Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Emerging computational tools and models for studying gut microbiota composition and function

Seo-Young Park[1,4], Arinzechukwu Ufondu[2,4], Kyongbum Lee[1] and Arul Jayaraman[2,3]

The gut microbiota and its metabolites play critical roles in human health and disease. Advances in high-throughput sequencing, mass spectrometry, and other omics assay platforms have improved our ability to generate large volumes of data exploring the temporal variations in the compositions and functions of microbial communities. To elucidate mechanisms, methods and tools are needed that can rigorously model the dependencies within time-series data. Longitudinal data are often sparse and unevenly sampled, and nontrivial challenges remain in determining statistical significance, normalization across different data types, and model validation. In this review, we highlight recent developments in models and software tools for the analysis of time series microbiome and metabolome data, as well as integration of these data.

Addresses
[1] Department of Chemical and Biological Engineering, Tufts University, Medford, MA, USA

[2] Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, USA
[3] Department of Biomedical Engineering, Texas A&M University, College Station, TX, USA

Corresponding authors:
Lee, Kyongbum (kyongbum.lee@tufts.edu), Jayaraman, Arul (arulj@tamu.edu)
[4] Equal contribution.

Current Opinion in Biotechnology 2020, **66**:301–311

This review comes from a themed issue on **Tissue, cell and pathway engineering**

Edited by **Li Tang**, **Peng Xu** and **Haoran Zhang**

For a complete overview see the Issue and the Editorial

Available online 25th November 2020

https://doi.org/10.1016/j.copbio.2020.10.005

Introduction

The gut microbiota is a complex system with dynamic interactions between the microorganisms, host, and environmental factors such as the host's diet. In humans, the gut microbiota comprises of about 10^{14} microorganisms belonging to more than 1000 species that perform diverse range of metabolic functions to impact physiology and health [1]. Alterations in the microbiota composition have been associated with inflammatory bowel disease (IBD) [2], nonalcoholic fatty liver disease [3], inflammatory arthritis [4], asthma [5], sepsis [6], autism spectrum disorder [7], and neurodegenerative diseases.

A common goal of microbiota studies has been to determine the impact of a treatment, host condition, or other factors on community composition at a few timepoints [8]. Increasingly, these studies have also included metabolomics assays as readouts on microbiota function. These 'snapshots' of the gut microbiota composition and function have provided valuable insights into what factors can perturb the microbiota, what compositional and metabolic changes occur in the microbiota due to the perturbation, and what consequences these changes have on the host's physiology. On the other hand, snapshots at a few timepoints offer limited information regarding the dynamics of compositional and metabolic changes, and thus are not well suited for inferring causal relationships between these changes and host physiology [9]. Gerber [10] provides germane scenarios, for example, a microbiome feature exhibiting periodicity, that illustrate the pitfalls of drawing conclusions from snapshot analysis.

To better understand the dynamics of gut microbiota and elucidate causal relationships, investigative approaches are needed that combine longitudinal study design with appropriate models and methods for time series data analysis. These approaches could, for example, determine whether a major functional changes of a microbiota (e.g. diminished capacity to metabolize conjugated bile salts [11]) can be driven by shifts in the abundances of a small set of taxa or whether this requires broader, community-wide dysbiosis. Another, related question is whether there is a dominant set of taxa that govern the dynamics of multiple metabolic functions at the community level. Importantly, studying the dynamics of a microbiota by integrating multi-omics (e.g. metagenomics and metabolomics) longitudinal data could distinguish between taxa that merely correlate in their abundance with a metabolic function versus taxa whose change in abundance causes the change in function. However, it is important to note that even with a longitudinal study design and rigorous statistical analysis, it is rare that causation can be unequivocally established without following up the omics analysis with hypothesis-driven experiments. Mathematical models can determine causality in a statistical sense, subject to the assumptions used to build the models. In this regard, longitudinal data and related analysis tools provide

Table 1

Tools for microbiome and metabolome time-series analysis

Techniques/model	Summary	Reference
Longitudinal differential abundance testing		
MEBA	Bayesian framework for evaluating the significance of longitudinal differences in metabolomic data profiles	[18,19]
GP2S test	Smooth splines are fitted to time-series data and is used to evaluate the difference between two sample groups	[21,22]
SS-ANOVA (Metaspline, MetaLonDA)	Smoothing spline ANOVA is used in conjunction with a permutation test to detect longitudinally differentially abundant microbiome features	[24]
Multitask learning/regularized multivariate regression	Machine learning approach to find longitudinal biomarkers from metabolomic time-series data	[28]
Temporal network generation		
gLV	Ecological modelling of interactions within a microbial community	[34]
Granger causality	Tests if one time-series depends on another time-series	[35,36••,47]
ARIMA modelling	Forecasting method based on autoregression	[38]
Dynamic Bayesian network (DBN)	Bayesian inference for time-series data	[39]
Seqtime	Classification scheme to determine fitness for time-series modelling of longitudinal microbial data	[41]
Temporal pattern clustering		
TCAP	Clustering method that takes into account inversions and time delays between different time-series profiles	[29]
DMM	Clustering algorithm for metagenomic data used to identify distinct community states	[44•]
MC-TIMME2	Bayesian framework for clustering microbiome features (e.g. taxa) having similar temporal dynamics	[45]
PCoA, MANOVA	Dimensionality reduction methods used to visualize OTUs grouped based on UniFrac distances	[46]
k-medoid clustering	Partitioning method for clustering; in the context of longitudinal microbiome data, the clusters group samples based on similarity measures of microbial dynamics	[47]
MetaboClust	An ensemble software platform with functions for clustering time series metabolomics data	[49]
MITRE	Bayesian supervised machine learning classifier; used to predict the host's physiological status from microbiome time-series data	[50•]

plausible candidate associations between metabolites, genes, and/or microbes that can be tested for cause-and-effect using carefully designed gain-of-function and loss-of-function experiments.

In this review, we highlight recent developments in tools for time-series analysis of metagenomic and metabolomics data and discuss approaches for integrating these and other omics data in the context of longitudinal microbiome studies.

Microbiome and metabolome time-series analysis

Single time point studies struggle to provide sufficient information on periodic patterns, interdependencies, or temporal variations present within the microbiome. In recent years, longitudinal study designs have gained in popularity, as continued improvement in DNA sequencing efficiency have made time-course studies more tractable [12–14]. Increasingly, these studies have also profiled metabolites produced within the intestine to obtain a more detailed picture of metabolic activities that have occurred [15–17].

Analyzing longitudinal data on microbiomes presents (Table 1) a unique set of challenges [18–20]. Microbiome data is characterized by sparsity. In a taxonomic profiling experiment, it is common that most taxa map to zero reads, but it is not clear if this is evidence that the taxa with zero reads is absent or instead indicates absence of evidence due to insufficient depth of coverage. Microbiome data is also compositional in that an increase in the relative abundance of a community member has to negatively correlate with the relative abundance of other members. Furthermore, the number of replicates is usually small relative to the number of variables (e.g. species) observed on a sample. For longitudinal studies, these challenges are compounded by statistical time dependencies (e.g. autocorrelations) between observations [21]. Furthermore, longitudinal studies are often complicated by limited time points, sample dropout, and uneven sampling frequency. As such, conventional approaches, (e.g. hypothesis testing methods from classical multivariate statistics), are insufficient to fully exploit the information content of longitudinal data. The remainder of this section describes three general approaches for significance testing and modeling of time series omics data from longitudinal studies: longitudinal differential

abundance testing, temporal network generation, and temporal pattern clustering.

Longitudinal differential abundance testing

One important question that can be addressed through a longitudinal study is whether an observed difference in microbial abundance or metabolite concentration between experimental groups having distinct phenotypes is continuous or transient. While there are a number of well-established univariate and multivariate methods to perform significance tests at individual time points within a time series, these methods cannot ascertain if/when significance is either gained or lost in between time points, and is likely to miss transient features [22]. Interpolation and smoothing methods have recently been proposed as options to overcome these limitations. These methods have been applied to both metataxonomic and metabolomic time-series data. To infer time periods where specific microbial features (e.g. taxa) are differentially abundant, *MetaDprof* uses smoothing spline ANOVA (ssANOVA) to fit a piecewise polynomial curve, that is, spline, through longitudinal data [22]. The splines are used to estimate a difference function that describes the time dependence of differential abundance for features. Integration of the difference function over a time interval of interest calculates an area statistic that is used to test if the differential abundance of a feature over the interval is significant.

Figure 1 illustrates the use of spline fitting to compare time-dependent changes in abundance of a bacterial species ('Feature 1' in the figure) in two experimental groups. MetaLonDA is a similar method, except that the feature abundances are assumed to follow a negative binomial distribution instead of a Gaussian distribution [23••]. Compared to MetaDprof, MetaLonDA showed better specificity and sensitivity in detecting time periods of differentially abundant features when evaluated on the same set of simulated data [23••]. Unlike earlier tools, MetaLonDA is able to handle uneven sample sizes and time intervals, which are common in human subject studies. Applied to infant stool microbiome data from the DIABIMMUNE project, MetaLonDA found significant differences between the Finnish and Russian cohorts in the relative abundances of Bacteroides and Bifidobacterium species during the first year of life, matching the conclusion of the original study [24]. Additionally, MetaLonDA showed that Bacteroides is significantly more abundant from day 96 to day 584 in Finnish infants, while Bifidobacterium is significantly more abundant in Russian infants from day 96 to day 720 [24]. Importantly, the smoothing spline method detected time periods of differential abundance for several genera that were not mentioned in the original study; these taxa could affect exposure of the infant to bacterial lipopolysaccharides and development of autoimmune diseases. It should be noted that ssANOVA methods are best suited to large data sets with frequent time points. When sample sizes are sparse over multiple time points, the resulting fitted splines have large variations. Another approach for longitudinal microbiome data analysis is to use generalized additive mixed effect models (GAMMs). These models typically compare the abundance of specific taxa across experimental groups. In addition to taxonomic abundance data, information about the subjects such as gender, age, and diet can also be incorporated into a GAMM. However, overfitting can be a concern, and it is recommended to use prior knowledge when determining which, if any, additional data should be included in the model. In a large longitudinal study, Fettweiss *et al.* [25] used a GAMM to investigate the differences in microbiota between women who gave preterm and full-term births. By incorporating subject information such as body mass index, ethnicity, and vaginal pH, the GAMM of this study accurately modeled the abundance differences of specific bacteria in the two groups.

Figure 1

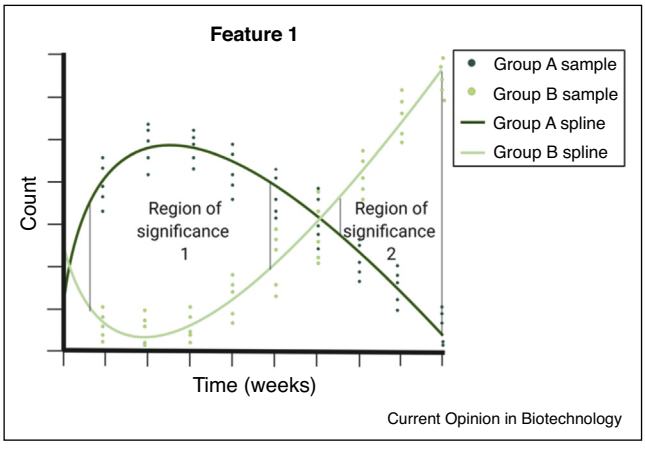


Illustration of using splines to model longitudinal microbiome or metabolome data. Splines, or piecewise polynomial parametric curves, are fit to the feature abundances from two experimental groups. The splines are used to calculate a difference function, which can be integrated over a time interval of interest. The integrated area is evaluated by a statistical test to determine if there is a significant difference between the two experimental groups over the time interval.

Compared to taxonomic profiling and metagenomic data, there have been fewer published works on metabolomic data. One common question that arises when analyzing metabolomic data is whether two or more metabolite profiles are significantly different. For a comparison of time-series data from two experimental groups, Hoteling-T^2 statistic can be used as a measure of global longitudinal dissimilarity. If there are more than two experimental conditions, multivariate empirical Bayes analysis of variance may be used to test for significance [26,27]. Another way to answer this question is to determine if there are any significant discriminatory biomarkers Lin *et al.* [28], for example using multivariate regression. Alternatively, smooth splines could be coupled with permutation tests

to perform differential abundance testing of time-series metabolomics data. A related, data fitting-based method was described in a recent study by Dickinson *et al.*, who used a Gaussian Process Two-Sample (GP2S) test to determine the combined effects of drought and infection on the metabolism of nitrogen fixing legumes [29,30]. Linear mixed models offer an attractive option to investigate the longitudinal trend of metabolites and identify metabolites that show significant concentration changes over time. Because of the correlations present in high dimensional longitudinal metabolomics data, dimension reduction techniques are needed that can be applied to time series data. One promising technique is dynamic probabilistic PCA (DPPCA), which models the correlations in multivariate data that occur due to repeated measurements in time [31].

For assessing longitudinal differential abundances, fitting splines is one of the most promising techniques that has broad applicability. By using statistical tests to infer time periods ('regions' in Figure 1) of significance, well-established procedures can be applied to control for false discovery, an important consideration for multivariate analysis. A non-trivial downside of using fitted splines is that the technique requires a relatively large number of time points to achieve statistical power.

Temporal network generation

Interactions in a microbial community are driven by a number of bidirectional mechanisms including competition for nutrients and attachment sites between microbiota community members, as well as direct interactions between community members and the host through metabolites [32,33]. This results in a complex network of bacterial interactions that can be altered by the host's age, diet, disease initiation or other exogenous perturbations. As such, models and computational tools are needed that can accurately infer the ecological relationships of a microbiota and its associated metabolic network. Temporal networks, which are network representations that depict how a connected system evolves with time, have emerged as useful tools for this analysis.

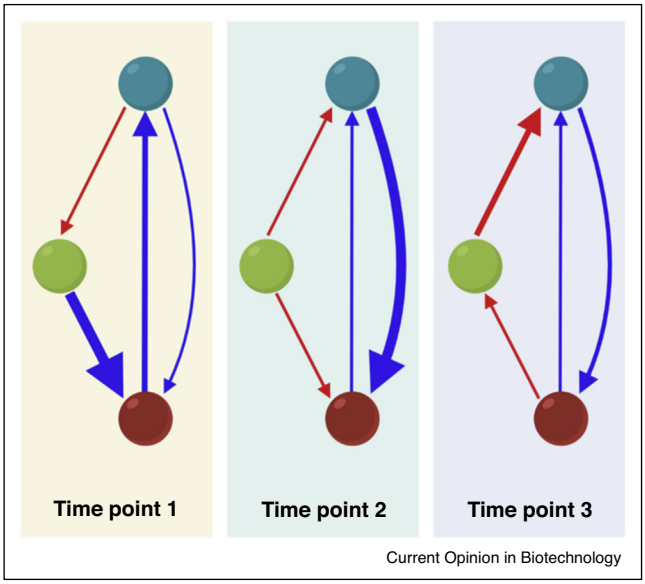
The generalized Lotka-Volterra (gLV) model of population dynamics has gained popularity in studies on microbial communities as a means to infer the time-dependent ecological structure of a microbiota. In a recent application, Venturelli *et al.* [34] used the gLV model and metabolomic data to determine the population dynamics of a synthetic microbial community and infer a resource interaction network between the microbes. This work suggests that multispecies community dynamics are primarily governed by growth parameters of individual species and pairwise interactions between species.

One limitation of gLV models is that incorporating mutualism into the models can predict indefinite growth. Mutualistic interactions could be studied by inferring statistical causality. Granger causality test is a statistical method that uses autoregression to test if a set of time-series variables adds explanatory power to the time-series of interest. Mainali *et al.* [35] utilized Granger causality to analyze time-series microbiome data from a previously published longitudinal study on different body site microbiomes comprising nearly 400 time points [36••]. They found that strong interactions (high values of Granger causality coefficient) within a genus are usually negative and short term (1–2 days), while strong interactions between different genera tend to be positive and longer term (up to 20 days). These results suggest that mutualism could be a necessary interaction mechanism to afford the persistence of multiple taxa in a microbiota, while also facilitating the formation of distinct microbial communities within gut microenvironments. Granger causality is also well-suited for analyzing time series metabolomics data. For example, Wang *et al.* used this test to identify time-shift correlations in metabolite and protein interactions in development of grape berries [37]. Regression methods have also been used to infer temporal networks. Ridenhour *et al.* utilized an autoregressive integrated moving average (ARIMA) model to predict ecological interactions and dynamics of the microbiomes of woodrats in response to oxalate feeding [38]. Typically, regression models such as ARIMA include penalties (e.g. elastic-net regularization) designed to reduce model complexity by biasing the regression against models that overfit the data.

Dynamic Bayesian networks (DBNs), which extend standard Bayesian networks to time series data, provide a probabilistic, non-linear alternative to ARIMA and other linear forecasting models. Lugo-Martinez *et al.* inferred a DBN model [39] using the Conditional Gaussian Bayesian Networks package [40] to detect time-dependent ecological interactions between the microbes within gut, vaginal, and oral microbiota. The authors found that microbiomes sampled from different individuals may have different rates of interactions even if the overall interactions are similar. This variation is a potentially confounding factor when analyzing longitudinal data across individuals. To address this issue, the authors utilized spline fitting and dynamic warping techniques to 'temporally align' the microbiome time-series data before the DBN is learned on the aligned data.

Faust *et al.* proposed a classification scheme to determine presence of temporal structure, test for neutrality (whether interactions between the taxa are necessary for the observed dynamics), and determine goodness of fit of a time-series data set to a deterministic model [41]. The scheme first tests for dependency between consecutive time points within a time-series, then determines if interactions between the microbes adequately explain the dynamics. If these two conditions are met, parameters to fit an interaction model (e.g. gLV or Ricker model) are

Figure 2



Schematic illustration of temporal networks used to analyze longitudinal microbiome and metabolome data. The example in the figure shows a directed interaction network that evolves over time. The blue and red arrows indicate a positive and negative association, respectively. The relative strength of the interaction is depicted by the thickness of the arrow. These associations can be determined through a number of the methods discussed in the text such as the gLV model and Granger causality test.

estimated. Goodness of fit of the chosen interaction model to the time-series data is determined based on the mean correlation of the original time-series to the time-series inferred by the interaction model. Together, these tests determine whether the data set is a good match for temporal network generation, and if so, whether the selected interaction model is well suited for the data. Figure 2 depicts the use of a temporal network to model associations between bacteria or metabolites in time series microbiome or metabolome data.

A network is often the most intuitive way to represent both reaction-mediated and regulatory interactions occurring between metabolites. Furthermore, subtle metabolic differences between experimental groups that are not captured by differential analysis of the abundances of metabolites can be discovered by comparing the connectivity of metabolic networks. On the other hand, conventional static networks are not well-suited for modeling time series data, because they do not explicitly account for time-dependent variations in network connectivity. To address this limitation, Huang *et al.* proposed a dynamic network construction method that links a pair of metabolites based on changes in the concentration ratios of these metabolites in consecutive time points [42].

Granger causality test is fast becoming the method of choice for generating temporal networks from longitudinal microbiome and metabolomic data. Although gLV models are more popular, and often easier to apply, their difficulty in appropriately modelling mutualism limits their effectiveness. Furthermore, only a modest number of features can be reliably modelled using gLV equations when there are few time points in the data set. This does not pose a problem if the goal is to perform a focused interrogation of a subset of microbes in a community for which there is extensive prior biological knowledge. However, deciding which subset should be modeled is a non-trivial task when analyzing less well characterized systems such as the gut microbiota.

Temporal pattern clustering

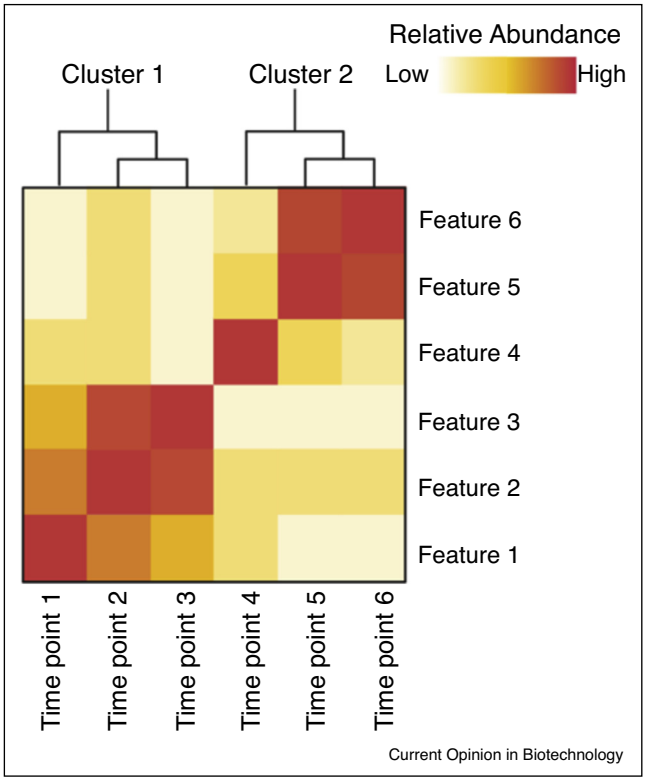
Shifts in the gut microbiota composition may occur as a result of aging, change in diet, disease, or some other factor, and can establish a series of stable microbiota clusters, or 'community states', that impacts how the microbiota interacts with the host [43]. Different community states possess different metabolic potentials, reflecting differences in alpha-diversity and beta-diversity. In this regard, accurately clustering time points having a similar microbial abundance distribution could provide valuable insights into community responses to perturbations, mechanisms of dysbiosis, and other aspects of microbiota dynamics under various stresses.

Dirichlet Multinomial Mixtures (DMM) is a probabilistic modeling scheme that uses a mixture of multivariate (beta) distributions as the prior. This scheme has been used to identify three distinct phases of gut microbiota development in infant cohorts from the The Environmental Determinants of Diabetes in the Young (TEDDY) study on environmental factors of type 1 diabetes [44•]. Another probabilistic approach, Microbial Counts Trajectories Infinite Mixture Model Engine 2.0 (MC-TIMME2) models microbiota community shifts using nonparametric Bayesian methods and stochastic (Dirichlet) processes to cluster time-series microbiome data [45]. Classical clustering approaches have also been utilized to identify temporal patterns in microbiome data. Fan *et al.* used a combination of principal coordinate analysis (PCoA), hierarchical clustering, and Linear discriminant analysis Effect Size to determine biomarkers associated with three distinct gut microbiota clusters in developing shrimp [46]. Similarly, temporal shifts in fecal microbiome were determined in IBD patients and healthy controls by comparing the inter-individual and intra-individual Bray-Curtis dissimilarity in the microbiome data between consecutive time points. This analysis showed that temporal shifts in the fecal microbiome are more frequent and more extreme in IBD patients compared to healthy controls. Furthermore, this analysis confirmed previous work showing IBD patients exhibit a reduction in the abundance of obligate anaerobes

accompanied by an expansion in facultative anaerobes [2]. An example of a top-down partitioning algorithm is k-medoid clustering, which can be used in conjunction with a dissimilarity (e.g. Jenson–Shannon divergence) matrix that models differences in abundance distributions as probability distribution distances [47]. This approach allows for the discovery of periodicity in time series microbiome data. Applied to vaginal microbiome data from a study by Gajer *et al.* [48], the clustering analysis detected two distinct, periodically alternating community states. Furthermore, the study determined that the periodicity was largely driven by the antagonistic and cyclical relationship between *Lactobacillus* iners and *Atopobium* species.

MetaboClust is a software package that clusters metabolites with similar temporal patterns [49]. Several clustering algorithms (such as *k*-means and Density-based spatial clustering of applications with noise) and methods for assessing the quality of clusters are available in this package, as well as pathway enrichment analysis, which can be used to determine how much overlap there is between biochemical pathways (e.g. KEGG pathway maps) and clusters. Temporal clustering by affinity propagation (TCAP) is an extension of *k*-means, and has been used to cluster metabolomics data on time-dependent responses to abiotic and biotic stresses in plants [29]. In addition to clustering, classification techniques (e.g. Microbiome Interpretable Temporal Rule Engine (MITRE)) has also been used to link changes in the time profiles of microbe clusters to host status [50•]. Although MITRE was primarily designed for microbiome data, the authors note that the software can be readily adapted to time-series metabolomics data. Figure 3 shows a graphic illustrating temporal clustering.

Figure 3



Graphic illustration of temporal clustering for analysis of microbiome or metabolome data. The heatmap shows a hypothetical data set with six time points and six features. Hierarchical clustering on the features (rows) over time (columns) identifies two clusters (clusters 1 and 2). This clustering indicates that features 1 through 3 are more abundant during time points 1 through 3, whereas features 4 through 6 are more abundant during time points 4 through 6. This analysis shows a significant shift in the microbiome or metabolome has undergone after time point 3.

Clustering algorithms such as k-medoid are an intuitive option for determining if a microbiome or metabolome data set has groups of bacteria or metabolites that trend similarly over time. A disadvantage of these top-down methods is the need to specify the number of clusters. It is usually difficult to know a priori the optimal number of clusters, especially if the research. Moreover, this number can depend on the similarity metric. However, there are a number of post-hoc methods to assess the 'goodness' of k (the number of cluster specified by the user). Typically, an iterative process is applied to determine the correct cluster number and similarity metric.

Integration of longitudinal multi-omics data

General pattern analysis

While changes in the composition of a microbial community can result in functional changes, their interrelationship is often non-additive and/or nonlinear, and it is difficult to predict shifts in community function over time solely based on taxonomic profiling data [51]. To better understand how a shift in community structure leads to changes in function, it is desirable to co-analyze the dynamics of both microbial abundance and metabolic products [52]. Previous efforts to integrate the analysis of different types of omics data sets have typically sought to identify correlations between pairs of observations. Another approach has been to apply multivariate methods for covariance analysis such as principal component analysis (PCA), canonical correlation analysis (CCA), and co-inertia analysis (CIA). A recent study by Ishii *et al.* on microbiota changes in mice fed a high-fat/high-fiber diet used a hierarchical clustering analysis (HCA), PCA, and orthogonal partial least squares discriminant analysis (OPLS-DA) to identify covariation patterns between intestinal metabolites and microbiota composition and function [53].

In recent years, a growing number of studies have presented extensions to the above methods for omics data integration that more explicitly take into account time

Table 2

Methods for multi-omics data integration and their advantages and disadvantages

Methods	Description	Advantages	Disadvantages
Dimension reduction and visualization methods (PCA, CCA, CIA, OPLS-DA)	Summarizing the variables in each feature type by using a small number of linear combinations to maximize the association between the feature types.	High dimensionality reduction. Can be used to construct networks linking modules of features. Can be applied to infer biological networks from multi-omics data.	Poor interpretability because each node represents multiple, if not all, features.
Clustering methods (HCA)	Describe the proximity between objects.	No a priori information about the number of clusters required. Easy to implement. Well suited to high-throughput microbiome and other omics data.	High time complexity. Not work well in the presence of noise and outlier.
Marginal correlation analysis (WGCNA)	Weight networks encode additional information in the form of connection strengths.	Easy to implement, nonparametric options available.	Subject to spurious findings due to confounding.
Classification methods (Random Forest)	Ensemble machine learning method based on the use of classification and decision regression trees.	Very high-quality models, user-friendly, no distributional assumptions about the predictor or response variables. Powerful alternatives to traditional parametric and semiparametric statistical methods for the analysis of microbiome and other omics data.	Not easy to understand predictions. Not explicitly perform feature (taxa, genes) selection.
Functional shifts' taxonomic contributors (FishTaco)	A tool for analysis of the taxonomic contributions to disease-associated functional shifts.	The total abundance of a certain gene can be viewed as a linear sum of all taxon-level contributions.	Less accurate about the information of the genomics content of each taxon and the total functional abundances.
Genome-scale metabolic model (AGORA)	Resource of semi-automatically generated genome-scale metabolic models for human gut bacteria, which is compatible with human genome-scale metabolic models for analysis of host-microbiome interactions.	A profound knowledge on enzyme kinetics is not necessary. Useful to understand possible routes.	Manual curation needed of many pathways outside central metabolism. Weak in polymeric metabolites with large heterogeneity.

dependence in the data (Tables 2 and 4). One approach is to build correlation networks where pairwise correlation coefficients between different omics variables (e.g. bacterial abundance and metabolite concentration) are calculated based on time profiles. A drawback of this analysis is that the result can be quite sensitive to the choice of significance threshold. This drawback could be mitigated by constructing a weighted network, where the connection strength between two variables is scaled using a power function reflecting the scale-free topology of biological networks. This approach was used by Hewavitharana *et al.* in a recent study on microbe-metabolite interactions during different phases of anaerobic soil disinfestation (ASD) [54].

Identification of significant species and their metabolic products

Another approach that more explicitly models the time dependence of omics data is to fit smoothing splines through time-series of individual data features (e.g. specific taxa and metabolites) [55]. Sparse multivariate ordination methods can then be used to cluster features with similar temporal trajectories irrespective of the type of omics data. In one application involving waste degradation in bioreactors, microbial abundance and metabolomics data were integrated with reactor performance data to study how particular taxa and their metabolic products drive overall reactor dynamics. Another emerging approach is to use machine learning-based classification methods. Tripathi *et al.* used a Random Forest classifier trained on time-series stool microbiome and metabolome data to predict intermittent hypoxia and hypercapnia in murine models of atherosclerosis [56].

An excellent example of a study integrating multiple strategies for multi-omics data analysis is the work by Zhou *et al.* [57••], who performed an extensive longitudinal analysis comparing the microbiota and metabolite profiles of healthy and prediabetic individuals. Leveraging the repeated observations afforded by the longitudinal design, this study evaluated the variations in omics measures (e.g. metabolites, taxa) within an individual over time as well as between individuals using intra-class correlation coefficients (ICC) from linear mixed-effect (LME) models.

An intriguing opportunity for longitudinal multi-omics data integration is provided by computational tools for functional profiling (e.g. gene abundance prediction) from microbiome (e.g. taxonomic abundance) data [58]. In principle, the metagenome of a microbiota should map to the reactions and metabolic products that can occur in the microbiota, and could thus be used to link metabolic products with source organisms having the required

Table 3

Comparison of 16s rRNA and shotgun sequencing for microbiome study

Factors	16S rRNA sequencing	Shotgun sequencing
Bacterial coverage	Broad	Narrow
Taxonomy resolution	Family or genus of bacteria	Species or strains, including non-bacterial organisms
Cross-domain coverage	No (bacteria only)	Yes (all 3 taxonomic domains)
Host DNA interference	Low	High
Minimum DNA input	10 copies of 16S rRNA	1 ng
Functional profiling	No	Yes
Accuracy	Low risk	High risk (over or mis-predictions)
Databases	Well-established	Relatively new and growing
Output	Table of OTU abundance	Tables of gene and species abundance

enzymes [59]. Further developing these metagenomic prediction frameworks could potentially add a new, biological knowledge-driven dimension to modeling longitudinal data sets. This in turn could enhance mechanistic understanding of time-dependent relationships between taxonomic abundance and functional dynamics.

Functional profiling

Because of its lower cost, 16S rRNA sequencing has been more commonly used for profiling the taxonomic abundance of gut microbiota. However, deeper, for example, whole genome shotgun (WGS), sequencing methods have become increasingly more accessible and popular. Table 3 summarizes the relative merits of these two sequencing methods [60].

Genome-scale metabolic (GSM) models derived from human gut metagenomic data provide another exciting opportunity for developing new tools that can infer multi-omics interaction networks. While flux balance analysis (FBA) and related constraint-based metabolic modeling techniques have been successfully utilized to study the metabolic capabilities of specific gut bacteria [61], challenges remain in applying these techniques to model entire communities [62]. These challenges include modeling metabolic exchanges between many taxa and selecting an appropriate objective function to account for potential trade-offs between species and community growth rates [63]. The intestinal microbiota of a human or animal host comprises a very large number of diverse species, many of which remain poorly characterized. Assembly of gut organisms through reconstruction and analysis (AGORA) is a recent effort by Magnúsdóttir *et al.* to semi-automatically reconstruct GSMs for hundreds of human gut bacteria [64]. To validate AGORA, the authors used the model in conjunction with constrained optimization methods (flux balance and flux variability analysis) to predict the growth of bacterial species in the human gut and to elucidate how availability of metabolites and species composition both influence community function. Collections of human gut bacterial GSMs such as AGORA and K-Base could facilitate the mapping of metagenomic reads to metabolic functions. In turn, these functional mappings could open a way for determining biochemical associations between microbial taxa that extend beyond empirical correlations. As an example, a GSM model with integration of longitudinal metabolome data and microbiomes was applied to demonstrate complex alternations in microbial sulfur metabolism in Parkinson's disease [65••].

Future directions

Integration of experimental design considerations into microbiome studies is important. Even when longitudinal data are collected, there is often a mismatch between microbial community dynamics and sampling frequency. Because of practical limitations, most *in vivo* studies collect samples infrequently, for example, over days or weeks whereas the characteristic times of microbial community dynamics in the gut microbiota can be on the scale of hours, Thus, microbiota studies would benefit not only from sampling at multiple time points, but also from a consideration of when samples are collected. One possible solution is to use in silico studies, for example, GSMs, to simulate the microbiome or metabolome dynamics and use it to identify optimal sampling windows for sample collection.

Similarly, computational and statistical methods that can rigorously address the challenges of microbiome time-series data analysis and facilitate integration of different types of omics data are vital in the effort to improve our understanding of the intestinal microbiota and their impacts on human health and diseases. In parallel with the growing emphasis on mechanistic investigation, a promising collection of software tools has become available to support the analysis of large omics data sets from longitudinal studies. However, the field is still nascent, and there is as yet no clear consensus regarding a standardized process for reporting and comparing outcomes from different analyses. As a result, the choice of a data analysis tool often relies on trial-and-error. In this regard, open-source platforms that provide options to conduct multiple analyses on consistently formatted and normalized data sets would be a powerful resource. Finally, any new tools need to be developed with a greater emphasis

Table 4

Recent key references for longitudinal multi-omics analysis

Method	Types of data	Description	References
Weighted correlation network analysis (WGCNA)	• Soil microbiome and metabolome • Time-series data	Identified groups of metabolites (short-chain organic acids, methyl sulfide compounds, hydrocarbons, and *p*-cresol) and microorganisms (mostly Firmicutes) that are associated with different phases of soil transformation	[54]
Fitted splines and sparse multivariate techniques (linear mixed model spline, sPCA, sPLS, block sPLS)	• Infant gut microbiome and metabolome (stool samples) • Time-series data	One of the first approaches to integrate different types of multi-omics time series data sets; identified multi-omic signatures characterizing waste degradation dynamics in bioreactors.	[55]
Random Forest classifier	• Mouse fecal microbiome and metabolome • 12–20 time points	Predicted key microbes and metabolites as biomarkers (*Mogibacteriaceae*, *Clostridiaceae*, bile acids, and fatty acids) for obstructive sleep apnea, which is associated with atherosclerosis.	[56]
Intra-class correlation (ICC) from linear mixed effect (LME) model, integrated canonical pathway analysis	• Human peripheral blood cell transcriptome • Human plasma metabolome, cytokines, and proteome • Human stool, nasal, tongue, and skin microbiomes (16S rRNA and WGS sequencing) • 1092 time points over 4 years	Modeled associations between host biomolecules and pathway changes over time in order to investigate early molecular signs of T2D; changes in both the richness and the diversity of microbiota over time differed between healthy controls and individuals with prediabetes; indolelactic acid and hippuric acid identified as markers of gut microbiota diversity.	[57••]
Functional shifts' taxonomic contributors (FishTaco)	• Human microbiomes from different body sites • Metabolomic data from subjects with type 2 diabetes (T2D) or inflammatory bowel disease (IBD)	A comparative analysis tool to characterize and quantify the relationship between taxonomic and functional shifts; for example, *Bifidobacterium* was the main driver of the enrichment of multiple sugar metabolizing pathways in T2D samples, whereas *Clostridium* and *Methanobrevibacter* species contributed to the enrichment of methane metabolism pathway in both T2D and IBD.	[58]
Community level metabolic model and correlation analysis	• Murine fecal metabolome and microbiome (16S rRNA sequencing data) • Three time points	Characterized the functional consequence of environmental chemical (phthalate) exposure on murine cecal microbiota; identified a link between chemical exposure and production of a toxic metabolite (*p*-cresol) by *Clostridium* species.	[59]
Genome-scale metabolic model	• Human gut microbiome • Metabolome • Three time points	Analyzed host-microbiome interactions; identified gut microbiota-specific metabolic functions and validated a predicted growth medium for *Bacteroides caccae.*	[64,65••]

on user-friendliness and interpretability. It is important that they are accessible to researchers without formal backgrounds in data science or informatics. Development of Graphical User Interfaces (GUIs) to visually interact with the data and analysis results would be an effective way to address this.

Conflict of interest statement

Nothing declared.

CRediT authorship contribution statement

Seo-Young Park: Conceptualization, Writing - original draft, Visualization. **Arinzechukwu Ufondu:** Conceptualization, Writing - original draft, Software. **Kyongbum Lee:** Supervision, Writing - review & editing. **Arul Jayaraman:** Supervision, Writing - review & editing.

Acknowledgements

This work was supported in part by grants from the National Center for Complementary and Integrative Health to AJ and KL (RO1-AT-010282) and the Ray Nesbitt Chair Endowment to AJ.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, Mele MC: **What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases**. *Microorganisms* 2019, **7**:14.

2. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ *et al.*: **Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases**. *Nature* 2019, **569**:655-662.

3. Safari Z, Gérard P: **The links between the gut microbiome and non-alcoholic fatty liver disease (NAFLD)**. *Cell Mol Life Sci* 2019, **76**:1541-1558.

4. Scher JU, Nayak RR, Ubeda C, Turnbaugh PJ, Abramson SB: **Pharmacomicrobiomics in inflammatory arthritis: gut microbiome as modulator of therapeutic response**. *Nat Rev Rheumatol* 2020, **16**:282-292.

5. Levan SR, Stamnes KA, Lin DL, Panzer AR, Fukui E, McCauley K, Fujimura KE, McKean M, Ownby DR, Zoratti EM *et al.*: **Elevated faecal 12,13-diHOME concentration in neonates at high risk for asthma is produced by gut bacteria and impedes immune tolerance**. *Nat Microbiol* 2019, **4**:1851-1861.

6. Singer JR, Blosser EG, Zindl CL, Silberger DJ, Conlan S, Laufer VA, DiToro D, Deming C, Kumar R, Morrow CD *et al.*: **Preventing dysbiosis of the neonatal mouse intestinal microbiome protects against late-onset sepsis**. *Nat Med* 2019, **25**:1772-1782.

7. Sharon G, Cruz NJ, Kang DW, Gandal MJ, Wang B, Kim YM, Zink EM, Casey CP, Taylor BC, Lane CJ *et al.*: **Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice**. *Cell* 2019, **177**:1600-1618.

8. Zmora N, Suez J, Elinav E: **You are what you eat: diet, health and the gut microbiota**. *Nat Rev Gastroenterol Hepatol* 2019, **16**:35-56.

9. Hagan T, Cortese M, Rouphael N, Khurana S, Golding H, Pulendran B, Hagan T, Cortese M, Rouphael N, Boudreau C *et al.*: **Antibiotics-driven gut microbiome perturbation alters immunity to vaccines in humans**. *Cell* 2019, **178**:1313-1328.

10. Gerber GK: **The dynamic microbiome**. *FEBS Lett* 2014, **588**:4131-4139.

11. Das P, Marcišauskas S, Ji B, Nielsen J: **Metagenomic analysis of bile salt biotransformation in the human gut microbiome**. *BMC Genomics* 2019, **20**:517.

12. Osman MA, Neoh HM, Mutalib NSA, Chin SF, Jamal R: **16S rRNA gene sequencing for deciphering the colorectal cancer gut microbiome: current protocols and workflows**. *Front Microbiol* 2018, **9**:767.

13. Barko PC, McMichael MA, Swanson KS, Williams DA: **The gastrointestinal microbiome: a review**. *J Vet Intern Med* 2018, **32**:9-25.

14. Castaner O, Goday A, Park YM, Lee SH, Magkos F, Shiow SATE, Schröder H: **The gut microbiome profile in obesity: a systematic review**. *Int J Endocrinol* 2018, **2018**.

15. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, Mohney RP, Small KS, Bell JT, Steves CJ *et al.*: **The fecal metabolome as a functional readout of the gut microbiome**. *Nat Genet* 2018, **50**:790-795.

16. Karu N, Deng L, Slae M, Guo AC, Sajed T, Huynh H, Wine E, Wishart DS: **A review on human fecal metabolomics: methods, applications and the human fecal metabolome database**. *Anal Chim Acta* 2018, **1030**:1-24.

17. Kdadra M, Höckner S, Leung H, Kremer W, Schiffer E: **Metabolomics biomarkers of prostate cancer: a systematic review**. *Diagnostics* 2019, **9**:21.

18. Kumar MS, Slud EV, Okrah K, Hicks SC, Hannenhalli S, Corrada Bravo H: **Analysis and correction of compositional bias in sparse sequencing count data**. *BMC Genomics* 2018, **19**:799.

19. Wang C, Hu J, Blaser MJ, Li H: **Estimating and testing the microbial causal mediation effect with high-dimensional and compositional microbiome data**. *Bioinformatics* 2020, **36**:347-355.

20. Galloway-Peña J, Guindani M: **Editorial: novel approaches in microbiome analyses and data visualization**. *Front Microbiol* 2018, **9**:2274.

21. Coenen AR, Hu SK, Luo E, Muratore D, Weitz JS: **A primer for microbiome time-series analysis**. *Front Genet* 2020, **11**:1-15.

22. Luo D, Ziebell S, An L: **An informative approach on differential abundance analysis for time-course metagenomic sequencing data**. *Bioinformatics* 2017, **33**:1286-1292.

23. •• Metwally AA, Yang J, Ascoli C, Dai Y, Finn PW, Perkins DL: **MetaLonDA: a flexible R package for identifying time intervals of differentially abundant features in metagenomic longitudinal studies**. *Microbiome* 2018, **6**:32.
A new tool for longitudinal microbiome data analysis to identify significant time intervals.

24. Vatanen T, Kostic AD, D'Hennezel E, Siljander H, Franzosa EA, Yassour M, Kolde R, Vlamakis H, Arthur TD, Hämäläinen AM *et al.*: **Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans**. *Cell* 2016, **165**:842-853.

25. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, Huang B, Arodz TJ, Edupuganti L, Glascock AL *et al.*: **The vaginal microbiome and preterm birth**. *Nat Med* 2019, **25**:1012-1021.

26. Gonzales GB, De Saeger S: **Elastic net regularized regression for time-series analysis of plasma metabolome stability under sub-optimal freezing condition**. *Sci Rep* 2018, **8**:1-10.

27. Tai YC, Speed TP: **A multivariate empirical Bayes statistic for replicated microarray time course data**. *Ann Stat* 2006, **34**:2387-2412.

28. Lin Z, Zhang Q, Dai S, Gao X: **Discovering temporal patterns in longitudinal nontargeted metabolomics data via group and nuclear norm regularized multivariate regression**. *Metabolites* 2020, **10**:33.

29. Dickinson E, Rusilowicz MJ, Dickinson M, Charlton AJ, Bechtold U, Mullineaux PM, Wilson J: **Integrating transcriptomic techniques and k-means clustering in metabolomics to identify markers of abiotic and biotic stress in *Medicago truncatula***. *Metabolomics* 2018, **14**:126.

30. Stegle O, Denby KJ, Cooke EJ, Wild DL, Ghahramani Z, Borgwardt KM: **A robust bayesian two-sample test for detecting intervals of differential gene expression in microarray time series**. *J Comput Biol* 2010, **17**:355-367.

31. Nyamundanda G, Gormley IC, Brennan L: **A dynamic probabilistic principal components model for the analysis of longitudinal metabolomics data**. *J R Stat Soc Ser C Applied Stat* 2014, **63**:763-782.

32. Parker A, Lawson MAE, Vaux L, Pin C: **Host-microbe interaction in the gastrointestinal tract**. *Environ Microbiol* 2018, **20**:2337-2353.

33. Meng X, Zhou HY, Shen HH, Lufumpa E, Li XM, Guo B, Li BZ: **Microbe-metabolite-host axis, two-way action in the pathogenesis and treatment of human autoimmunity**. *Autoimmun Rev* 2019, **18**:455-475.

34. Venturelli OS, Carr AV, Fisher G, Hsu RH, Lau R, Bowen BP, Hromada S, Northen T, Arkin AP: **Deciphering microbial interactions in synthetic human gut microbiome communities**. *Mol Syst Biol* 2018, **14**:e8157.

35. Ai D, Li X, Liu G, Liang X, Xia LC: **Constructing the microbial association network from large-scale time series data using granger causality**. *Genes (Basel)* 2019, **10**:216.

36. •• Mainali K, Bewick S, Vecchio-Pagan B, Karig D, Fagan WF: **Detecting interaction networks in the human microbiome with conditional Granger causality**. *PLoS Comput Biol* 2019, **15**: e1007037.
One of the first studies to apply a causal model to investigate inter-species and intra-species interactions.

37. Wang L, Sun X, Weiszmann J, Weckwerth W: **System-level and granger network analysis of integrated proteomic and metabolomic dynamics identifies key points of grape berry development at the interface of primary and secondary metabolism**. *Front Plant Sci* 2017, **8**:1066.

38. Ridenhour BJ, Brooker SL, Williams JE, Van Leuven JT, Miller AW, Dearing MD, Remien CH: **Modeling time-series data from microbial communities**. *ISME J* 2017, **11**:2526-2537.

39. Lugo-Martinez J, Ruiz-Perez D, Narasimhan G, Bar-Joseph Z: **Dynamic interaction network inference from longitudinal microbiome data**. *Microbiome* 2019, **7**:54.

40. McGeachie MJ, Chang HH, Weiss ST: **CGBayesNets: conditional Gaussian Bayesian network learning and inference with mixed discrete and continuous data**. *PLoS Comput Biol* 2014, **10**.

41. Faust K, Bauchinger F, Laroche B, de Buyl S, Lahti L, Washburne AD, Gonze D, Widder S: **Signatures of ecological processes in microbial community time series**. *Microbiome* 2018, **6**:1-13.

42. Huang X, Zeng J, Zhou L, Hu C, Yin P, Lin X: **A new strategy for analyzing time-series data using dynamic networks: identifying prospective biomarkers of hepatocellular carcinoma**. *Sci Rep* 2016, **6**:32448.

43. Shetty SA, Hugenholtz F, Lahti L, Smidt H, de Vos WM: **Intestinal microbiome landscaping: insight in community assemblage and implications for microbial modulation strategies**. *FEMS Microbiol Rev* 2017, **41**:182-199.

44. • Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, Ross MC, Lloyd RE, Doddapaneni HV, Metcalf GA *et al.*: **Temporal development of the gut microbiome in early childhood from the TEDDY study**. *Nature* 2018, **562**:583-588.
A longitudinal study design provided a robust analysis of gut microbiome development in infants.

45. Creswell R, Tan J, Leff JW, Brooks B, Mahowald MA, Thieroff-Ekerdt R, Gerber GK: **High-resolution temporal profiling of the human gut microbiome reveals consistent and cascading alterations in response to dietary glycans**. *Genome Med* 2020, **12**:1-16.

46. Fan J, Chen L, Mai G, Zhang H, Yang J, Deng D, Ma Y: **Dynamics of the gut microbiota in developmental stages of *Litopenaeus vannamei* reveal its association with body weight**. *Sci Rep* 2019, **9**:1-10.

47. Baksi KD, Kuntal BK, Mande SS: **"TIME": a web application for obtaining insights into microbial ecology using longitudinal microbiome data**. *Front Microbiol* 2018, **9**:36.

48. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, Koenig SSK, Fu L, Ma ZS, Zhou X *et al.*: **Temporal dynamics of the human vaginal microbiota**. *Sci Transl Med* 2012:a52.

49. Rusilowicz MJ, Dickinson M, Charlton AJ, O'Keefe S, Wilson J: **MetaboClust: using interactive time-series cluster analysis to relate metabolomic data with perturbed pathways**. *PLoS One* 2018, **13**:e0205968.

50. • Bogart E, Creswell R, Gerber GK: **MITRE: inferring features from microbiota time-series data linked to host status**. *Genome Biol* 2019, **20**:1-15.
One of the first studies that used machine learning to predict impact of microbiome changes over time on host physiology.

51. Heintz-Buschart A, Wilmes P: **Human gut microbiome: function matters**. *Trends Microbiol* 2018, **26**:563-574.

52. Wang Q, Wang K, Wu W, Giannoulatou E, Ho JWK, Li L: **Host and microbiome multi-omics integration: applications and methodologies**. *Biophys Rev* 2019, **11**:55-65.

53. Ishii C, Nakanishi Y, Murakami S, Nozu R, Ueno M, Hioki K, Aw W, Hirayama A, Soga T, Ito M *et al.*: **A metabologenomic approach reveals changes in the intestinal environment of mice fed on American diet**. *Int J Mol Sci* 2018, **19**:4079.

54. Hewavitharana SS, Klarer E, Reed AJ, Leisso R, Poirier B, Honaas L, Rudell DR, Mazzola M: **Temporal dynamics of the soil metabolome and microbiome during simulated anaerobic soil disinfestation**. *Front Microbiol* 2019, **10**:2365.

55. Bodein A, Chapleur O, Droit A, Lê Cao K-A: **A generic multivariate framework for the integration of microbiome longitudinal studies with other data types**. *Front Genet* 2019, **10**:1-18.

56. Tripathi A, Xu ZZ, Xue J, Poulsen O, Gonzalez A, Humphrey G, Meehan MJ, Melnik AV, Ackermann G, Zhou D *et al.*: **Intermittent hypoxia and hypercapnia reproducibly change the gut microbiome and metabolome across rodent model systems**. *mSystems* 2019, **4** e00058-19.

57. •• Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, Zhang MJ, Rao V, Avina M, Mishra T *et al.*: **Longitudinal multi-omics of host–microbe dynamics in prediabetes**. *Nature* 2019, **569**:663-671.
Used linear mixed model regression and constraint-based metabolic modeling for extensive longitudinal analyses comparing microbiota and metabolite profiles of healthy and prediabetic individuals.

58. Manor O, Borenstein E: **Systematic characterization and analysis of the taxonomic drivers of functional shifts in the human microbiome**. *Cell Host Microbe* 2017, **21**:254-267.

59. Lei M, Menon R, Manteiga S, Alden N, Hunt C, Alaniz RC, Lee K, Jayaraman A: **Environmental chemical diethylhexyl phthalate alters intestinal microbiota community structure and metabolite profile in mice**. *mSystems* 2019, **4**:1-16.

60. Ye SH, Siddle KJ, Park DJ, Sabeti PC: **Benchmarking metagenomics tools for taxonomic classification**. *Cell* 2019, **178**:779-794.

61. Devika NT, Raman K: **Deciphering the metabolic capabilities of Bifidobacteria using genome-scale metabolic models**. *Sci Rep* 2019, **9**:1-9.

62. Sen P, Orešič M: **Metabolic modeling of human gut microbiota on a genome scale: an overview**. *Metabolites* 2019, **9**.

63. Diener C, Gibbons SM, Resendis-Antonio O: **MICOM: metagenome-scale modeling to infer metabolic interactions in the gut microbiota**. *mSystems* 2020, **5**:1-17.

64. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jäger C, Baginska J, Wilmes P *et al.*: **Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota**. *Nat Biotechnol* 2017, **35**:81-89.

65. •• Hertel J, Harms AC, Heinken A, Baldini F, Thinnes CC, Glaab E, Vasco DA, Pietzner M, Stewart ID, Wareham NJ *et al.*: **Integrated analyses of microbiome and longitudinal metabolome data reveal microbial-host interactions on sulfur metabolism in Parkinson's disease**. *Cell Rep* 2019, **29**:1767-1777.e8.
Combined longitudinal metabolomics and metabolic modeling of gut microbiomes to link gut microbiota-dependent metabolism of sulfur and taurine-conjugated bile acids with severity of Parkinson's disease.