Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Using systems biology approaches to elucidate cause and effect in host–microbiome interactions

Elizabeth A. Witherden[a,1], David L. Moyes[a,1], Kenneth D. Bruce[a,1], S. Dusko Ehrlich[a,b,1] and Saeed Shoaie[a,c,1]

Abstract

The human microbiome is a diverse and complex ecosystem integral for healthy human development. Recent advances in next-generation sequencing technology have paved the way for a 'multi-omics' era of microbiome research, uncovering associations between microbial dysbiosis and disease. Our ability to harness the full potential of these 'multi-omics' datasets are currently constrained by several technical, analytical, computational and bioinformatics factors. However, it may be possible to overcome such limitations through the use of novel systems biology thinking and approaches, to integrate and analyse these large 'multi-omics' datasets. Thus, the question arises - can systems biology approaches pave the way to a new era in microbiome research; determining underlying mechanisms in health and disease, and identifying key microbial interactions and causalities?

Addresses

[a] Centre for Host–Microbiome Interactions, Dental Institute, King's College London, London, SE1 9RT, United Kingdom

[b] Metagenopolis, Institut National de La Recherché Agronomique, 78350, Jouy en Josas, France

[c] Centre for Translational Microbiome Research, Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, SE-171 77 Stockholm, Sweden

Corresponding author: Shoaie, Saeed, Centre for Host–Microbiome Interactions, Dental Institute, King's College London, London, SE1 9RT, United Kingdom. (saeed.shoaie@kcl.ac.uk)

[1] All authors contributed equally.

Current Opinion in Systems Biology 2017, **3**:141–146

This review comes from a themed issue on **Clinical and translational systems biology (2017)**

Edited by **Jesper Tegnér** and **David Gomez-Cabrero**

For a complete overview see the Issue and the Editorial

Available online 12 May 2017

http://dx.doi.org/10.1016/j.coisb.2017.05.003

Keywords

Host–microbiome, Genome-scale model, Metagenomics, Gene catalogue, Integrative analysis, Metabolism.

Introduction

The past decade has been a golden age for microbiome research. Advances in next-generation sequencing and bioinformatics techniques have set the stage for 'multi-omics' approaches for studying the human microbiome in both health and disease [1,2]. Multi-omics approaches extend beyond "traditional" microbial diversity and composition analysis as generated by 16S rRNA data sets, advancing into metagenomics, host–microbial interactions, and functional modelling with the aim of elucidating disease causalities [1,3–6]. These advances are all underpinned by bacterial ecology and systems biology concepts, which have been adapted to characterize and fully elucidate the role of the human microbiome in health and disease.

To-date, systems-level approaches have focused on genome reconstructions, where genome-scale models have been built to model the functional relationships of highly abundant microorganisms within an ecosystem [7,8]. In these models, whole-genome assembly data is used in an attempt to link annotated genes to functional categories, functional gene networks, host–microbial interactions, and microbial–microbial interactions [3]. Such approaches, however, rely heavily on the quality of genome sequences and the availability of curated genome databases, as well as the quality of gene and genome annotation data. Here we discuss the current state of metagenomics research in the context of 'multi-omics' analysis and systems biology.

Main text

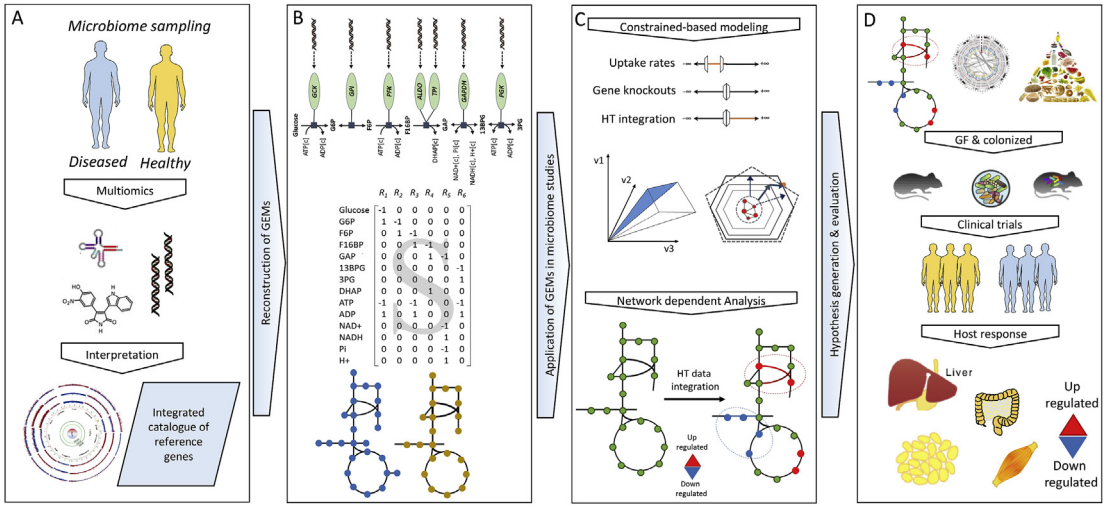
Advances in next-generation sequencing technology, have led to the establishment of the field of metagenomics. In its simplest form, metagenomics refers to the study of the genetic material recovered directly from the totality of organisms present in an environmental sample or microbial community [9]. In metagenomic studies, genomic DNA is isolated from the sample of interest, and randomly sheared before being shot-gun sequenced. The resulting output is a mass of short sequencing reads that need to be "trimmed" for quality, assembled, and mapped to gene databases allowing identification of the microbial population structure (taxonomy) and function (gene annotations). Despite there being numerous platforms available for metagenomic sequencing (Ie. Illumina, Ion Proton), there is a bottle neck for metagenomic studies in the lack of downstream resources for read mapping and subsequent bioinformatics analysis of the generated datasets.

Developing reference gene catalogues

In the early 2000's, international initiatives; the Human Microbiome Project (HMP) [10,11], and the International Human Microbiome Consortium (IHMC) [2], were established to generate sequencing resources that would aid in the characterisation of the human microbiome. One of the main aims of these initiatives was to generate and curate genome databases for taxonomic discrimination of microbial communities, as well as mapping and annotating a large number of entire microbial genomes (Figure 1A) [10]. To date, even though there are several well established 16S rRNA gene databases (ie. SILVA [12], Greengenes [13]) which can be used for taxonomic binning of sequencing reads, the curated resources available for analysing metagenomics and 'multi-omics' datasets have been limited. Reference gene catalogues are, however, becoming increasingly available — albeit primarily focused on the bacterial constituents of the gut microbiome in humans [10,14,15], and other mammals [16–18].

In 2010, the first of these microbial gene catalogues for the human gut metagenome was published [14]. This catalogue contained 3.3 million non-redundant microbial genes, 99.91% of which represented genes of bacterial origin, with the remainder of archaeal, eukaryotic or viral origin [14]. Generated from data obtained by sequencing faecal samples from 124 European individuals, this gene catalogue was estimated to cover the entire genomes of up to ~1000 of the dominant bacterial species identified in the human gut [14]. Li and colleagues [15] built on this work by curating a human gut reference catalogue containing 9,879,896 genes [15]. This Integrated Gene Catalogue (IGC) composes near complete sets of genes from the most abundant gut microbes identified in individuals from three continents [15]. Although, this catalogue is considerably more complete than the previous gut catalogues of Qin [14] and HMP [10,11], it is still primarily focused on the bacterial constituents of the gut microbiome. Until a curated gene catalogue representing gut bacterial, archaeal, viral and fungal genes and genomes is established, the full potential of microbiome research will not be realised [19]. Analogously, there is a need to establish reference gene catalogues specific for other body sites including the oral cavity, skin, and vagina [19]. The

Figure 1



A proposed framework for the integrative analysis of multi-omics microbiome data using genome-scale modelling to understand causality of the ecosystem and elucidate the interactions. After microbiome sampling of healthy and diseased individuals, different high-throughput (HT) analysis can be applied to the samples (A). Metagenomic outputs assist in the construction of catalogues for reference genes at different human microbiome sites. All of the 'multi-omics' data sets generated are interpreted individually and the results will depict any associations between the microbiome and health and disease. Most microbiome studies focus on this particular area and their data can be used as an input to the GEMs reconstruction process. (B). Based on availability of whole genome sequence data for the target microorganism, a GEM can be generated. The high-quality reads can be used to construct gene and pathway summaries [45], and this needs to be implemented in the process of GEMs generation. Since the individual phenotypic knowledge for most of these microbes is missing, omics data is used to compile a set of metabolic tasks for evaluation and validation of GEMs functionality. (C). To perform simulations with GEMs, it is necessary to introduce an objective function and maximizing biomass yield is the most relevant one for microbes metabolic modelling. The steps for high quality GEMs reconstruction has been extensively reviewed in different articles [46]. FBA is applied to simulated-ready GEMs for microbiome to predict the target organism phenotype under certain constraints. GEMs, as fully connected and functional networks are a great platform to perform integrative analysis of clinical data for identification of relevant predictive biomarkers as well as novel therapeutic targets for microbiome associated diseases. (D). The GEMs' generated hypothesis can be in the form of probiotic and prebiotic design or gene knock. *In-vivo* and *in-vitro* experiments would assist in evaluating the GEMs predictions at the first stage and the confirmed could be used for clinical trials. Using the generated GEMs on human tissue/cells, one can explore the effect of a generated hypothesis on human host physiology using the simulated-ready tissue/cell GEMs [47]. Overall, this proposed pipeline can effectively speed up the generation of specific diagnosis and treatments in microbiome studies, although it requires more dedicated data generation for constructing high quality models.