ANNUAL REVIEWS

Annual Review of Genetics

The Relationship Between the Human Genome and Microbiome Comes into View

Julia K. Goodrich,[1,2] Emily R. Davenport,[2] Andrew G. Clark,[2] and Ruth E. Ley[1,2]

[1]Department of Microbiome Science, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; email: ruth.ley@tuebingen.mpg.de

[2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA

Annu. Rev. Genet. 2017. 51:413–33

First published as a Review in Advance on September 20, 2017

The *Annual Review of Genetics* is online at genet.annualreviews.org

https://doi.org/10.1146/annurev-genet-110711-155532

Keywords

microbiome, genome-wide association studies, heritability, microbiota composition, human studies

Abstract

The body's microbiome, composed of microbial cells that number in the trillions, is involved in human health and disease in ways that are just starting to emerge. The microbiome is assembled at birth, develops with its host, and is greatly influenced by environmental factors such as diet and other exposures. Recently, a role for human genetic variation has emerged as also influential in accounting for interpersonal differences in microbiomes. Thus, human genes may influence health directly or by promoting a beneficial microbiome. Studies of the heritability of gut microbiotas reveal a subset of microbes whose abundances are partly genetically determined by the host. However, the use of genome-wide association studies (GWASs) to identify human genetic variants associated with microbiome phenotypes has proven challenging. Studies to date are small by GWAS standards, and cross-study comparisons are hampered by differences in analytical approaches. Nevertheless, associations between microbes or microbial genes and human genes have emerged that are consistent between human populations. Most notably, higher levels of beneficial gut bacteria called Bifidobacteria are associated with the human lactase nonpersister genotype, which typically confers lactose intolerance, in several different human populations. It is time for the microbiome to be incorporated into studies that quantify interactions among genotype, environment, and the microbiome in order to predict human disease susceptibility.

INTRODUCTION

Microbes coat the body's surfaces. In the human gut, microbial cells reach densities of 10^{12} per mL and, in aggregate, form a mass of up to a kilogram, constituting what amounts to an additional organ whose genome vastly expands the host's in size and metabolic function. The host's profound dependence on the microbiome for both establishment and maintenance of a normal phenotype is illustrated most vividly by comparisons of animals raised with and without a microbiome. Germfree animals, which are born, raised, and maintained aseptically, are devoid of microbial cells and therefore lack the many cues expected and necessary for their postnatal development and subsequent normal functioning. The abnormalities of germfree animals range across organ systems, from the immune system to the cardiovascular system, and include basic functions such as lipid cycling, energy balance, and behavior (57). Many of the unusual phenotypes exhibited by germfree animals reverse upon colonization with a microbiome (27), yet others require microbial exposure for reversal at critical points in the animal's development (1, 24, 52, 59). Given the clear importance of the microbiome for host normalcy, key goals are to understand factors that determine colonization and abundance of commensal microbes and the impact of specific constituents on host health. The host's microbiome is acquired from birth onwards through contact with others and the surrounding world, so unsurprisingly, environmental factors strongly influence its composition. The microbiome can contribute to fitness and illness (6, 23, 26, 71); therefore, the host should have a strong interest in shaping the microbiome in such a way to promote its own fitness.

For its part of the symbiosis, the microbiota performs functions beneficial to the host, from enhancing digestion to protecting against the invasion of pathogens. Natural selection acts on individual bacterial species to enhance their fitness and to improve the function of the microbiota as a stable community. Selection pressure on the host itself can also result in selection of a microbiota that performs functions beneficial to the host (39, 49). If members of the microbiota enhance host fitness (49), this enhancement could have the effect of ensuring the presence of host habitat for the microbiota over the longer term. Indeed, a microbiota allows its host to exploit specific niches, for instance through detoxification of plant secondary compounds. As an example, goats can safely consume the toxic plant *Leucaena* only when they harbor gut bacteria able to degrade 3,4-dihydroxypyridine, a breakdown product of the *Leucaena* amino-acid mimosine (33). Furthermore, certain mammals exhibit behavioral or other traits to ensure the beneficial microbiota transfers to the next generation (49). Mechanisms for selecting, retaining, and transferring key elements of the microbiome are likely to be genetically encoded by the host, and the discovery of these genetic factors will point to mechanisms underlying host–microbe symbioses.

One way to uncover potentially new host–microbe interactions is to search for human genes with alleles that covary across a population with traits in the microbiome. It is likely that human alleles critical in maintaining essential microbial functions have been permanently established in the gene pool. Indeed, the genetic underpinnings of human gut physiology and function that help maintain the microbial habitat may not present much variation that can be associated with differences in the microbiome across a population. However, microbes and/or functions beneficial only in a specific context may show a signal of association with human genetic variation. For example, the strongest evidence of recent selection on the human genome is seen in geographically restricted areas that present specific environmental challenges, such as high altitude, high pathogen load, and high toxicity (54). It is likely that genetic evidence for selection on attributes of the microbiome may also be linked to specific challenges that humans have faced in recent evolution.

Indeed, one of the strongest signals of recent selection on humans consists of the genetic changes that enabled lactase persistence in adulthood and thereby the drinking of nonhuman milk. Remarkably, the most consistent signal to emerge from genome-wide association studies

(GWASs) of the microbiome is related—it consists of an association between host genotype, milk consumption, and Bifidobacteria (4, 5, 18, 68). In this instance, Bifidobacteria are more highly abundant in the gut microbiome of hosts who ingest milk postweaning and who lack the lactase–persister genotype (see the section titled *Bifidobacterium* and Human Lactase Persistence for an expanded discussion of this association). Other examples of human genetic variation associated with variation in the microbiome are starting to emerge (**Table 1** and **Figure 1**). Here, we review the recent findings from human microbiome heritability analyses and GWASs and the challenges emerging from the marriage of microbiome and human genetics.

THE MICROBIOME AS A COMPLEX TRAIT IN HUMAN GENETICS

The microbiome is a complex community of organisms, and many of its attributes can be modeled in studies that examine the role of host genetics. Typically, microbiomes are characterized either with 16S rRNA gene sequencing or through metagenome sequencing (20). These data sets allow the quantification of taxa or gene functions across samples and can also form the basis of various ecological metrics that characterize diversity in a sample or within a population. The microbiota in the human gut, for instance, can be described as hundreds of operational taxonomic units (OTUs) per individual, with tens of thousands of OTUs represented across a population. These OTUs can be collapsed into higher taxonomic levels along their phylogeny (e.g., genus, family, order). Host genetics may influence the total number and composition of the taxa present (alpha diversity). Meanwhile, genetic effects could become apparent when the extent of OTU sharing between individuals of varying relatedness (beta diversity) is considered. Beyond taking a census approach to determine which microbes are present, shotgun sequencing can be used to characterize the functional metagenomic landscape of the microbiome. Microbial genes can be grouped into functional categories or pathways, and the abundances and presence or absence of those groups could be targets of modulation through host genetics. Any and all of these attributes can be characterized and modeled as quantitative traits for which heritability can be estimated and quantitative trait loci identified. Each of these microbiome attributes should be considered because it is unclear a priori how the host genome might influence the microbiome.

HERITABLE TAXA OF THE HUMAN GUT MICROBIOME ARE INCREASINGLY VALIDATED ACROSS STUDIES

The identification of heritable taxa from comparisons of related individuals preceded GWAS for two reasons: (*a*) It motivated GWASs, ensuring there were genetic determinants of microbiome composition, and (*b*) it reduced the number of traits ascertained in a given GWAS by constraining them to the heritable list. The first unbiased search for heritable taxa among the human gut microbiota was conducted by Goodrich and colleagues (21), who studied genotyped twins from the TwinsUK registry. Stool samples were obtained from over 1,000 twin pairs (3,261 samples total), and the modeled data consisted of 16S rRNA gene sequences (18). Genotyped twins allowed for both twin-based heritability analysis and GWASs to identify host genes and metabolic pathways associated with these heritable taxa. Heritability analysis revealed that approximately 10% of the 945 taxa identified by 16S and shared by a minimum of 50% of the subjects had a heritability greater than 0.2 with 95% confidence intervals that did not overlap zero. Remarkably, of the 26 heritable taxa identified, 13 were nominally replicated in the Canadian Genetic Environmental Microbial Project cohort of 270 related individuals from 123 families (65) (**Figure 1**). Six of the 26 were not addressed for technical reasons, implying that more than half of the heritable taxa that could be compared were heritable in a second population. Given that the human gut

Table 1 Summary of human microbiome genome-wide association studies

Reference	Body site	Sample size	Subject cohort	Subject population	Data type	Microbiome attributes examined	Number of variants	Significance threshold
4	15 sites within the oral and nasal cavities, gastrointestinal tract, and on skin[a]	93	Human Microbiome Project	United States	16S	Beta diversity (first five PCs); 615 taxa (genus to phylum)	33,814	Beta-diversity pathway-based analysis: $P \leq 10^{-6}$; Taxa: Genome-wide; Q-value < 0.1
Main findings: Beta diversity: enrichment of genes involved in leptin signaling in obesity, several other immunity-related pathways, and Kyoto Encyclopedia of Genes and Genomes pathway primary bile acid biosynthesis. Taxa: 83 associations including *HLA-DRA* (*Selenomonas* in the throat), *TLR1* (*Lautropia* in the tongue dorsum), and *LCT* (*Bifidobacterium* in the gastrointestinal tract).								
10	Stool	127	NA	North American Hutterites	16S	Alpha diversity; 102 taxa (genus to phylum)[b]	212,153[b]	Genome-wide; Q-value < 0.2
Main findings: Significant associations with at least eight bacterial taxa including *Akkermansia* with a variant near *PLD1*. Gene set enrichment analysis identified enrichment of variants involved in olfactory receptor activity with five taxa. Stomach and intestines were identified as candidate tissues where host genetic variation may be acting to influence bacterial abundance of genus *Faecalibacterium*.								
28	Lung	147	Environment And Genetics in Lung cancer Etiology study	Italy	16S	Alpha and beta diversity; bacterial taxa (number not specified)	383,263	$P < 5 \times 10^{-8}$
Main findings: No significant associations after correcting for skewness and kurtosis of beta-diversity distributions.								
18	Stool	2,139	TwinsUK	United Kingdom	16S	Alpha and beta diversity; 782 OTUs; 163 taxa (genus to phylum)	1,300,091	Genome-wide BH adjusted $P < 0.1$

(*Continued*)

Table 1 (*Continued*)

Reference	Body site	Sample size	Subject cohort	Subject population	Data type	Microbiome attributes examined	Number of variants	Significance threshold
Main findings: 92 significant associations with OTUs and taxa. Includes an association between the heritable taxon of unclassified Clostridiaceae with *SLIT3* and *Bifidobacterium* with *LCT*. Beta diversity was associated with *UHRF2*. Association between *Akkermansia* and predicted tissue-specific expression of *SIGLEC15*.								
5	Stool	1,514: 984 (discovery) and 530 (replication)	LifeLines Deep (discovery); 500 Functional Genomics Project and Maastricht IBS cohort (replication)	Netherlands	Metagenomics	219 microbial taxa; 636 MetaCyc pathways; 661 GO terms	8.1 million	Discovery: $P < 5 \times 10^{-5}$; Replication: $P < 0.01$ (same direction); Meta-analysis: $P < 5 \times 10^{-8}$
Main findings: Meta-analysis: 9 loci associated to bacteria, 21 loci associated to MetaCyc pathways, and 12 loci associated to GO terms Strongest association with taxa: Blautia with SNPs near *LINGO2* and Methanobacteriaceae with an extended lncRNA. Strongest association with MetaCyc pathways: pathway involved in plant-derived steroid degradation with *SORCS2* and *SLIT3*. Two GO terms were associated with SNPs near clusters of C-type lectin domain family 4 genes. GG genotype at SNP rs4988235 (near *LCT* gene) was associated with high *Bifidobacterium*.								
65	Stool	1,561: 1,098 (discovery) and 463 (replication)	Genetic Environmental Microbial Project	Canada and United States (discovery); Canada, United States, and Israel (replication)	16S	Alpha diversity; 166 taxa (genus to phylum)	3,727,707	Genome-wide $P < 5 \times 10^{-8}$; Study-wise correction for number of effective tests: $P < 4.13 \times 10^{-10}$
Main findings: 58 genome-wide significant associations with taxa, 6 of which achieved study-wise significance. Four loci were replicated: Rikenellaceae with SNPs near *UBR3*, *Faecalibacterium* with *CNTN6*, *Lachnospira* with *DMRTB1*, and *Eubacterium* with *SALL3*.								

(*Continued*)

Table 1 (*Continued*)

Reference	Body site	Sample size	Subject cohort	Subject population	Data type	Microbiome attributes examined	Number of variants	Significance threshold
68	Stool	1,812 (discovery); 371 (replication)	PopGen biobank and Food Chain Plus project (discovery); Food Chain Plus project–obesity (replication)	Germany	16S	Beta diversity; 40 OTUs; 58 taxa (genus to phylum)	6,344,846	$P < 5 \times 10^{-8}$
Main findings: 42 loci associated with beta diversity including variants in the *VDR* gene. 21 of the 42 beta diversity–associated loci replicate. 54 significant associations with taxa; unclassified Porphyromonadaceae was associated with *SLC2A9*. Several lncRNA including *LINC01192* are associated with Lactobacillales.								
30	Nasopharynx; nasal vestibule	144	NA	North American Hutterites	16S	90 genera (nasopharynx)[b]; 76 genera (nasal vestibule)	148,653[b]	Genome-wide; Q-value < 0.05
Main findings: 37 associations with genera; these genetic loci are enriched for genes in mucosal immunity pathways. Associations included *Dermacoccus* with a SNP near *TINCR* and an unclassified genus of family Micrococcaceae with *PGLYRP4*.								

Abbreviations: BH, Benjamini–Hochberg; GO, gene ontology; IBS, inflammatory bowel syndrome; lncRNA, long noncoding RNA; OTUs, operational taxonomic units; PCs, principal components; SNP, single-nucleotide polymorphism.

[a]Human Microbiome Project–specific sites tested: attached keratinized gingiva (gums), buccal mucosa (cheek), hard palate, palatine tonsils, saliva, subgingival plaque, supragingival plaque, throat, tongue dorsum, anterior nares (nostrils), left and right antecubital fossa (inner elbow), left and right retroauricular crease (behind the ear), and stool.

[b]Number reported is from the analysis combining summer and winter samples.

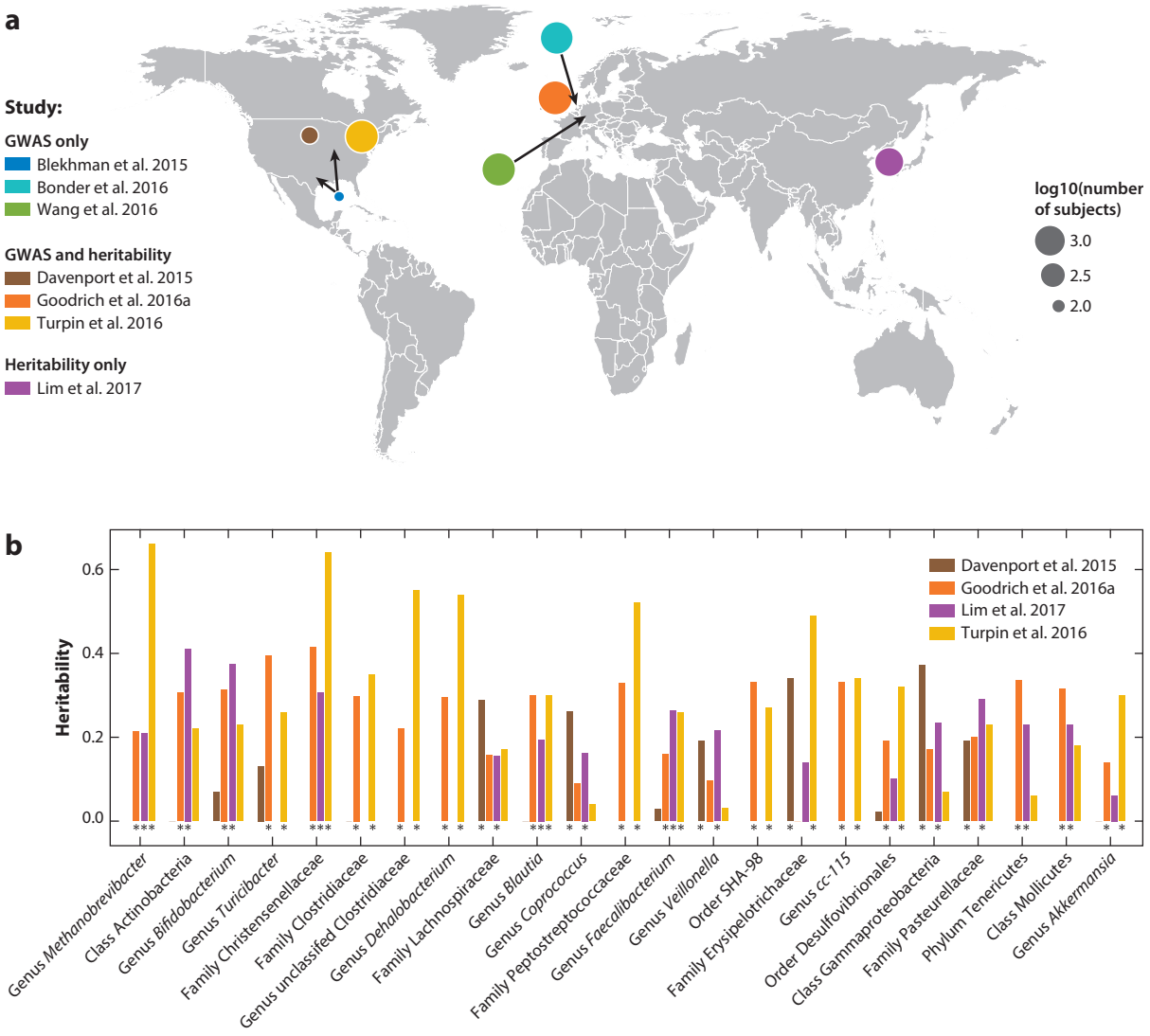


Figure 1

Heritability studies and GWASs of the human gut microbiome. (*a*) A world map indicates the locations and relative sample size of the currently published human gut microbiome heritability studies and GWASs. Each color represents a single study, and the size of the circle indicates the study's sample size. Studies shown include Blekhman et al. (4), Bonder et al. (5), Wang et al. (68), Davenport et al. (10), Goodrich et al. (18), Turpin et al. (65), and Lim et al. (40). (*b*) A comparison of taxon heritability across studies. Only taxa found to have nominally significant heritability estimates [$P < 0.05$ or "chip heritability" estimates (10) with a standard error not overlapping 0; marked with an asterisk] in at least two of the four heritability studies are shown in the bar chart (bars are colored by study). Abbreviation: GWAS, genome-wide association study.

microbiome is variable across subjects and highly influenced by environmental factors such as diet, this congruence between studies argues strongly that specific, identifiable taxa are responsive to host genotype across populations and warrant mechanistic follow-up.

The list of heritable taxa that have passed significance testing is only a small subset of the nominally heritable list within each study. The nominally heritable list is generally assumed to

be more likely to contain false positives and is usually not discussed in a study's main findings. However, an expansion of the population size and reanalysis of heritability in UK twins showed that the list of nominally heritable taxa might be quite valuable. In the study of UK twins, the list of heritable taxa was generated twice, first with 416 twin pairs (21) and then again with an expanded set of 1,126 twin pairs (18). This tripling of the data set revealed the following: (*a*) The list of taxa stayed constant, with minimal reshuffling of the heritability rankings, and (*b*) the confidence intervals around the heritability estimates were reduced. This had the effect that taxa formerly excluded from the heritable list due to confidence intervals overlapping with zero were considered heritable through the expanded analysis. Furthermore, these observations underscored that the types of heritable taxa are not dependent on the specific set of individuals studied within the population. These results also demonstrated the expected increase in power to detect heritable microbes with larger sample sizes and the gain of confidence in the results from smaller sample sets. Small-scale analyses may therefore yield valuable insights into the heritability of taxa even when underpowered, and the nominally heritable taxa may be interesting to pursue further.

Most estimates of microbial abundance stem from 16S rRNA gene sequence data, which provide phylogenetic information but generally little functional information. Furthermore, many taxa are functionally redundant in the human microbiome, and this is thought to contribute to the stability of the system, particularly in the gut (loss of a taxon does not lead to loss of function if the function is widely shared). So it is interesting that some taxa are indeed heritable, and this observation implies (*a*) that some attribute of the taxa is under selection and (*b*) that that attribute is phylogenetically restricted.

Overall, heritability estimates of components of the gut microbiota are generally low compared with other heritable traits (53). Heritability estimates calculated for the UK twin fecal microbiome ranged from 0 to approximately 0.40 (18, 21). Heritability estimates have also been obtained using Korean twins and their families ($h^2 = 0$–0.46) (40), Canadian families consisting of mostly siblings ($h^2 = 0$–0.67) (65), and the North American Hutterites (using seasons combined chip-heritability, $h^2 = 0$–0.37) (10). The low values of the heritability estimates may be linked to the fact that the data are derived from stool, which is a mix of mucosal and luminal contents. If a heritable microbe can be quantified in its original habitat (e.g., the mucosal surface), true heritability estimates may be higher. The low heritability should be considered not just a first pass but also a worst-case scenario, because more focused studies are bound to yield higher values.

Christensenellaceae

Goodrich et al. (21) reported the most highly heritable taxon to be the family Christensenellaceae. Subsequently, the heritability of Christensenellaceae has been validated in Canadians of European descent ($h^2 = 0.64$) and in Koreans ($h^2 = 0.31$) (40, 65) (**Figure 1**). Christensenellaceae is a family within Firmicutes that is relatively small (i.e., has less branch length compared with a family such as the Ruminococcaceae), which might explain why the whole family is heritable. The heritability estimate for the whole family is driven by taxa that constitute branches of the phylogeny lacking cultured representatives at this time. *Christensenella minuta*, the first laboratory isolate, which lent its name to the family (47), has lower, nonsignificant heritability in the TwinsUK data set ($h^2 = 0.27$) (21), but was found to be heritable ($h^2 = 0.54$) in the study of Turpin et al. (65).

Goodrich et al. (21) reported that the family Christensenellaceae constitutes the hub (i.e., most interconnected node) of a co-occurrence network consortium that includes the families Methanobacteriaceae, Dehalobacteriaceae, SHA-98, RF39 (Tenericutes), and ML615J-28 (Tenericutes), all of which are heritable. This consortium was also present in the data of Yatsunenko et al. (72), derived from young adult twins from Missouri, USA.

This Christensenellaceae consortium was, in addition to being heritable, enriched in lean versus obese individuals in the UK twins (21). It also positively correlated with alpha diversity, which was higher in lean subjects compared with obese subjects in the TwinsUK population. The association of Christensenellaceae with a lean phenotype was also observed in Missouri twins (64), the Dutch LifeLines Deep population (17), Koreans (40), and Japanese individuals (51). Christensenellaceae was subsequently linked to visceral fat phenotypes in the TwinsUK cohort (3), as well as to healthy levels of triglycerides in the Dutch LifeLines Deep cohort (17). Furthermore, the Christensenellaceae increased in relative abundance in the stool of subjects consuming resistant starch and correlated with levels of specific short-chain fatty acids in stool (66).

Taken together, these studies point to an interaction between the Christensenellaceae and its consortium members with diet and host lipid metabolism and adiposity. Goodrich and colleagues (21) tested the causality of the association between a lean host phenotype and the relative abundance of the Christensenellaceae experimentally using fecal transplants into germfree mice. Amendment of an obese microbiome known to be extremely low in Christensenellaceae with live *C. minuta* cells protected germfree mouse recipients from the levels of adiposity gains observed in controls (i.e., same obese-derived microbiome with no addition of *C. minuta* or with addition of heat-killed *C. minuta*). This finding linked the Christensenellaceae functionally to the lean phenotype, and the underlying mechanisms are currently under investigation.

Methanogens

The co-occurrence of Christensenellaceae with methanogens (i.e., members of the domain Archaea that produce methane) observed in UK twins was also reported by Hansen et al. (25) prior to the renaming of the family, and more recently another group reported co-occurrence of these taxa in a study of North Americans (66). Methanogens correlate with leanness in several studies (2, 38, 46, 56). *Methanobrevibacter smithii* (the dominant human gut methanogen) carriage was first shown to be heritable in Missouri twins (25). Corroborating this early finding, methanogen abundance was shown to be heritable in UK twins using both 16S rRNA data ($h^2 = 0.21$) (18) and metagenomic data ($h^2 = 0.38$) (69), as well as in Canadians of European descent ($h^2 = 0.66$) (65) and in a cohort of Korean twins ($h^2 = 0.21$) (40). In a study using fecal metagenomic data on 1,514 individuals, Bonder et al. (5) showed that methanogen abundance was associated with single-nucleotide polymorphisms (SNPs) located within a long noncoding RNA. Why methanogens are heritable and/or linked to this specific region of the genome remains unclear. Because methanogen abundances are typically correlated with other facets of the microbiome including specific taxa and alpha diversity (see the section titled Measures of Richness), the association with the gene region may be driven by any of these co-occurring taxa, complicating the task of understanding any mechanisms underlying the association.

Measures of Richness

Alpha diversity, expressed as various measures (e.g., number of observed OTUs, Shannon index, Faith's Phylogenetic Diversity) (20), is heritable in at least three populations (10, 18, 65). Although moderate heritability has been observed for some alpha-diversity metrics, none of these studies reported significant associations with genetic variants (10, 18, 65). Alpha diversity is commonly negatively associated with several chronic inflammatory diseases such as inflammatory bowel disease (IBD) and obesity (48). For example, alpha diversity (assessed from 16S rRNA gene diversity analysis of fecal samples) has been observed to be lower in patients with metabolic syndrome compared with controls (38, 40). The reasons why microbiomes exhibit lower alpha diversity may differ between disease states.

One important factor that shapes the gut microbiome habitat and has been associated with alpha diversity is gut transit time, which relates to stool consistency. The Bristol Stool Scale (BSS) is often used as a proxy for colonic transit time because the two tend to be negatively associated (i.e., lower BSS is indicative of longer transit time and harder stool). BSS has been negatively correlated with species richness and with the abundances of *Methanobrevibacter* and *Akkermansia* (i.e., these taxa and alpha diversity are higher in hard stool) (67). In contrast, Tigchelaar et al. (62) did not find a significant correlation between BSS and species richness; however, they did report that decreasing BSS score (i.e., harder stools) was significantly associated with Archaea (i.e., methanogens) and the bacterial families Christensenellaceae and Dehalobacteriaceae (all members of the heritable co-occurring consortium in the TwinsUK population). Roager et al. (55) also reported that species richness was positively associated with colonic transit time. In addition, this group reported that three OTUs belonging to Christensenellaceae and one OTU classified as *Methanobrevibacter* positively associated with colonic transit time and with protein degradation products. In accord, in a study of Japanese subjects, the Christensenellaceae family negatively associated with bowel movement frequency (51). With longer transit time, microbes have longer to work on substrates and to liberate additional substrates (which can increase niche space and diversity), and potentially slower-growing microbes have the necessary gut retention time to reach measurable levels. All of these processes could lead to greater richness. Apart from associations between gut transit times and diseased states such as cystic fibrosis and IBD, little work has been done on the genetics of gut transit time.

IDENTIFYING MICROBIOME–HOST GENOTYPE ASSOCIATIONS

Attributes of the microbiome that are used as traits in GWAS (**Table 1**) include both (*a*) individual-level measurements, such as alpha diversity, relative abundances of specific taxa, and functional pathways or gene ontology terms; and (*b*) cross-sample traits, such as beta-diversity metrics. Alpha diversity and microbial or functional pathway abundances can simply be treated as individual quantitative traits, and standard GWAS methods can be applied to each trait (7). The microbiome GWASs to date have used standard additive genetic modeling approaches (4, 10, 18, 30, 68), rank-based correlations (5), or combination models, where common taxa are modeled as quantitative traits and rare taxa are modeled as binary traits (65).

Another avenue being explored for GWAS of microbiome attributes is the use of variable selection methods for high-dimensional data. Recently, Lynch et al. (43) developed a pipeline called HOMINID, which uses a penalized regression method called Lasso. This pipeline performs a single regression for each genetic variant with all taxa as predictors. When HOMINID was applied to data from 93 participants in the Human Microbiome Project, six genetic variants remained significant following multiple testing correction. Application of this method to some of the recent microbiome GWASs that include thousands of individuals could be useful in identifying more associations.

Association with beta diversity is more complex because it is a measure of similarity or dissimilarity between two samples, resulting in a value for each pair of individuals. In the first GWAS of beta diversity, Blekhman et al. (4) performed principal coordinates analysis (PCoA) on the pairwise beta-diversity matrix and ran a GWAS for each of the first five principal coordinates (PCs). This amounted to looking for human genetic variants that were associated with the majority of the microbiome variation in the data set and allowed for a reduction in the dimensionality of the data compared with testing the association of each taxon with all genetic variants. Recently, Wang et al. (68) used the function envfit in the vegan R package to fit each genotype onto the main axes of the beta-diversity PCoA (by default the first two PCs). In this method, genotype was

treated as a categorical variable, and researchers identified the SNPs associated with community composition by determining if the centroids for the three genotypes (with respect to the main axes of the PCoA) were significantly different. Hua et al. (28) developed a tool called microbiome GWAS for associating beta diversity with each genetic variant. Microbiome GWAS is based on the intuition that if a variant is associated with the microbiome, any two individuals with more alleles in common at a given locus (e.g., individuals have two alleles in common if both individuals are AA, and none in common if one is AA and the other is GG) will have more similar microbiotas and therefore smaller beta-diversity distances.

STATISTICAL CHALLENGES OF MICROBIOME GENOME-WIDE ASSOCIATION STUDIES

Treating the microbiome as a complex trait in GWASs is relatively new, and published studies have had small sample sizes (in the low thousands) by GWAS standards (tens or hundreds of thousands). As such, it has been challenging for associations of specific alleles with microbiome traits to reach study-wide significance due to the burden of multiple testing in these small studies. Indeed, study-wide significance is a high bar when the number of tests is based on the total number of SNPs (hundreds of thousands to millions) combined with the total number of traits (typically in the high hundreds to thousands).

There are a number of ways researchers reduce the number of tests performed in order to mitigate the large multiple testing burden. Some studies begin by focusing their analysis on metrics of overall community composition (alpha and beta diversity). Although this is an important initial step, its main limitation is that it does not provide information about which specific microbes are influenced by host genetics. Additionally, given the significant impact of environmental factors on the microbiome, any signal from an association with only a subset of the community will likely be drowned out when the community as a whole is examined.

Of the thousands of taxa inhabiting the gut, relatively few are shared among all or most individuals in a population (29). The presence of large numbers of rare taxa leads to the problem of zero inflation in attempts to model all taxa in the gut (70). Most of the previously published studies reduced the number of tests by excluding taxa and microbial functions with low abundance or prevalence, limiting the traits of interest to those that are more widely shared in the population. This strategy mitigates both issues relating to modeling zero-inflated data, for which there would be low power to detect associations, and the multiple testing burden. Even after this filtering, the number of remaining traits is typically in the hundreds, and performing an association for each taxon and microbial function with each genetic variant results in a very large multiple testing burden. A strict Bonferroni correction for multiple testing would require studies to reach P values of 5×10^{-10} to 5×10^{-11}.

In addition to filtering microbiome attributes, researchers can also reduce the number of tests by restricting which host genetic variants are examined. Constraining the SNPs tested to candidate gene sets is one strategy (19, 35). The drawback of this approach is that relevant genes not on the candidate gene list may be missed. Alternative approaches limit testing to SNPs only in genic regions (4). Although this approach focuses on functional regions of the genome, it likely misses much of the signal, as human GWAS hits are often identified in intergenic regions thought to be regulatory in nature (12).

Replication cohorts can be used within a study to provide confidence in suggestive associations that do not pass a strict study-wide significance threshold in the discovery cohort (5, 65, 68). For instance, Bonder et al. (5) used a three-step approach where all associations meeting a relaxed significance threshold ($P < 5 \times 10^{-5}$) in a discovery cohort were then examined in an

independent cohort. If an association replicated in the independent cohort (with the same direction of association and a $P < 0.01$), it was tested in a meta-analysis using both the discovery and replication cohorts. The estimated study-wide false discovery rate for associations that passed the meta-analysis significance ($P = 5 \times 10^{-8}$) was 12%.

Despite the challenges in modeling the microbiome with a true GWAS, there is immense value in reporting the results of an unbiased discovery approach. Results from separate studies can then be compared with each other to identify taxa that are reproducibly associated with variants in the genome (9, 19, 68). This brings to light the critical importance of validating suggestive associations across studies until larger, more powerful studies reduce the numbers of false positives.

BIFIDOBACTERIUM AND HUMAN LACTASE PERSISTENCE

The most consistent signal to have emerged from human gut microbiome GWASs to date is the association between *Bifidobacterium* in the fecal microbiota and SNPs near the *LCT* gene on chromosome 2, first reported by Blekhman et al. (4) in the Human Microbiome Project subjects. Remarkably, this association has since been replicated in twins from the United Kingdom (18), North American Hutterites (18), a Dutch cohort (5), and individuals from northern Germany (68). The CC genotype of the SNP rs4988235 at this locus is associated with lactase nonpersistence (13, 63) and elevated abundance of Bifidobacteria compared with the TT or TC genotype. Bifidobacteria can utilize the milk sugar lactose as an energy source. These observations led Goodrich et al. (19) to suggest that Bifidobacteria break down lactose and increase in abundance in hosts who are lactase nonpersistent yet nevertheless consume lactose. This scenario implied a host genotype by diet interaction (**Figure 2**). This prediction was verified by Bonder et al. (5), who, by characterizing the microbiome using shotgun metagenomics, observed the same association of the lactase nonpersister genotype with Bifidobacteria in 1,514 samples derived from three cohorts. Bonder et al. also had dietary information on the subjects and observed the association only in those consuming milk. The exact nature of the association between *Bifidobacterium*, lactose in the diet, and the lactase nonpersister genotype still requires experimental confirmation. More work is needed to decipher which species and strains of Bifidobacteria are implicated in this association. It is possible that the presence of the Bifidobacteria confers a degree of lactose tolerance to lactase nonpersisters.

To date, the lactase persistence genotype and Bifidobacteria association has been detected in persons of European descent only. Persons of African descent may exhibit independently acquired lactase persistence via a different genetic mechanism (63). However, it is intriguing to note that the phenotype (lactose tolerance or intolerance) predicted by the genotype is not always accurate (54), indicating that the microbiome may be mediating the phenotype. An association with Bifidobacteria may be expected in African populations as well, or it could be that the African equivalent occurs through lactase activity of a different gut microbe.

GENOME-WIDE ASSOCIATION STUDIES REVEAL TISSUES, GENES, AND PATHWAYS CONSISTENTLY ASSOCIATED WITH MICROBIOME ATTRIBUTES

In addition to the replication of the *Bifidobacterium–LCT* association across multiple studies of the gut microbiome, a number of other broad trends have emerged from cross-study comparisons of microbiome GWASs. First, select host tissues and pathways have been implicated across studies. Additionally, specific human genes repeatedly associate with the microbiome, although the corresponding taxa vary. Finally, multiple lines of evidence point to human genetic influence on the abundance of distinct microbial pathways and functions.

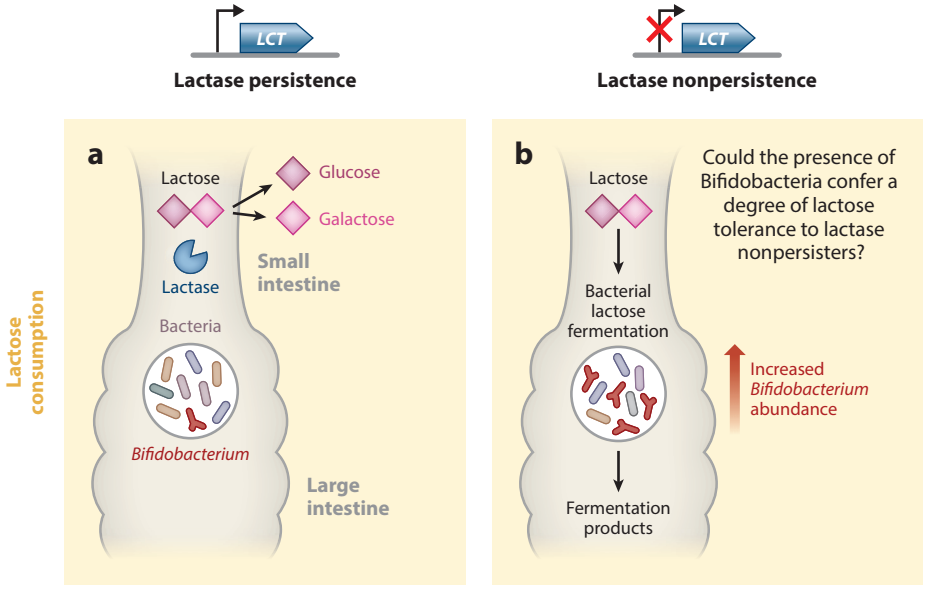


Figure 2

Interaction between the human lactase nonpersister genotype, *Bifidobacterium* abundance, and lactose in the host's diet. Genetic variants near the *LCT* gene are associated with lactase persistence and in several microbiome GWASs were recently associated with *Bifidobacterium* relative abundance in fecal samples. The association may function according to the following scenario: (*a*) If an individual is a lactase persister and consumes lactose, the lactose is typically broken down into glucose and galactose in the small intestine by host lactase. (*b*) However, if the individual is a lactase nonpersister and consumes lactose, it travels to the large intestine, where it is fermented by lactose-utilizing bacteria, which includes *Bifidobacterium*. If Bifidobacteria are present, then the presence of lactose promotes their abundance. In individuals who do not consume lactose, *Bifidobacterium* abundance remains unaffected by their lactase persistence status (*not shown*). Figure adapted from Reference 19.

Host Tissues and Pathways Implicated

The most consistent finding between reports to date is that the regions containing variants associated with the microbiome are enriched for genes related to immunity. Through a pathway enrichment analysis, Blekhman et al. (4) identified genes involved in the following immunity-related pathways: leptin signaling in obesity, melatonin signaling, JAK/STAT signaling, chemokine signaling, CXCR4 signaling, and role of pattern recognition receptors in recognition of bacteria and viruses. Genes associated with the microbiome of nasal, oral, and skin body sites drove most of the enrichment in these pathways. In a study of the nasal microbiota in a Hutterite population, Igartua et al. (30) used the Ingenuity Pathway Analysis Knowledge Base to identify protein–protein interaction (PPI) networks from genes near nasal microbiome–associated loci. Both of the significant PPI networks identified contain highly connected proteins that play important roles in modulating mucosal immunity. These included hubs at IgA, IgG, IL12/IL12RA, TCR, and STAT5A/B.

Immunity-related genes are also implicated in many of the gut microbiome GWASs. A targeted gene analysis by Bonder et al. (5) revealed several significant associations of microbial and functional abundances with immune response genes. The strongest signal in the targeted analysis was between the GO2000 term cell–cell signaling and a SNP in the *C11orf30–LRRC32* locus, which has been associated with multiple immunity-related phenotypes. Other associations include genes

implicated in IBD risk (*CCL2*, *DAP2*, *IL23R*), nucleotide-binding oligomerization domain genes *NOD1* and *NOD2*, two *CLEC* loci, and two genetic variants in the major histocompatibility complex region. Additionally, the most significant association found by Turpin et al. (65) that was also validated in their replication cohort was between the abundance of the family Rikenellaceae and a locus containing the gene *UBR3*, which encodes for a protein involved in the protein ubiquitination pathway. The authors note that ubiquitination plays many crucial roles in the immune system.

In the gut microbiome studies, evidence suggests genetic variation may act in digestive tract tissues to affect microbiome composition. For instance, Wang et al. (68) reported enrichment for genes expressed in the digestive tract. In North American Hutterites, genetic variants associated with *Faecalibacterium* were enriched in the DNase hypersensitivity sites of intestine and stomach tissues (10). Goodrich et al. (18) did not perform an enrichment analysis to identify candidate tissues, but when searching for an association between taxa and predicted gene expression across tissues, the authors found significant associations only with expression in the transverse colon.

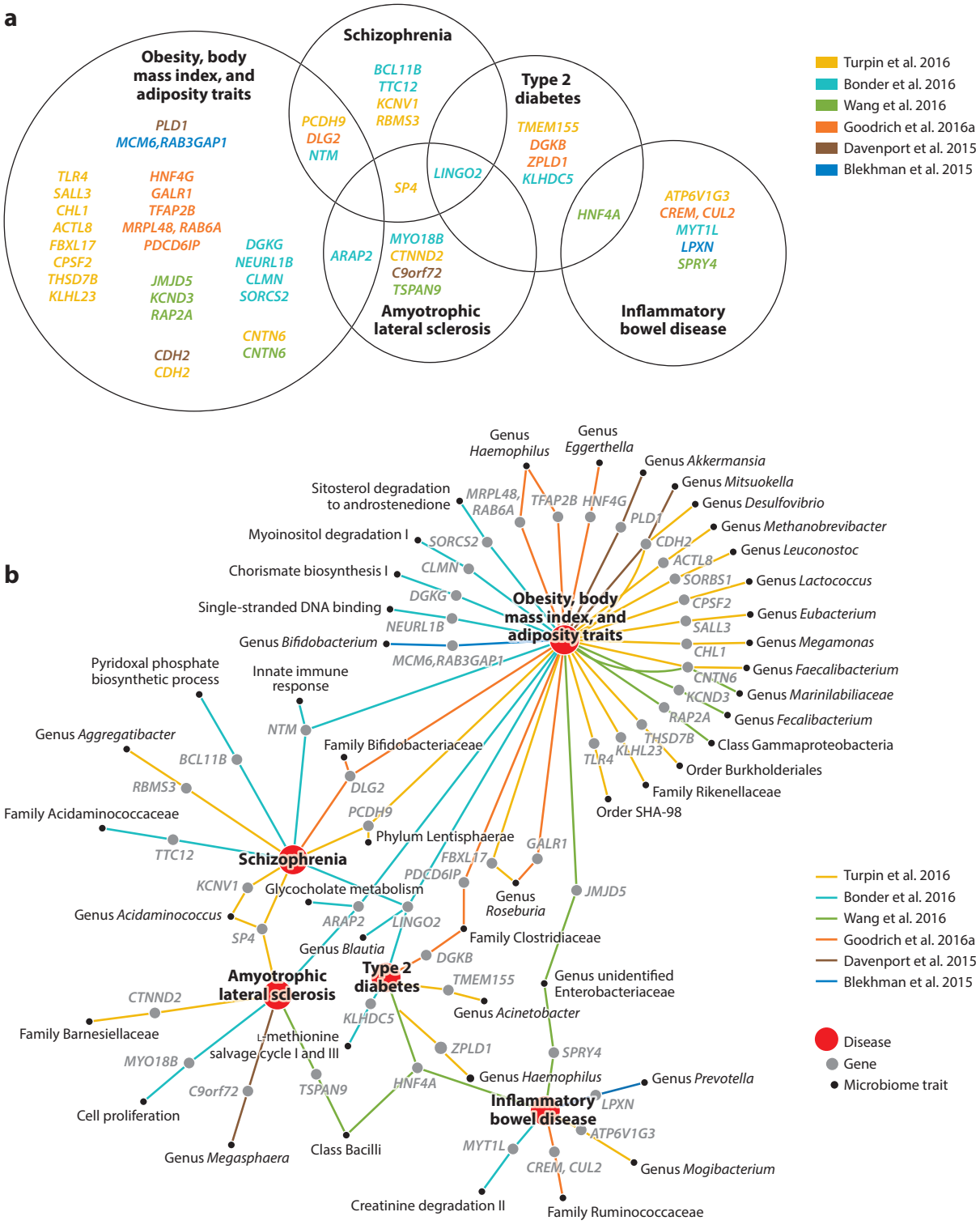
Many microbiome GWASs report genes associated with several of the same complex diseases (**Figure 3**). This includes IBD (4, 5, 65, 68), obesity (4, 5, 10, 65, 68), and type 2 diabetes (5, 65, 68), all of which are also associated with alterations in the gut microbiota (16, 37, 64). IBD risk genes are also repeatedly linked to gut microbiota composition in targeted association analyses (5, 15, 31, 35). Genetic variants near the genes *PLD1* and *LINGO2*, which have been implicated in obesity GWASs (42, 50), are associated with the abundance of *Akkermansia* (10) and *Blautia* (5), respectively. Both *Akkermansia* and *Blautia* have been linked to obesity-related phenotypes (3, 14, 21). The overlap observed between genetic variants associated with both microbiome attributes and complex diseases motivates further investigation to better understand how human genetic variation impacts the microbiome in the context of these diseases.

Specific Human Genes and Proteins Implicated

The gene *SLIT3* has been reported by three studies as having an association with some aspect of the microbiome. The most significant microbial pathway association observed by Bonder et al. (5) was between *SLIT3* and the sitosterol degradation to androstenedione pathway (involved in plant-derived steroid degradation). Goodrich et al. (18) also found an association with a variant in this gene and the abundance of unclassified Clostridiaceae ($h^2 = 0.32$). The nasal microbiome GWAS performed by Igartua et al. (30) identified a significant association between the abundance

Figure 3

Many microbiome-associated loci across studies are also implicated in disease susceptibility by GWASs. (*a*) Venn diagram of genes within or near microbiome-associated variants that have been previously associated in disease case-control GWASs. Text color indicates the microbiome GWAS that reported the association (*see key in figure*). (*b*) A network illustrating the microbiome-gene-disease associations. Nodes represent the microbiome traits (*black dots*), their associated genes (*gray circles*), and diseases (*red circles*) that are also associated with those genes. Lines indicate an association either between a microbiome trait and a gene or between a disease and a gene (colored according to the study where the association was identified—*see key in figure*). Lines do not indicate any causal direction, only an association. For both panels *a* and *b*, all microbiome-associated SNPs were downloaded from the supplements of each stool microbiome GWAS [stool associations in table S5 from Blekhman et al. (4), combined seasons associated loci in table S7 from Davenport et al. (10), genus- to phylum-level taxa associations in table S5 from Goodrich et al. (18), and all associations in supplementary table 7 from Wang et al. (68), supplementary table 6 from Turpin et al. (65), and supplementary table 3 from Bonder et al. (5)]. To determine the nearest and neighboring genes for each associated locus, the software DEPICT (**https://www.broadinstitute.org/depict**) with the 1,000 Genomes phase 3 CEU data was used. The GWAS catalog was then queried to obtain diseases that have been associated with each gene, and the five diseases that have the most overlap with microbiome-associated genes are displayed in the Venn diagram and network.


a
Obesity, body mass index, and adiposity traits
Schizophrenia
Type 2 diabetes
Amyotrophic lateral sclerosis
Inflammatory bowel disease
PLD1
MCM6,RAB3GAP1
TLR4
SALL3
CHL1
ACTL8
FBXL17
CPSF2
THSD7B
KLHL23
HNF4G
GALR1
TFAP2B
MRPL48, RAB6A
PDCD6IP
JMJD5
KCND3
RAP2A
CDH2
CDH2
DGKG
NEURL1B
CLMN
SORCS2
CNTN6
CNTN6
BCL11B
TTC12
KCNV1
RBMS3
PCDH9
DLG2
NTM
SP4
LINGO2
ARAP2
MYO18B
CTNND2
C9orf72
TSPAN9
TMEM155
DGKB
ZPLD1
KLHDC5
HNF4A
ATP6V1G3
CREM, CUL2
MYT1L
LPXN
SPRY4
Turpin et al. 2016
Bonder et al. 2016
Wang et al. 2016
Goodrich et al. 2016a
Davenport et al. 2015
Blekhman et al. 2015
b
Obesity, body mass index, and adiposity traits
Schizophrenia
Type 2 diabetes
Amyotrophic lateral sclerosis
Inflammatory bowel disease
Genus Haemophilus
Genus Eggerthella
Genus Akkermansia
Genus Mitsuokella
Genus Desulfovibrio
Genus Methanobrevibacter
Genus Leuconostoc
Genus Lactococcus
Genus Eubacterium
Genus Megamonas
Genus Faecalibacterium
Genus Marinilabiliaceae
Genus Fecalibacterium
Class Gammaproteobacteria
Order Burkholderiales
Family Rikenellaceae
Order SHA-98
Sitosterol degradation to androstenedione
Myoinositol degradation I
Chorismate biosynthesis I
Single-stranded DNA binding
Genus Bifidobacterium
Pyridoxal phosphate biosynthetic process
Innate immune response
Genus Aggregatibacter
Family Bifidobacteriaceae
Family Acidaminococcaceae
Phylum Lentisphaerae
Glycocholate metabolism
Genus Acidaminococcus
Genus Blautia
Family Clostridiaceae
Genus Roseburia
Genus unidentified Enterobacteriaceae
Genus Acinetobacter
Family Barnesiellaceae
L-methionine salvage cycle I and III
Cell proliferation
Genus Megasphaera
Class Bacilli
Genus Haemophilus
Genus Prevotella
Genus Mogibacterium
Creatinine degradation II
Family Ruminococcaceae
MRPL48, RAB6A
TFAP2B
HNF4G
PLD1
CDH2
ACTL8
SORBS1
CPSF2
SALL3
CHL1
CNTN6
KCND3
RAP2A
THSD7B
KLHL23
TLR4
SORCS2
CLMN
DGKG
NEURL1B
MCM6,RAB6GAP1
NTM
BCL11B
RBMS3
DLG2
PCDH9
TTC12
KCNV1
SP4
ARAP2
LINGO2
FBXL17
PDCD6IP
GALR1
JMJD5
DGKB
TMEM155
KLHDC5
ZPLD1
SPRY4
HNF4A
TSPAN9
C9orf72
MYO18B
CTNND2
LPXN
ATP6V1G3
MYT1L
CREM, CUL2
Turpin et al. 2016
Bonder et al. 2016
Wang et al. 2016
Goodrich et al. 2016a
Davenport et al. 2015
Blekhman et al. 2015
Disease
Gene
Microbiome trait

of *Dermacoccus* in the nasal vestibule and another variant in this gene. SLIT3 is a secreted protein expressed in several tissues including skin, stomach, small intestine, and colon (11). Hypermethylation at the *SLIT3* 5′ CpG island occurs in colorectal cancers (11). SLIT3 likely plays a role in inflammation: The expression of *SLIT3* increases after lipopolysaccharide stimulation of mouse macrophages (60), and this gene has been associated with body mass index in a GWAS for obesity (41).

Another emerging theme from the human microbiome GWASs is a link between host genetics, the microbiome, and bile acids. One of the strongest signals of association with overall community composition reported by Wang et al. (68) was with SNPs located in the gene that encodes for the vitamin D receptor (VDR). Further exploration into this association revealed that *Parabacteroides* was the taxon most highly associated with *VDR* and that *Parabacteroides* abundance was also significantly higher in *VDR* knockout mice compared with wild-type mice. VDR is a known receptor for secondary bile acids (45), and activation of VDR can inhibit bile acid synthesis (22). This led Wang et al. (68) to profile serum bile acids: They reported significant correlations between the bile acid measurements, gut microbiome composition, and genetic variation at *VDR* as well as other loci. Interestingly, Blekhman et al. (4) observed an enrichment of microbiome-associated genes in the primary bile acid biosynthesis pathway in the Kyoto Encyclopedia of Genes and Genomes. In addition, Xie et al. (69) reported that the abundance of bile salt hydrolase genes was significantly heritable in UK twins ($h^2 = 0.29$), and Bonder et al. (5) identified an association between the MetaCyc bacterial bile acid metabolism pathway and SNPs in the *ARAP2* gene. All of these studies support an interaction between host genetics and the microbiome through the regulation of bile acid metabolism.

One last example suggests a link between a specific gene family, members of the gut microbiome, and transit time (see the section titled Measures of Richness for further discussion of transit time and the microbiome). Jankipersadsing et al. (32) conducted a GWAS using stool frequency as a trait in the LifeLines Deep population ($n = 1{,}546$) and reported that the second strongest association is with the gene *ALDH1A1*, which plays a role in xenobiotic metabolism. Goodrich et al. (18) identified an association between another member of the aldehyde dehydrogenases gene family (*ALDH1L1*) and SHA-98, a member of the heritable Christensenellaceae consortium that includes the methanogens. *ALDH1L1* is involved in one-carbon metabolism. Whether and how these findings may be related remain to be clarified.

Microbial Pathways Implicated

To date, only Bonder et al. (5) have used shotgun metagenomics to investigate the relationship between microbial pathways and host genetic variation genome-wide, preventing a cross-study comparison of microbial pathways that are associated with genetic variants. However, Xie et al. (69) recently reported estimates of heritability for gut microbial pathways using 127 twin pairs from the TwinsUK cohort. As a result, it is possible to search for overlap between the significantly heritable pathways reported in this study and the pathways with a genetic association reported by Bonder et al. (5).

In addition to the bile acid metabolism example mentioned above, both studies suggested that host genetics could have some influence on the abundance of microbial genes involved in riboflavin biosynthesis ($h^2 = 0.51$). Humans acquire riboflavin (vitamin B_2) both through their diet and from riboflavin-producing gut microbes. The machinery required for riboflavin synthesis has been found in the genomes of most of the Bacteroidetes, Fusobacteria, and Proteobacteria examined, whereas a complete riboflavin operon was present in only about half of the Firmicutes and almost no Actinobacteria (44, 61). Riboflavin can be used as a redox mediator by *Faecalibacterium*

prausnitzii to facilitate extracellular electron transfer, which consequently promotes its growth (34). Increased riboflavin metabolism has been observed in individuals with ulcerative colitis (36), whereas *F. prausnitzii* abundance is reduced in individuals with IBD (58). Riboflavin biosynthesis was correlated with SNPs near the gene *CLEC4A*, which encodes a C-type lectin (5). Members of this gene family have a wide range of functions including important roles in inflammation and immunity (8). As more studies use metagenomics to investigate which microbial genes and functions are influenced by human genetics, comparisons across studies will be important for validating these initial findings.

CURRENT DIFFICULTIES WHEN COMPARING ACROSS STUDIES

Although common themes emerge in comparisons of genetic factors that influence the human microbiome across GWASs, direct validation of specific associations is largely still lacking. Each study curates the uncovered associations and chooses to highlight only a subset, making comparisons across studies problematic. An association highlighted in one study may be present in another, but if it falls just under the significance threshold, it might not be reported. For example, the *Bifidobacterium–LCT* finding does not reach genome-wide significance in most studies, and it likely would not have been reported as a main finding without Blekhman et al. (4) previously pointing it out. However, when specifically targeting the association between *LCT* and *Bifidobacterium*, most of the published studies observed this association. This is again related to power issues in untargeted studies that make it difficult to differentiate between false positives and real signals.

There may be several more cases like the *LCT* example that are overlooked in comparisons of the highlighted results of the reported findings. Additionally, any differences in the data analysis pipeline—for example, in the filtering of less abundant and less prevalent taxa—make comparisons across studies extremely difficult. To properly compare studies requires all data to be analyzed in the same way, which implies a laborious reprocessing of all data sets. Aspects of the analysis that are important to standardize include the method for OTU picking, the database and algorithm for taxonomy classification, cutoffs for taxa inclusion, the transformation method used on the microbiome data (if any), and the test for association (e.g., recent studies have used linear/logistic mixed models, negative binomial generalized linear models, log-normal generalized estimating equation models, and rank-based Spearman correlations). Only after this standardization is completed can there be a more reliable comparison of associations across all studies for the same taxon $\times$ SNP pairs.

PROSPECTUS

The microbiome is a complex trait. Initial forays into the identification of genes that covary with aspects of the microbiome are promising and have highlighted the role of immunity and diet in shaping the microbiome, although more direct comparisons are needed between studies where all data are similarly processed. Heritable taxa are so far remarkably consistent across studies, and many are health-associated. How the microbiome interacts with genotype to influence disease phenotype is an open frontier. Acquisition of SNP genotype information on common gut microbes based on deep metagenomic data will open opportunities to examine the evolutionary tuning of the microbiome to the human gut. Larger studies—across multiple populations and in the context of disease susceptibility—should continue to shed light on human–microbiome interaction and coevolution.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Angela Poole for comments on the manuscript. Support was provided by the Max Planck Society and by National Institutes of Health (NIH) grant R01 DK093595. E.R.D. is funded by NIH grant F32 DK109595.

LITERATURE CITED

1. An D, Oh SF, Olszak T, Neves JF, Avci FY, et al. 2014. Sphingolipids from a symbiotic microbe regulate homeostasis of host intestinal natural killer T cells. *Cell* 156(1–2):123–33
2. Armougom F, Henry M, Vialettes B, Raccah D, Raoult D. 2009. Monitoring bacterial community of human gut microbiota reveals an increase in *Lactobacillus* in obese patients and methanogens in anorexic patients. *PLOS ONE* 4(9):e7125
3. Beaumont M, Goodrich JK, Jackson MA, Yet I, Davenport ER, et al. 2016. Heritable components of the human fecal microbiome are associated with visceral fat. *Genome Biol.* 17(1):189
4. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, et al. 2015. Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* 16:191
5. Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, et al. 2016. The effect of host genetics on the gut microbiome. *Nat. Genet.* 48(11):1407–12
6. Budden KF, Gellatly SL, Wood DLA, Cooper MA, Morrison M, et al. 2017. Emerging pathogenic links between microbiota and the gut–lung axis. *Nat. Rev. Microbiol.* 15(1):55–63
7. Bush WS, Moore JH. 2012. Chapter 11: Genome-wide association studies. *PLOS Comput. Biol.* 8(12):e1002822
8. Dambuza IM, Brown GD. 2015. C-type lectins in immunity: recent developments. *Curr. Opin. Immunol.* 32:21–27
9. Davenport ER. 2016. Elucidating the role of the host genome in shaping microbiome composition. *Gut Microbes* 7:178–84
10. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y. 2015. Genome-wide association studies of the human gut microbiota. *PLOS ONE* 10(11):e0140301
11. Dickinson RE, Dallol A, Bieche I, Krex D, Morton D, et al. 2004. Epigenetic inactivation of *SLIT3* and *SLIT1* genes in human cancers. *Br. J. Cancer* 91(12):2071–78
12. Edwards SL, Beesley J, French JD, Dunning AM. 2013. Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* 93(5):779–97
13. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Järvelä I. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 30(2):233–37
14. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, et al. 2013. Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *PNAS* 110(22):9066–71
15. Frank DN, Robertson CE, Hamm CM, Kpadeh Z, Zhang T, et al. 2011. Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm. Bowel Dis.* 17(1):179–84
16. Frank DN, St. Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. 2007. Molecular–phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *PNAS* 104(34):13780–85
17. Fu J, Bonder MJ, Cenit MC, Tigchelaar EF, Maatman A, et al. 2015. The gut microbiome contributes to a substantial proportion of the variation in blood lipids. *Circ. Res.* 117(9):817–24
18. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, et al. 2016a. Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* 19(5):731–43

19. Goodrich JK, Davenport ER, Waters JL, Clark AG, Ley RE. 2016b. Cross-species comparisons of host genetic associations with the microbiome. *Science* 352(6285):532–35
20. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, et al. 2014a. Conducting a microbiome study. *Cell* 158(2):250–62
21. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, et al. 2014b. Human genetics shape the gut microbiome. *Cell* 159(4):789–99
22. Han S, Chiang JYL. 2009. Mechanism of vitamin D receptor inhibition of cholesterol 7α-hydroxylase gene transcription in human hepatocytes. *Drug Metab. Dispos.* 37(3):469–78
23. Hand TW, Vujkovic-Cvijin I, Ridaura VK, Belkaid Y. 2016. Linking the microbiota, chronic disease, and the immune system. *Trends Endocrinol. Metab.* 27(12):831–43
24. Hansen CHF, Nielsen DS, Kverka M, Zakostelska Z, Klimesova K, et al. 2012. Patterns of early gut colonization shape future immune responses of the host. *PLOS ONE* 7(3):e34043
25. Hansen EE, Lozupone CA, Rey FE, Wu M, Guruge JL, et al. 2011. Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *PNAS* 108(Suppl. 1):4599–606
26. Honda K, Littman DR. 2016. The microbiota in adaptive immune homeostasis and disease. *Nature* 535(7610):75–84
27. Hooper LV. 2004. Bacterial contributions to mammalian gut development. *Trends Microbiol.* 12(3):129–34
28. Hua X, Song L, Yu G, Goedert JJ, Abnet CC, et al. 2015. MicrobiomeGWAS: a tool for identifying host genetic variants associated with microbiome composition. bioRxiv 031187. **https://doi.org/10.1101/031187**
29. Huse SM, Ye Y, Zhou Y, Fodor AA. 2012. A core human microbiome as viewed through 16S rRNA sequence clusters. *PLOS ONE* 7(6):e34242
30. Igartua C, Davenport ER, Gilad Y, Nicolae DL, Pinto J, Ober C. 2017. Host genetic variation in mucosal immunity pathways influences the upper airway microbiome. *Microbiome* 5(1):16
31. Imhann F, Vich Vila A, Bonder MJ, Fu J, Gevers D, et al. 2016. Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut*. Oct. 8. **https://doi.org/10.1136/gutjnl-2016-312135**
32. Jankipersadsing SA, Hadizadeh F, Bonder MJ, Tigchelaar EF, Deelen P, et al. 2017. A GWAS meta-analysis suggests roles for xenobiotic metabolism and ion channel activity in the biology of stool frequency. *Gut* 66:756–58
33. Jones RJ, Megarrity RG. 1986. Successful transfer of DHP-degrading bacteria from Hawaiian goats to Australian ruminants to overcome the toxicity of *Leucaena*. *Aust. Vet. J.* 63(8):259–62
34. Khan MT, Browne WR, van Dijl JM, Harmsen HJM. 2012. How can *Faecalibacterium prausnitzii* employ riboflavin for extracellular electron transfer? *Antioxid. Redox Signal.* 17(10):1433–40
35. Knights D, Silverberg MS, Weersma RK, Gevers D, Dijkstra G, et al. 2014. Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med.* 6(12):107
36. Kostic AD, Xavier RJ, Gevers D. 2014. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 146(6):1489–99
37. Larsen N, Vogensen FK, van den Berg FWJ, Nielsen DS, Andreasen AS, et al. 2010. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLOS ONE* 5(2):e9085
38. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, et al. 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature* 500(7464):541–46
39. Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124(4):837–48
40. Lim MY, You HJ, Yoon HS, Kwon B, Lee JY, et al. 2017. The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. *Gut* 66:1031–38
41. Liu Y-J, Liu X-G, Wang L, Dina C, Yan H, et al. 2008. Genome-wide association scans identified *CTNNBL1* as a novel gene for obesity. *Hum. Mol. Genet.* 17(12):1803–13
42. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, et al. 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518(7538):197–206
43. Lynch J, Tang K, Sands J, Sands M, Tang E, et al. 2016. HOMINID: a framework for identifying associations between host genetic variation and microbiome composition. bioRxiv 081323. **https://doi.org/10.1101/081323**

44. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, et al. 2017. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* 35(1):81–89
45. Makishima M, Lu TT, Xie W, Whitfield GK, Domoto H, et al. 2002. Vitamin D receptor as an intestinal bile acid sensor. *Science* 296(5571):1313–16
46. Million M, Maraninchi M, Henry M, Armougom F, Richet H, et al. 2012. Obesity-associated gut microbiota is enriched in *Lactobacillus reuteri* and depleted in *Bifidobacterium animalis* and *Methanobrevibacter smithii*. *Int. J. Obes.* 36(6):817–25
47. Morotomi M, Nagai F, Watanabe Y. 2012. Description of *Christensenella minuta* gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order Clostridiales, and proposal of Christensenellaceae fam. nov. *Int. J. Syst. Evol. Microbiol.* 62(1):144–49
48. Mosca A, Leclerc M, Hugot JP. 2016. Gut microbiota diversity and human diseases: Should we reintroduce key predators in our ecosystem? *Front. Microbiol.* 7:455
49. Mueller UG, Sachs JL. 2015. Engineering microbiomes to improve plant and animal health. *Trends Microbiol.* 23(10):606–17
50. Ng MCY, Hester JM, Wing MR, Li J, Xu J, et al. 2012. Genome-wide association of BMI in African Americans. *Obesity* 20(3):622–27
51. Oki K, Toyama M, Banno T, Chonan O, Benno Y, Watanabe K. 2016. Comprehensive analysis of the fecal microbiota of healthy Japanese adults reveals a new bacterial lineage associated with a phenotype characterized by a high frequency of bowel movements and a lean body type. *BMC Microbiol.* 16(1):284
52. Olszak T, An D, Zeissig S, Vera MP, Richter J, et al. 2012. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* 336(6080):489–93
53. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, et al. 2015. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* 47(7):702–9
54. Ranciaro A, Campbell MC, Hirbo JB, Ko W-Y, Froment A, et al. 2014. Genetic origins of lactase persistence and the spread of pastoralism in Africa. *Am. J. Hum. Genet.* 94(4):496–510
55. Roager HM, Hansen LBS, Bahl MI, Frandsen HL, Carvalho V, et al. 2016. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat. Microbiol.* 1(9):16093
56. Schwiertz A, Taras D, Schäfer K, Beijer S, Bos NA, et al. 2010. Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* 18(1):190–95
57. Smith K, McCoy KD, Macpherson AJ. 2007. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin. Immunol.* 19(2):59–69
58. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, et al. 2009. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm. Bowel Dis.* 15(8):1183–89
59. Sudo N, Sawamura S, Tanaka K, Aiba Y, Kubo C, Koga Y. 1997. The requirement of intestinal bacterial flora for the development of an IgE production system fully susceptible to oral tolerance induction. *J. Immunol.* 159(4):1739–45
60. Tanno T, Fujiwara A, Sakaguchi K, Tanaka K, Takenaka S, Tsuyama S. 2007. Slit3 regulates cell motility through RAC/CDC42 activation in lipopolysaccharide-stimulated macrophages. *FEBS Lett.* 581(5):1022–26
61. Thakur K, Tomar SK, De S. 2016. Lactic acid bacteria as a cell factory for riboflavin production. *Microb. Biotechnol.* 9(4):441–51
62. Tigchelaar EF, Bonder MJ, Jankipersadsing SA, Fu J, Wijmenga C, Zhernakova A. 2016. Gut microbiota composition associated with stool consistency. *Gut* 65(3):540–42
63. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39(1):31–40
64. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, et al. 2009. A core gut microbiome in obese and lean twins. *Nature* 457(7228):480–84
65. Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, et al. 2016. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* 48(11):1413–17
66. Upadhyaya B, McCormack L, Fardin-Kia AR, Juenemann R, Nichenametla S, et al. 2016. Impact of dietary resistant starch type 4 on human gut microbiota and immunometabolic functions. *Sci. Rep.* 6:28797

67. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. 2016. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* 65(1):57–62
68. Wang J, Thingholm LB, Skiecevičienė J, Rausch P, Kummen M, et al. 2016. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat. Genet.* 48(11):1396–406
69. Xie H, Guo R, Zhong H, Feng Q, Lan Z, et al. 2016. Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst.* 3(6):572–84.e3
70. Xu L, Paterson AD, Turpin W, Xu W. 2015. Assessment and selection of competing models for zero-inflated microbiome data. *PLOS ONE* 10(7):e0129606
71. Yamamoto M, Matsumoto S. 2016. Gut microbiota and colorectal cancer. *Genes Environ.* 38:11
72. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, et al. 2012. Human gut microbiome viewed across age and geography. *Nature* 486(7402):222–27