

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

Omer Weissbrod^{1,2}, Daphna Rothschild^{1,2}, Elad Barkan^{1,2} and Eran Segal^{1,2*}

¹Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 7610001, Israel

²Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 7610001, Israel

*Correspondence: eran.segal@weizmann.ac.il

Abstract

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 genotypes 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: (a) Adopting uniform data and reporting formats to facilitate replication and meta-analysis efforts; (b) enforcing stringent statistical criteria to reduce the number of false positive findings; and (c) 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.

Introduction

In recent years the importance of the gut microbiome in human metabolism and health is increasingly gaining recognition [1–12]. 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 [8,11,20,22–37]. However, to date there is no consensus regarding how and to what extent does host genetics shapes 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.

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. Recent twin studies have demonstrated that the level of similarity is almost the same for MZ and DZ twins [24], that both MZ and DZ twins have extremely similar gut microbiomes compared with non-twin pairs [24], and that this similarity decreases when twins start living apart [32]. 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 [8,20,22,35]. Second, over 20% of gut microbiome β -diversity variance can be inferred via several measured environmental factors, such as answers to food frequency and drug use questionnaires [8,33,34], whereas no statistically significant result was obtained when applying a similar methodology to

genetic variants [8]. 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 [36]). 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 [38]. A recent re-analysis of the largest reported twin study to date (2,252 twins) found that the average heritability of gut bacterial taxa likely lies between 1.9% and 8% [8]. Taken together, these results indicate that there are several genetically heritable bacterial taxa, but that 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 (PCo) 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 [39], 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 [28,30,39]. In contrast, a recent analysis of 1,046 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 [20,40], but it is not possible to tease apart the roles of genetics and environment in such studies. 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 [11,36,37]. The first such studies in humans focused on specific genes and pathways, and have identified several significant microbiome-associated variants [41–45]. 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 93-1,812 individuals [8,27–31].

A substantial difficulty of mGWAS is the high 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 different studies, even when allowing SNPs up to 1Mb apart and associations with different bacterial taxa to be considered as overlapping [8]. This lack of consistency could originate either from differences in the underlying analysis methods or from lack of reproducibility, necessitating further investigation of the reported associations. The only genetic variants consistently shown to be microbiome-associated in multiple mGWAS are located in close proximity to the LCT gene, which is associated with lactase persistence [8,27,30,31,46]. However, while important, this association may be confounded by lactose consumption [46].

Several recent studies alleviated the multiple testing burden by testing for association with the entire microbiome composition rather than individual taxa, and identified genetic variants located in the vitamin D receptor gene and in several genes associated with health disorders [8,29,30]. A recent study further argued that a small number of genetic variants can infer over 10% of the microbiome β -diversity composition [29]. However, the results found in one study could not be replicated in others, with the exception of LCT related variants [8].

Other than the LCT variants, the most consistently reported host-microbiome associations involve immunity-related variants, although no two studies reported an association with the same variant (see ref. [37] and references therein for a comprehensive review). It has also been observed that many mGWAS hits are found near genes associated with complex diseases [24,27–29,31,39,41,47–50], and that multiple studies have implicated variants residing in the same genes, though the exact loci differed between studies [11,36,41].

The above results demonstrate that certain bacterial taxa are clearly heritable, but that the variants underpinning this heritability have not been reliably identified. This contradiction suggests that the heritability of bacterial taxa arises due to the aggregated effects of multiple genetic variants, each having an individually weak effect that cannot be reliably identified with existing sample sizes. This property has long been recognized as being common to most complex human traits, and has been extensively studied in GWAS, as elaborated below.

Contrasting microbiome and traditional genome-wide association studies

It is beneficial to reflect on the current state of mGWAS by drawing parallels with the history of GWAS [51–53]. The key idea behind GWAS is to associate genetic variants with traits of interest using large cohorts of unrelated individuals. Since 2005, over 3,200 GWAS with unique PubMed IDs have been reported in the GWAS Catalog [54], compared with 6 published mGWAS [8,27–31]. The initial motivation for GWAS arose due to the observation that common traits, such as height or BMI, are associated with a large number of genetic variants with small effect sizes, thus requiring large cohorts to be identified

reliably [51]. The small effect sizes reported in existing mGWAS suggests that the same pattern holds for host genome-microbiome associations.

The very first GWAS, which became possible thanks to the advent of low cost genotyping arrays, were met with high hopes. However, it soon became apparent that most reported associations failed to replicate [55]. The GWAS community consequently took actions to encourage reproducibility [56], chief among which was the adoption of stringent requirements for reporting associations. The same process seems to occur in current mGWAS, which also became possible due to the declining costs of the required technologies, and whose reported associations typically fail to replicate. Unfortunately, the number of hypotheses tested in a typical mGWAS is orders of magnitude larger than in a typical GWAS, suggesting that even more stringent statistical criteria need to be enforced.

A second important development in the history of GWAS was the adoption of common data formats, data processing techniques, analysis workflows and reporting guidelines [57]. These developments helped streamline, reduce the technical burden, and facilitate replication efforts in GWAS. Notably, the adoption of the common plink format [58] helped method developers to release software that could be used across different research groups in a unified manner. Future mGWAS would greatly benefit from such a standardization effort, as existing mGWAS were carried out using vastly different statistical methods, which hinders replication efforts. Some of the technical mGWAS aspects deserving additional investigation include the taxonomic levels of taxa to test, whether to also test for associations with the functional composition of the microbiome (e.g. bacterial genes), and statistical modeling of zero-inflation.

As more and more GWASs were being published, several common threads began to emerge. First, it became apparent that virtually all common traits are extremely polygenic, to the degree that most studied traits have to date been associated with dozens or hundreds of genetic variants with small effect sizes [54]. Second, most associated loci do not reside within coding regions, and there is often an excellent correlation between the number of associations on a chromosome and the chromosome length [59], suggesting that associated variants are spread uniformly throughout the genome. Third, pleiotropy was found to be extremely common, as a great number of loci were independently implicated with multiple traits. These observations suggested that the genetic architecture of common traits was far more complicated than initially thought, with some researchers hypothesizing that almost all genetic variants are associated with every trait [60]. As the gut microbiome is a highly complex organism [61], we believe that it is quite likely for the same patterns to emerge in mGWAS.

In response to the perceived complexity of common traits, GWAS gradually became larger and larger. While typical GWAS in 2007 spanned 3,000 individuals, many GWAS today are two orders of magnitude larger (**Figure 2**), with the recently released UK Biobank spanning approximately 500,000 individuals [62], and with plans underway to genotype 1,000,000 individuals for the Million Veteran Program [63]. These developments suggest that mGWAS sample sizes will similarly have to increase by at least two orders of magnitude to uncover the underlying biology behind gut microbiome and host genome interactions. In recent years many GWAS began releasing publicly available summary statistics that describe the association of each genetic variant to the studied trait, which enable combining results across multiple studies without the logistic and legal complications required to access private genetic data [64]. The mGWAS community would likely benefit from such data sharing ethics as well. Other potential approaches to increase power include oversampling of individuals with extreme microbiome-associated phenotypes, such as obesity [65,66], and restricting the analysis to taxa previously established to be heritable.

A second response of the GWAS community to the perceived complexity of common traits was phrasing of new research questions, which treat the genome in a global manner rather than a local one. Instead of trying to find genetic associations with tiny effects, many researchers began investigating the genetic architecture of common traits as a distinct entity [67,68]. This line of research arguably began with genotype-based heritability estimation [69], and has since opened the door to many new research questions, which have provided us with a new understanding of common traits. For example, recent developments enable us to study the degree of polygenicity of a given trait [70–72], the genetic similarity between pairs of traits [73–75], and the degree to which different traits are affected by various functional elements in the genome [76–80]. Such global approaches have arguably provided more insights into the underlying biology of common traits than direct genetic associations. We anticipate that similar global approaches could be carried over to mGWAS, as elaborated below.

Global versus Local Approaches

We encourage future mGWAS to adopt a holistic approach towards studying microbiome and host genome interactions. Specifically, we distinguish between local approaches, which consider the microbiome as a collection of taxa and the host genome as a collection of variants, and global approaches, which consider them as distinct entities (**Table 1**).

Global approaches are arguably more suitable for microbiome analyses because they can capture complex dynamics involving several taxa, in line with the view of the microbiome as a complex organism. In addition, global approaches can be substantially more powerful because they involve fewer tested hypotheses, leading to a less severe multiple testing correction, and because they aggregate multiple effects that may individually be too weak to be noticed. Finally, global approaches can jointly investigate the dynamics of the microbiome, the host genome, and additional factors, such as dietary habits.

Global approaches in statistical genetics are often carried out using linear mixed models [69,81–83], whereas in statistical ecology they are typically performed via ordination methods [84], Mantel tests [84] or multivariate analysis of variance [85]. To date, relatively little work has been done on combining the two frameworks together. One example is the recently proposed microbiome-association index, which measures the association of a host phenotype with the entire genomes of the microbiome and of the host in a single analysis [8]. We believe that this emerging field is a fertile ground for future developments.

Concluding remarks

The first mGWAS made many interesting discoveries, but have largely raised interesting questions rather than providing conclusive findings. It is our view that mGWAS would greatly benefit from adopting the lessons learned by the GWAS community over the several years. We specifically advocate adopting stringent statistical criteria, standard data formats, and a holistic approach towards studying microbiome and host genome interactions. Such approaches will require the development of new statistical methods, that will likely combine state of the art techniques from statistical genetics and statistical ecology. We anticipate that the combination of such approaches, along with larger sample sizes and with the integration of an increasing number of lifestyle and diet related factors, will lead to exciting new discoveries.

Acknowledgements

We thank the Segal group members for fruitful discussions. E.S. is supported by the Crown Human Genome Center; the Else Kroener Fresenius Foundation; Donald L. Schwarz, Sherman Oaks, CA; Jack N. Halpern, New York, NY; Leesa Steinberg, Canada; and grants funded by the European Research Council and the Israel Science Foundation. D.R. received a Levi Eshkol PhD Scholarship for Personalized Medicine by the Israeli Ministry of Science.

References

1. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: **An obesity-associated gut microbiome with increased capacity for energy harvest.** *Nature* 2006, **444**:1027–131.
2. Ivanov II, Frutos R de L, Manel N, Yoshinaga K, Rifkin DB, Sartor RB, Finlay BB, Littman DR: **Specific Microbiota Direct the Differentiation of IL-17-Producing T-Helper Cells in the Mucosa of the Small Intestine.** *Cell Host Microbe* 2008, **4**:337–349.
3. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, et al.: **Gut Microbiota from Twins Discordant for Obesity Modulate Metabolism in Mice.** *Science* 2013, **341**:1241214–1241214.
4. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, et al.: **Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis.** *Nat Med* 2013, **19**:576–585.
5. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, et al.: **Personalized Nutrition by Prediction of Glycemic Responses.** *Cell* 2015, **163**:1079–1094.
6. Hand TW, Vujkovic-Cvijin I, Ridaura VK, Belkaid Y: **Linking the Microbiota, Chronic Disease, and the Immune System.** *Trends Endocrinol Metab* 2016, **27**:831–843.
7. Budden KF, Gellatly SL, Wood DLA, Cooper MA, Morrison M, Hugenholtz P, Hansbro PM: **Emerging pathogenic links between microbiota and the gut–lung axis.** *Nat Rev Microbiol* 2016, **15**:55–63.
8. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, et al.: **Environment dominates over host genetics in shaping human gut microbiota.** *Nature* 2018.
9. Zmora N, Zeevi D, Korem T, Segal E, Elinav E: **Taking it Personally: Personalized Utilization of the Human Microbiome in Health and Disease.** *Cell Host Microbe* 2016, **19**:12–20.
10. Honda K, Littman DR: **The microbiota in adaptive immune homeostasis and disease.** *Nature* 2016, **535**:75–84.
11. Hall AB, Tolonen AC, Xavier RJ: **Human genetic variation and the gut microbiome in disease.** *Nat Rev Genet* 2017, **18**:690–699.
12. Gao R, Gao Z, Huang L, Qin H: **Gut microbiota and colorectal cancer.** *Eur J Clin Microbiol Infect Dis* 2017, **36**:757–769.
13. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI: **Human nutrition, the gut microbiome and the immune system.** *Nature* 2011, **474**:327–336.
14. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, et al.: **A core gut microbiome in obese and lean twins.** *Nature* 2009, **457**:480–484.

15. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, et al.: **A metagenome-wide association study of gut microbiota in type 2 diabetes.** *Nature* 2012, **490**:55–60.
16. Clemente JC, Ursell LK, Parfrey LW, Knight R: **The Impact of the Gut Microbiota on Human Health: An Integrative View.** *Cell* 2012, **148**:1258–1270.
17. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, et al.: **Richness of human gut microbiome correlates with metabolic markers.** *Nature* 2013, **500**:541–546.
18. Antonopoulos DA, Huse SM, Morrison HG, Schmidt TM, Sogin ML, Young VB: **Reproducible Community Dynamics of the Gastrointestinal Microbiota following Antibiotic Perturbation.** *Infect Immun* 2009, **77**:2367–2375.
19. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, et al.: **Moving pictures of the human microbiome.** *Genome Biol* 2011, **12**:R50.
20. Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, et al.: **Human gut microbiome viewed across age and geography.** *Nature* 2012, **486**:222.
21. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, Knight R, Ley RE: **Conducting a Microbiome Study.** *Cell* 2014, **158**:250–262.
22. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielnny S, et al.: **Cohabiting family members share microbiota with one another and with their dogs.** *eLife* 2013, **2**:e00458.
23. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, et al.: **Diet rapidly and reproducibly alters the human gut microbiome.** *Nature* 2013, **505**:559–563.
24. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, et al.: **Human Genetics Shape the Gut Microbiome.** *Cell* 2014, **159**:789–799.
25. Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI: **Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome.** *Cell Host Microbe* 2008, **3**:213–223.
26. Dey N, Wagner VE, Blanton LV, Cheng J, Fontana L, Haque R, Ahmed T, Gordon JI: **Regulators of Gut Motility Revealed by a Gnotobiotic Model of Diet-Microbiome Interactions Related to Travel.** *Cell* 2015, **163**:95–107.
27. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, Spector TD, Keinan A, Ley RE, Gevers D, et al.: **Host genetic variation impacts microbiome composition across human body sites.** *Genome Biol* 2015, **16**:191.
28. Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI, Guttman DS, Griffiths A, Panaccione R, Otley A, et al.: **Association of host genome with intestinal microbial composition in a large healthy cohort.** *Nat Genet* 2016, **48**:1413–1417.

29. Wang J, Thingholm LB, Skiecevičienė J, Rausch P, Kummen M, Hov JR, Degenhardt F, Heinsen F-A, Rühlemann MC, Szymczak S, et al.: **Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota.** *Nat Genet* 2016, **48**:1396–1406.
30. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, Spector TD, Bell JT, Clark AG, Ley RE: **Genetic Determinants of the Gut Microbiome in UK Twins.** *Cell Host Microbe* 2016, **19**:731–743.
31. Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, Deelen P, Vatanen T, Schirmer M, Smeekens SP, et al.: **The effect of host genetics on the gut microbiome.** *Nat Genet* 2016, **48**:1407–1412.
32. Xie H, Guo R, Zhong H, Feng Q, Lan Z, Qin B, Ward KJ, Jackson MA, Xia Y, Chen X, et al.: **Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome.** *Cell Syst* 2016, **3**:572–584.e3.
33. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, et al.: **Population-level analysis of gut microbiome variation.** *Science* 2016, **352**:560–564.
34. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, et al.: **Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity.** *Science* 2016, **352**:565–569.
35. Abeles SR, Jones MB, Santiago-Rodriguez TM, Ly M, Klitgord N, Yooseph S, Nelson KE, Pride DT: **Microbial diversity in individuals and their household contacts following typical antibiotic courses.** *Microbiome* 2016, **4**:39.
36. Goodrich JK, Davenport ER, Clark AG, Ley RE: **The Relationship Between the Human Genome and Microbiome Comes into View.** *Annu Rev Genet* 2017, **51**:413–433.
37. Kurilshikov A, Wijmenga C, Fu J, Zhernakova A: **Host Genetics and Gut Microbiome: Challenges and Perspectives.** *Trends Immunol* 2017, **38**:633–647.
38. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D: **Meta-analysis of the heritability of human traits based on fifty years of twin studies.** *Nat Genet* 2015, **47**:702–709.
39. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y: **Genome-Wide Association Studies of the Human Gut Microbiota.** *PLoS ONE* 2015, **10**:e0140301.
40. Zhao P, Irwin DM, Dong D: **Host genetics is associated with the gut microbial community membership rather than the structure.** *Mol Biosyst* 2016, **12**:1676–1686.
41. Knights D, Silverberg MS, Weersma RK, Gevers D, Dijkstra G, Huang H, Tyler AD, van Sommeren S, Imhann F, Stempak JM, et al.: **Complex host genetics influence the microbiome in inflammatory bowel disease.** *Genome Med* 2014, **6**:107.

42. Khachatryan ZA, Ktsoyan ZA, Manukyan GP, Kelly D, Ghazaryan KA, Aminov RI: **Predominant Role of Host Genetics in Controlling the Composition of Gut Microbiota.** *PLoS ONE* 2008, **3**:e3064.
43. Tong M, McHardy I, Ruegger P, Goudarzi M, Kashyap PC, Haritunians T, Li X, Graeber TG, Schwager E, Huttenhower C, et al.: **Reprogramming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism.** *ISME J* 2014, **8**:2193–2206.
44. Li D, Achkar J-P, Haritunians T, Jacobs JP, Hui KY, D'Amato M, Brand S, Radford-Smith G, Halfvarson J, Niess J-H, et al.: **A Pleiotropic Missense Variant in SLC39A8 Is Associated With Crohn's Disease and Human Gut Microbiome Composition.** *Gastroenterology* 2016, **151**:724–732.
45. Imhann F, Vich Vila A, Bonder MJ, Fu J, Gevers D, Visschedijk MC, Spekhorst LM, Alberts R, Franke L, van Dullemen HM, et al.: **Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease.** *Gut* 2018, **67**:108–119.
46. Goodrich JK, Davenport ER, Waters JL, Clark AG, Ley RE: **Cross-species comparisons of host genetic associations with the microbiome.** *Science* 2016, **352**:532–535.
47. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, Guiot Y, Derrien M, Muccioli GG, Delzenne NM, et al.: **Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity.** *Proc Natl Acad Sci* 2013, **110**:9066–9071.
48. Frank DN, Robertson CE, Hamm CM, Kpadeh Z, Zhang T, Chen H, Zhu W, Sartor RB, Boedeker EC, Harpaz N, et al.: **Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases.** *Inflamm Bowel Dis* 2011, **17**:179–184.
49. Ng MCY, Hester JM, Wing MR, Li J, Xu J, Hicks PJ, Roh BH, Lu L, Divers J, Langefeld CD, et al.: **Genome-Wide Association of BMI in African Americans.** *Obesity* 2012, **20**:622–627.
50. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al.: **Genetic studies of body mass index yield new insights for obesity biology.** *Nature* 2015, **518**:197–206.
51. Visscher PM, Brown MA, McCarthy MI, Yang J: **Five years of GWAS discovery.** *Am J Hum Genet* 2012, **90**:7–24.
52. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J: **10 years of GWAS discovery: biology, function, and translation.** *Am J Hum Genet* 2017, **101**:5–22.
53. Price AL, Spencer CCA, Donnelly P: **Progress and promise in understanding the genetic basis of common diseases.** *Proc R Soc B Biol Sci* 2015, **282**.
54. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, et al.: **The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.** *Nucleic Acids Res* 2014, **42**:D1001–D1006.
55. Kraft P, Zeggini E, Ioannidis JPA: **Replication in Genome-Wide Association Studies.** *Stat Sci* 2009, **24**:561–573.

56. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, et al.: **Replicating genotype–phenotype associations.** *Nature* 2007, **447**:655–660.
57. Bush WS, Moore JH: **Chapter 11: Genome-Wide Association Studies.** *PLoS Comput Biol* 2012, **8**:e1002822.
58. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al.: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**:559–75.
59. Shi H, Kichaev G, Pasaniuc B: **Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data.** *Am J Hum Genet* 2016, **99**:139–53.
60. Boyle EA, Li YI, Pritchard JK: **An Expanded View of Complex Traits: From Polygenic to Omnigenic.** *Cell* 2017, **169**:1177–1186.
61. Benson AK: **The gut microbiome—an emerging complex trait.** *Nat Genet* 2016, **48**:1301–1302.
62. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O’Connell J, et al.: **Genome-wide genetic data on 500,000 UK Biobank participants.** *bioRxiv* 2017,
63. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, et al.: **Million Veteran Program: A mega-biobank to study genetic influences on health and disease.** *J Clin Epidemiol* 2016, **70**:214–223.
64. Pasaniuc B, Price AL: **Dissecting the genetics of complex traits using summary association statistics.** *Nat Rev Genet* 2017, **18**:117–127.
65. Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D: **Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies.** *Genet Epidemiol* 2011, **35**:790–799.
66. Sham PC, Purcell SM: **Statistical power and significance testing in large-scale genetic studies.** *Nat Rev Genet* 2014, **15**:335–346.
67. Dudbridge F: **Polygenic Epidemiology.** *Genet Epidemiol* 2016, **40**:268–272.
68. Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ: **Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture.** *Proc R Soc B Biol Sci* 2016, **283**:20160569.
69. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, et al.: **Common SNPs explain a large proportion of the heritability for human height.** *Nat Genet* 2010, **42**:565–569.
70. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, Patterson N, Daly MJ, Price AL, Neale BM: **LD Score regression**

- distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015, **47**:291–295.
71. Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM: **Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model.** *PLoS Genet* 2015, **11**:e1004969.
 72. Zhou X, Carbonetto P, Stephens M: **Polygenic modeling with Bayesian sparse linear mixed models.** *PLoS Genet* 2013, **9**:e1003264.
 73. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, ReproGen C, Psychiatric Genomics C, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control C, Duncan L, et al.: **An atlas of genetic correlations across human diseases and traits.** *Nat Genet* 2015, **47**:1236–1241.
 74. Cross-Disorder Group of the Psychiatric Genomics Consortium: **Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs.** *Nat Genet* 2013, **45**:984–994.
 75. Loh P-R, Bhatia G, Gusev A, Finucane HK, Bulik-Sullivan BK, Pollack SJ, Schizophrenia Working Group of the Psychiatric Genomics C, de Candia TR, Lee SH, Wray NR, et al.: **Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis.** *Nat Genet* 2015, **47**:1385–1392.
 76. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K: **Partitioning heritability by functional annotation using genome-wide association summary statistics.** *Nat Genet* 2015, **47**:1228–1235.
 77. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjalmsdottir BJ, Xu H, Zang C, Ripke S, Bulik-Sullivan B, Stahl E, et al.: **Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases.** *Am J Hum Genet* 2014, **95**:535–552.
 78. Pickrell JK: **Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits.** *Am J Hum Genet* 2014, **94**:559–573.
 79. Calderon D, Bhaskar A, Knowles DA, Golan D, Raj T, Fu AQ, Pritchard JK: **Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression.** *Am J Hum Genet* 2017, **101**:686–699.
 80. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, Gilad Y, Pritchard JK: **RNA splicing is a primary link between genetic variation and disease.** *Science* 2016, **352**:600–604.
 81. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL: **Advantages and pitfalls in the application of mixed-model association methods.** *Nat Genet* 2014, **46**:100–106.
 82. Golan D, Rosset S: **Effective genetic-risk prediction using mixed models.** *Am J Hum Genet* 2014, **95**:383–393.
 83. Weissbrod O, Geiger D, Rosset S: **Multikernel linear mixed models for complex phenotype prediction.** *Genome Res* 2016, **26**:969–79.

84. Legendre P, Legendre L: *Numerical Ecology*. Elsevier; 2012.
85. Anderson MJ: **A new method for non-parametric multivariate analysis of variance**. *Austral Ecol* 2001, **26**:32–46.

Table 1: Overview of global and local analyses. Shown are common types of local and global analyses of the microbiome and of the host genome. Also shown is the approximate number of tests required for every type of analysis, under the assumption that there are one million host genetic variants and several hundred taxa.

		<i>Microbiome</i>	
		Local	Global
<i>Host Genome</i>	Local	<p>SNP-taxon association test: Associating a specific host SNP with a specific bacterial taxon.</p> <p>Requires hundreds of millions of tests</p>	<p>SNP-microbiome association tests: Associating a specific host SNP with microbiome β-diversity.</p> <p>Requires one million tests</p>
	Global	<p>Taxon heritability test: Estimating the fraction of variance of the abundance of a specific taxon inferred by the host genome</p> <p>Requires hundreds of tests</p>	<p>Microbiome-host association tests Measuring the correspondence between host-genome similarity and microbiome β-diversity (e.g. via a Mantel test);</p> <p>Microbiome-association index: Measuring the fraction of variance of a host phenotype that can be jointly inferred by the gut microbiome and host genome contents</p> <p>Requires a single test</p>

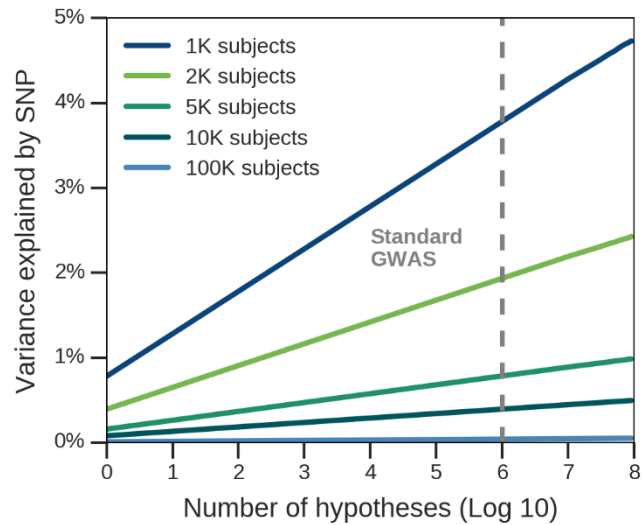


Figure 1: Plot of the fraction of variance a tested SNP needs to explain in order to be identified with 80% power, as a function of the number of tested hypotheses, for various sample sizes (based on standard derivations [52]). Increasing the number of hypotheses leads to reduced power to identify variants with small effect sizes, due to the severe multiple testing correction. Variants previously implicated in GWAS often explain less than 0.1% of the trait variance [52].

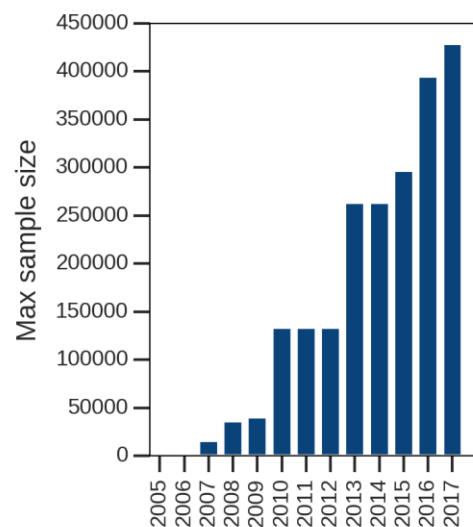


Figure 2: A bar chart depicting the largest GWAS performed up to every year between 2005-2017, as reported in the GWAS catalog [54]. GWAS sizes increased by almost 500-fold over 12 years.