) MicrobiomeGWAS: a tool for identifying host genetic variants associated with microbiome composition. <https://doi.org/10.1101/031187>
- Huang Y, Ma S-F, Espindola MS et al (2017) Microbes are associated with host innate immune response in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 196:208–219. <https://doi.org/10.1164/rccm.201607-1525OC>
- Igartua C, Davenport ER, Gilad Y et al (2017) Host genetic variation in mucosal immunity pathways influences the upper airway microbiome. *Microbiome* 5:16. <https://doi.org/10.1186/s40168-016-0227-5>
- iHMP Research Network Consortium (2014) The integrative human microbiome project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. *Cell Host Microbe* 16:276–289. <https://doi.org/10.1016/j.chom.2014.08.014>
- Imhann F, Vila AV, Bonder MJ et al (2018) Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* 67:108–119. <https://doi.org/10.1136/gutjnl-2016-312135>
- Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. *Phil Trans R Soc A* 374:20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Kellermayer R (2017) Challenges for epigenetic research in inflammatory bowel diseases. *Epigenomics* 9:527–538. <https://doi.org/10.2217/epi-2016-0155>
- Knights D, Silverberg MS, Weersma RK et al (2014) Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med* 6:107. <https://doi.org/10.1186/s13073-014-0107-1>
- Kostic AD, Gevers D, Siljander H et al (2015) The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* 17:260–273. <https://doi.org/10.1016/j.chom.2015.01.001>
- Kreznar JH, Keller MP, Traeger LL et al (2017) Host genotype and gut microbiome modulate insulin secretion and diet-induced metabolic phenotypes. *Cell Rep* 18:1739–1750. <https://doi.org/10.1016/j.celrep.2017.01.062>
- Kroemer G, Zitvogel L (2018) Cancer immunotherapy in 2017: the breakthrough of the microbiota. *Nat Rev Immunol* 18:87–88
- Kumar H, Lund R, Laiho A et al (2014) Gut microbiota as an epigenetic regulator: pilot study based on whole-genome methylation analysis. *mBio* 5:e02113–e02114. <https://doi.org/10.1128/mBio.02113-14>
- Kurilshikov A, Wijmenga C, Fu J, Zhernakova A (2017) Host genetics and gut microbiome: challenges and perspectives. *Trends Immunol* 38:633–647. <https://doi.org/10.1016/j.it.2017.06.003>
- Lam KC, Vyshenska D, Hu J et al (2018) Transkingdom network reveals bacterial players associated with cervical cancer gene expression program. *PeerJ* 6:e5590
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* 9:559. <https://doi.org/10.1186/1471-2105-9-559>
- Larsen PE, Collart FR, Field D et al (2011) Predicted relative metabolomic turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. *Microb Inform Exp* 1: 4. <https://doi.org/10.1186/2042-5783-1-4>
- Li D, Xie Z, Le Pape M, Dye T (2015) An evaluation of statistical methods for DNA methylation microarray data analysis. *BMC Bioinf* 16:217. <https://doi.org/10.1186/s12859-015-0641-x>

- Lippert C, Listgarten J, Liu Y et al (2011) FaST linear mixed models for genome-wide association studies. *Nat Methods* 8:833–835. <https://doi.org/10.1038/nmeth.1681>
- Lowe R, Shirley N, Bleackley M et al (2017) Transcriptomics technologies. *PLoS Comput Biol* 13:e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>
- Lozupone CA, Knight R (2008) Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev* 32:557–578. <https://doi.org/10.1111/j.1574-6976.2008.00111.x>
- Lynch J, Tang K, Priya S et al (2017) HOMINID: a framework for identifying associations between host genetic variation and microbiome composition. *GigaScience* 6:1–7. <https://doi.org/10.1093/gigascience/gix107>
- McGeachie MJ, Sordillo JE, Gibson T et al (2016) Longitudinal prediction of the infant gut microbiome with dynamic Bayesian networks. *Sci Rep* 6:20359. <https://doi.org/10.1038/srep20359>
- McHardy IH, Goudarzi M, Tong M et al (2013) Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome* 1:17. <https://doi.org/10.1186/2049-2618-1-17>
- Meng C, Zeleznik OA, Thallinger GG et al (2016) Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief Bioinform* 17:628–641. <https://doi.org/10.1093/bib/bbv108>
- Molyneux PL, Willis-Owen SAG, Cox MJ et al (2017) Host–microbial interactions in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 195:1640–1650. <https://doi.org/10.1164/rccm.201607-1408OC>
- Moore LD, Le T, Fan G (2013) DNA methylation and its basic function. *Neuropsychopharmacology* 38:23–38. <https://doi.org/10.1038/npp.2012.112>
- Morgan XC, Kabackchiev B, Waldron L et al (2015) Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biol* 16:67. <https://doi.org/10.1186/s13059-015-0637-x>
- Morgun A, Dzutsev A, Dong X et al (2015) Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks. *Gut* 64:1732–1743. <https://doi.org/10.1136/gutjnl-2014-308820>
- Nakatsu G, Li X, Zhou H et al (2015) Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun* 6:8727. <https://doi.org/10.1038/ncomms9727>
- Nguyen N-P, Warnow T, Pop M, White B (2016) A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *Npj Biofilms and Microbiomes* 2:16004. <https://doi.org/10.1038/npjbiofilms.2016.4>
- Noecker C, Eng A, Srinivasan S et al (2016) Metabolic model-based integration of microbiome taxonomic and metabolomic profiles elucidates mechanistic links between ecological and metabolic variation. *mSystems* 1:e00013–e00015. <https://doi.org/10.1128/mSystems.00013-15>
- Nugent JL, McCoy AN, Addamo CJ et al (2014) Altered tissue metabolites correlate with microbial dysbiosis in colorectal adenomas. *J Proteome Res* 13:1921–1929. <https://doi.org/10.1021/pr4009783>
- Org E, Parks BW, Joo JWW et al (2015) Genetic and environmental control of host-gut microbiota interactions. *Genome Res* 25:1558–1569. <https://doi.org/10.1101/gr.194118.115>
- Pan W (2001) Akaike's information criterion in generalized estimating equations. *Biometrics* 57:120–125
- Pedersen HK, Gudmundsdottir V, Nielsen HB et al (2016) Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 535:376–381. <https://doi.org/10.1038/nature18646>
- Price AL, Patterson NJ, Plenge RM et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909. <https://doi.org/10.1038/ng1847>
- Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. <https://doi.org/10.1086/519795>
- Qin Y, Wade PA (2018) Crosstalk between the microbiome and epigenome: messages from bugs. *J Biochem (Tokyo)* 163:105–112. <https://doi.org/10.1093/jb/mvx080>
- Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 12:529–541. <https://doi.org/10.1038/nrg3000>
- Ramanan D, Bowcutt R, Lee SC et al (2016) Helminth infection promotes colonization resistance via type 2 immunity. *Science* 352:608–612. <https://doi.org/10.1126/science.aaf3229>
- Ritchie ME, Phipson B, Wu D et al (2015) limma powers differential expression analyses for RNAsequencing and microarray studies. *Nucleic Acids Res* 43:e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Rodrigues RR, Shulzhenko N, Morgun A (2018) Transkingdom networks: a systems biology approach to identify causal members of host-microbiota interactions. *Methods Mol Biol* 1849:227–242. https://doi.org/10.1007/978-1-4939-8728-3_15
- Rothschild D, Weissbrod O, Barkan E et al (2018) Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555:210–215. <https://doi.org/10.1038/nature25973>
- Schaefer J, Opgen-Rhein R, Strimmer K (2015) GeneNet: Modeling and Inferring Gene Networks. R package version 1.2.13 <https://CRAN.R-project.org/package=GeneNet>. Accessed 28 Dec 2018
- Schwartz S, Friedberg I, Ivanov IV et al (2012) A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response. *Genome Biol* 13:r32. <https://doi.org/10.1186/gb-2012-13-4-r32>
- Shabalin AA (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28:1353–1358. <https://doi.org/10.1093/bioinformatics/bts163>
- Shen Y, Rahman M, Piccolo SR et al (2015) ASSIGN: context-specific genomic profiling of multiple heterogeneous biological pathways. *Bioinformatics* 31:1745–1753. <https://doi.org/10.1093/bioinformatics/btv031>
- Shoae S, Karlsson F, Mardinoglu A et al (2013) Understanding the interactions between bacteria in the human gut through metabolic modeling. *Sci Rep* 3:2532. <https://doi.org/10.1038/srep02532>
- Steinway SN, Biggs MB, Jr TPL et al (2015) Inference of network dynamics and metabolic interactions in the gut microbiome. *PLoS Comput Biol* 11:e1004338. <https://doi.org/10.1371/journal.pcbi.1004338>
- Tahara T, Hirata I, Nakano N et al (2017) Potential link between Fusobacterium enrichment and DNA methylation accumulation in the inflammatory colonic mucosa in ulcerative colitis. *Oncotarget* 8: 61917–61926. <https://doi.org/10.18632/oncotarget.18716>
- Tian Y, Nichols RG, Cai J et al (2018) Vitamin A deficiency in mice alters host and gut microbial metabolism leading to altered energy homeostasis. *J Nutr Biochem* 54:28–34. <https://doi.org/10.1016/j.jnutbio.2017.10.011>
- Trygg J, Wold S (2003) O2-PLS, a two-block (X–Y) latent variable regression (LVR) method with an integral OSC filter. *J Chemom* 17: 53–64. <https://doi.org/10.1002/cem.775>
- Tsay J-CJ, Wu BG, Badri MH et al (2018) Airway microbiota is associated with up-regulation of the PI3K pathway in lung cancer. *Am J Respir Crit Care Med*. <https://doi.org/10.1164/rccm.201710-2118OC>
- Turnbaugh PJ, Ley RE, Hamady M et al (2007) The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* 449:804–810. <https://doi.org/10.1038/nature06244>
- Turpin W, Espin-Garcia O, Xu W et al (2016) Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat Genet* 48:1413–1417. <https://doi.org/10.1038/ng.3693>
- Wang W, Baladandayuthapani V, Morris JS et al (2013) iBAG: integrative Bayesian analysis of high-dimensional multiplatform genomics

- data. *Bioinformatics* 29:149–159. <https://doi.org/10.1093/bioinformatics/bts655>
- Wang J, Thingholm LB, Skiecevičienė J et al (2016) Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat Genet* 48:1396–1406. <https://doi.org/10.1038/ng.3695>
- Wang J, Kurilshikov A, Radjabzadeh D et al (2018) Meta-analysis of human genome-microbiome association studies: the MiBioGen consortium initiative. *Microbiome* 6:101. <https://doi.org/10.1186/s40168-018-0479-3>
- Weir TL, Manter DK, Sheflin AM et al (2013) Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS One* 8:e70803. <https://doi.org/10.1371/journal.pone.0070803>
- Xie H, Guo R, Zhong H et al (2016) Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst* 3:572–584.e3. <https://doi.org/10.1016/j.cels.2016.10.004>
- Zhang X-S, Li J, Krautkramer KA et al (2018) Antibiotic-induced acceleration of type 1 diabetes alters maturation of innate intestinal immunity. *ELife Sci* 7:1–37. <https://doi.org/10.7554/eLife.37816>
- Zhao Y, Johnson WE (2018) Exploring host-microbe interactions in lung cancer. *Am J Respir Crit Care Med*. <https://doi.org/10.1164/rccm.201807-1225ED>
- Zhao N, Chen J, Carroll IM et al (2015) Testing in microbiome-profiling studies with MiRKAT, the microbiome regression-based kernel association test. *Am J Hum Genet* 96:797–807. <https://doi.org/10.1016/j.ajhg.2015.04.003>
- Zhernakova A, Kurilshikov A, Bonder MJ et al (2016) Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352:565–569. <https://doi.org/10.1126/science.aad3369>
- Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44:821–824. <https://doi.org/10.1038/ng.2310>
- Zierer J, Jackson MA, Kastenmüller G et al (2018) The fecal metabolome as a functional readout of the gut microbiome. *Nat Genet* 50:790–795. <https://doi.org/10.1038/s41588-018-0135-7>