Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Systems biology of the human microbiome

Beatriz Peñalver Bernabé[1], Lauren Cralle[1,2] and Jack A Gilbert[1,2,3]

Recent research has shown that the microbiome — a collection of microorganisms, including bacteria, fungi, and viruses, living on and in a host — are of extraordinary importance in human health, even from conception and development in the uterus. Therefore, to further our ability to diagnose disease, to predict treatment outcomes, and to identify novel therapeutics, it is essential to include microbiome and microbial metabolic biomarkers in Systems Biology investigations. In clinical studies or, more precisely, Systems Medicine approaches, we can use the diversity and individual characteristics of the personal microbiome to enhance our resolution for patient stratification. In this review, we explore several Systems Medicine approaches, including Microbiome Wide Association Studies to understand the role of the human microbiome in health and disease, with a focus on 'preventive medicine' or P4 (i.e., personalized, predictive, preventive, participatory) medicine.

Addresses
[1] The Microbiome Center, Department of Surgery, University of Chicago, Chicago, USA
[2] Biosciences Division, Argonne National Laboratory, Lemont, IL, USA
[3] Marine Biology Laboratory, Woods Hole, MA, USA

Corresponding author: Gilbert, Jack A (gilbertjack@uchicago.edu)

Current Opinion in Biotechnology 2018, **51**:146–153

This review comes from a themed issue on **Systems biology**

Edited by **Nathan Price** and **Eran Segal**

https://doi.org/10.1016/j.copbio.2018.01.018

0958-1669/© 2018 Elsevier Ltd. All rights reserved.

Characterizing the human microbiome

In recent years, Systems Biology has revolutionized our discovery of biomarkers to prevent, diagnose, and treat diseases. For example, the personalized diagnosis of HER2 breast cancer is one of the first examples implemented at the clinical level [1]. Systems Biology approaches allow us to make sense of the vast amount of data generated by '-omics' technologies, such as genomics, transcriptomics, metabolomics, and proteomics, through statistical, computational, and mathematical approaches that enable us to reveal the emergent properties of studied systems.

The Human Microbiome is heterogeneous between body sites (e.g. skin, gut, vagina), is distinctly personal [2], evolves over our life span [3], and has been implicated in, among other conditions, obesity [4] and depression [5•]. Clinical studies to characterize the microbiome must consider numerous elements [6•], including cohort selection, participant attrition, sample size, experimental design, sample collection, transportation and preservation, and more. Sample size is crucial to achieve statistical power, though few methods are currently available to establish *a priori* sample size for microbiome studies [7]. Many microbiome studies suffer from small sample sizes that may not capture the variability of the system, and we possess limited understanding of how to calculate sample size for longitudinal investigations. These limitations likely result from a lack of information about variability, which has led to a number of large scale efforts aimed at characterizing data from groups of participants in an attempt to quantify the variance in different traits [8••,9••]. For the microbiome, crowdsourcing efforts, such as American Gut (www.americangut.org), provide a unique opportunity to create data resources that can be used to predict statistical power for clinical studies.

To perform a Microbiome Wide Association Study (MWAS) [6•], it is necessary to profile the microbiome to identify biomarkers that can be associated with host traits. The microbiome can be characterized using 16S/18S/ITS rRNA amplicon sequencing to identify the relative abundances of the different species, shotgun metagenomic sequencing to identify the organisms functional potential, metatranscriptomics (RNA-seq) to determine their functional response to change, metabolomics to identify microbial products, meta-proteomics [10] (UPLC-MS) to identify the enzymes being produced, and imaging (e.g. 3D cartography [11•]) to visualize the spatial structure of the microbiome. The most common method is amplicon sequencing, usually using 16S rRNA [2,3] amplicons to describe bacterial and archaeal diversity, community structure, and composition of the microbiota. The benefit of amplicon sequencing is that it is inexpensive (<$20 a sample), is fast, and provides easy-to-interpret biomarker units. Traditionally, these biomarkers have been known as operational taxonomical units (OTUs) and were clusters of similar taxa (e.g. QIIME [12], Mothur [13]); however, new computational techniques have enabled this data to be probed at a greater taxonomic resolution [14••,15•,16•], enabling the identification of biomarkers potentially at strain-level resolution (Table 1). Once that amplicon sequencing has

Table 1

Summary of the most common experimental and computational methods employed in Systems Microbiome Medicine

MOLECULAR GENOTYPING	MICROBIOME FEATURES	STATIC DATA	DYNAMIC DATA
OTU identification (16S rRNA) QIIME Mothur	**Microbiome structure (e.g., alpha and beta diversity)** Phyloseq QIIME Mothur	**Co-abundance networks** SparCc CCLasso SPICE-EASI	**Causation networks** Dynamic Bayesian Networks SVAR MDSINE (Lotka-Voltera) LIMITS (LOTKA-Voltera)
Sub-OTU identification (16S rRNA) DADA2 Deblur Oligotyping		**Host-microbiome interactions & Multi-omics interactions** DESeq2 matagenomeSeq LefSE LinearLog-Contrast with l1 penalization Random Forest Gradient bootstrapping tree Neural networks Spearman correlation Procrustes Co-inertia Spearman correlation	**Host-microbiome interactions & Multi-omics interactions** Dynamic hierarchical clustering & LefSE Dynamic Bayesian Networks Neural networks SVAR Lotka-Voltera Agent Based models
Genome reconstruction (shotgun) Anvi'o MetaAMOS Xander	**Microbiome metabolism** PICRUST HUMAnN CASINO (GEMs)		

been processed and annotated to known bacterial taxa, amplicon sequence data needs to be treated or normalized to avoid experimental and technical artifacts [12,17,18]. Subsequently, normalized amplicon data can be processed through computational pipelines (Table 1) to study the community structure (e.g., alpha and beta diversity) and to perform the statistical analysis that will link these biomarkers to host traits (e.g. phyloseq [19], QIIME [12], Mothur [13]).

Amplicon sequencing is limited, however, by the taxonomic resolution (i.e. you cannot usually identify microbes to the species or strain level), and it provides no information on the functional capacity of the microbes, although techniques exist to computationally predict microbial function for members of the ecological community that have a known sequence (e.g. PICRUSt [20]). Therefore, to characterize microbial biomarkers such as genomic strain or functional gene, shotgun metagenomics is used, whereby the total genomic DNA of a sample is randomly sequenced [21,22••]. While this provides less coverage of the total community composition, it does provide greater taxonomic resolution and potential functional information, which improves the ability to identify associations with host traits and patient stratification. However, shotgun metagenomic sequencing is expensive ($300–500 a sample), and analysis is more labor intensive than human genomics, mostly because there are no reference genomes for a majority of the organisms in a sample, which makes it harder to interpret the sequencing data [23•]. However, there are a number computational pipelines, such as MetAMOS [24], Xander [25•], and Anvi'o [26•], that reduce the workload (Table 1).

Importantly, metagenomic analysis only describes the genetics and functional potential of the microbiome, as it does not characterize the genes that are actively transcribed and translated into proteins. Metatranscriptomics [27] and metaproteomics [28] can be used to explore these phenomena, but they are more expensive than amplicon or metagenomic sequencing — metatranscriptomics can cost more than $500 per sample, while metaproteomics can cost more than $1000 per sample. Metatranscriptomics is easier to implement experimentally and computationally [29•] than meta-proteomics; in the latter, the cells have to be isolated and the extracted proteins must be analyzed using LC-MS methods [30•]. Metaproteomics provides useful biomarkers, as these are the active proteins and enzymes that are influencing host traits, but cost and difficulty of sample preparation limit the application of this approach.

The culmination of genetics, transcriptomics, and proteomics is of course the metabolome, which represents the small molecules generated by the individual cell or community of microbes. The influence of microbial metabolites of human health is well recognized [31•]. In fact, metabolite biomarkers can often show the strongest association with host traits, likely because they have direct influence on host function [31•,32]. Microbial metabolites, such as short chain fatty acids, have been shown to have a significant influence on local inflammation [32], hormonal balance [9••], and even on mitochondrial activity [33•]. The presence of microbe-related metabolites is commonly determined by gas or liquid chromatography followed by mass spectrophotometry [29•], and the cost can vary from a few dollars for single metabolites to more than $100 per sample for an untargeted analysis of the

Download English Version:

https://daneshyari.com/en/article/6487315

Download Persian Version:

https://daneshyari.com/article/6487315

Daneshyari.com