King's Research Portal

[Link to publication record in King's Research Portal](#)

Accepted Manuscript

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

Elizabeth A. Witherden, David L. Moyes, Kenneth D. Bruce, Stanislav D. Ehrlich, Saeed Shoaie

Please cite this article as: Witherden EA, Moyes DL, Bruce KD, Ehrlich SD, Shoaie S, Using systems biology approaches to elucidate cause and effect in host-microbiome interactions, *Current Opinion in Systems Biology* (2017), doi: 10.1016/j.coisb.2017.05.003.

Title:

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

Authors:

Elizabeth A. Witherden[1], David L. Moyes[1], Kenneth D. Bruce[1], Stanislav D. Ehrlich[1,2] and Saeed Shoaie[1,3,*]

Author Affiliations:

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

[2] Metagenopolis, Institut National de la recherché Agronomique, 78350, Jouy en Josas, France.

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

* Corresponding author

All authors contributed equally.

Corresponding Author Details:

Email: saeed.shoaie@kcl.ac.uk, phone: (+44) 0207 188 4362

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?

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 ~1,000 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 paucity of genes and genomes from non-bacterial origins in these catalogues means that despite covering many of the genes present in the microbiome, current catalogues under-represent many distinct gene families from entirely different evolutionary paths - e.g. eukaryotic genes. Further, although these catalogues are now being created and curated, the next big question is what do we do with this data to achieve its maximum potential? As we begin to answer this, we need to develop a variety of systems biology tools and platforms to take us in to the next phase of analyses.

< Application of Systems Biology Approaches to study Host-Microbiome Interactions >

Microbiome association studies have shown links between certain microorganisms and chronic conditions including Type 2 Diabetes [20,21], liver cirrhosis [22] and colon cancer [23]. None, however, have moved beyond this and elucidated the causalities [20-23]. There is now great interest in using systems biology approaches to elucidate the causalities between microbial species and their individual contribution to the overall ecosystem's

phenotype and interactions with diet and host [24]. Key in the systems biology are genome scale models.

Genome-scale models are the common denominator of systems biology, and have been applied as a powerful scaffold to identify the genotype-phenotype relationships in both individual bacterial species and microbial communities [25]. These models are set up to describe the complex cellular functions through the integration of 'multi-omics' data and specific objective-functions. Reconstruction of Genome-scale metabolic Models (GEMs) has become well-established over the last decade for a variety of microbial species and host tissues/cell line interactions [26-28]. Since most of the bacterial species in the human microbiome are typically challenging to culture *in vitro*, mainly bottom-up reconstruction methods have been used. Genome sequences and similarity based annotations are the main data input required for such reconstructions. As Figure 1B illustrates, GEMs consist of biochemical reactions with their gene-protein associations of the target organism. For mathematical representations of GEMs, the stoichiometric coefficients are used to construct a stoichiometric (S) matrix. Rows and columns in the S matrix consist of "all" the involved metabolites and reactions in the network. The S matrix plays a key role in different systems biology tools and applications since it enables the mathematical formulation of the different biological networks. Several GEMs have been reconstructed in bottom-up approaches for bacteria that are constituents of different human microbiomes. In this regard, having a well-established gene catalogue and integrating it with 'multi-omics' data forms an important complement to high-quality systems biology models. Unfortunately, GEMs are currently only applied to microbial networks, although a limited number of studies are beginning to integrate microbial datasets with both host and fungal networks.

Figure 1C shows a GEM as a well-connected network that can be applied for network dependent analysis using 'multi-omics' data or through constraints implementation assisting in determining the phenotypic potential of a target organism. GEMs are widely applied in constraint-based modelling, referred to as Flux Balance Analysis (FBA), to predict and interpret physiological data and moreover, used in design and discovery [29,30]. Like host modelling, the application of microbiome GEMs has been mainly evolved in two paths. Using the network properties and contextualizing of high-throughput data through mapping 'multi-omics' data to GEM to identify reporter metabolites and/or sub-networks [31,32]. Another path is applying constraint-based modelling to predict the cellular phenotypes[33]. Using both applications in microbiome studies, has made it possible to elucidate the interactions between different microbial species, and the overall contribution of individual microorganisms to microbiome metabolism, host phenotype and nutrients uptake [34]. Generating GEMs for bacteria from the predominant taxa identified in the human microbiome and subsequently performing FBA to predict interactions, demonstrated how the gut microbiome and diet interact and influence amino acid profiles seen in the plasma [34,35]. This modelling approach can be validated using the data from mono and co-colonized bacteria in germ free mice. Several optimization algorithms have been created to

predict the interactions between individual, while the overall microbial community and each bacterium are optimized [33,34,36]. This type of community and systems-level optimization has been applied to the human gut microbiome and successfully predicted the profile of key metabolites in faeces and plasma [34]. Further, this approach has been applied to determine the best diet to "improve" a host phenotype using the abundances of an individual's gut microbiome. In addition, compartmentalizing the microbiome metabolism based on each species, enables integrative analyses using transcript data and the investigation of how transcriptional responses between microorganisms within the community vary in different conditions. Such analyses allow for the identification of different diagnostic biomarkers and novel therapeutic targets for metabolic diseases that are associated with the microbiome (Figure 1D)[37].

Nowhere is the new world of possibilities being opened up by GEMs and other modelling analysis techniques more evident than in the genesis of explorations of the antimicrobial resistance gene (ARG) profile of a microbiome, otherwise known as the "resistome" [38]. Between 2005 and 2010, there were 28 papers mentioning resistome on Pubmed. In contrast, the next 5 years (2011 - 2015) had 150 papers with almost two thirds of these investigating the resistome in the environment. In 2016 alone, there were 92 papers with half now relating to the resistome in host organisms.

Analysis of the resistome relies on the gene sequence detail that is increasingly available via metagenomics to define the presence and abundance of specific gene sets representing the resistome. This provides an ability to track the development and spread of specific antimicrobial resistance genes through different communities and habitats [39-41]. Whilst it may seem counter-intuitive to be using GEMs in analysing the resistome, given the involvement of these genes in resistance to antimicrobial drugs, however, if we consider what many of these genes are predominantly involved in, the use of GEMs becomes more obvious when we realise the primary function for many of these genes. Although they are important for resistance to antimicrobial drugs, many of these genes in their unmutated form are either direct targets for these drugs, or involved in the cellular pathways targeted by these drugs. Likewise, many of the targets for antimicrobial drugs are either directly or indirectly involved in metabolic pathways or other essential cellular processes. Thus, they are amenable to analysis in two different ways - both through analysis of the development and movement of these genes, as well as the pattern of the genes. Using systems biology approaches such as GEMs to analyse the resistome has immense predictive power, for example in defining how our microbiota will affect our responses to xenobiotics (drugs, dietary compounds and toxins) [42]. As such, this moves well beyond a simple understanding of gene presence and abundance, and provides an unprecedented opportunity to examine the resistome transmission from the environment to human or animal hosts, and even between individuals, as well as the selection pressures and mechanisms of evolution of these genes within a community. Given the global concern

relating to antimicrobial resistance and its rapid spread, this represents a particularly important tool set [43]. As such, we are now beginning to explore the concepts of functional pools of genes within microbial communities, along with the potential for the transfer of these genes between different species and communities, as well as the way these pools of genes may change in dysbiotic conditions associated with host disease states.

Conclusions:

The current focus of human host-microbiome studies is centred around generating 'multi-omics' datasets to investigate the role of the microbiome in human health and diseases. Generating such data is remarkably important in the microbiome field for understanding the interactions between microbes and their host. A recent study has used GEMs of host and microbiome data on conventionally raised and germ-free mice and showed that global metabolic differences in mice tissues was influenced by the gut microbiome [44]. This therefore highlights the benefits of systems biology approaches and its capabilities for describing mechanistic relationships in the microbiome and host-microbial interactions. This is a necessary step-forward in microbiome research allowing for a better explanation of the role of the microbiome in associated diseases. In this concept, Genome-scale metabolic science is a great platform to understand causalities, perform integrative analysis, simulations, design, discovery, clinical interventions.

Acknowledgments:

References:

1. Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC, Huttenhower C: **Sequencing and beyond: integrating molecular 'omics' for microbial community profiling**. *Nat Rev Microbiol* 2015, **13**:360-372.
2. Group NHW, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, et al.: **The NIH Human Microbiome Project**. *Genome Res* 2009, **19**:2317-2323.
3. Greenblum S, Chiu HC, Levy R, Carr R, Borenstein E: **Towards a predictive systems-level model of the human microbiome: progress, challenges, and opportunities**. *Curr Opin Biotechnol* 2013, **24**:810-820.
4. Thiele I, Heinken A, Fleming RM: **A systems biology approach to studying the role of microbes in human health**. *Curr Opin Biotechnol* 2013, **24**:4-12.
5. Borenstein E: **Computational systems biology and in silico modeling of the human microbiome**. *Brief Bioinform* 2012, **13**:769-780.
6. Manor O, Borenstein E: **Systematic Characterization and Analysis of the Taxonomic Drivers of Functional Shifts in the Human Microbiome**. *Cell Host Microbe* 2017.
7. Karlsson FH, Nookaew I, Petranovic D, Nielsen J: **Prospects for systems biology and modeling of the gut microbiome**. *Trends Biotechnol* 2011, **29**:251-258.
8. Oberhardt MA, Palsson BO, Papin JA: **Applications of genome-scale metabolic reconstructions**. *Mol Syst Biol* 2009, **5**:320.
9. Roumpeka DD, Wallace RJ, Escalettes F, Fotheringham I, Watson M: **A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data**. *Front Genet* 2017, **8**:23.
10. Human Microbiome Project C: **A framework for human microbiome research**. *Nature* 2012, **486**:215-221.
11. Human Microbiome Project C: **Structure, function and diversity of the healthy human microbiome**. *Nature* 2012, **486**:207-214.
12. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO: **The SILVA ribosomal RNA gene database project: improved data processing and web-based tools**. *Nucleic Acids Res* 2013, **41**:D590-596.
13. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL: **Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB**. *Appl Environ Microbiol* 2006, **72**:5069-5072.
14. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al.: **A human gut microbial gene catalogue established by metagenomic sequencing**. *Nature* 2010, **464**:59-65.
15. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, et al.: **An integrated catalog of reference genes in the human gut microbiome**. *Nat Biotechnol* 2014, **32**:834-841.
16. Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, Li X, Long H, Zhang J, Zhang D, et al.: **A catalog of the mouse gut metagenome**. *Nat Biotechnol* 2015, **33**:1103-1108.
17. Xiao L, Estelle J, Kiilerich P, Ramayo-Caldas Y, Xia Z, Feng Q, Liang S, Pedersen AO, Kjeldsen NJ, Liu C, et al.: **A reference gene catalogue of the pig gut microbiome**. *Nat Microbiol* 2016:16161.
18. Almeida M, Hebert A, Abraham AL, Rasmussen S, Monnet C, Pons N, Delbes C, Loux V, Batto JM, Leonard P, et al.: **Construction of a dairy microbial genome catalog opens new perspectives for the metagenomic analysis of dairy fermented products**. *BMC Genomics* 2014, **15**:1101.
19. Wu H, Tremaroli V, Backhed F: **Linking Microbiota to Human Diseases: A Systems Biology Perspective**. *Trends Endocrinol Metab* 2015, **26**:758-770.
20. Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, Nielsen J, Backhed F: **Gut metagenome in European women with normal, impaired and diabetic glucose control**. *Nature* 2013, **498**:99-103.

21. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Krogh Pedersen H, et al.: **Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota**. *Nature* 2015, **528**:262-266.

22. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L, et al.: **Alterations of the human gut microbiome in liver cirrhosis**. *Nature* 2014, **513**:59-64.

23. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Bohm J, Brunetti F, Habermann N, et al.: **Potential of fecal microbiota for early-stage detection of colorectal cancer**. *Mol Syst Biol* 2014, **10**:766.

24. Sonnenburg JL, Backhed F: **Diet-microbiota interactions as moderators of human metabolism**. *Nature* 2016, **535**:56-64.

25. Lewis NE, Nagarajan H, Palsson BO: **Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods**. *Nat Rev Microbiol* 2012, **10**:291-305.

26. Mardinoglu A, Agren R, Kampf C, Asplund A, Nookaew I, Jacobson P, Walley AJ, Froguel P, Carlsson LM, Uhlen M, et al.: **Integration of clinical data with a genome-scale metabolic model of the human adipocyte**. *Mol Syst Biol* 2013, **9**:649.

27. Ghaffari P, Mardinoglu A, Asplund A, Shoaie S, Kampf C, Uhlen M, Nielsen J: **Identifying anti-growth factors for human cancer cell lines through genome-scale metabolic modeling**. *Sci Rep* 2015, **5**:8183.

28. Shoaie S, Nielsen J: **Elucidating the interactions between the human gut microbiota and its host through metabolic modeling**. *Front Genet* 2014, **5**:86.

29. Agren R, Liu LM, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J: **The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum**. *Plos Computational Biology* 2013, **9**.

30. O'Brien EJ, Monk JM, Palsson BO: **Using Genome-scale Models to Predict Biological Capabilities**. *Cell* 2015, **161**:971-987.

31. Greenblum S, Turnbaugh PJ, Borenstein E: **Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease**. *Proc Natl Acad Sci U S A* 2012, **109**:594-599.

32. Patil KR, Nielsen J: **Uncovering transcriptional regulation of metabolism by using metabolic network topology**. *Proc Natl Acad Sci U S A* 2005, **102**:2685-2689.

33. Zomorrodi AR, Maranas CD: **OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities**. *Plos Computational Biology* 2012, **8**.

34. Shoaie S, Ghaffari P, Kovatcheva-Datchary P, Mardinoglu A, Sen P, Pujos-Guillot E, de Wouters T, Juste C, Rizkalla S, Chilloux J, et al.: **Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome**. *Cell Metabolism* 2015, **22**:320-331.

35. Shoaie S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S, Nielsen J: **Understanding the interactions between bacteria in the human gut through metabolic modeling**. *Scientific Reports* 2013, **3**.

36. Chowdhury A, Zomorrodi AR, Maranas CD: **Bilevel optimization techniques in computational strain design**. *Computers & Chemical Engineering* 2015, **72**:363-372.

37. Levy R, Borenstein E: **Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules**. *Proc Natl Acad Sci U S A* 2013, **110**:12804-12809.

38. Wright GD: **The antibiotic resistome: the nexus of chemical and genetic diversity**. *Nat Rev Microbiol* 2007, **5**:175-186.

39. Willmann M, Peter S: **Translational metagenomics and the human resistome: confronting the menace of the new millennium**. *J Mol Med (Berl)* 2017, **95**:41-51.

40. Dos Santos DF, Istvan P, Quirino BF, Kruger RH: **Functional Metagenomics as a Tool for Identification of New Antibiotic Resistance Genes from Natural Environments**. *Microb Ecol* 2017, **73**:479-491.

41. Sukumar S, Roberts AP, Martin FE, Adler CJ: **Metagenomic Insights into Transferable Antibiotic Resistance in Oral Bacteria**. *J Dent Res* 2016, **95**:969-976.

42. Spanogiannopoulos P, Bess EN, Carmody RN, Turnbaugh PJ: **The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism**. *Nat Rev Microbiol* 2016, **14**:273-287.

43. Manaia CM: **Assessing the Risk of Antibiotic Resistance Transmission from the Environment to Humans: Non-Direct Proportionality between Abundance and Risk**. *Trends Microbiol* 2016.

44. Mardinoglu A, Shoaie S, Bergentall M, Ghaffari P, Zhang C, Larsson E, Backhed F, Nielsen J: **The gut microbiota modulates host amino acid and glutathione metabolism in mice**. *Mol Syst Biol* 2015, **11**:834.

45. Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, Rodriguez-Mueller B, Zucker J, Thiagarajan M, Henrissat B, et al.: **Metabolic reconstruction for metagenomic data and its application to the human microbiome**. *PLoS Comput Biol* 2012, **8**:e1002358.

46. Thiele I, Palsson BO: **A protocol for generating a high-quality genome-scale metabolic reconstruction**. *Nature Protocols* 2010, **5**:93-121.

47. Mardinoglul A, Nielsen J: **New paradigms for metabolic modeling of human cells**. *Current Opinion in Biotechnology* 2015, **34**:91-97.

Reference Annotations:

Papers of particular interest have been annotated and highlighted as:

* of special interest.

**of outstanding interest.

** 3. Greenblum S, Chiu HC, Levy R, Carr R, Borenstein E: **Towards a predictive systems-level model of the human microbiome: progress, challenges, and opportunities**. *Curr Opin Biotechnol* 2013, **24**:810-820.
This opinion piece highlights the idea that microbiome studies using next-generation sequencing technology should aim to study the microbiome as a distinct community of microbial species whose metabolism is tightly intertwined. They propose a shift toward using predictive modelling networks to study the human microbiome in health and disease and discuss the expectations and limitations of doing so.

* 10. Human Microbiome Project C: **Structure, function and diversity of the healthy human microbiome**. *Nature* 2012, **486**:207-214.
This paper describes the large-scale Human Microbiome Project, that aimed to map the healthy human microbiome. A total of 242 healthy American adults had their microbiomes mapped across 18 distinct body sites. This study highlighted the diversity of the community composition and metabolic function of the microbiome at distinct body sites.

** 14. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al.: **A human gut microbial gene catalogue established by metagenomic sequencing**. *Nature* 2010, **464**:59-65.
This multi-centre study characterised the gut microbiome of 124 European individuals and established a 3.3 million non-redundant microbial gene catalogue. This gene catalogue identified a core set of microbial functions within the human gastrointestinal tract.

** 15. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, et al.: **An integrated catalog of reference genes in the human gut microbiome**. *Nat Biotechnol* 2014, **32**:834-841.
This multi-cohort study combines metagenomic sequence data from 249 individuals from the MetaHIT project with data from 1,018 individuals from previous HMP, MetaHIT and Chinese diabetes studies, to develop the largest integrated gene catalogue (IGC) for the gastrointestinal tract to date. This IGC comprises 9,879,896 genes that represent almost complete sets of genes from the majority of microorganisms that live in the gut.

** 33. Zomorrodi AR, Maranas CD: **OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities**. *Plos Computational Biology* 2012, **8**.
OptCom is a platform for flux balance analysis of microbial communities generated in this study. A multi-level optimization has been used to quantify the metabolic interactions between the individuals as well as individuals and the community are optimized.

** 34. Shoaie S, Ghaffari P, Kovatcheva-Datchary P, Mardinoglu A, Sen P, Pujos-Guillot E, de Wouters T, Juste C, Rizkalla S, Chilloux J, et al.: **Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome**. *Cell Metabolism* 2015, **22**:320-331.

Here the authors have developed system-level optimization for microbial community modelling useings GEMS. This modelling concept predicts the profile of key metabolites in plasma and faecal samples influenced by gut microbes.

** 38. Wright GD: **The antibiotic resistome: the nexus of chemical and genetic diversity**. *Nat Rev Microbiol* 2007, **5**:175-186.

This review explores the diversity of antimicrobial resistance genes present within the microbiome, exploring how they are selected for and how they evolve from other genes more normally involved in regular metabolic processes.

*42. Spanogiannopoulos P, Bess EN, Carmody RN, Turnbaugh PJ: **The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism**. *Nat Rev Microbiol* 2016, **14**:273-287.

This review delineates how the microbiome and its global metabolism affects the presence of different 'xenobiotics' in the human gut, showing how different microbial metabolic processes can interact with each other and host processes to deliver new compounds affecting efficacy of drugs or with toxic effects.

** 44. Mardinoglu A, Shoaie S, Bergentall M, Ghaffari P, Zhang C, Larsson E, Backhed F, Nielsen J: **The gut microbiota modulates host amino acid and glutathione metabolism in mice**. *Mol Syst Biol* 2015, **11**:834.

The authors of this study showed how the gut microbiome impacts metabolism of the small intestine, liver and adipose tissue using conventionally raised and germ free mice gene expression data and generating GEMs for tissue specific microbial strains.

Figure 1 Caption:

< 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. >
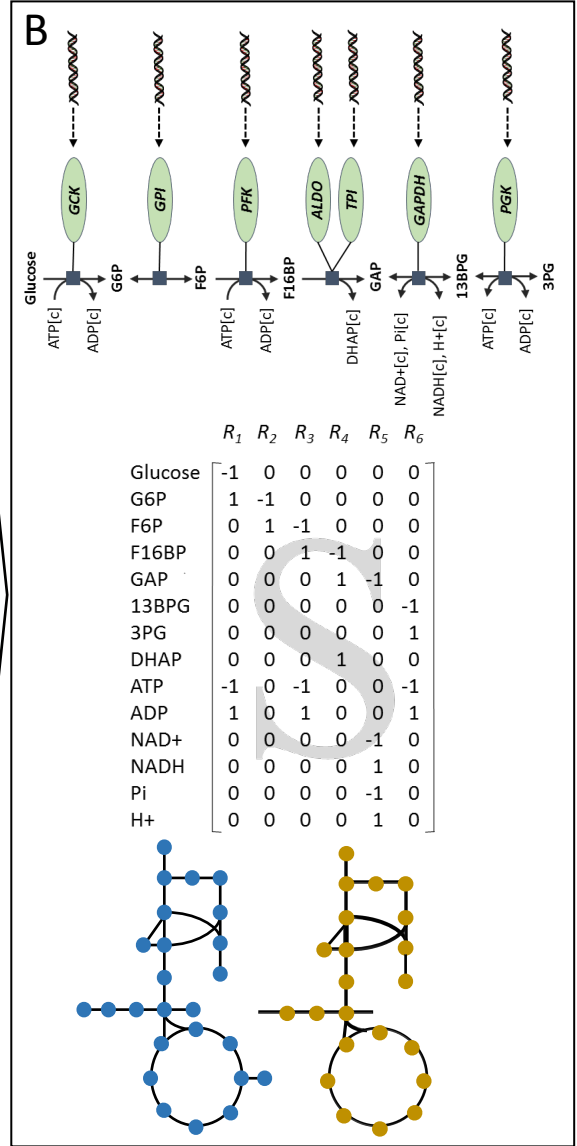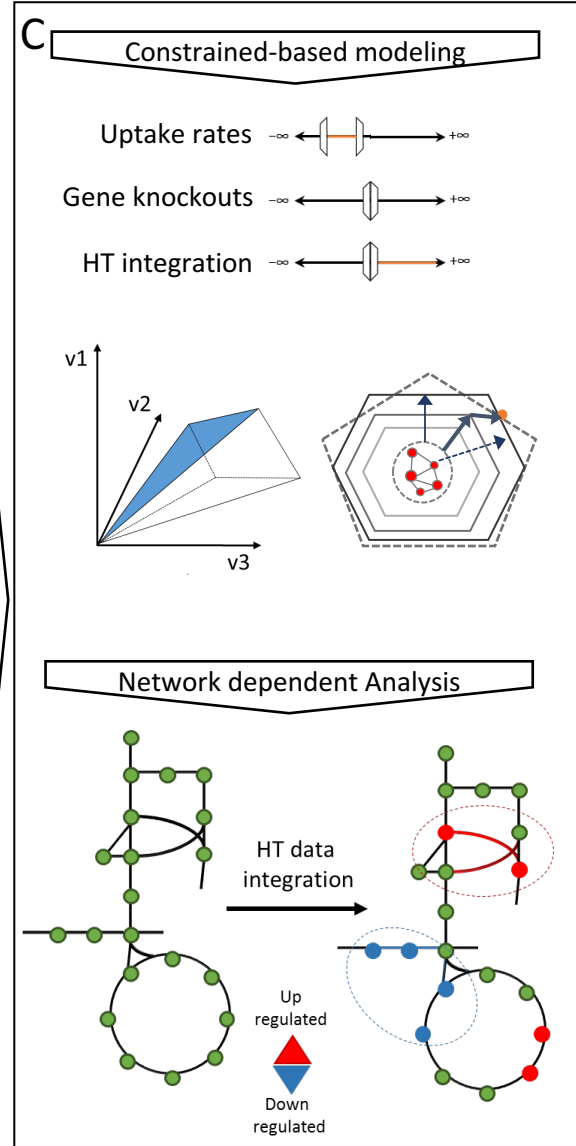
After microbiome sampling of healthy and diseased individuals, different high-throughput (HT) analysis can be applied to the samples (A). Metagenomic outputs assist in the construction of catalogues for reference genes at different human microbiome sites. All of the 'multi-omics' data sets generated 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.
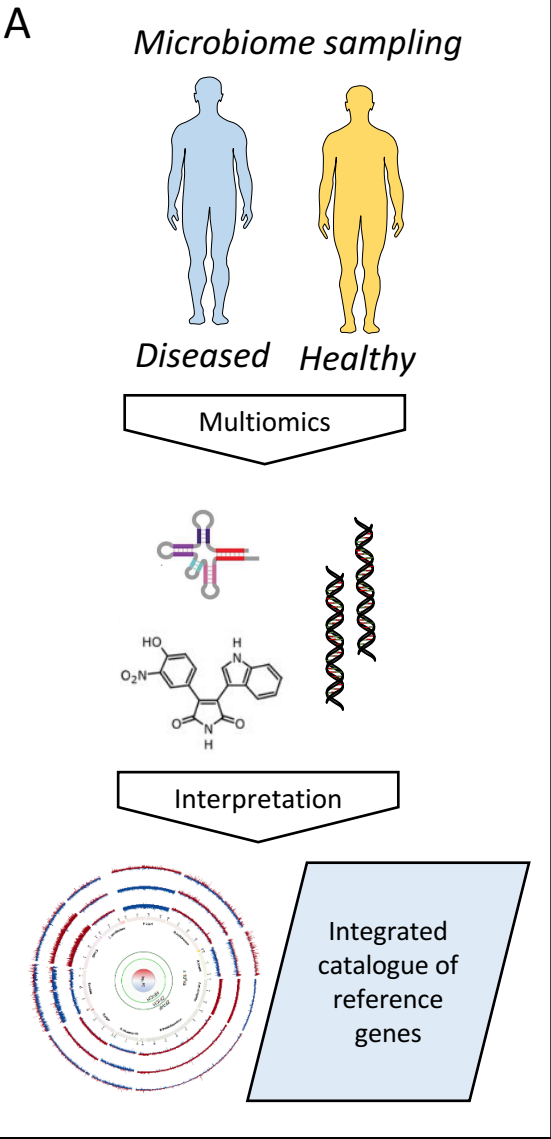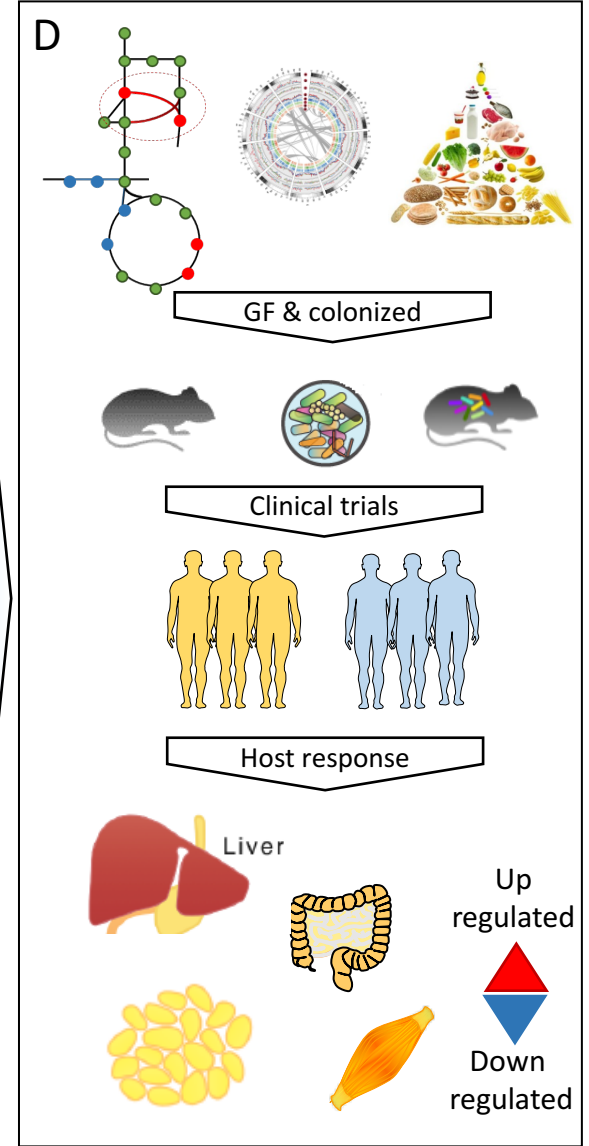
A

Microbiome sampling

Diseased *Healthy*

Multiomics

Interpretation

Integrated catalogue of reference genes

B

GCK   GPI   PFK   ALDO   TPI   GAPDH   PGK

Glucose G6P F6P F16BP GAP 13BPG 3PG

ATP[c] ADP[c]   ATP[c] ADP[c]   DHAP[c]   NAD+[c], Pi[c] NADH[c], H+[c]   ATP[c] ADP[c]

	R_1	R_2	R_3	R_4	R_5	R_6
Glucose	-1	0	0	0	0	0
G6P	1	-1	0	0	0	0
F6P	0	1	-1	0	0	0
F16BP	0	0	1	-1	0	0
GAP	0	0	0	1	-1	0
13BPG	0	0	0	0	0	-1
3PG	0	0	0	0	0	1
DHAP	0	0	0	1	0	0
ATP	-1	0	-1	0	0	-1
ADP	1	0	1	0	0	1
NAD+	0	0	0	0	-1	0
NADH	0	0	0	0	1	0
Pi	0	0	0	0	-1	0
H+	0	0	0	0	1	0

C

Constrained-based modeling

Uptake rates $-\infty$ ◄———► $+\infty$

Gene knockouts $-\infty$ ◄———► $+\infty$

HT integration $-\infty$ ◄———► $+\infty$

v1
v2
v3

Network dependent Analysis

HT data integration

Up regulated

Down regulated

D

GF & colonized

Clinical trials

Host response

Liver

Up regulated

Down regulated

Title:

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

Highlights:

- Metagenomic gene catalogues to include archaea, fungi and bacteria proposed.
- System-level framework for processing multi-omics data sets developed.
- Review of Systems Biology approaches for studying host-microbiome interactions.