Recent advances that allow us to collect more data on DNA sequences and metabolites have increased our understanding of connections between the intestinal microbiota and metabolites at a whole-systems level. We can also now better study the effects of specific microbes on specific metabolites. Here, we review how the microbiota determines levels of specific metabolites, how the metabolite profile develops in infants, and prospects for assessing a person’s physiological state based on their microbes and/or metabolites. Although data acquisition technologies have improved, the computational challenges in integrating data from multiple levels remain formidable; developments in this area will significantly improve our ability to interpret current and future data sets.

Keywords: Metabolomics; Systems Biology; Data Analysis; Bacteria.
the entire microbial community on the whole metabolite repertoire are just beginning to be understood.

Metabolomics and metatolite profiling analyses have been widely used to identify disease biomarkers. For example, quantification of triglyceride, glucose, and cholesterol levels in the blood can be used to determine the risk of heart disease. Similarly, the first microbiome studies sought to identify taxa that correlated with disease, physiological state, drug use, or dietary intake. However, not all exposures can alter the composition of the microbial community or its gene content; some can affect gene expression.13,14

Humanized mice (created by transplanting human fecal microbiota into the mouse gut) have metabolomes distinct from those of conventionally raised mice.15 This observation indicates that different gut microbes can produce changes in metabolites throughout their host. This shift in focus from determining “who is there” toward understanding “what are they doing” is driving current studies of the human microbiota. Metabolic studies will allow us to move from observing patterns to understanding mechanisms.

Metabolic analyses also help researchers understand the effects of rare taxa and taxa with genomic variations that affect function. Organisms are considered to be of the same species if they have >97% identity in the 16S ribosomal RNA gene. However, genomes from the same species can have large differences in DNA sequences outside the 16S ribosomal RNA gene. Importantly, they often have different sets of gene clusters that regulate production of specialized metabolites (eg, antibiotics, virulence factors, siderophores, and so on) and the composition of the microbial communities can encode many antibiotic resistance genes.16 Rasko et al determined that among 17 Escherichia coli isolates, the average genome size of a single isolate was 5020 nucleotides, although the pan-genome was ~13,000 nucleotides.17 Furthermore, rare taxa might have a large effect on the overall community metabolome if they have important metabolic activities, perhaps acting as keystone species.

Although definitions of what constitutes a core microbiome in terms of membership is elusive, there does seem to be at least a core functional profile for the gut microbiota.10 Identifying biologically important variations against this core remains a challenge. Metabolomic analyses provide a partial picture of metabolism rather than the potential for metabolism, as is provided by genome analysis, and the expression of this core set of functions can change with alterations in available substrates, such as xenobiotics, even if the microbial species membership and abundance remain constant.13 We review the intimate connections among animal hosts, their microbiota, and the metabolites produced by either one.

Different microbial communities metabolize xenobiotic agents and dietary components in different ways to produce variable effects on many tissues in the host, including the brain18 (Figure 1). We discuss general metabolomic technologies and their implementation for the study of human health, assess cases in which changes in the gut microbiota alter host metabolic profiles, examine the ways in which the gut microbiota processes xenobiotics and nutritional inputs, and examine the analytical limitations of associating microbial abundances with metabolic profiles.

Metabolomics in Assessment of Metabolic Status

Metabolomic studies analyze complex systems, including the repertoire of small molecule metabolites in the gut, using high-throughput analytical methods. Mass spectrometry and nuclear magnetic resonance spectroscopy allow robust and sensitive identification of metabolites produced by microbes and host cells in samples such as feces, urine, and tissue (see comprehensive reviews by Dettmer et al19 and Slupsky20). These tools allow researchers to determine the effects that treatments or perturbations have on the host’s metabolic profile by analyzing the presence and quantity of thousands of metabolites simultaneously. Although it is a challenge to assign spectral features, spectral networking platforms,21,22 aided by open-source metabolome databases such as HMDB,23 METLIN,24 LIPID MAPS,25 MassBank,26 and NIST,27 allow for faster identification and annotation of known and unknown metabolites.28 By comparing perturbation and postperturbation metabolic profiles using multivariate statistics, metabolites that are significantly affected by experimental variables can be identified and placed into the larger context of how the host was affected overall.

Effects of the Microbiome on the Metabolome

Metabolic analyses allow for the metabolism of the gut microbiota to be directly compared with metabolic outcomes in the host. Wikoff et al29 directly tested the effect of the gut microbiota on the host by comparing the plasma metabolomic profile, obtained via untargeted mass spectrometry, between germ-free and conventionally raised mice. They found that concentrations of more than 10% of all metabolites detected in the plasma differed by at least 50% between mice with and without gut microbes. Furthermore, many metabolites were detected only in serum from conventionally raised mice (not germ-free mice). For example, serum levels of tryptophan decreased 40% in serum from conventional mice compared with germ-free mice, likely because of the presence of bacteria that produce tryptophanases.29

Another detailed study evaluated the systemic effects of probiotics, prebiotics, and their combination (termed “synbiotics”) in initially germ-free mice colonized with a combination of microbes representing those found in a human infant (Bacteroides distasonis, Clostridium perfringens, Escherichia coli, Bifidobacterium breve, Bifidobacterium longum, Staphylococcus aureus, and Staphylococcus epidermidis).30 Dietary supplementation with the probiotic Lactobacillus rhamnosus NCC4007 and the probiotic galactosyl-oligosaccharides significantly altered the relative proportions of the 7-member community and led to systemic changes in the metabolic profiles of different tissues from mice. For example, a prebiotic increased proportions of B breve, B longum, and B distasonis; decreased proportions of E coli and C perfringens; and altered lipid metabolism by
reducing plasma levels of glucose and hepatic levels of triglycerides. Probiotics also had systemic effects, lowering plasma levels of lipoprotein, hepatic levels of glutamine, and glycogen levels. Overall, prebiotics significantly altered the metabolome in the plasma, urine, feces, liver, pancreas, renal cortex, renal medulla, and adrenal glands; probiotics produced differences in all these compartments except the pancreas.

Interestingly, another study that evaluated the effects of probiotics and prebiotics in adults found that neither significantly affected the proportions of microbes in fecal samples, but RNA sequencing data showed altered expression of microbial genes that control carbohydrate metabolism. It is possible that the relatively simpler communities that reside in infants are more susceptible to probiotic and prebiotic manipulation than the more diverse and complex communities found in adults. Prebiotics and probiotics might therefore have the largest effects when administered early in life. However, this hypothesis requires testing in animal models.

The dietary components that escape digestion in the upper gastrointestinal tract provide most of the substrates for the intestinal microbiota. Fermentation of carbohydrates by the intestinal microbiota leads to the production of short-chain fatty acids such as butyrate, propionate, and acetate. Studies have shown that patients with inflammatory bowel diseases such as ulcerative colitis have fewer butyrate-producing bacteria (eg, *Roseburia hominis* and *Faecalibacterium prausnitzii*) in their intestine, resulting in lower levels of butyrate. In addition to butyrate, propionate...
can potenti ate de novo generation of T regulatory cells in the peripheral immune system. Modulation of butyrate- and propionate-producing microbes might therefore be used to treat inflammatory bowel diseases such as ulcerative colitis. However, the anti-inflammatory mechanisms of butyrate and other short-chain fatty acids remain poorly defined.

**Predictive Microbial Metagenomes**

Metagenomic information can help determine how metabolism is affected by different disease states linked to dysbiosis. Studies of obesity have shown that subjects with increased adiposity have lower microbial diversity than lean subjects.3,4 The more diverse microbiota of lean subjects contains significantly higher proportions of microbes correlated with anti-inflammatory responses, such as *Faecalibacterium prausnitzii*. The less diverse microbiota of obese subjects contains higher proportions of *Bacteroides* species and *Ruminococcus gnavus*, each of which could have inflammatory effects.3,4 Gene content analysis of these groups revealed that the less diverse microbiota appeared to produce lower levels of butyrate, have increased potential for production of hydrogen sulfide, and have reduced capability for management of oxidative stress. One poorly understood aspect of the microbiome, and its potential to produce a variety of metabolites, is whether microbial diversity itself has protective effects for the host or whether low diversity is a side effect of specific disorders (rather than a cause).5,6 This relationship can best be resolved in humans by prospective longitudinal studies.

Although it would be ideal to obtain metabolomic and metagenomic data for every sample for which a 16S amplicon profile has been collected, these techniques are currently far more expensive than 16S amplicon profiling. Fully matched data sets are therefore prohibitively expensive and time consuming to produce. However, recent advances in software, including Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt),6 that exploit the strong association between phylogeny and function now allow researchers to estimate the metabolomic functional profile of a community using 16S amplicon sequences. Briefly, PICRUSt takes a phylogenetic tree in which the gene profile of a subset of nodes is known and then uses ancestral state reconstruction to estimate the functional gene content for other uncharacterized nodes. PICRUSt was able to make strong predictions (average Spearman r = 0.82) for inferred metagenomes from 16S marker genes compared with fully sequenced metagenomes obtained from the Human Microbiome Project.

Another powerful computational tool is Predicted Relative Metabolic Turnover, which uses gene counts to predict the relative consumption and production of metabolites in a system; it can be used for modeling and hypothesis generation.5,7 Tools such as PICRUSt and Predicted Relative Metabolic Turnover could be cost-effective methods to determine whether additional resources should be used for more comprehensive metabolic profiling and metagenomic sequencing. However, findings must be validated with matched data sets to assess the limits of their performance.

**Metabolomic Profiles of Infants**

Changes to the microbiome and immune system during infancy may have lasting effects, such as contributing to the development of allergies.14,38,39 Distinct changes in the microbiota occur during the first 2 years of life and correlate with changes in environment and diet; these can be tracked by studying changes in infants’ fecal metabolomes. A study of infants at risk for celiac disease showed that the metabolomes of infants younger than 6 months of age were dominated by sugars, including lactose and glucose. However, after 6 months, their metabolomes shifted, increasing concentrations of amino acids and short-chain fatty acids. Principal coordinates analysis showed that the metabolome of children at 2 years of age resembles more closely that of adults because of increased levels of acetate and butyrate.40 These findings are supported by 16S amplicon studies showing that the infant microbiota comes to resemble that of adults from the same community at 2 years of age.41 It is also apparent that the intestinal microbiota of infants is specifically adapted to metabolize the infant’s earliest nutrient source: breast milk. Specific *Bifidobacterium* species have genomes enriched in genes that regulate the processing of human milk-derived oligosaccharides. These might have a competitive advantage that places them among the first colonizers of the human intestine.42

**Xenobiotic Metabolism**

In addition to diet-derived macronutrients, the microbes residing in the gastrointestinal tract may be exposed to a variety of xenobiotic compounds (antibiotics, other drugs, and diet-derived bioactive compounds). Because the gut microbiome encodes so many enzymes with different activities, it is not surprising that many of the xenobiotic compounds are often metabolized by the gut microbiota. It has been at least 40 years since we began to appreciate the contribution of microbes to xenobiotic metabolism.53–55 However, we are only beginning to uncover the mechanisms of this process. Adding to the complexity of these interactions, xenobiotics can also modulate the expression and activity of the gut microbiome.12 Metabolites of microbial origin may interfere with host metabolism of xenobiotics, and diet-derived nutrients can regulate microbial metabolism of xenobiotics.

One of the first studies to provide detailed evidence for the interaction between the gut microbiota and metabolism of xenobiotics was from Clayton et al in 2009.46 Their study leveraged a powerful metabolomic analysis pipeline to correlate the presence of the microbial metabolite p-cresol with a reduction in the ratio of sulfonated to glucuronidated acetylamino phen. Increased production of p-cresol reduces the capacity of the liver to properly metabolize this widely used analgesic drug, presumably because p-cresol competes with sulfotransferases.46 Subsequent studies from this group showed that metabolites of microbial origin could modulate expression and activities of a range of host enzymes, including those of major xenobiotic-metabolizing cytochrome enzymes.57 These seminal observations are
beginning to lay the foundation for a metagenomic approach to selection of therapy based on microbial and host metabolism.

In addition to its interactions with metabolite production, the gut microbiota can also have a more direct role in xenobiotic metabolism by catalyzing a multitude of reactions that influence the fate of these compounds. Recent reviews have summarized the many processes by which microbes metabolize xenobiotics. Although these activities are largely catalogue, there are only a few for which the exact mechanisms have been characterized. For example, it has been known for decades that the cardiac drug digoxin can be inactivated by *Eggerthella lenta*, a common gut bacterium within the Actinobacteria. Researchers have recently identified a cytochrome-encoding operon that is up-regulated by digoxin and other cardiac glycosides and is unique to strains capable of inactivating digoxin. Inactivation of digoxin was blocked by increasing dietary protein intake by mice monoassociated with *E. lenta*, likely due to the inhibitory effect of arginine.

Wallace et al. studied how the microbiota can determine the effects of the colorectal cancer drug irinotecan. Enzymes produced by microbes have long been known to deconjugate an irinotecan metabolite in the gut, causing inflammation, diarrhea, and anorexia. After a successful screen for a small molecule inhibitor of the microbial β-glucuronidase enzyme that mediates this deconjugation, Wallace et al showed that the side effects of irinotecan could be greatly reduced by coadministration with this β-glucuronidase inhibitor. Interestingly, recent studies have shown that the presence of the microbiota increases the efficacy of chemotherapeutic drugs, indicating that the microbiota has a previously unappreciated but integral role in mediating responses to these drugs.

**Computational Challenges to Discovering Correlations**

Identifying statistically meaningful patterns in metabolite contingency tables (tables recording the abundance of each metabolite count in each sample) is straightforward in theory but often conducted with mathematically unfounded techniques in practice. For instance, analysis of variance and Student t test are frequently used to identify significant differences in abundances of metabolites among sample groups without establishing that the underlying data meet the distribution requirements. Normality, equality of variance, and homogeneous population characteristics are required for proper calculations of statistical significance (either P values or false discovery rates). Although nonparametric tests can be substituted to deal with the non-normality of the data, these approaches still do not resolve 2 fundamental computational challenges: extraction of biologically significant results from the mass of statistically significant results and the fact that multivariate biological data are typically normalized to a sum; the simplex constraints this imposes violate the Euclidean space models assumed by most test statistics (see the following text).

The most widely applied method to reduce biologically irrelevant but statistically significant results is to remove features (taxa, KEGG Orthology groups, Enzyme Commission numbers, and so on) from the contingency table before testing on the basis of a metric that assigns expected biological relevance to a feature. This “metric” is usually as simple as overall table abundance (eg, remove feature if i is in <1% of all observations) or overall sample representation (eg, remove feature if i is in <20% of samples). This filtering approach is motivated by the intuition that more widely shared features will be more biologically important and has the additional attraction of reducing the severity of multiple hypothesis test correction factors. Unfortunately, although widespread, this approach has not yet been

![Figure 2](image-url)
systematically benchmarked or evaluated for sensitivity, specificity, or even false discovery control, particularly in fields combining microbiome and metabolomic data sets.

A complementary approach to identifying differential representation of features among groups is to look for interactions among features via co-occurrence analysis (Figure 2). Traditional co-occurrence detection methods, including Spearman or Pearson correlations, between feature vectors are not reliable when the data are “compositional” (i.e., lie in a simplex rather than Euclidean space). Because compositionality is a feature of much -omics data (16S amplicon surveys are inherently compositional because normalization for unequal sampling effort in any contingency table introduces compositionality), methods such as SparCC and CoNet have been developed to capture true correlations. Although these methods are well founded in mathematics, they have been benchmarked and validated in only limited circumstances and their performance has not yet been characterized for metabolomic data in general.

Conclusion

The overall diversity and plasticity of the gut microbiota, in comparison to our human genomes, provides exciting new prospects for personalized medicine, particularly for studies to determine the mechanisms by which microbes affect production of metabolites from drugs and diet. Although there is much work to be done, especially in terms of computational methods, the experimental frameworks of metabolomics and microbial community analysis that have emerged should allow for rapid host characterization followed by subsequent analyses of clinical potential.

References


Reprint requests
Address requests for reprints to: Jennie Smoly Caruthers, Biotechnology Building, 596 UCB Boulder, Colorado 80309-0596 Rob Knight, PhD. e-mail: rob.knight@colorado.edu.

Conflicts of interest
The authors disclose no conflicts.

Funding
Supported by a National Institutes of Health Signaling and Cellular Recognition Training Grant (T32 GM08759 to L.K.U.), by the Canadian Institutes of Health Research (MFE-112991 to H.J.H.), by National Research Initiative Grant 2009-55200-05197 from the USDA National Institute for Food and Agriculture (to J.V. and L.R.), and in part by the Howard Hughes Medical Institute.