Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Host genetics and microbiome associations through the lens of genome wide association studies

Omer Weissbrod[1,2], Daphna Rothschild[1,2], Elad Barkan[1,2] and Eran Segal[1,2]

Recent studies indicate that the gut microbiome is partially heritable, motivating the need to investigate microbiome–host genome associations via microbial genome-wide association studies (mGWAS). Existing mGWAS demonstrate that microbiome–host genotype associations are typically weak and are spread across multiple variants, similar to associations often observed in genome-wide association studies (GWAS) of complex traits. Here we reconsider mGWAS by viewing them through the lens of GWAS, and demonstrate that there are striking similarities between the challenges and pitfalls faced by the two study designs. We further advocate the mGWAS community to adopt three key lessons learned over the history of GWAS: firstly, adopting uniform data and reporting formats to facilitate replication and meta-analysis efforts; secondly, enforcing stringent statistical criteria to reduce the number of false positive findings; and thirdly, considering the microbiome and the host genome as distinct entities, rather than studying different taxa and single nucleotide polymorphism (SNPs) separately. Finally, we anticipate that mGWAS sample sizes will have to increase by orders of magnitude to reproducibly associate the host genome with the gut microbiome.

Addresses
[1] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 7610001, Israel
[2] Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 7610001, Israel

Corresponding author: Segal, Eran (eran.segal@weizmann.ac.il)

Current Opinion in Microbiology 2018, **44**:9–19

This review comes from a themed issue on **Microbiotia**

Edited by **Jeroen Raes**

https://doi.org/10.1016/j.mib.2018.05.003

Introduction

In recent years the importance of the gut microbiome in human metabolism and health is increasingly gaining recognition [1–7,8••,9,10,11•,12]. Recent studies have associated the microbiome with various health parameters including obesity, diabetes mellitus, cancer, and inflammatory, metabolic and neurodegenerative disorders [13–17].

A fundamental question is how strongly the microbiome is genetically inherited as opposed to being shaped by the environment. The microbiome evolves during childhood, and then becomes relatively stable and robust to perturbations [18–20]. This apparent host-adaptation evokes the classic question of 'nature versus nurture': Does the microbiome adapt to its host due to shared early environmental exposure, or are certain microbiome compositions inherently more suitable to specific host genomes?

The recent advent of 16S rRNA gene sequencing and metagenomic sequencing technologies enable carrying out gut microbiome studies with thousands of individuals [21]. Recent studies employing these technologies have uncovered evidence for both environmental and host genetic association with the microbiome composition [8••,11•,20,22–27,28••,29••,30••,31••,32,33•,34•,35•,36•,37•,38•]. However, to date there is no consensus regarding how and to what extent host genetics shape the gut microbiome, as compared to environmental factors.

In this article, we first review recent studies of environment and host genome associations with the human gut microbiome. We show that existing evidence suggests that the gut microbiome is predominantly shaped by environmental factors, and that host genotype–microbiome associations are weak, spread across multiple sites across the host genome, and together explain a relatively small fraction of the microbiome configuration of individuals. We then draw parallels between existing mGWAS and early GWAS, and use these to demonstrate how some of the pitfalls encountered in early GWAS, and their respective solutions, could be applied to mGWAS.

A short history of mGWAS

The microbiome is predominantly shaped by non-genetic factors

Recent studies have provided strong evidence that environmental factors play a much greater role than host genetics in shaping the gut microbiome. It can be difficult to tease apart environmental from genetic inheritance in humans, since children typically live with their parents. However, twin studies can tease these factors apart by comparing microbiome similarity among monozygotic (MZ) and dizygotic (DZ) twins, under the assumption

that significant differences in the degree of similarity are attributed solely to genetic effects. A recent large scale study of 416 twin pairs demonstrated that while several bacterial taxa are significantly heritable, the overall degree of similarity between MZ and DZ twins is only borderline significant ($P = 0.032$ under an unweighted UniFrac dissimilarity, $P > 0.05$ under Bray-Curtis and weighted UniFrac dissimilarity) [24]. In contrast, the same study showed that both MZ and DZ twins have significantly similar gut microbiomes compared with non-twin pairs ($P < 0.009$ under all similarity measures) [24]. A subsequent study showed that this similarity decreases when twins start living apart [33•]. These results indicate that environment overshadows host genetics in shaping the gut microbiome.

Recent non-twin studies provide additional support for the dominant role of environment in shaping the gut microbiome. First, there is an excessive bacterial similarity among individuals sharing a household, but no such similarity was observed across family members without a history of household sharing [8••,20,22,36•]. Second, over 20% of the variability of gut microbiome β-diversity (a measure of microbiome dissimilarity between pairs of individuals) can be inferred via several measured environmental factors, such as answers to food frequency and drug use questionnaires [8••,34•,35•], whereas no statistically significant result was obtained when applying a similar methodology to genetic variants [8••]. Third, several environmental factors have been robustly associated with both gut microbiome β-diversity and with individual taxa across multiple studies [34•,35•]. These results further demonstrate that the gut microbiome is predominantly shaped by environmental factors.

Twin studies identify significantly heritable gut microbiome taxa

Despite the strong role of non-genetic factors in shaping the gut microbiome, recent twin studies identified 33 significantly heritable bacterial taxa (most notably the family Christensenellaceae [37•]). The estimated heritability of these taxa was typically 10–30%, which is substantially lower than several well-known human complex traits, such as height, body mass index (BMI), and even education attainment [39]. A recent re-analysis of the largest reported twin study to date (2252 twins) found that the average heritability of gut bacterial taxa likely lies between 1.9% and 8.1% [8••]. Taken together, these results indicate that while there are several genetically heritable bacterial taxa, the overall gut microbiome heritability is relatively small.

Limited evidence for gut microbiome–host genotype associations from non-twins data

A potential shortcoming of twin studies is the difficulty of assembling large cohorts. Genotyping of unrelated individuals with a relatively common environment facilitates the assembly of much larger cohorts. These cohorts enable directly associating the gut microbiome with the host genotype, by searching for a greater co-presence of bacterial taxa among genetically closer individuals. However, the results from such studies have been inconclusive and mostly failed to replicate.

One of the first studies to employ the above approach identified a significant correlation between the top microbiome principal coordinate and top host-genome principal component (PC), based on human DNA residues extracted from stool samples [27]. An analysis of 127 Hutterites reported several heritable taxa [40], but the statistical significance of these results after multiple testing correction has not been reported. Additionally, several recent studies have identified a significant heritability of bacterial α-diversity (a measure of diversity of a bacterial community) [28••,30••,40]. In contrast, a recent analysis of 1046 Israeli individuals from different ancestral origins but a relatively shared environment did not replicate any of the above results, and did not identify statistically significant host-genomics associations with either the overall microbiome composition or individual taxa [8••]. Another recent study identified significant co-occurrence of bacterial taxa among 270 family members [28••], and several other studies identified a significantly different microbiome composition between individuals from different populations [20,32,41]. However, the interpretation of these results is unclear because unlike twin studies, it is not possible to tease apart the roles of genetics and environment in such studies [42]. Overall, these inconclusive results again suggest that the heritable component of the gut microbiome is small.

Limited power of microbiome genome wide association studies

Microbiome association studies attempt to not only identify heritable taxa, but also to pinpoint the host genetic variants that underlie this heritability [11•,37•,38•]. The first such studies in humans focused on specific genes and pathways, and have identified several significant microbiome-associated variants [43–47]. However, a potential shortcoming of the above studies is that they require previous knowledge of associated genes, and thus cannot discover new associations. Thus, recent studies have performed unbiased microbiome-genome wide association studies (mGWAS) spanning 93–1812 individuals [8••,27,28••,29••,30••,31••,32].

A substantial difficulty of mGWAS is the large number of tested hypotheses, which is equal to the number of genetic variants multiplied by the number of tested taxa, genes and pathways. This leads to a severe multiple testing correction and to reduced power (Figure 1). Consequently, most mGWAS findings are not statistically significant after multiple testing correction. A recent analysis demonstrated that there is almost no overlap between the loci reported in