

# Microbiome-wide association studies link dynamic microbial consortia to disease

Jack A. Gilbert<sup>1</sup>, Robert A. Quinn<sup>2-4</sup>, Justine Debelius<sup>5</sup>, Zhenjiang Z. Xu<sup>5</sup>, James Morton<sup>6</sup>, Neha Garg<sup>2,3</sup>, Janet K. Jansson<sup>7</sup>, Pieter C. Dorrestein<sup>2-5</sup> & Rob Knight<sup>4-6</sup>

**Rapid advances in DNA sequencing, metabolomics, proteomics and computational tools are dramatically increasing access to the microbiome and identification of its links with disease. In particular, time-series studies and multiple molecular perspectives are facilitating microbiome-wide association studies, which are analogous to genome-wide association studies. Early findings point to actionable outcomes of microbiome-wide association studies, although their clinical application has yet to be approved. An appreciation of the complexity of interactions among the microbiome and the host's diet, chemistry and health, as well as determining the frequency of observations that are needed to capture and integrate this dynamic interface, is paramount for developing precision diagnostics and therapies that are based on the microbiome.**

The role of individual species of microbes in infectious disease has been known since the work of microbiologists Robert Koch and Louis Pasteur in the nineteenth century. Yet the part played by complex communities of microbes (known as microbiotas) in providing fertile ground for infections and in setting the stage for non-communicable diseases has been appreciated only in the past decade. The gut microbiota, for example, has been linked to a variety of conditions, some of which are predictable (irritable bowel syndrome<sup>1</sup> and inflammatory bowel disease (IBD) in adults<sup>2</sup> and children<sup>3</sup>), whereas others are intriguing (obesity<sup>4,5</sup>, cardiovascular disease<sup>6</sup>, colon cancer<sup>7</sup> and rheumatoid arthritis<sup>8</sup>) or truly surprising (major depression<sup>9</sup>, Parkinson's disease<sup>10</sup> and autism spectrum disorder<sup>11</sup>).

Many ways in which the microbiota might drive disease have been identified, but their relative importance is yet to be determined. For instance, the taxonomic composition of the microbiota might be most important, and this could be influenced by the overall diversity of species or by the presence of particular taxa, either of which can distinguish healthy individuals from those with disease states. If the collective genes of the microbiota (the microbiome) are more important, the overall genetic diversity or genetic composition, or even specific genetic lineages or metabolic pathways, might play a crucial part in shaping a disease. How such genes are expressed as transcripts and proteins could also have an effect. If the metabolome — the set of chemicals produced by the microbiota and host — is of overriding concern, whether different communities of microbes could lead to the same metabolic and immunological consequences should be considered. Overall, the molecular states of the microbiome probably interact through myriad feedback mechanisms that constantly respond and react to one another to produce the observed disease outcomes.

This Review describes the ways in which the microbiota and the microbiome, as well as specific functions of both, have been linked to various diseases. It also looks at some of the technical and conceptual pitfalls that must be avoided when designing studies that investigate these links. Such issues become compounded when studies are scaled up to cover tens of thousands of people over time and when they are designed to understand subtle and systems-level effects that result from the interactions of many

factors. Microbiome-wide association studies (MWAS)<sup>12</sup>, which capture this scale and the multidimensional interactions, and provide a means of capturing complex interactions to predict practicable links between microbial systems and disease states. MWAS can link whole microbiomes or their features to phenotypes such as disease, with appropriate controls for composition of the microbiota and unusual statistical characteristics of microbiome data sets. Although MWAS are somewhat analogous to genome-wide association studies (GWAS), the microbiome contains many more genes than does the host genome, and its composition changes over time within a person (Box 1). MWAS are useful for untangling the mechanisms that link communities of microbes and their functions to disease, although most clinical applications are yet to be fully realized. To achieve this, model systems should be devised and implemented that allow the testing of hypotheses on isolated and combinatorial functions of microbes and interventions for capturing mechanisms of action. Such systems should also enable these ideas to be applied more generally to the complex communities of microbes that inhabit the body.

## Microbial biomarkers

The human microbiota is the collection of microscopic organisms that live in the body, and it contains representatives from all domains of life: the archaea, the bacteria and the eukarya. Viruses, including bacteriophages, are not always encompassed by the definition of the microbiota. They probably should be, however, because they can shape the structure of the community through top-down ecological control and they have their own effects on the immune system of the host<sup>13</sup>. Most approaches to identifying microbial changes have applied biomarker discovery to test for differences between people with the conditions of interest and controls. Changes in the structure of the microbiota that are associated with disease states can occur at any taxonomic rank and along any relevant branch of the phylogenetic tree. For example, changes at the phylum level have been reported in human obesity<sup>5</sup> and IBD<sup>2</sup>, and strain-level associations have been made with the metabolism of drugs in humans<sup>14</sup>. For instance, the risk of colon cancer in mice increases in the presence of particular strains of *Escherichia coli* that express a gene cluster that produces a genotoxic secondary metabolite called colibactin<sup>15</sup>. Between these extremes, changes

<sup>1</sup>Department of Surgery, University of Chicago, Chicago, Illinois 60637, USA. <sup>2</sup>Department of Pharmacology, University of California San Diego, La Jolla, California 92093, USA. <sup>3</sup>Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, USA. <sup>4</sup>Center for Microbiome Innovation, Jacobs School of Engineering, University of California, San Diego, La Jolla, California 92093, USA. <sup>5</sup>Department of Pediatrics, University of California, San Diego School of Medicine, La Jolla, California 92093, USA. <sup>6</sup>Department of Computer Science and Engineering, Jacobs School of Engineering, University of California San Diego, La Jolla, California 92093, USA. <sup>7</sup>Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington 99354, USA.

## BOX 1

# Principles of microbiome-wide association studies

Microbiome-wide association studies (MWAS) are similar in concept to GWAS: the goal of both is to link a complex collection of features (for example, species or genes) to phenotype. However, there are important differences between the two. First, there are many more microbial genes than human ones, with some studies estimating that there are more than 100 microbial genes for every human gene<sup>24,111–113</sup>. Consequently, the issue of multiple comparisons is of greater importance to MWAS. Second, all individuals share almost the same collection of human genes but their dissimilarity in microbial species and microbial genes is much greater<sup>24,112</sup>. Third, genes in the human genome can be counted easily but most microbiome data comes in the form of relative abundance. Compositional statistics therefore apply and the data cannot be represented in familiar Euclidean spaces. As a result, microbiome analyses are very prone to misinterpretation. For instance, it is impossible to infer the growth or decay of microbes purely on the basis of relative abundance data because the growth of one species could also be explained by the decline of all other species. Last, whereas the human genome is essentially fixed within an individual (except in special cases such as the immune system and cancer), the microbiome of each person changes profoundly throughout his or her lifetime. Several designs for MWAS link the overall microbiome to specific phenotypes. A number of important questions must therefore be asked when designing MWAS.

## ● At what level will the microbiome or microbiota be assessed?

MWAS can be carried out using species, genes, functional categories of genes or, less frequently, transcripts and proteins as features. Metabolome-wide association studies are also possible, and they can be carried out at the level of individual spectra, groups of related spectra or pathways. These analyses often give different results; for example, in the Human Microbiome Project, pathway-level analysis of the shotgun metagenomic data suggested that much less variability existed between people than did taxon-level analysis.

● **Will the microbiome be examined in terms of overall variation or as a collection of individual features?** Techniques for reducing the dimensionality of the microbiome include: clustering, principal coordinates analysis (PCoA) with a variety of distance metrics, principal component analysis, correspondence analysis, factor analysis and discriminant analysis. In clustering analyses, which include enterotyping, samples are grouped into clusters. The resulting clusters are then tested for association with a phenotype (for example, whether the resting levels of blood glucose are identical in each cluster). During dimensionality reduction, one or more axes are discovered through a supervised or an unsupervised approach, and the dependence of

phenotype on locations along these axes is tested, for example, by correlation approaches. Supervised approaches such as discriminant analysis make use of phenotype labels and provide the projection of the data that best separates these class labels. Statistical tests of location on the resulting axes must therefore be used with caution because even small departures from the random model can lead to apparent separation when there is none. Unsupervised approaches such as PCoA use only the intrinsic similarities and differences in the samples; however, they may not reveal separation by phenotypic state even when it exists (because it could come only in later principal axes).

Techniques for associating individual features of the microbiome with phenotype, including appropriate statistics for repeated measures, are Metastats<sup>114</sup>, DESeq2 (ref. 115) and ANCOM (analysis of composition of microbiomes)<sup>116</sup>, as well as various machine-learning approaches such as Random Forests<sup>117</sup>. Unfortunately, it is also challenging to infer differentially significant species in compositional data sets. Many state-of-the-art tools make assumptions about the underlying data to identify significantly different species. Analysts need to gauge the assumptions given by each tool before applying them to their data sets because these assumptions are typically not true of real-world data.

● **What corrections will be performed for multiple statistical comparisons, sparsity and compositionality of the data, and other features of microbiome and related data sets?** Often, associations will be sought between the microbiome (as a whole or as a collection of features) and measures of phenotype. In many of these studies, the differences between phenotypes can be described by a select few features. Conventional statistical tests can be confounded by the underlying ecology. For instance, multiple microbes can share the same functional roles. As a result, differences in microbial abundances might yield the same phenotype. Analyses should be separated into planned analyses (those chosen before the analysis) and ad hoc analyses (those performed after); ad hoc analyses should be considered to be exploratory rather than formal statistical tests.

● **How will causality be established?** Causality can be approached in a number of ways: through prospective longitudinal studies that demonstrate that a microbial or metabolic change precedes the disease phenotype; the demonstration that a clinical manipulation of the microbiome affects the disease process; preclinical work in mice or other animal models that demonstrates the plausibility of a mechanism; or establishment of the activity of chemical products of the microbiome that are linked to specific microbes or the genes that produce them. Studies that combine animal models with proof-of-relevance in people are especially effective, although they are rare.

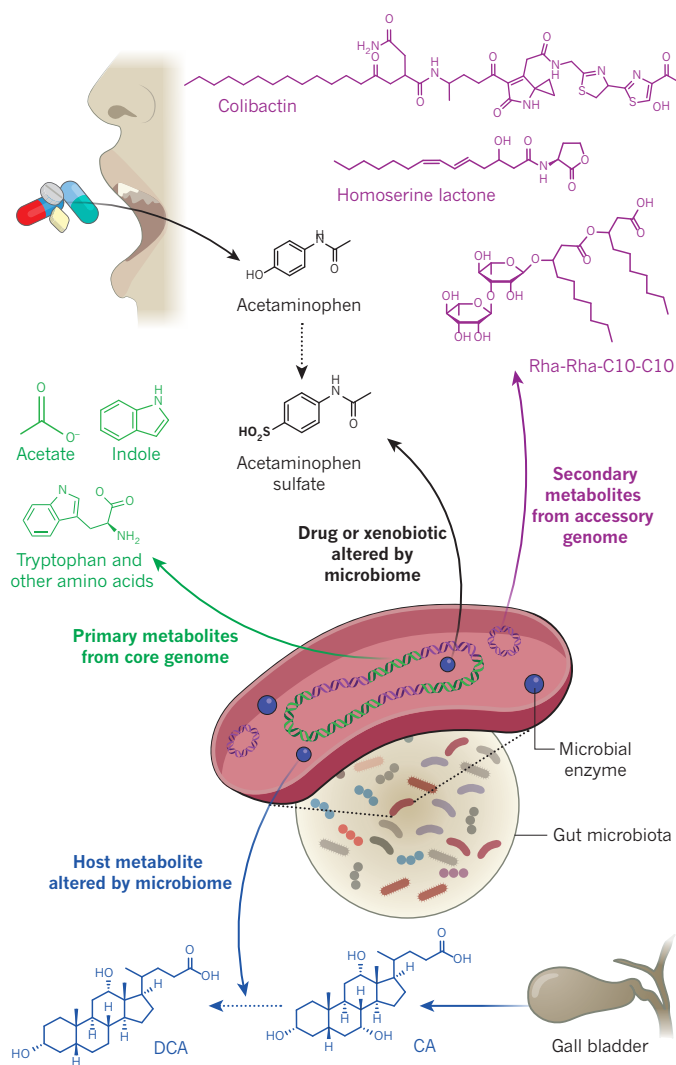
at the genus level are useful for many applications, including microbial source tracking<sup>16</sup> and, more controversially, defining enterotypes, which are classifications of types of microbial communities in the gut<sup>17</sup>.

## Taxonomic biomarkers

Most studies have focused on identifying single organisms as biomarkers, but separating collections of samples on the basis of similarity between communities has also been useful for a wide range of diseases, including IBD<sup>18</sup>. However, the extent to which the choice of metric for pairwise comparisons of communities can influence the result is not widely appreciated. The fit between the data and a statistical model is often used to assess the validity of the technique. But when a collection of samples is highly heterogeneous, which includes situations as simple as the collection of skin samples from different people, models that better fit the data in the original data set might provide no clear biological interpretation.

Importantly, this problem cannot be overcome by collecting more data because using the incorrect statistical model can obscure results that can be clearly determined, even with limited numbers of DNA or RNA sequences<sup>19</sup>. The choice of distance metric, level of taxonomic resolution or a particular taxon to focus on can involve dozens of further, implicit, comparisons that also must be accounted for statistically.

The identification of interactions between microbes is essential for microbial ecology. Correlation networks have proved useful for distilling relevant links from a morass of potential interactions. However, interpretation is still complex for two reasons. First, the abundance of specific microorganisms in each microbiota is sampled through a multinomial distribution, which leads to large numbers of negative correlations and induces a substantial bias in network topology. Second, taxonomic data are extremely sparse: most samples have zero abundance of a particular organism. Because of these correlation problems, network analyses can



**Figure 1 | Sources of metabolites from the human microbiome.** The core physiology of the microbial cells that make up the microbiome can produce by-products and intermediates that affect health, including short-chain fatty acids (such as acetate) and tryptophan metabolites. Secondary (or specialized) metabolites are produced from accessory genetic elements that are often transferred horizontally between microbes. Some of these metabolites, including colibactin<sup>15</sup> and rhamnolipids<sup>109</sup> (Rha-Rha-C10-C10), are known to cause disease. Microbes can also alter metabolites that are produced by the host, such as bile acids<sup>110</sup> (CA, cholic acid) and even drugs that are consumed, such as acetaminophen (paracetamol)<sup>61</sup>. DCA, deoxycholic acid; Rha, rhamnose.

be inherently flawed<sup>20,21</sup>. Despite such limitations, taxonomic correlation networks have identified microbial interactions that are linked to disease, including beneficial and harmful networks of microbes that are associated with Crohn's disease<sup>18</sup>.

These examples of successful biomarker discovery have yet to provide standard guidelines; however, they have produced interesting findings. For example, a higher level of taxonomic resolution is not always better. 16S ribosomal RNA operational taxonomic units (OTUs), which are clusters of sequences that are defined by sequence identity, at the species level are best for matching samples, yet this taxonomic level actually decreases the accuracy of classifying individuals as lean or obese<sup>22</sup>. The level of resolution is therefore dependent on the context.

### Functional biomarkers

Shotgun metagenomics, the sequencing of fragments from total DNA rather than of specific genes, provides more-complete information about the microbial community and enables many powerful analyses, although

the choice can be bewildering, even to experienced researchers in the field. As well as identifying taxa down to the level of strains or genomic single-nucleotide polymorphisms (SNPs), DNA sequences can be grouped into many functional classifications using databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes), COG (Clusters of Orthologous Groups of Proteins), GO (Gene Ontology) and EggNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups). Metagenomics studies<sup>23,24</sup> commonly show a surprising consistency in functional profiles, although the limited variation that does exist can often be explained by taxonomy. Studies that separate samples of interest from controls at different functional resolutions, are yet to be adequately performed, however. Shotgun metagenomics seems to outperform amplicon-based taxonomic analysis in the identification of individuals (compare ref. 25 with ref. 26). Re-analysis of 16S rRNA amplicon data using oligotyping<sup>27</sup>, a technique that is based on the fine detail of polymorphisms, has improved resolution, and this is demonstrated by its ability to identify sexual partners through shared sequences. No examples are thought to exist in which shotgun metagenomics has been able to identify a medically relevant trait that could not have been revealed through taxonomic analysis alone, although the potential for doing so is high.

Integrating human metagenomic and metabolomic profiles has great potential for discriminating between disease traits (Fig. 1). The ability to systematically link the variance in metabolomic data between samples with changes in the composition and structure of communities of microbes from the same samples enables not only improved resolution but also the potential to infer the mechanisms that produce observed trends<sup>28</sup>. This potential is highlighted by a study<sup>29</sup> that shows how the microbiome alters bile-acid metabolite profiles during the establishment of *Clostridium difficile* in mice. Similarly, the ability to link metabolite profiles in urine and blood serum to microbial metabolism in the gut can help to synthesize links between dysbiosis (an imbalance of microbes in the body) and the onset of neurological symptoms that are associated with conditions such as autism spectrum disorder in a mouse model<sup>30</sup>. Metaproteomics is also enabling the identification of new biomarkers. Proteins such as L-lactate dehydrogenase and arginine deiminase, as well as those that are involved in the synthesis of exopolysaccharides, iron metabolism and the immune response, seem to be indicative of a healthy human oral cavity<sup>31</sup>. The combination of microbial community profiling with metabolomics and proteomics has precipitated understanding of how the microbiota responds to specific disease states, including IBD<sup>32–34</sup>. The combined findings reveal specific species (for example, *Faecalibacterium prausnitzii*), proteins and metabolites that are involved in the metabolism of butyrate and bile acids, which can be used to differentiate between individuals with inflammation of the ileum that is the result of Crohn's disease and those with inflammation of the colon and a healthy gut. In another example<sup>35</sup>, children with non-alcoholic fatty liver disease show a significant increase in Gammaproteobacteria and *Prevotella* as well as in levels of ethanol and certain short-chain fatty acids (SCFAs), which leads to an increase in energy production and a decrease in the metabolism of carbohydrates and amino acids and in the activity of the urea cycle and urea transport systems.

### From correlation to causation

A crucial challenge for the field is to move beyond associations between the microbiome and specific clinical states towards the establishment of causality. The importance of MWAS with large cohorts in determining causality should not be underestimated. The limitation of the case-control model is that it is impossible to distinguish whether the microbiome drives the disease, the disease drives the microbiome or if both are modified by a confounding factor. For example, a lack of replication of the microbiota differences that separate people with type 2 diabetes from controls in Chinese and European cohorts was found to be due to variation in the levels of usage of the drug metformin, which is used only in the disease state and with different frequencies in the two populations and which had a large and unanticipated effect on their microbiotas<sup>36</sup>. Consequently, the effect that had been attributed to the disease was



actually the result of the treatment.

Several popular methods exist for identifying causality, each of which has specific strengths and weaknesses. Prospective longitudinal studies, such as of the CHILD (Canadian Healthy Infant Longitudinal Development) birth cohort, allow researchers to test whether changes in the microbiome precede or follow the development of disease. Such studies are expensive, however, and can require large populations to capture rare events. If it is difficult to continue to collect samples, the study population can also be affected by attrition. Intervention studies, in which a deliberate clinical event such as the administration of a drug is used to drive change in the microbiome and phenotypes, are useful, but it is often unethical to withhold treatment from a control group to isolate the effect of the specific intervention. Interventions such as faecal microbiota transplantation also face substantial regulatory hurdles, especially in the United States. The comparison of identical and non-identical twins can be valuable for unravelling genetic differences in the host: causality can be established because the microbiome is not known to modify the inheritable host genome. However, such cohorts are difficult to assemble and privacy issues can be considerable, especially when the same twins are used in many studies. Animal models can be helpful for establishing mechanisms, but the quantitative importance of these mechanisms for human disease is often less clear. For example, the demonstration that faecal microbiota transplantation from people who are lean or obese to germ-free mice can confer differences in adiposity indicates that microbes can affect this phenotype, but it does not establish that transplantation can affect the weight of obese people<sup>37</sup>.

### The metabolome reveals important microbial activities

Metabolomic biomarkers are especially useful for diagnostics because changes in metabolism can be rapid and can reveal the physiological state of both the host and its microbiota. Such biomarkers are also the end products of the metabolism of microbes and they can provide mechanistic explanations for particular associations between microbes and disease. The metabolome is being characterized through metabolomics (the study of the complete repertoire of molecules in the body, which is analogous to genomics, the study of the complete repertoire of genes in the genome), metabonomics (the comparison of general metabolomics profiles with their many unidentified compounds, rather than the comparison of specific metabolites within profiles) and exposomics (the study of cumulative exposures to molecules from the environment)<sup>38–40</sup>. A crucial challenge for the characterization of these molecules is that only about 1.8% of the chemical data that can be collected with mass spectrometry can be annotated<sup>39,41</sup>. Unlike the genomics community, the mass-spectrometry community lacks adequate mechanisms of knowledge dissemination that enable data reuse. To overcome this challenge, the community is developing a plethora of resources to store data from mass spectrometry, including databases such as MassBank, METLIN, MetaboLights and the Human Metabolome Database (HMDB), the Metabolomics Workbench platform and the software OpenMS. Its efforts have also led to GNPS (Global Natural Products Social Molecular Networking), the first crowdsourced platform that enables the community-driven curation of mass-spectrometry data and dissemination of existing knowledge of mass spectrometry in the public domain<sup>40,41</sup>. Ultimately, these databases and infrastructures for analysis will allow the estimation of metabolite flux from the genomes to enable prediction of the overall function of communities of microbes<sup>42</sup>. Although a few strains of gut microbes are pathogenic, most are harmless or beneficial to health; similarly, some molecules that are produced by microbes are detrimental to the health of the host<sup>43</sup>, but most are innocuous or even beneficial<sup>44</sup>. Metabolites are particularly important agents of the human microbiome. This is because molecules that are produced by the microbiota can cross epithelial barriers more freely than the microbes to cause systemic effects at distant sites in the body.

The small-molecule repertoire of the human microbiome consists of four groups. The first is composed of primary metabolites, which are molecules produced by the catabolic and anabolic reactions that are required for cellular growth and homeostasis. The second group comprises

specialized metabolites, which includes virulence factors, secondary metabolites and natural products. These compounds are produced by accessory genetic elements that are often acquired through horizontal gene transfer, and they are designed to directly influence the cells of the host and other microbes (Fig. 1). Knowledge of changes in the secondary metabolites can be useful for understanding toxins, quorum sensing and beneficial secondary metabolites of food such as lycopeneoids and carotenoids. The third group is composed of metabolites produced by cells of the host or from exogenous sources that are directly modified by microbial enzymes to create unique chemical products. Knowledge of changes in this group can be useful for understanding how microbes modify products of the host's metabolism. The final group is the exposome, which describes the chemistry and metabolites that are encountered through exposure to personal-care products, medical intervention, food or the environment. Knowledge of changes in this group is especially useful for understanding how compounds that are applied to the body, whether intentionally or unintentionally, can trigger toxic responses or can be modified into forms that differ in activity from the originally applied compound. Although decades of research on primary metabolism have led to a good understanding of these four groups of metabolites, the specialized metabolome of microbes is a veritable sea of unknown chemistry<sup>45,46</sup>.

### Linking metabolomes to health and disease

Evidence is accumulating that the metabolic output of the microbial metabolome has a direct impact on human health. Significant opportunities exist to elucidate the mechanisms that result in this effect. However, current methods of chemical annotation can identify only a small fraction of detected metabolites within the metabolome<sup>41</sup>, and models for testing hypotheses about the interactions among microbes, their molecules and the host are challenging to use.

The best-known examples of microbiome-derived primary metabolites that affect human health are probably the SCFAs. SCFAs such as acetate, propionate and butyrate are produced through the fermentation of dietary fibre by gut microbes and then absorbed by epithelial cells, which provides them with energy<sup>44,47</sup>. Defects in the production of SCFAs have been linked to many conditions, including IBD<sup>48,49</sup>, although it is unclear whether testing for SCFA levels *per se* has clinical value.

The development of germ-free animal models has been very useful for identifying primary metabolites produced or altered by the microbiota of the host. A comparison of metabolomes from germ-free and colonized mice revealed that indole-3-propionic acid and other products of tryptophan metabolism are found only in mice with an intact microbiota and are associated with the presence of *Clostridium sporogenes*<sup>50</sup>. These tryptophan metabolites are thought to affect neuronal signalling in the gut and brain<sup>51</sup>. But their role in human health remains elusive.

Some specialized metabolites from the human microbiota are known to cause disease. For instance, colibactin induces double-strand breaks in the DNA of human cells<sup>52</sup>. The genetic machinery for the production of colibactin can be transferred from pathogenic to non-pathogenic strains of *E. coli*. Colibactin is associated with colorectal cancer in mouse models<sup>15</sup> and provides an example of how commensal microbes of the gut can harbour or acquire specialized metabolites that can result in disease<sup>53</sup>. The true pathogenic elements within the human microbiota might be the genetic islands encoding specialized metabolites that circulate within the microbial ecosystem, rather than the core genomes of pathogenic species (Fig. 1). The prevalence of these genetic islands could be associated with the prevalence of microbiome-associated diseases; it is here that the interface between GWAS and MWAS can best be understood. Tests at the level of single genes, such as for genes that are necessary for colibactin production, might prove more useful for identifying preventive treatments than would tests for the presence or absence of specific taxa, akin to the way that levels of glucose or insulin are measured for diabetes.

### Integration of multi-omics studies

To comprehensively understand the role of the human microbiome and

its metabolome in health and disease, integrative analyses are needed that apply 'omics' techniques to animal or other empirical models. Integrative analysis can help to identify the effect of treatment with antibiotics on the gut microbiota during infection with *Clostridium difficile* in both mice and people<sup>54</sup>. A multidisciplinary approach that employs mathematical modelling, 16S rRNA gene sequencing, metagenome sequencing and animal models identified how microbiotas can help the hosts to resist *C. difficile* infection, which led to the identification of *Clostridium scindens* as a candidate for resistance to infection in mouse models. In a study of 24 people who took antibiotics while undergoing chemotherapy<sup>54</sup>, half had active *C. difficile* infections, which suggests that there is an association between *C. scindens* and resistance to infection with *C. difficile*. Prevention of *C. difficile* infections through transfer of *C. scindens* to animals that were undergoing treatment with antibiotics confirmed this role<sup>55</sup>. Metagenomic and metabolomic-based findings<sup>56</sup> have been used to identify the importance of bile acids in this resistance to infection, and subsequent experiments showed that certain levels of specific bile acids were associated with resistance to *C. difficile* during treatment with antibiotics. This work is an excellent example of how a comprehensive approach to microbiome analysis can link the microbiome to disease. The next step is to translate such findings into clinically useful tests.

The resident microbiota of the human gut has an important role in modulating the efficacy and toxicity of pharmaceuticals<sup>14,57</sup>. Variability in the microbiomes of individuals<sup>58</sup> leads to differences in the metabolism of drugs and therefore in effective dose availability and side effects. Simultaneous measurement of variability in both the microbiome and the metabolome will play an important part in identifying causative mechanisms of xenobiotic metabolism. The role of microbiome-associated drug toxicity is exemplified in the treatment of colon cancer with irinotecan, which resulted in decreased efficacy of the drug in 40% of treated individuals<sup>59</sup>. Irinotecan is reactivated in the gut by microbial  $\beta$ -glucuronidase enzymes, which leads to diarrhoea and prevents administration of the appropriate dose. Inhibitors that modulate the activity of the commensal microbiota by specifically inhibiting the  $\beta$ -glucuronidase enzyme in bacteria are in clinical trials<sup>59</sup>. This represents a precedent for the translation of metabolic mechanisms of the human microbiota into clinical applications and highlights the importance of investigations in the fields of pharmacomicrobiomics<sup>60</sup> and pharmacometabonomics<sup>61,62</sup>. In conjunction with testing for the genes that encode these enzymes, inhibitor-based therapies could increase the efficacy of irinotecan, although the diagnostic and therapeutic system this approach requires has yet to be demonstrated. Advanced data acquisition and computing and streamlined analysis pipelines are enabling multi-omics analysis to be performed on clinically relevant timescales, and the adaptation of multi-omics microbiome analysis in the clinic will probably emerge within the next decade<sup>63</sup>. Future prospects should also reflect on how inhibiting specific enzymes of commensal microbes affects the overall activity and structure of the gut microbiota in the longer term.

### Dynamics of the microbiome

Although many MWAS take a case-control approach, understanding how the microbiome as a whole changes remains a challenge. Relatively few studies have assessed the whole microbiome at many time points; such studies point towards using dynamic — rather than static — features as the input for MWAS. It is challenging to capture the dynamics of an invisible microbial world through snapshots of its current state. However, the situation has vastly improved in the past 15 years, during which DNA sequencing costs dropped by a factor of about one million. By increasing the frequency and depth of observations, the rate and directionality of the transfer of bacteria between ecosystems is starting to be inferred.

### Assessing the transfer of microbes between environments

The application of microbial survey techniques to built environments and the people who inhabit those spaces has shown the utility of high spatiotemporal resolution for inferring interactions between people and surfaces in the environment at the microbiome level<sup>64</sup>. But even with daily

sampling and observations at multiple sites on each individual (such as the nose, the hand and the foot), as well as their pets and surfaces in their home, it is still difficult to make more than comparative statements about the microbial similarity of surfaces and changes in this similarity over time. Higher-resolution temporal analysis, such as hourly sampling<sup>65</sup>, can improve appreciation of the successional dynamics of these communities. These tools have not yet been applied to understanding how consistently specific components of the microbiota are transferred to or from people, let alone within the body. Alternative approaches, such as using differential coverage of parts of the microbial genome to infer activity in a single sample<sup>66</sup>, have great promise for directly revealing activity, but samples still need to be assessed across time points because the activity of microbes can change rapidly in response to conditions. Direct monitoring of the transfer of microbes between environments and of the rapid dynamics in those environments will require a substantial improvement in the determination of genotypic resolution and temporal and spatial sampling. Near real-time microbial epidemiology is being demonstrated with genotypic resolution (at the strain level) through the rapid genomic sequencing of individual species of pathogenic microbes in hospital settings<sup>67</sup>. It is essential that this technology is developed to be more applicable to entire communities of microbes, especially because the most important inputs to MWAS might not be the relative abundance of each microbe or gene at a single time point but rather the variations in particular species over time, as well as their co-variations in linked environments.

### Tracking pathogenic infections

Clinical application of MWAS inspires a vision of a future in which the studies are used to track entire communities of microorganisms involved in the complex 'pathobiomes' that are associated with different disease states. For example, the transfer of bacteria from mother to child might be tracked and augmented by personalized microbial therapies that range from vaginal inoculation<sup>68</sup> to customized prebiotic and probiotic supplements that are based on breast milk<sup>69</sup>. This would require automated approaches to quantify the abundance and composition (at serotype resolution) of whole communities of bacteria, as well as rapid deployment of MWAS techniques to determine current health status or to predict future health status from the trajectories. Although such sensors are not yet available, key platforms are being developed that will provide a substantial improvement on existing systems<sup>70</sup>. However, real-time interpretation of the vast quantity of data that are produced by these sensors will require a radical improvement in automated data processing. This will demand the integration of statistical modelling, high-performance computing and engineering to enable high-throughput transfer, interpretation and visualization of spatiotemporal data. Despite the limitations of existing correlation network techniques (that is, their sparsity and compositionality), network analysis has helped to uncover real associations in complex data. One useful example of the prediction of interactions, and subsequent validation of the prediction through empirical observation, comes from marine microbial ecology: network analysis has been applied to microbial-community sequence data to predict an interaction between an acoel flatworm (*Symsagittifera* sp.) and a green microalga (*Tetraselmis* sp.), and the finding was subsequently validated using microscopy<sup>71</sup>. An example in people is the use of correlation network analysis to demonstrate the connectivity of organisms in the microbiota of human milk. Cooperative and opportunistic subgroups have been identified in which the opportunistic pathogenic species could, in principle, be suppressed through competitive exclusion<sup>72</sup>, pointing to therapeutic approaches based on probiotics (that directly introduce beneficial competing microorganisms) or prebiotics (that encourage the growth of beneficial microorganisms). The movement of microbes between environments cannot be captured by these methods, but the ability of these microorganisms to establish themselves and proliferate on arrival can be inferred through an understanding of the ecological network of their destination and their ability to incorporate. In early life, the shifting sands of an infant's microbiota can lead to an increase or decrease in the colonization success of particular microorganisms<sup>68</sup>. These dynamics have been tracked using longitudinal

characterization in work that has demonstrated the correlations between the microbiota of mother and child, especially after vaginal delivery, as well as the influence of this interaction on the transitional succession of microbial ecology in the child's gut<sup>73</sup>. The application of longitudinal design to MWAS would significantly improve the ability to understand the complex linkage between the microbiome and disease, and it would also improve knowledge of the link between environmental exposure and health outcomes through MWAS-enabled epidemiological investigations. Projects such as the Integrative Human Microbiome Project (iHMP)<sup>74</sup> are beginning to apply these approaches to larger populations.

### The visualization of complex longitudinal data

Visualization improves the interpretation of data and could help to guide clinical decision-making<sup>75</sup>. For example, temporal dynamics could be observed in the human gut following a faecal microbiota transplant to treat a *C. difficile* infection<sup>76</sup>, and the successional dynamics of infant microbial development could be explored<sup>77,78</sup>. Better visualization can also help to define the stability, resistance to perturbation and resilience to change of microbial communities; however, the quality of the initial experimental design is important. Healthy adults have unique microbial dynamics, yet patterns of stability and resistance show elements of similarity, which hints at the potential for universal ecological rules that define these relationships between individuals<sup>79</sup>. Determining the frequency at which longitudinal samples should be taken to capture the dynamics that are relevant to a specific disease state is an open problem<sup>80</sup>. For instance, two studies with different sampling intervals<sup>81,82</sup> found conflicting results with regards to the stability of the microbiota during pregnancy, although differences in dietary intervention could have confounded the patterns. Capturing the temporal dynamics of specific characteristics, such as the level of glucose in the blood or behavioural traits, also presents a constant challenge. The frequency at which various types of data show patterns that enable the integration and mechanistic prediction of microbial interactions should be considered. In an era of precision medicine, an understanding of when and how often different sources of information must be acquired to enable the appropriate integration of data is paramount. At best, inappropriate sampling frequencies fail to produce correlations even when mechanistic interactions exist; at worst, they produce misleading information, which might lead to the identification of incorrect biomarkers or therapeutic targets.

### From explanation to prediction

The microbiome, or even the microbiota, could be used to predict the onset of disease before it occurs and to guide individualized therapies.

### Stratification on the basis of the microbiome

The stratification of people for treatment holds considerable promise. For example, variation in the toxicity of acetaminophen (paracetamol) in the liver is largely caused by differences in how the drug, which is an analogue of the naturally occurring amino acid tyrosine, is metabolized through the tyrosine sulfonation pathway<sup>61</sup>. Similarly, the effectiveness of digoxin depends on whether the gut of an individual contains specific strains of *Eggerthella lenta*, the plasmids of which encode an enzyme that rapidly degrades digoxin and renders it ineffective<sup>14</sup>. Similar stories are emerging for many other classes of drug, which suggests that incorporating the gut microbiome into the stratification of participants in clinical trials and the prescription of medication could be of great value. An especially interesting example is the emerging relationship between trimethylamine *N*-oxide (TMAO) and cardiovascular disease. People can metabolize choline, which is found in dietary sources such as red meat and cheese, in a variety of ways. One such pathway is catalysed by groups of bacteria that are found only in some individuals: choline is metabolized to trimethylamine, which is then oxidized to TMAO, a compound that contributes to the formation of atherosclerotic plaques through mechanisms not yet well understood<sup>83</sup>, although work in mice suggests a possible pathway<sup>6</sup>. Inhibition of the enzyme that produces TMAO or targeting relevant bacteria could therefore provide potent weapons against heart disease.

### BOX 2

## Integrating the host genome into MWAS

Although, intuitively, the host genome was thought to be important in shaping the microbiome, evidence to support this had been lacking. Single genes have been known to exert large effects on the gut microbiome in mice; for example, the *ob/ob*<sup>5,118</sup> and Toll-like receptor 5 knockout models<sup>119</sup> of obesity have been well studied, and the changes in the microbiota that are induced by a single-allele mutation can even confer part of the adiposity phenotype when transmitted by oral gavage to a genetically normal mouse. Consequently, it is well established that a genetic change can trigger an aberrant microbial community that is transmissible and can transmit the phenotype. Studies of panels of mice have shown that diet has a much larger effect on the microbiota than does the host genotype<sup>120</sup>, and this is consistent with the observation that studies consisting of only dozens of people are unable to demonstrate that monozygotic twins are more similar in composition and function of their microbiota than are dizygotic twins<sup>23</sup>. However, larger studies composed of hundreds of individuals are able to find a small association between host genetics and the overall microbial community<sup>121,122</sup>. Intriguingly, a few taxa seem to be highly heritable, notably *Christensenella*, which is associated with leanness and even leads to weight reduction when fed to germ-free mice inoculated with the gut microbiota of obese people<sup>122</sup>.

Conversely, it might be possible to predict whether a particular diet has adverse consequences for the heart at the level of the individual rather than the population. A study<sup>84</sup> involving hundreds of people was able to demonstrate this potential for diabetes; it used continuous monitoring of blood-glucose levels to understand the effects of standardized meals and their dietary components. Remarkably, ice cream was less deleterious than white rice for some people's blood glucose, and differences such as these could largely be predicted by the microbiota (and not by other factors). Consequently, using the microbiota to reduce the immense variability experienced by those who receive dietary therapies holds much promise.

Several studies have substantially advanced the field towards the goal of using the microbiome or the microbiota to predict disease before it occurs. Fascinatingly, different diseases have different dynamics. Gingivitis, an inflammation of the gums that can be reversed with thorough cleaning of the teeth, shows relapse trends that are specific to individuals, which indicates that a person's unique gingivitis-causing community of microbes returns in a predictable way<sup>85</sup>. By contrast, many individuals carry the same community of dental-caries-causing bacteria, yet the emergence of caries can be predicted months in advance of observable clinical symptoms by monitoring changes in the microbiota<sup>86</sup>. Similarly, the development of rheumatoid arthritis can be predicted using both oral and gut microbial biomarkers<sup>87</sup>. The potential use of oral biomarkers to predict disease that emerges at less accessible sites in the body is exciting. The oral microbiota and gut microbiota share many community members, yet the structures of communities are highly distinct, and only weak associations have been found between them<sup>24,88</sup>. The oral cavity provides an ideal site for non-invasive sampling and biomarker testing; the ability to use the oral microbiota to predict disease, following MWAS, therefore has tremendous promise. Predictive models are also being applied to many other sites in the body and to many other conditions, including obesity<sup>22,89</sup>, IBD<sup>18,90</sup> and acne<sup>91</sup>.

### An evidence scale for microbiome studies

Although many studies have reported links between the microbiome and disease, technical variation between the studies, the effects of which often



exceed those of the underlying biology, makes it difficult to compare and interpret MWAS findings<sup>78</sup>. Efforts to quantify methodological effects enable considerable progress to be made towards performing large-scale epidemiological studies of the microbiome<sup>92,93</sup>, but the ability to determine when specific biases require studies to be analysed separately rather than together still relies on intuition.

Crohn's disease represents one of the best-studied links between the microbiome and disease. Multiple studies<sup>18,32–34,94</sup>, including investigations of a cohort of Swedish twins ( $n = 40$  pairs of twins)<sup>32–34</sup>, have revealed the depletion of beneficial members of the microbiota (for example, *F. prausnitzii*, a producer of butyrate) in people with inflammation in the small intestine that is associated with Crohn's disease, known as ileal Crohn's disease, compared with those who have inflammation of the colon or with healthy individuals. Increases in Proteobacteria have also been seen in these and many other studies<sup>3,18,32</sup>. Analysis of faecal samples from the Swedish twin cohort<sup>33</sup> also revealed the depletion in ileal Crohn's disease of proteins required for the metabolism of butyrate, whereas metabolite analysis<sup>34</sup> revealed an increase in the amounts of some bile-acid metabolites and pancreatic enzymes, as well as thousands of unidentified metabolites, that could be used to differentiate people with Crohn's disease from healthy individuals.

Type 1 diabetes has been studied in many disparate but small cohorts ( $n < 20$  per group per study) of newly diagnosed children<sup>95,96</sup>. These studies identified an elevated relative abundance of *Bacteroides* and a reduced relative abundance of *Prevotella* in those with the disease compared with controls. A longitudinal study<sup>97</sup> of children with a high risk of developing diabetes determined that increases in  $\alpha$  diversity (the diversity of species at a particular site in the body) during development were slowed in children who went on to develop diabetes ( $n = 4$ ) but not in seroconverters without clinical symptoms ( $n = 7$ ) or in healthy children ( $n = 22$ ). A metabolic study in a different cohort found that children who developed diabetes ( $n = 50$ ) had lower levels of triglycerides compared with controls ( $n = 67$ ). Seroconversion was associated with a transient increase in 2-hydroxybutyrate and a decrease in ketoleucine. Some of these metabolites might have microbial origins.

Rheumatoid arthritis, a disease not typically thought of as being associated with the gut or the mouth, has been linked to the microbiomes of both. People with rheumatoid arthritis demonstrate consistently increased relative abundances of species of *Prevotella* in their oral and gut microbiotas<sup>8,87,98</sup>. Those with newly diagnosed ( $n = 31$ ) and chronic ( $n = 32$ ) rheumatoid arthritis have higher rates of periodontal disease than do healthy controls ( $n = 18$ ), even when other risk factors such as age and smoking are taken into account<sup>98</sup>. Amplicon sequencing has shown that *Prevotella* and *Leptotrichia* OTUs are increased in individuals with rheumatoid arthritis, independent of their periodontal disease status<sup>98</sup>. Metagenomic profiling of oral and gut microbiomes has identified elevated levels of *Prevotella copri* in people with rheumatoid arthritis ( $n = 115$ ) compared with controls ( $n = 97$ ), as well as an enrichment in Gram-positive microorganisms, including members of the family Veillonellaceae<sup>87</sup>. The presence of *Lactobacillus salivarius* in the oral cavity and faeces correlates positively with antibody titres, and this microorganism was more likely to be present in active cases of rheumatoid arthritis than in controls. Treatment with disease-modifying antirheumatic drugs can partially restore characteristics of the control microbiome, including decreased levels of *Prevotella*, in individuals with rheumatoid arthritis<sup>98</sup>.

Cardiovascular disease<sup>99,100</sup> has been linked to high levels of TMAO, a metabolite of phosphatidylcholine, and TMAO is strongly correlated with both atherosclerotic plaques in a mouse model<sup>6,83</sup> and adverse cardiovascular outcomes in people<sup>99,100</sup>. TMAO has been implicated in other conditions that involve the vascular system, including renal disease<sup>101</sup> and colon cancer<sup>102</sup>. Treatment with antibiotics attenuates the production of TMAO in both mice<sup>6</sup> and people<sup>103</sup> after challenge with phosphatidylcholine. Alterations have been seen in 16S rRNA amplicon sequencing profiles of adults from Sweden<sup>99</sup> and China<sup>100</sup> who have experienced cardiovascular events, although the same OTUs were not identified in

both cohorts. TMAO might also modulate platelet function and the risk of developing thrombosis in people<sup>104</sup>. Subsequent experiments in conventional mice have confirmed that TMAO has a role in thrombosis, whereas germ-free mice seem to be protected from developing this phenotype<sup>104</sup>. In conventional mice, long-term exposure to dietary choline altered the composition of the microbiome, and several candidate taxa, including the families Lachnospiraceae and Mogibacteriaceae, were negatively associated with thrombosis<sup>104</sup>. Interestingly, the identification of the role of TMAO in cardiovascular disease began in a study of serum metabolites, and only later moved to studies of the microbiome.

The links between autism spectrum disorder and the microbiome remain controversial; although studies in people have provided statistically significant associations, they can be confounded by factors that include the diet, gastrointestinal issues and drugs. A 16S rRNA amplicon study showed that people with autism spectrum disorder ( $n = 20$ ) had a lower  $\alpha$  diversity than did neurotypical individuals ( $n = 20$ ) (ref. 11). Autism spectrum disorder was associated with higher levels of *Akkermansia* and fewer species of fermenter bacteria, including *Prevotella*, *Coprococcus* and *Veillonellaceae*<sup>11</sup>. A study of the offspring of mice who had undergone maternal immune activation (MIA) showed that alterations occur in the microbiomes and metabolomes of such mice, including a reduction in the levels of members of the family Lachnospiraceae, which ferment SCFAs<sup>30</sup>. The introduction of *Bacteroides fragilis*, a common commensal microbe, led to decreased expression of 4-ethylphenylsulfate and corrected behavioural symptoms. The administration of 4-ethylphenylsulfate was sufficient to transmit symptoms of anxiety to wild-type mice<sup>30</sup> and led to permanent immune dysfunction<sup>105</sup>.

Despite the emergence of some common themes such as the presence of specific taxa, overall trends in  $\alpha$  diversity and the ability to separate cases and controls using metrics of  $\beta$  diversity (the differences in community composition between different samples), it is impossible to determine whether a particular condition has a smaller or larger effect on the diversity of the microbiota than another, owing to the way that individual studies are conducted. A set of standardized protocols would enable many different biological and technical effects to be placed on a scale that compares common effect sizes. The Microbiome Quality Control Project is beginning to do this for technical effects by comparing the specific effects of sample storage, DNA extraction, PCR amplification and bioinformatics pipelines, all of which can have surprisingly large effects; for example, methods and databases used in the assignment of taxonomy can have much larger effects on the apparent profile of a microbiome than does which biological specimen was examined<sup>106</sup>. Large-scale efforts such as the Earth Microbiome Project<sup>107</sup> and American Gut are beginning to address these issues by studying tens of thousands of samples using common methods. The dream would be to provide quantitative information that indicates which biological effects are larger than specific technical effects (to facilitate a rational choice for which studies to compare) and describes the directionality of effects, which would enable the use of generalized linear models to detrend for specific variables so that subtle effects can be seen against the background. For example, American Gut has observed that the age of an individual and their self-reported frequency of alcohol consumption have approximately equal statistically significant effects on the diversity of the gut microbiota: to measure the influence of one variable accurately, it is therefore necessary to detrend for the other (American Gut, unpublished observations). By contrast, body mass index (BMI) has a much smaller, although still detectable effect, on the gut microbiota, which means that controls for age and alcohol use must be applied (or the data detrended) to understand the specific effects of BMI. The development of a scale for this type of effect size would also be enormously useful for scoping out new studies: it would enable an educated guess to be made about the expected effect size of an intervention or condition from a large database of past studies of similar phenomena, and the number of participants and longitudinal sampling design (if applicable) could be scoped out rationally on the same basis, to the relief of both investigators and their institutional review boards.

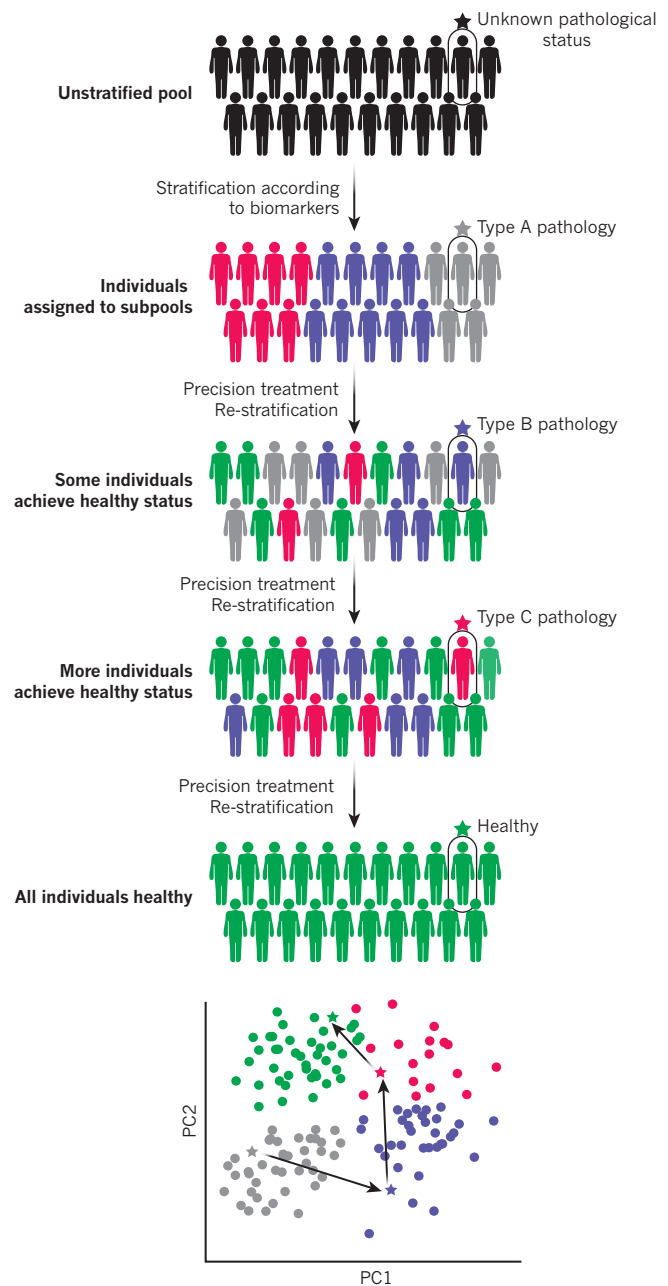
## Developing a microbial Global Positioning System

An important challenge for the field is to move beyond abstract maps of the microbiome, which enables multivariate samples to be placed in the context of other samples. It is important to understand which factors, including the host genome (Box 2), can change the microbiome from a given starting point on a 'map', as well as where the ideal endpoint would be. Such a microbial Global Positioning System (GPS) would comprise a defined start point, a defined end point and directions for how to get from one to the other and would depend on the standardization of results from microbiome studies so that each participant can be located accurately on the map and their progress tracked. It also relies on well-defined clinical cohorts that enable desirable and undesirable endpoints to be assessed. Unstratified patients have the potential to be placed anywhere on the map, and their initial location is determined, for example, by principal coordinates analysis (PCoA) of UniFrac distances between samples<sup>108</sup>, as performed by the Human Microbiome Project and American Gut. Stratification is then performed to identify certain groups of people in different parts of the map, according to specific biomarkers, such as genes, functions, metabolites or networks of these features, and perhaps by crossing different levels of analysis. These biomarkers are then used to relocate study participants to appropriate regions of the map, which helps to suggest specific treatments that would move them from their present location to another (Fig. 2). For example, a small change in the diet might provide a subtle shift in location on the map and treatment with antibiotics might produce a larger shift whereas faecal transplantation could be considered 'teleportation'. Readout of biomarkers over time would allow the progress of each participant to be tracked from unhealthy to healthier regions of the map. Overall, many more participants would be expected to reach a healthy location on the map than would be possible with unstratified treatment, although genetic defects, intractable microbiome states or other factors might prevent the recovery of some. This vision requires a substantially faster, cheaper and more accurate readout of the microbiome across multiple levels than is possible at present, although it will provide an exceptionally powerful and clinically relevant model after it has been subjected to the appropriate regulatory processes.

## Perspective

The considerable power of using the microbiome, or even the inexpensively assayed microbiota, to separate cases from controls, as well as to predict responses to treatment or the development of diseases in the absence of treatment, has already been demonstrated through carefully controlled MWAS in research settings. To further develop these techniques for robust clinical use, MWAS must be validated in larger and more diverse populations. Methodologies must also be standardized so that differences in the size of technical effects between laboratories do not outweigh differences in the size of biological effects, which can make studies difficult to combine<sup>78,90</sup>. This problem remains a crucial challenge to overcome and prevents findings from being developed into clinical tests.

Longitudinal studies have been especially informative in revealing microbiome dynamics that cannot be observed through a before–after model. In infants, where profound changes in the microbiota and microbiome occur in the first three years of life, a more detailed understanding of the developmental process, and deviations from it, is required to understand whether changes introduced by diet, environmental exposures, antibiotics and other factors in early life keep the microbiome on track or divert it towards danger. Similarly, moving away from taxonomic inventories towards an understanding of the genes, transcripts, proteins and metabolites of the microbiome in a multi-omics, systems-biology context is crucial for generalizing our understanding of a wide range of diseases in which the microbiome is involved, as well as for developing biomarkers that could be the basis of useful clinical tests. However, these are conflicting imperatives: multi-omics studies greatly increase the cost of analysing each sample, which means that longitudinal studies on large populations quickly become infeasible and tests are too expensive and slow to apply on clinically relevant timescales. Consequently, even higher-throughput and cheaper methods to process samples for multi-omics studies, as well



**Figure 2 | Developing a microbial Global Positioning System to stratify individuals and to guide their treatment.** An unstratified pool of individuals (black), all of whom have the same disease but with different underlying states (red, blue and grey), are stratified according to a biomarker from the microbiota, the microbiome or the metabolome (differentiated on a PCoA plot (bottom) or other analysis). This enables treatments to be chosen for each subpool, which facilitates movement from an 'unhealthy' region to a 'healthy' region of the microbial 'map'. The position of an individual in the main pool indicates the same person over time. The microbial Global Positioning System therefore enables determination of the current location of an individual in terms of their microbiome configuration, as well as a prediction of their final destination and directions for how to get there. Ideally, this moves all individuals in the pool to a healthy status (green) and microbiome, although in real-world situations no treatment will work perfectly. PC, principal coordinate.

as improved modelling techniques that derive systems-level dynamic parameters from fewer samples, are urgently required. These advances will rapidly bring us nearer to the dream of a microbial GPS. The Human Microbiome Project, the Earth Microbiome Project, American Gut and other large-scale efforts have already, and very effectively, provided a microbial 'map' that enables healthy and diseased samples to be placed



in context, provided that consistent laboratory and bioinformatics methods are used. In the next few years, data that are collected using consistent protocols will enable intervention studies from many investigators to be aggregated to build a general picture of how the microbiome can change in specific directions in multivariate space. This understanding will facilitate the provision of 'turn-by-turn' directions that enable individuals to use their microbiome and perhaps even their genotype to understand where they might want to go on this map and how they can get there most effectively, in a way that preserves their lifelong health. ■

Received 7 December 2015; accepted 6 May 2016.

- Manichanh, C. *et al.* Anal gas evacuation and colonic microbiota in patients with flatulence: effect of diet. *Gut* **63**, 401–408 (2014).
- Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104**, 13780–13785 (2007).  
**This study linked the microbiota to IBD and also demonstrated that various forms of the condition have distinct signatures of microbiota.**
- Lewis, J. D. *et al.* Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host Microbe* **18**, 489–500 (2015).
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
- Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
- Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Med.* **19**, 576–585 (2013).
- Kostic, A. D. *et al.* Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res.* **22**, 292–298 (2012).  
**This study identified high levels of *Fusobacterium nucleatum* in tissue from human tumours; the bacterium was later confirmed to cause tumours in experiments in animals.**
- Scher, J. U. *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2**, e01202 (2013).  
**This paper provided the first evidence to directly link the gut microbiota to rheumatoid arthritis in people.**
- Naseribafrouei, A. *et al.* Correlation between the human fecal microbiota and depression. *Neurogastroenterol. Motil.* **26**, 1155–1162 (2014).
- Scheperjans, F. *et al.* Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov. Disord.* **30**, 350–358 (2015).
- Kang, D. W. *et al.* Reduced incidence of *Prevotella* and other fermenters in intestinal microflora of autistic children. *PLoS ONE* **8**, e68322 (2013).
- Kostic, A. D., Howitt, M. R. & Garrett, W. S. Exploring host–microbiota interactions in animal models and humans. *Genes Dev.* **27**, 701–718 (2013).
- Barr, J. J. *et al.* Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl Acad. Sci. USA* **110**, 10771–10776 (2013).
- Haiser, H. J. *et al.* Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* **341**, 295–298 (2013).  
**This study provided a mechanism to underpin the high variation between individuals in efficacy of the cardiac drug digoxin, which was suspected (but not yet proven) to be linked to its metabolism by *Eggerthella lenta*.**
- Arthur, J. C. *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120–123 (2012).
- Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* **8**, 761–763 (2011).
- Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
- Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**, 382–392 (2014).  
**This study of treatment-naive children who had been freshly diagnosed with Crohn's disease enabled the effects of treatment to be separated from those of the condition.**
- Kuczynski, J. *et al.* Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nature Methods* **7**, 813–819 (2010).
- Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
- Lovell, D., Pawlowsky-Glahn, V., Egozcue, J. J., Marguerat, S. & Bahler, J. Proportionality: a valid alternative to correlation for relative data. *PLoS Comput. Biol.* **11**, e1004075 (2015).
- Knights, D., Parfrey, L. W., Zaneveld, J., Lozupone, C. & Knight, R. Human-associated microbial signatures: examining their predictive value. *Cell Host Microbe* **10**, 292–296 (2011).
- Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
- The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- Fierer, N. *et al.* Forensic identification using skin bacterial communities. *Proc. Natl Acad. Sci. USA* **107**, 6477–6481 (2010).
- Franzosa, E. A. *et al.* Identifying personal microbiomes using metagenomic codes. *Proc. Natl Acad. Sci. USA* **112**, E2930–E2938 (2015).
- Eren, A. M. *et al.* Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* **4**, 1111–1119 (2013).  
**This paper demonstrated how careful analysis of exact 16S rRNA sequences that avoids clustering into OTUs can reveal fine-grained information that can be useful for forensic matching.**
- Noecker, C. *et al.* Metabolic model-based integration of microbiome taxonomic and metabolomic profiles elucidates mechanistic links between ecological and metabolic variation. *mSystems* **13**, <http://dx.doi.org/10.1128/mSystems.00013-15> (2015).
- Koenigsnecht, M. J. *et al.* Dynamics and establishment of *Clostridium difficile* infection in the murine gastrointestinal tract. *Infect. Immun.* **83**, 934–941 (2015).
- Hsiao, E. Y. *et al.* Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155**, 1451–1463 (2013).  
**This study showed that the phenotype of a mouse model of autism spectrum disorder could be traced, in part, to a single molecule (4-ethylphenylsulfate) and a shift in the microbiota that can be partially restored using a probiotic.**
- Belda-Ferre, P. *et al.* The human oral metaproteome reveals potential biomarkers for caries disease. *Proteomics* **15**, 3497–3507 (2015).
- Willing, B. P. *et al.* A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**, 1844–1854 (2010).
- Erickson, A. R. *et al.* Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS ONE* **7**, e49138 (2012).
- Jansson, J. *et al.* Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE* **4**, e6386 (2009).
- Michail, S. *et al.* Altered gut microbial energy and metabolism in children with non-alcoholic fatty liver disease. *FEMS Microbiol. Ecol.* **91**, 1–9 (2015).
- Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262–266 (2015).  
**This paper compared two discordant studies of microbiomes in type 2 diabetes and showed that the alleged effect of diabetes could be attributed mostly to differences in use of metformin, which has an unexpectedly large effect on the microbiome, between the two populations.**
- Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214 (2013).  
**This study demonstrated that phenotypes such as increased adiposity could be transferred from people to mice using personalized culture collections.**
- Li, H. & Jia, W. Cometabolism of microbes and host: implications for drug metabolism and drug-induced toxicity. *Clin. Pharmacol. Ther.* **94**, 574–581 (2013).
- Nicholson, J. K., Lindon, J. C. & Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189 (1999).
- Wishart, D. S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nature Rev. Drug Discov.* <http://dx.doi.org/10.1038/nrd.2016.32> (2016).
- da Silva, R. R., Dorrestein, P. C. & Quinn, R. A. Illuminating the dark matter in metabolomics. *Proc. Natl Acad. Sci. USA* **112**, 12549–12550 (2015).
- Gilbert, J. A. & Henry, C. Predicting ecosystem emergent properties at multiple scales. *Environ. Microbiol. Rep.* **7**, 20–22 (2015).
- Allen, L. *et al.* Pyocyanin production by *Pseudomonas aeruginosa* induces neutrophil apoptosis and impairs neutrophil-mediated host defenses *in vivo*. *J. Immunol.* **174**, 3643–3649 (2005).
- Puertollano, E., Kolida, S. & Yaqoob, P. Biological significance of short-chain fatty acid metabolism by the intestinal microbiome. *Curr. Opin. Clin. Nutr. Metab. Care* **17**, 139–144 (2014).
- Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421 (2014).
- Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402–1414 (2014).  
**This paper showed that the human microbiome harbours many biosynthetic gene clusters, including those required for the production of antibiotics.**
- Cummings, J. H. Fermentation in the human large intestine: evidence and implications for health. *Lancet* **1**, 1206–1209 (1983).
- Huda-Faujan, N. *et al.* The impact of the level of the intestinal short chain fatty acids in inflammatory bowel disease patients versus healthy subjects. *Open Biochem. J.* **4**, 53–58 (2010).
- Ríos-Covián, D. *et al.* Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* **7**, 185 (2016).
- Wikoff, W. R. *et al.* Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl Acad. Sci. USA* **106**, 3698–3703 (2009).
- Donia, M. S. & Fischbach, M. A. Small molecules from the human microbiota. *Science* **349**, 1254766 (2015).
- Nougayrède, J. P. *et al.* *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848–851 (2006).
- Putze, J. *et al.* Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infect. Immun.* **77**, 4696–4703 (2009).
- Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* **517**, 205–208 (2015).
- Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnol.* **31**, 814–821 (2013).
- Allegretti, J. R. *et al.* Recurrent *Clostridium difficile* infection associates with distinct bile acid and microbiome profiles. *Aliment. Pharmacol. Ther.* **43**, 1142–1153 (2016).
- Maurice, C. F., Haiser, H. J. & Turnbaugh, P. J. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* **152**, 39–50 (2013).
- Mani, S., Boelsterli, U. A. & Redinbo, M. R. Understanding and modulating

- mammalian-microbial communication for improved human health. *Annu. Rev. Pharmacol. Toxicol.* **54**, 559–580 (2014).
59. Wallace, B. D. *et al.* Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330**, 831–835 (2010).  
**This paper demonstrated that the cancer therapeutic drug irinotecan causes severe diarrhoea because of its reactivation and metabolism by bacterial  $\beta$ -glucuronidases; inhibiting these enzymes with a drug that targets the bacteria, rather than the host, reduces toxicity.**
60. ElRakaiby, M. *et al.* Pharmacomicrobiomics: the impact of human microbiome variations on systems pharmacology and personalized therapeutics. *OMICS* **18**, 402–414 (2014).
61. Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R. & Nicholson, J. K. Pharmacometabonomic identification of a significant host–microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl Acad. Sci. USA* **106**, 14728–14733 (2009).  
**This study provided the first link between the toxicity of a drug (in this case, acetaminophen, a widely used analgesic) and microbial metabolism.**
62. Wilson, I. D. Drugs, bugs, and personalized medicine: pharmacometabonomics enters the ring. *Proc. Natl Acad. Sci. USA* **106**, 14187–14188 (2009).
63. Quinn, R. A. *et al.* From sample to multi-omics conclusions in under 48 hours. *mSystems* <http://dx.doi.org/10.1128/mSystems.00038-16> (2016).
64. Lax, S. *et al.* Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* **345**, 1048–1052 (2014).
65. Gibbons, S. M. *et al.* Ecological succession and viability of human-associated microbiota on restroom surfaces. *Appl. Environ. Microbiol.* **81**, 765–773 (2015).
66. Korem, T. *et al.* Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101–1106 (2015).
67. Quick, J. *et al.* Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* **16**, 114 (2015).
68. Dominguez-Bello, M. G. *et al.* Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nature Med.* **22**, 250–253 (2016).
69. Goyal, M. S., Venkatesh, S., Milbrandt, J., Gordon, J. I. & Raichle, M. E. Feeding the brain and nurturing the mind: linking nutrition and the gut microbiota to brain development. *Proc. Natl Acad. Sci. USA* **112**, 14105–14112 (2015).
70. Biteen, J. S. *et al.* Tools for the microbiome: nano and beyond. *ACS Nano* **10**, 6–37 (2016).
71. Lima-Mendez, G. *et al.* Determinants of community structure in the global plankton interactome. *Science* **348**, 1262073 (2015).
72. Sam Ma, Z. *et al.* Network analysis suggests a potentially ‘evil’ alliance of opportunistic pathogens inhibited by a cooperative network in human milk bacterial communities. *Sci. Rep.* **5**, 8275 (2015).
73. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015); erratum **17**, 852 (2015).
74. The Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project: dynamic analysis of microbiome–host omics profiles during periods of human health and disease. *Cell Host Microbe* **16**, 276–289 (2014).
75. Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A. & Knight, R. EMPor: a tool for visualizing high-throughput microbial community data. *Gigascience* **2**, 16 (2013).
76. Weingarden, A. *et al.* Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent *Clostridium difficile* infection. *Microbiome* **3**, 10 (2015).  
**This paper introduced animation techniques that revealed the transformation of the whole microbiota during faecal microbiota transplantation for *C. difficile* infection.**
77. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4578–4585 (2011).
78. Lozupone, C. A. *et al.* Meta-analyses of studies of the human microbiota. *Genome Res.* **23**, 1704–1714 (2013).
79. Flores, G. E. *et al.* Temporal variability is a personalized feature of the human microbiome. *Genome Biol.* **15**, 531 (2014).
80. Shade, A. *et al.* Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio* **5**, e01371-14 (2014).
81. DiGiulio, D. B. *et al.* Temporal and spatial variation of the human microbiota during pregnancy. *Proc. Natl Acad. Sci. USA* **112**, 11060–11065 (2015).
82. Koren, O. *et al.* Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150**, 470–480 (2012).
83. Wang, Z. *et al.* Prognostic value of choline and betaine depends on intestinal microbiota-generated metabolite trimethylamine-N-oxide. *Eur. Heart J.* **35**, 904–910 (2014).
84. Zeevi, D. *et al.* Personalized nutrition by prediction of glycemic responses. *Cell* **163**, 1079–1094 (2015).  
**This study showed that individual glycaemic responses could be predicted using the microbiome; it also revealed that although population averages match conventional glycaemic-index values, the responses of individuals are highly idiosyncratic and dependent on the microbiome.**
85. Teng, F. *et al.* Prediction of early childhood caries via spatial-temporal variations of oral microbiota. *Cell Host Microbe* **18**, 296–306 (2015).
86. Huang, S. *et al.* Predictive modeling of gingivitis severity and susceptibility via oral microbiota. *ISME J.* **8**, 1768–1780 (2014).
87. Zhang, X. *et al.* The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature Med.* **21**, 895–905 (2015).
88. Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the human body. *Nature* **509**, 357–360 (2014).
89. Cotillard, A. *et al.* Dietary intervention impact on gut microbial gene richness. *Nature* **500**, 585–588 (2013).
90. Walters, W. A., Xu, Z. & Knight, R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett.* **588**, 4223–4233 (2014).
91. Kang, D., Shi, B., Erfe, M. C., Craft, N. & Li, H. Vitamin B<sub>12</sub> modulates the transcriptome of the skin microbiota in acne pathogenesis. *Sci. Transl. Med.* **7**, 293ra103 (2015).
92. Sinha, R., Abnet, C. C., White, O., Knight, R. & Huttenhower, C. The microbiome quality control project: baseline study design and future directions. *Genome Biol.* **16**, 276 (2015).
93. Sinha, R. *et al.* Collecting fecal samples for microbiome analyses in epidemiology studies. *Cancer Epidemiol. Biomarkers Prev.* **25**, 407–416 (2016).
94. Sokol, H. *et al.* *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105**, 16731–16736 (2008).
95. de Goffau, M. C. *et al.* Fecal microbiota composition differs between children with  $\beta$ -cell autoimmunity and those without. *Diabetes* **62**, 1238–1244 (2013).
96. Giongo, A. *et al.* Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J.* **5**, 82–91 (2011).
97. Kostic, A. D. *et al.* The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* **17**, 260–273 (2015).
98. Scher, J. U. *et al.* Periodontal disease and the oral microbiota in new-onset rheumatoid arthritis. *Arthritis Rheum.* **64**, 3083–3094 (2012).
99. Koren, O. *et al.* Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4592–4598 (2011).
100. Yin, J. *et al.* Dysbiosis of gut microbiota with reduced trimethylamine-N-oxide level in patients with large-artery atherosclerotic stroke or transient ischemic attack. *J. Am. Heart Assoc.* **4**, e002699 (2015).
101. Tang, W. H. *et al.* Gut microbiota-dependent trimethylamine N-oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. *Circ. Res.* **116**, 448–455 (2015).
102. Xu, R., Wang, Q. & Li, L. A genome-wide systems analysis reveals strong link between colorectal cancer and trimethylamine N-oxide (TMAO), a gut microbial metabolite of dietary meat and fat. *BMC Genomics* **16** (suppl. 7), S4 (2015).
103. Tang, W. H. *et al.* Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368**, 1575–1584 (2013).
104. Zhu, W. *et al.* Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell* **165**, 111–124 (2016).
105. Hsiao, E. Y., McBride, S. W., Chow, J., Mazmanian, S. K. & Patterson, P. H. Modeling an autism risk factor in mice leads to permanent immune dysregulation. *Proc. Natl Acad. Sci. USA* **109**, 12776–12781 (2012).
106. Liu, Z., DeSantis, T. Z., Andersen, G. L. & Knight, R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res.* **36**, e120 (2008).
107. Gilbert, J. A. *et al.* Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. *Stand. Genomic Sci.* **3**, 243–248 (2010).
108. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
109. Quinn, R. A. *et al.* Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome. *ISME J.* **10**, 1483–98 (2015).
110. Ridlon, J. M., Kang, D. J., Hylemon, P. B. & Bajaj, J. S. Bile acids and the gut microbiome. *Curr. Opin. Gastroenterol.* **30**, 332–338 (2014).
111. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359 (2006).  
**This study provided the first metagenomic gene catalogue of the human gut.**
112. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
113. Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* **449**, 804–810 (2007).
114. White, J. R., Nagarajan, N. & Pop, M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* **5**, e1000352 (2009).
115. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
116. Mandal, S. *et al.* Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb. Ecol. Health Dis.* **26**, 27663 (2015).
117. Knights, D., Costello, E. K. & Knight, R. Supervised classification of human microbiota. *FEBS Microbiol. Rev.* **35**, 343–359 (2011).
118. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
119. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
120. Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab.* **17**, 141–152 (2013).
121. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
122. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).

**Acknowledgements** This work and the work in the authors’ laboratories that it describes was supported in part by awards from the US National Institutes of Health, the US Department of Energy, the US National Science Foundation, the Alfred P. Sloan Foundation, the Crohn’s and Colitis Foundation of America and the US Office of Naval Research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at [go.nature.com/28jf0c5](http://go.nature.com/28jf0c5). Correspondence should be addressed to R.K. ([robknight@ucsd.edu](mailto:robknight@ucsd.edu)).