



Metabolic modeling with Big Data and the gut microbiome



Jaeyun Sung^{a,1}, Vanessa Hale^{b,c,1}, Annette C. Merkel^b, Pan-Jun Kim^{a,d}, Nicholas Chia^{b,c,e,*}

^a Asia Pacific Center for Theoretical Physics, Pohang, Gyeongbuk 37673, Republic of Korea

^b Center for Individualized Medicine, Microbiome Program, Mayo Clinic, Rochester, MN 55905, USA

^c Department of Surgery, Mayo Clinic, Rochester, MN 55905, USA

^d Department of Physics, Pohang University of Science and Technology, Pohang, Gyeongbuk 37673, Republic of Korea

^e Department of Biomedical Engineering, Mayo College, Rochester, MN 55905, USA

ARTICLE INFO

Article history:

Received 10 August 2015

Received in revised form 19 January 2016

Accepted 2 February 2016

Keywords:

Gut microbiome

Metabolic modeling

Microbial community

Data integration

Big Data

ABSTRACT

The recent advances in high-throughput omics technologies have enabled researchers to explore the intricacies of the human microbiome. On the clinical front, the gut microbial community has been the focus of many biomarker-discovery studies. While the recent deluge of high-throughput data in microbiome research has been vastly informative and groundbreaking, we have yet to capture the full potential of omics-based approaches. Realizing the promise of multi-omics data will require integration of disparate omics data, as well as a biologically relevant, mechanistic framework – or metabolic model – on which to overlay these data. Also, a new paradigm for metabolic model evaluation is necessary. Herein, we outline the need for multi-omics data integration, as well as the accompanying challenges. Furthermore, we present a framework for characterizing the ecology of the gut microbiome based on metabolic network modeling.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Introduction	10
2. The Big Data paradox	11
3. Metabolic models for Big Data synthesis	11
4. The challenges of Big Data	13
5. Future direction	13
Acknowledgments	14
References	14

1. Introduction

The promise of the Big Data revolution has yielded an ever-increasing array of data and data types in many fields. In the medical field, the sequencing of the human genome in 2003 opened the door to truly individualized medicine, tailored to our genetic predispositions and risk factors (Collins et al., 2003). The first manifestations of the Big

Data promise in medicine were necessarily surveys to identify biological markers of disease risk. While this resulted in databases upon databases of genetic events that explained risk behind hundreds of diseases, we quickly learned that genetics alone was not able to provide a full understanding of many health conditions (Lander, 2011). Researchers began to examine other factors, including the role of environmental influences such as the microbiome (Bultman, 2013; Zackular et al., 2013).

In 2008, the Human Microbiome Project (HMP) was established to characterize the role of human-associated microbial communities in human health and disease (Méthé et al., 2012; The Human Microbiome Project, 2012). Efforts led by the HMP consortium thus far have yielded numerous insights regarding the microbial composition of the human body and the ecological structure and function of the

Abbreviations: GEMs, genome scale metabolic models; COMM, community-scale metabolic modeling; HMP, Human Microbiome Project.

* Corresponding author at: 200 First St. SW, Rochester, MN 55905, USA.

E-mail address: chia.nicholas@mayo.edu (N. Chia).

¹ These authors contributed equally.

human microbiome. However, a shift from this “profiling” paradigm to one of mechanistic examination is now both warranted and feasible through the integration of multi-omics data onto a framework based on biomolecular pathways and networks.

The gut microbial community is increasingly well-characterized by various omics technologies – metagenomics, metatranscriptomics, metabolomics, metaproteomics – and offers much promise for data integration within a mechanistic framework (Erickson et al., 2012; Haiser et al., 2013a, 2013b; Weir et al., 2013). Gut microbes act as chemical transformers, converting host-acquired or host-produced nutrients into a milieu of metabolites (Lee and Hase, 2014). At the same time, the structure and function of the microbial community respond to changes in host diet or physiology (David et al., 2014; Kashyap et al., 2013; Liou et al., 2013), making microbes both modulators and reflections of the gut environment.

The gut microbiome contains over 3 million genes, or approximately 150-fold more than the human genome (Qin et al., 2010); thus, it becomes virtually impossible to obtain more independent samples than there are measured values within one individual's microbiome. The large data sets generated by the most recent omics technologies call for new methods of analysis. No longer can we afford to use a paradigm of statistical power where our insight dwindles with the amount of data we collect. Instead, we should rely on the fact that these variables are not independent of one another and therefore establish a more practical model for assessing the role of the microbiome.

A systems approach that utilizes metabolic networks may offer a potential solution. Network reconstruction is one such means of creating a scaffold for synthesizing multiple data types (Feist and Palsson, 2008; Lee et al., 2012; Reed et al., 2006; Töpfer et al., 2015). Metabolic models are composed of a collection of individual chemical reactions that are governed by the fundamental laws of mass conservation and thermodynamics. These models represent large-scale complex cellular dynamics and imply a network whose mechanistic chain of events can be computed to produce an outcome. Models are capable of converting large amounts of data – genetic, metabolic, biochemical – into phenotypes and interactions. The value of metabolic modeling for understanding the complex environment of the gut microbiome is in resolving biochemical relationships within and between microbial species and potentially predicting the effect of ecosystem-wide perturbations, such as antibiotics or pathogen invasion. There have been many recent efforts to model metabolic processes within microbial communities (Heinken and Thiele, 2015; Henry et al., 2009). However, the wealth of data available through multiple omics technologies remains underutilized in these models.

In this review, we discuss the promises and limitations offered by current mathematical paradigms for integrating disparate, yet complementary omics data, while pointing out the challenges that remain to be resolved. Finally, we offer our viewpoint on the need for an updated network-aware mathematical framework for statistical power – one that synthesizes multiple channels of information into a biological picture.

2. The Big Data paradox

The mathematical formalization of our knowledge is one of the most important aspects of any scientific study or clinical trial. As a practical tool, math is a means for taking pattern recognition and systematizing it. It is also a way for us to provide some form of communication and standard for comparing the results of different studies, and in the case of statistical significance, is meant to provide a measure of certainty against a null hypothesis.

Historically, clinical trials were developed around randomized treatment arms that were designed to answer the question, “Which treatment (A, B, or C) is better?” By selecting a straight forward metric, such as survival outcomes, statisticians could compare the efficacy of different treatments (Marubini and Valsecchi, 2004); however, this

precluded our ability to ask what would happen if we combined treatment A and B, or B and C, or all 3 treatments, except by running yet another clinical trial. At the center of these often long and laborious trials was the notion of statistical power (Lachin, 1981). Just how many cases and controls does one need to ensure we can achieve significance? It is a simple question, but an important one that has been the subject of many sophisticated refinements. Here, there is a fundamental clash between Big Data science and classic clinical trial statistics. Paradoxically, the more data we collect on each subject, the more we decrease our likelihood of identifying statistically significant parameters as a result of multiple hypothesis correction. This is a fundamental flaw in the way that current statistical power calculations deal with large datasets.

Approaches to obtaining information from Big Data are different. Big Data is characterized by high volume, variety, and velocity of data generation (Costa, 2013). The strength of multi-omics is not merely the observation of many data points, but the discovery of biological mechanism through observation. Multi-omic Big Data grants us the power to examine disease in a human biological context, rather than extensively relying on murine models, which are limited in relevance to the human gut microbiome (Nguyen et al., 2015). In order to succeed, the Big Data movement in individualized medicine will require a holistic merger between large-scale data and biological mechanism.

3. Metabolic models for Big Data synthesis

To identify specific biological markers of disease, many studies utilize statistical correlations, which fall short of identifying underlying mechanisms (de Vos and de Vos, 2012). In the past, using Big Data to elucidate a biological mechanism involved generating a limited set of hypotheses that were then tested in the lab. While this approach has great value, it becomes less tenable as the number of measurements grows. The massive data sets generated from high-throughput omics technologies guarantee us more correlations arising purely from random chance. In the gut microbiome, this is especially problematic. The number of potential correlations increases with the hundreds of species and thousands of genes. Furthermore, the number of identified factors contributing to microbial composition including diet (David et al., 2014; De Filippo et al., 2010; Turnbaugh et al., 2009), sex (Chen et al., 2016), and even preservation method of the sample (Sinha et al., 2015), continue to grow and make it more difficult to differentiate the confounding from the causal.

A metabolic network provides a global picture of how metabolites and biochemical reactions are interconnected within a particular organism (Thiele and Palsson, 2010). Flux balance analysis on genome-scale metabolic models (GEMs) can be used to simulate microbial growth or to predict the production rate of a particular metabolite (Palsson, 2015). The power of this approach is not only that it recapitulates the mechanistic chemical flow through an entire organism, but also that it has the potential to integrate multiple data types. As the example shown in Fig. 1 indicates, reactions can be linked to genes, which are informed by DNA or RNA sequencing. RNA expression informs the amount of flux a reaction can carry, and metabolomics is a direct measurement of the metabolites. This makes metabolic models an ideal platform for organismal and community-scale data synthesis.

Increasing evidence suggests that integrating disparate, but complementary, data types can increase the power of one's analysis. Examples of this include the use of whole genome sequencing as a scaffold for RNA data (J. Wang et al., 2013; K. Wang et al., 2013) and the use of phosphorylation data to understand changes in metabolite concentration (Yugi et al., 2014). Within the microbiome field, 16S rRNA data is combined with metagenomics to identify representative genomes and genome characteristics (PICRUSt: Langille et al., 2013; HUMAnN: Abubucker et al., 2012). Recent microbiome studies have also combined metagenome and metatranscriptome data to enable comparison between functional potential (metagenomic abundance/gene copy number) and functional activity (transcriptome level) (Franzosa et al.,

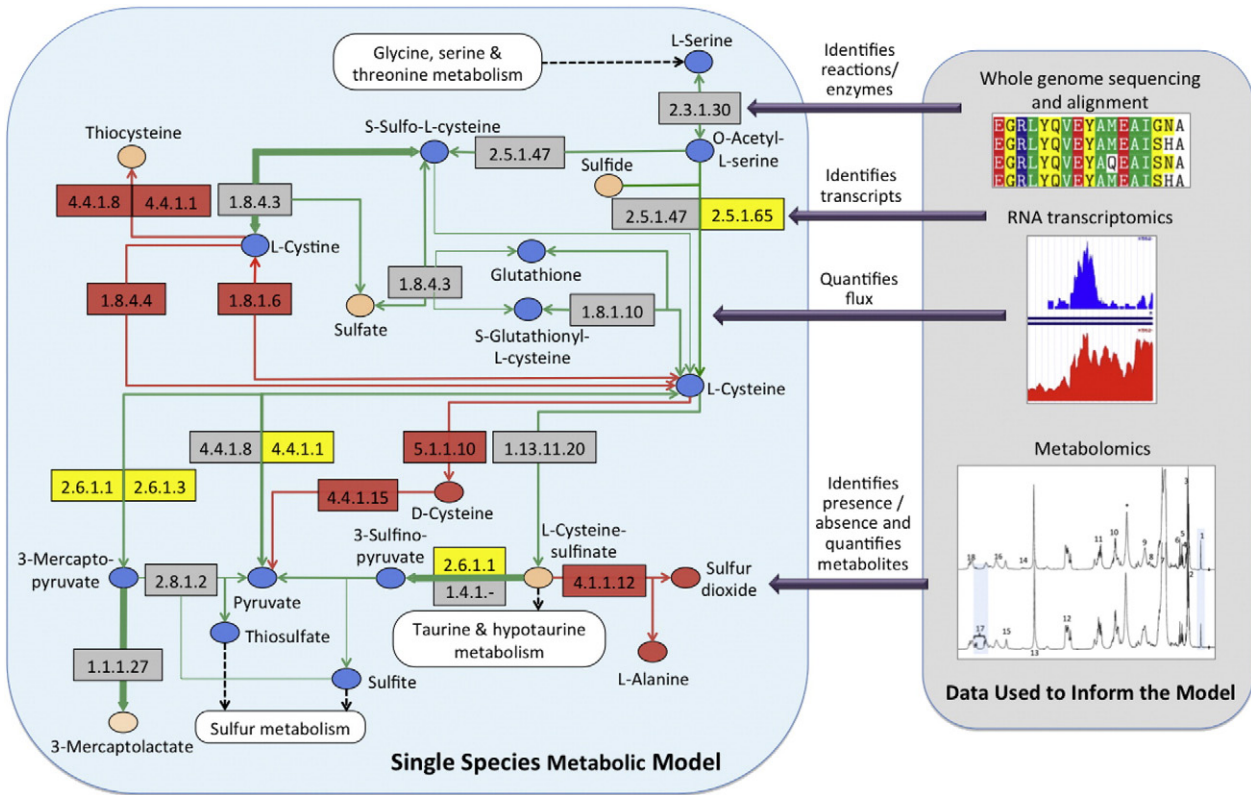


Fig. 1. Subset of a microbial metabolic network with integrated genome, metabolomics, and RNA data. This network is one portion of a cysteine/methionine metabolic network for one bacterial species. The model is constructed based on the bacterial genome. Each box represents a reaction. The numbers within the boxes are KEGG Enzyme Commission (EC) number and code for specific enzymes present in each reaction. Gray boxes represent reactions that occur in this bacteria, as predicted by its genome. Red boxes denote reactions that are not predicted by the genome. Circles represent metabolites consumed and produced within the reaction network. Arrows represent reaction pathways that do (green) or do not (red) occur in this bacteria, as predicted by the model. Black dashed arrows indicate input or output from or to other metabolic networks. Synthesis of omics data is used to inform and improve the model. For example, RNA transcriptomic data reveal what enzymes are being transcribed. In pathways that contain 2 possible enzymes that carry out the same reaction, RNA transcripts help us distinguish which of the enzyme(s) are active. In pathways catalyzed by more than 1 enzyme, yellow boxes indicate reactions/enzymes supported by RNA data. RNA data also quantifies flux which allows us to weight the reaction pathways accordingly: in this model, pathways with the greatest flux have the thickest arrows. Metabolomic data is also used to inform the model. Blue circles represent metabolites present and quantified through metabolomics. Red circles indicate metabolites that were not present or quantifiable. Peach circles represent metabolites that cannot currently be identified using metabolomics.

2015). This provides insights regarding host–microbial dynamics (Franzosa et al., 2014), highlights functional changes in the microbial community in response to diet (McNulty et al., 2011), and suggests potential disease mechanisms, as in the case with periodontitis (J. Wang et al., 2013; K. Wang et al., 2013). Metagenome or metatranscriptome data alone would not yield these insights. Other examples of omic

data integration include the combination of microbiome and metabolome data in the study of colon cancer (Weir et al., 2013), proteome and metagenome data related to Crohn's disease (Erickson et al., 2012), and metabolome, metagenome, and metatranscriptome data to examine the relationship between the gut microbiome and the xenobiotic metabolism of digoxin (Haider et al., 2013a, 2013b).

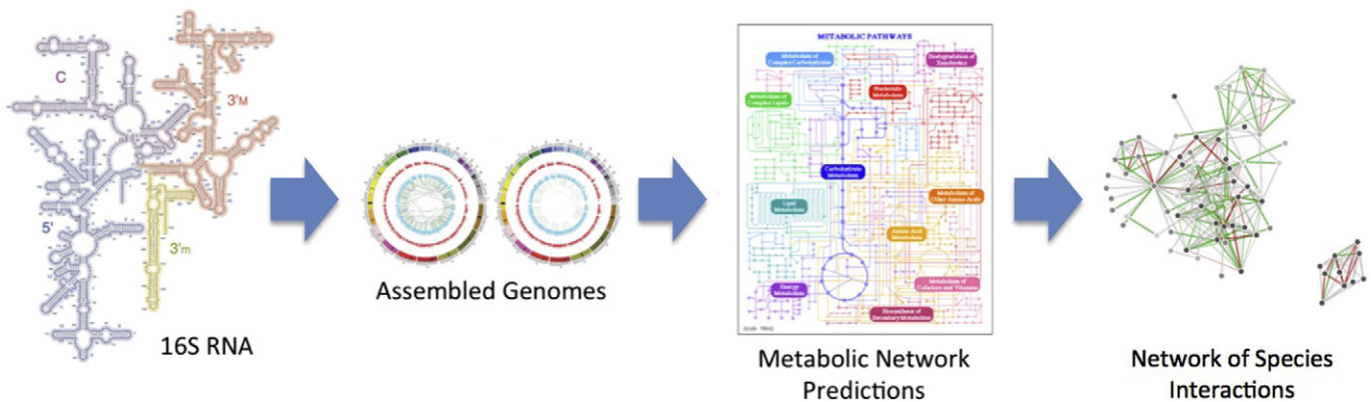


Fig. 2. The who, what, and how of microbial community metabolism. 16S rRNA deep microbial community profiling lets us rapidly and cheaply survey which microbes are present and in what abundances. Metagenome sequencing and genome assembly tells us the biological functions each microbial species can potentially perform. Metabolomics and metabolic network reconstructions allow us to understand the biochemical mechanisms of each microbe, and to make quantitative predictions regarding its metabolic activity. Finally, species interaction networks can be used to identify relationships between microbes within the same community.

The use of network-based approaches offers a promising avenue to extend beyond the omics integration strategies discussed above. In particular, overlaying high-throughput data onto a mechanistic framework—notably a metabolic model—can serve as a platform for making data integration more biologically meaningful. To fully understand the global picture of how organisms modulate the biochemical environment within our gut, especially in health or disease conditions, it is essential to obtain a clear evaluation of the complex interactions characterizing microbial ecology.

Recently, there has been growing interest in using metagenomic information for characterizing community-wide metabolic interactions (Levy et al., 2015; Zelezniak et al., 2015). Fig. 2 provides an overview of such modeling approaches. In the past, assembly of microbial genomes relied exclusively on culture of individual microbes. While high-throughput culture methods are rapidly developing, the majority of gut microbial species remain uncultured (Lagier et al., 2015). Deep metagenomic sequencing allows assembly of full microbial genomes without the need for culture (Jeraldo et al., 2015). Notably, not all cultured microbes have fully-assembled genomes, and not all fully-assembled genomes are completely annotated. However, culture-free metagenomic assembly provides the most complete picture of gastrointestinal microbial genomes currently available. These assemblies allow us to infer metabolic functions of particular organisms, enabling us to model the metabolic activity of an individual microbe. Genome scale models (GEMs) can then be used as the building blocks for community-scale metabolic modeling (COMM) to determine the microbial interactions between members of a community. COMMs provide both a map of the microbial interactions and the global community dynamics, allowing us to examine ecologically relevant traits such as robustness and stability (Proulx et al., 2005). Metabolic models have already demonstrated great potential for modeling the metabolism of the gut microbes (Bauer et al., 2015). However, current COMM reconstruction is based on pre-existing GEMs and not on data from specific microbial communities. Efforts are needed to incorporate new data onto these models as they become available.

4. The challenges of Big Data

While the value of Big Data synthesis is readily evident, implementation is not simple. Issues with data synthesis generally fall into one of three categories: identifying entities to include in the model, integrating the myriad databases, and statistically assessing the final networks (Fondi and Liò, 2015; Imam et al., 2015; Samal and Martin, 2011).

The process of entity identification is a seemingly straightforward process whereby one compares data to what is already known. The problem is that all sequence alignment and pattern-matching algorithms will always produce a result, even if the match is poor. How does one know a metabolite/reaction/organism truly belongs in a GEM or COMM? False inclusions result in errors that will potentially propagate and lead to false results while false exclusions may leave a network reconstruction incomplete and therefore unusable from a computational modeling perspective. These sorts of errors are unavoidable, especially in the gut microbiome, where there are millions of signals from hundreds of potential sources.

Database integration requires that the entities from one database can be related to the entities in another through a set of meaningful relationships, gene to protein, protein to metabolite, regulon to gene. The centrality of this to any Big Data synthesis can be seen in the growing number of calls for data standardization in the omics sciences (Alivisatos et al., 2015; Dräger and Palsson, 2014; Dubilier et al., 2015). The lack of centralized storage and management of multi-omics data has led to increasingly large hurdles and analytical bottlenecks as studies, now capable of measuring DNA, RNA, proteomics, and metabolomics, must struggle within individual labs for their own individual solutions. The recent scientific call for a more worldwide view of data gathering has been gaining traction in the microbiome

field; this is also a call for more global data management and data unity for future integration (Dubilier et al., 2015).

Lastly, one of the biggest shifts will be in how we identify statistically significant results when reconstructing or simulating a GEM or COMM. In other words, how do we determine if a network is statistically different from expected? Most methods utilize graph randomizations to generate networks for statistical comparison. This process randomly exchanges edges within a network, without regard to biochemical structure; thus, network significance is often grossly overestimated (Samal and Martin, 2011). Network significance must be assessed from a set of plausible or at least possible structures if we are to be able to assess the true significance of the results from a network model.

While these three hurdles may be unavoidable, they can be favorably embraced. For example, uncertainty can be incorporated through measures of statistical significance or likelihood (Benedict et al., 2014; Chia and Bundschuh, 2006). Instead of asking which reactions are to be included, the better question may be what is the likelihood of a reaction being present in a GEM? By ranking inclusion based on certainty we allow for alternate inclusions when supporting data, from the rest of the potential pathways, is weak. Similarly, errors in translating between the numerous data sources can be tolerated as long as data types can be correctly merged into the model. Finally, a statistical test of network accuracy or the resulting predictions needs to be measured against a set of realistic “random” networks or predictions. One potential way to do this is to limit “random” networks to real biochemical reactions that functionally produce biomass components necessary for cell growth (Samal and Martin, 2011). This eliminates “random” networks that violate mass, charge, or atomic element balance (Samal and Martin, 2011). Assessing the likelihood of the metabolic network predictions versus “random” network predictions, could allow us to assess the reliability of our results.

5. Future direction

By synthesizing multiple data types onto metabolic networks, we can better capture and elucidate the emergent, macro-level complexity within a microbial world. In Fig. 1, we described absolute values for the sake of simplicity. The complex reality is that each type of omics technology not only measures different entities, but also propagates some amount of error or ambiguity. To circumvent this, a likelihood-based approach would allow us to incorporate a measure of certainty as weighting in a network reconstruction (Benedict et al., 2014). These could, in principle, come from a variety of data types and be combined so as to produce the maximum-likelihood metabolic model. Such a framework would allow us to prioritize consistency between different data types and the overall network structure. This type of platform is already capable of improved gene annotation (Benedict et al., 2014) and gap-filling of metabolic networks (Benedict et al., 2014). In the future, one could use such a platform for improved metabolic modeling. Data synthesis allows us to maximize information to gain mechanistic insight into microbial community and host–microbe dynamics.

A multi-omics modeling approach has the potential for elucidating the intricate relationships between host and microbe. Mechanistic models at this front are critical to understanding how therapeutics like diet modulation or probiotics could impact diseases such as inflammatory bowel disease (Nell et al., 2010) or autism (Kang et al., 2013), or diseases linked with long-term environmental/microbial exposures such as colon or pancreatic cancer (Louis et al., 2014; Zambirinis et al., 2014). Moreover, a holistic and multi-omics microbial metabolic model provides the ideal scaffold for the addition of other host systems, such as the immune system or the nervous system. Metabolic modeling of microbial communities is a rapidly emerging research field with a wide range of approaches (Abubucker et al., 2012; Borenstein, 2012; Heinken et al., 2013; Klitgord and Segrè, 2010) that would all benefit from multi-omics synthesis.

Acknowledgments

The authors thank all members of the Chia Laboratory for discussions and critical reading of the manuscript. This work was supported by the Mayo Clinic Center for Individualized Medicine (V.H., A.M., N.C.), the National Institutes of Health under award number R01CA179243 (V.H., N.C.), and the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIP) (no. 2015R1C1A1A02037045) (J.S. and P.-J.K.).

References

- Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S.T., Methé, B., Schloss, P.D., Gevers, D., Mitreva, M., Huttenhower, C., 2012. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* 8, e1002358.
- Alivisatos, A.P., Blaser, M.J., Brodie, E.L., Chun, M., Dangi, J.L., Donohue, T.J., Dorrestein, P.C., Gilbert, J.A., Green, J.L., Jansson, J.K., Knight, R., Maxon, M.E., McFall-Ngai, M.J., Miller, J.F., Pollard, K.S., Ruby, E.G., Taha, S.A., Consortium, U.M., 2015. A unified initiative to harness Earth's microbiomes. *Science* 350, 507–508.
- Bauer, E., Laczny, C.C., Magnusdottir, S., Wilmes, P., Thiele, I., 2015. Phenotypic differentiation of gastrointestinal microbes is reflected in their encoded metabolic repertoires. *Microbiome* 3, 55.
- Benedict, M.N., Mundy, M.B., Henry, C.S., Chia, N., Price, N.D., 2014. Likelihood-based gene annotations for gap filling and quality assessment in genome-scale metabolic models. *PLoS Comput. Biol.* 10.
- Borenstein, E., 2012. Computational systems biology and in silico modeling of the human microbiome. *Brief. Bioinform.*
- Bultman, S.J., 2013. Emerging roles of the microbiome in cancer. *Carcinogenesis*.
- Chen, J., Ryu, E., Hathcock, M., Ballman, K., Chia, N., Olson, J.E., Nelson, H., 2016. Impact of demographics on human gut microbial diversity in a US Midwest population. *Peer J.* 4, e1514.
- Chia, N., Bundschuh, R., 2006. A practical approach to significance assessment in alignment with gaps. *J. Comput. Biol.* 13, 429–441.
- Collins, F.S., Morgan, M., Patrinos, A., 2003. The Human Genome Project: lessons from large-scale biology. *Science* 300, 286–290.
- Costa, F.F., 2013. Big data in biomedicine. *Drug Discov. Today* 00, 1–8.
- David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J., Turnbaugh, P.J., 2014. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563.
- De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., Lionetti, P., 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci.* 107, 14691–14696.
- de Vos, W.M., de Vos, E.A., 2012. Role of the intestinal microbiome in health and disease: from correlation to causation. *Nutr. Rev.* 70, S45–S56.
- Dräger, A., Palsson, B.Ø., 2014. Improving collaboration by standardization efforts in systems biology. *Front Bioeng. Biotechnol.* 2, 61.
- Dubilier, N., McFall-Ngai, M.J., Zhao, L., 2015. Microbiology: create a global microbiome effort. *Nature* 526, 631–634.
- Erickson, A.R., Cantarel, B.L., Lamendella, R., Darzi, Y., Mongodin, E.F., Pan, C., Shah, M., Halfvarson, J., Tysk, C., Henrissat, B., Raes, J., Verberkmoes, N.C., Fraser, C.M., Hettich, R.L., Jansson, J.K., 2012. Integrated metagenomics/metaproteomics reveals human host–microbiota signatures of Crohn's disease. *PLoS ONE* 7, e49138.
- Feist, A.M., Palsson, B.Ø., 2008. The growing scope of applications of genome-scale metabolic reconstructions: the case of *E. coli*. *Nat. Biotechnol.* 26, 659–667.
- Fondi, M., Liò, P., 2015. Multi-omics and metabolic modelling pipelines: challenges and tools for systems microbiology. *Microbiol. Res.* 171, 52–64.
- Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafiq, A., Abu-Ali, G., Morgan, X.C., Huttenhower, C., 2015. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat. Rev. Microbiol.* 13, 360–372.
- Franzosa, E.A., Morgan, X.C., Segata, N., Waldron, L., Reyes, J., Earl, A.M., Giannoukos, G., Boylan, M.R., Ciulla, D., Gevers, D., Izard, J., Garrett, W.S., Chan, A.T., Huttenhower, C., 2014. Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci. U. S. A.* 111, E2329–E2338.
- Haiser, H.J., Gootenberg, D.B., Chatman, K., Sirasani, G., Balskus, E.P., Turnbaugh, P.J., 2013a. Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* 341, 295–298.
- Haiser, H.J., Gootenberg, D.B., Chatman, K., Sirasani, G., Balskus, E.P., Turnbaugh, P.J., 2013b. Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science (New York, N.Y.)* 341, 295–298.
- Heinken, A., Thiele, I., 2015. Systematic prediction of health-relevant human–microbial co-metabolism through a computational framework. *Gut Microbes* 6, 120–130.
- Heinken, A., Sahoo, S., Fleming, R.M.T., Thiele, I., 2013. Systems-level characterization of a host–microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* 4, 28–40.
- Henry, C.S., Zinner, J.F., Cohoon, M.P., Stevens, R.L., 2009. iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biol.* 10, R69.
- Imam, S., Schäuble, S., Brooks, A.N., Baliga, N.S., Price, N.D., 2015. Data-driven integration of genome-scale regulatory and metabolic network models. *Front. Microbiol.* 6, 409.
- Jeraldo, P., Hernández, Á., White, B.A., O'Brien, D., Ahlquist, D., Boardman, L., Chia, N., 2015. Draft genome sequences of 24 microbial strains assembled from direct sequencing from 4 stool samples. *Genome Announc.* 3.
- Kang, D.-W., Park, J.G., Ilhan, Z.E., Wallstrom, G., LaBaer, J., Adams, J.B., Krajmalnik-Brown, R., 2013. Reduced incidence of *Prevotella* and other fermenters in intestinal microflora of autistic children. *PLoS ONE* 8, e68322.
- Kashyap, P.C., Marcobal, A., Ursell, L.K., Larauche, M., Duboc, H., Earle, K.A., Sonnenburg, E.D., Ferreyra, J.A., Higginbottom, S.K., Million, M., Tache, Y., Pasricha, P.J., Knight, R., Farrugia, G., Sonnenburg, J.L., 2013. Complex interactions among diet, gastrointestinal transit, and gut microbiota in humanized mice. *Gastroenterology* 144, 967–977.
- Klitgord, N., Segrè, D., 2010. Environments that induce synthetic microbial ecosystems. *PLoS Comput. Biol.* 6, e1001002.
- Lachin, J.M., 1981. Introduction to sample size determination and power analysis for clinical trials. *Control. Clin. Trials* 2, 93–113.
- Lagier, J.-C., Hugon, P., Khelafifa, S., Fournier, P.-E., La Scola, B., Raoult, D., 2015. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clin. Microbiol. Rev.* 28, 237–264.
- Lander, E.S., 2011. Initial impact of the sequencing of the human genome. *Nature* 470, 187–197.
- Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., Huttenhower, C., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821.
- Lee, D., Smallbone, K., Dunn, W.B., Murabito, E., Winder, C.L., Kell, D.B., Mendes, P., Swainston, N., 2012. Improving metabolic flux predictions using absolute gene expression data. *BMC Syst. Biol.* 6, 1–9.
- Lee, W.-J., Hase, K., 2014. Gut microbiota-generated metabolites in animal health and disease. *Nat. Chem. Biol.* 10, 416–424.
- Levy, R., Carr, R., Kreimer, A., Freilich, S., Borenstein, E., 2015. NetCooperate: a network-based tool for inferring host–microbe and microbe–microbe cooperation. *BMC Bioinform.* 16, 164.
- Liou, A.P., Paziuk, M., Luevano, J.-M., Machineni, S., Turnbaugh, P.J., Kaplan, L.M., 2013. Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Sci. Transl. Med.* 5 (178ra141).
- Louis, P., Hold, G.L., Flint, H.J., 2014. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.* 12, 661–672.
- Marubini, E., Valsecchi, M.G., 2004. Analysing Survival Data From Clinical Trials and Observational Studies. John Wiley & Sons Ltd, England.
- McNulty, N.P., Yatsunenko, T., Hsiao, A., Faith, J.J., Muegge, B.D., Goodman, A.L., Henrissat, B., Oozeer, R., Cools-Portier, S., Gobert, G., Chervaux, C., Knights, D., Lozupone, C.A., Knight, R., Duncan, A.E., Bain, J.R., Muehlbauer, M.J., Newgard, C.B., Heath, A.C., Gordon, J.I., 2011. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci. Transl. Med.* 3 (106ra106–106ra106).
- Méthé, B.A., Nelson, K.E., Pop, M., Creasy, H.H., Giglio, M.G., Huttenhower, C., Gevers, D., Paton, J.F., Abubucker, S., Badger, J.H., Chinwalla, A.T., Earl, A.M., FitzGerald, M.G., Fulton, R.S., Hallsworth-Pepin, K., Lobos, E.A., Madupu, R., Magrini, V., Martin, J.C., Mitreva, M., Muzny, D.M., Sodergren, E.J., Versalovic, J., Wollam, A.M., Worley, K.C., Wortman, J.R., Young, S.K., Zeng, Q., Aagaard, K.M., Abolude, O.O., Allen-Vercoe, E., Alm, E.J., Alvarado, L., Andersen, G.L., Anderson, S., Appelbaum, E., Arachchi, H.M., Armitage, G., Arze, C.A., Ayvaz, T., Baker, C.C., Begg, L., Belachew, T., Bhonagiri, V., Bihan, M., Blaser, M.J., Bloom, T., Vivien Bonazzi, J., Brooks, P., Buck, G.A., Buhay, C.J., Busam, D.A., Campbell, J.L., Canon, S.R., Cantarel, B.L., Chain, P.S., Chen, I.M.A., Chen, L., Chhibba, S., Chu, K., Ciulla, D.M., Clemente, J.C., Clifton, S.W., Conlan, S., Crabtree, J., Cutting, M.A., Davidovics, N.J., Davis, C.C., DeSantis, T.Z., Deal, C., Delehaunty, K.D., Dewhirst, F.E., Deych, E., Ding, Y., Doering, D.J., Dugan, S.P., Dunne, W.M., Durkin, A.S., Edgar, R.C., Erlich, R.L., Farmer, C.N., Farrell, R.M., Faust, K., Feldgarden, M., Felix, V.M., Fisher, S., Fodor, A.A., Forney, L., Foster, L., Di Francesco, V., Friedman, J., Friedrich, D.C., Fronick, C.C., Fulton, L.L., Gao, H., Garcia, N., Giannoukos, G., Giblin, C., Giovanni, M.Y., Goldberg, J.M., Goll, J., Gonzalez, A., Griggs, A., Gujja, S., Haas, B.J., Hamilton, H.A., Harris, E.L., Hepburn, A.A., Hertz, B., Hoffmann, D.E., Holder, M.E., Howarth, C., Huang, K.H., Huse, S.M., Izard, J., Jansson, J.K., Jiang, H., Jordan, C., Joshi, V., Katancik, J.A., Keitel, W.A., Kelley, S.T., Kells, C., Kinder-Haake, S., King, N.B., Knight, R., Knights, D., Kong, H.H., Koren, O., Koren, S., Kota, K.C., Kovar, C.L., Kyrpides, N.C., La Rosa, P.S., Lee, S.L., Lemon, K.P., Lennon, N., Lewis, C.M., Lewis, L., Ley, R.E., Li, K., Liolios, K., Liu, B., Liu, Y., Lo, C.-C., Lozupone, C.A., Lunsford, R.D., Madden, T., Mahurkar, A., Mannon, P.J., Mardis, E.R., Markowitz, V.M., Mavrommatis, K., McCorrison, J.M., McDonald, D., McEwen, J., McGuire, A.L., McInnes, P., Mehta, T., Mihindukulasuriya, K.A., Miller, J.R., Minx, P.J., Newsham, I., Nusbaum, C., O'Laughlin, M., Orvis, J., Pagani, I., Palaniappan, K., Patel, S.M., Pearson, M., Peterson, J., Podar, M., Pohl, C., Pollard, K.S., Priest, M.E., Proctor, L.M., Qin, X., Raes, J., Ravel, J., Reid, J.G., Rho, M., Rhodes, R., Riehle, K.P., Rivera, M.C., Rodriguez-Mueller, B., Rogers, Y.-H., Ross, M.C., Russ, C., Sanka, R.K., Pamela Sankar, J., Sathirapongasuti, F., Schloss, J.A., Schloss, P.D., Schmidt, T.M., Scholz, M., Schriml, L., Schubert, A.M., Segata, N., Segre, J.A., Shannon, W.D., Sharp, R.R., Sharpton, T.J., Shenoy, N., Sheth, N.U., Simone, G.A., Singh, I., Smillie, C.S., Sobel, J.D., Sommer, D.D., Spicer, P., Sutton, G.G., Sykes, S.M., Tabbaa, D.G., Thiagarajan, M., Tomlinson, C.M., Torralba, M., Treangen, T.J., Truty, R.M., Vishnivetskaya, T.A., Walker, J., Wang, L., Wang, Z., Ward, D.V., Warren, W., Watson, M.A., Wellington, C., Wetterstrand, K.A., White, J.R., Wilczek-Boney, K., Wu, Y.Q., Wylie, K.M., Wylie, T., Yandava, C., Ye, L., Ye, Y., Yooshep, S., Youmans, B.P., Zhang, L., Zhou, Y., Zhu, Y., Zoloth, L., Zucker, J.D., Birren, B.W., Gibbs, R.A., Highlander, S.K., Weinstock, G.M., Wilson, R.K., White, O., 2012. A framework for human microbiome research. *Nature* 486, 215–221.
- Nell, S., Suerbaum, S., Jensenhans, C., 2010. The impact of the microbiota on the pathogenesis of IBD: lessons from mouse infection models. *Nat. Rev. Microbiol.* 8, 564–577.

- Nguyen, T.L.A., Vieira-Silva, S., Liston, A., Raes, J., 2015. How informative is the mouse for human gut microbiota research? *Dis. Model. Mech.* 8, 1–16.
- Palsson, B.O., 2015. *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press.
- Proulx, S.R., Promislow, D.E.L., Phillips, P.C., 2005. Network thinking in ecology and evolution. *Trends Ecol. Evol.* 20, 345–353.
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S.D., Wang, J., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65.
- Reed, J.L., Famili, I., Thiele, I., Palsson, B.O., 2006. Towards multidimensional genome annotation. *Nat. Rev. Genet.* 7, 130–141.
- Samal, A., Martin, O.C., 2011. Randomizing genome-scale metabolic networks. *PLoS ONE* 6, e22295.
- Sinha, R., Chen, J., Amir, A., Vogtmann, E., Shi, J., Inman, K.S., Flores-Munguia, R., Sampson, J.N., Knight, R., Chia, N., 2015. Collecting fecal samples for microbiome analyses in epidemiology studies. *Cancer Epidemiol. Biomark. Prev.*
- The Human Microbiome Project, C., 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214.
- Thiele, I., Palsson, B.O., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121.
- Töpfer, N., Kleessen, S., Nikoloski, Z., 2015. Integration of metabolomics data into metabolic networks. *Front Plant Sci.* 6, 49.
- Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., Gordon, J.I., 2009. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* 1 (6ra14–16ra14).
- Wang, J., Qi, J., Zhao, H., He, S., Zhang, Y., Wei, S., Zhao, F., 2013a. Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Sci. Rep.* 3, 1843.
- Wang, K., Kim, C., Bradfield, J., Guo, Y., Toskala, E., Otieno, F.G., Hou, C., Thomas, K., Cardinale, C., Lyon, G.J., Golhar, R., Hakonarson, H., 2013b. Whole-genome DNA/RNA sequencing identifies truncating mutations in RBCK1 in a novel Mendelian disease with neuromuscular and cardiac involvement. *Genome Med.* 5, 67–67.
- Weir, T.L., Manter, D.K., Sheflin, A.M., Barnett, B.A., Heuberger, A.L., Ryan, E.P., 2013. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS ONE* 8.
- Yugi, K., Kubota, H., Toyoshima, Y., Noguchi, R., Kawata, K., Komori, Y., Uda, S., Kunida, K., Tomizawa, Y., Funato, Y., Miki, H., Matsumoto, M., Nakayama, Keiichi I., Kashikura, K., Endo, K., Ikeda, K., Soga, T., Kuroda, S., 2014. Reconstruction of insulin signal flow from phosphoproteome and metabolome data. *Cell Rep.* 8, 1171–1183.
- Zackular, J.P., Baxter, N.T., Iverson, K.D., Sadler, W.D., Petrosino, J.F., Chen, G.Y., Schloss, P.D., 2013. The gut microbiome modulates colon tumorigenesis. *mBio* 4, 13.
- Zambirinis, C.P., Pushalkar, S., Saxena, D., Miller, G., 2014. Pancreatic cancer, inflammation and microbiome. *Cancer J.* 20, 195–202.
- Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D.R., Bork, P., Patil, K.R., 2015. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci.* 112, 6449–6454.