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**

For a complete overview see the Issue and the Editorial

Available online 14th June 2018

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.

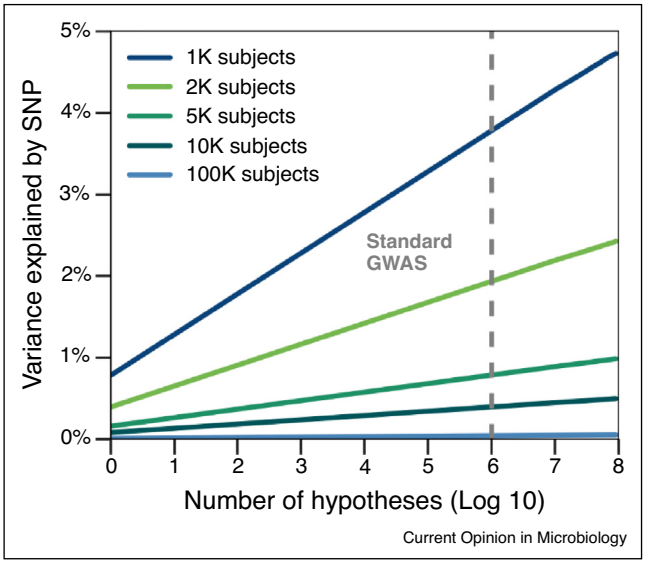
One of the first studies to employ the above approach identified a significant correlation between the top microbiome principal coordinate and top host-genome principal component (PC), based on human DNA residues extracted from stool samples [27]. An analysis of 127 Hutterites reported several heritable 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

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 [54•]). 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 [54•]. The fraction of variance a SNP needs to explain to be identified with 80% power in the presence of 10^6 hypotheses is 3.8% for $N = 1000$, 1.9% for $N = 2000$, 0.9% for $N = 5000$, 0.5% for $N = 10\,000$, and 0.1% for $N = 100{,}000$.

different studies, even when allowing SNPs up to 1 Mb 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••,32,48]. However, while important, this association may be confounded by lactose consumption [48].

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••,32]. 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 of 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. [38•] and references therein for a comprehensive review). It has also been observed that many mGWAS hits are found near host genes associated with complex diseases [24,27,28••,29••,31••,40,43,49–52], and that multiple studies have implicated variants residing in the same genes, though the exact loci differed between studies [11•,37•,43].

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.

A short history of GWAS

It is beneficial to reflect on the current state of mGWAS by drawing parallels with the history of GWAS [53,54•,55]. 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 [56], compared with seven published mGWAS [8••,27,28••,29••,30••,31••,32]. 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 [53]. 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 [57]. The GWAS community consequently took actions to encourage reproducibility [58], 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 [59]. These developments helped streamline, reduce the technical burden, and facilitate replication efforts in

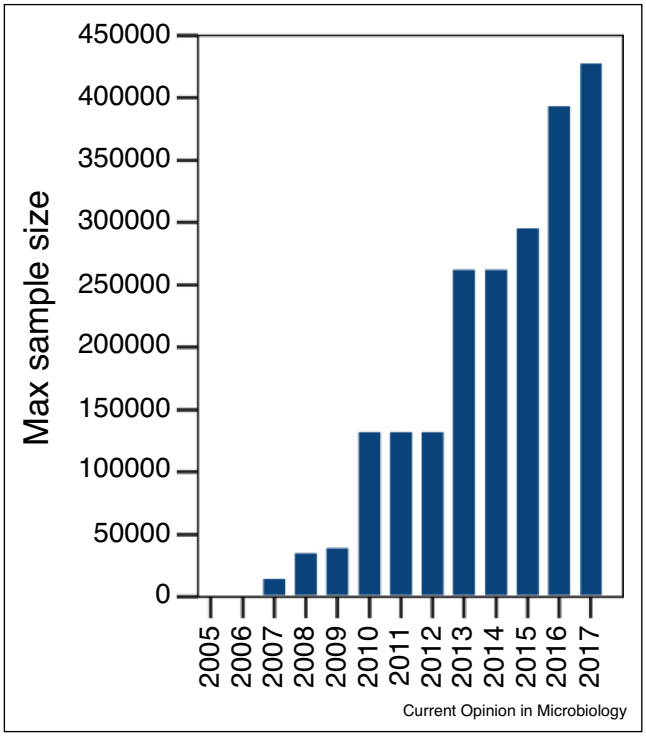
GWAS. Notably, the adoption of the common plink format [60] helped method developers to release software that could be used across different research groups in a unified manner. Future mGWAS would greatly benefit from such a standardization effort, as existing mGWAS were carried out using vastly different statistical methods, which hinders replication efforts.

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 [56]. 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 [61], 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 [62••]. As the gut microbiome can be seen as a highly complex organism [63], 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 [64], and with plans underway to genotype 1,000,00 individuals for the Million Veteran Program [65]. 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 of variant-trait associations, which enable combining results across multiple studies without the logistic and legal complications required to access private genetic data [66].

One class of GWAS that bears similarities to mGWAS are quantitative trait loci (QTL) studies. Such studies investigate the genetic determinants of molecular phenotypes such as gene expression [67] and DNA methylation [68]. QTL studies face several challenges similar to those of mGWAS: they investigate tens or hundreds of thousands of phenotypes, are relatively small compared to standard GWAS, and are often confounded by technical artifacts [69,70]. However, QTL studies often circumvent power limitations by only testing variants in close proximity to the molecular phenotype (cis associations) rather than across the entire genome (trans associations). To date, only a few trans-QTL studies have been published [71–78], and there was little overlap between the results reported in different studies [76]. Unfortunately, the strategy of testing only for cis associations cannot be carried over to mGWAS, because there is no analogue for '*cis*' and 'trans' in host genetics-microbiome associations.

Figure 2



A bar chart depicting the largest GWAS performed up to every year between 2005–2017, as reported in the GWAS catalog [56]. GWAS sizes increased by almost 500-fold over 12 years.

Applying the lessons of GWAS to mGWAS

As discussed above, two important developments in the history of GWAS were the enforcement of stringent statistical criteria for reporting associations, and the adoption of common data formats and analysis workflows. We strongly advocate that the mGWAS community adopt these practices to facilitate replication efforts.

Several approaches to increase power could also be carried from GWAS to mGWAS. First, publicly available summary association statistics will facilitate the aggregation of results across studies [66]. Second, power may be increased by oversampling of individuals with extreme microbiome-associated phenotypes, such as obesity [79,80]. Third, restricting the analysis to significantly heritable taxa can substantially increase power by decreasing the number of hypotheses. Finally, explicit

modeling of gene–environment interactions can increase power if environmental factors and variants have a non-additive effect on the abundance of bacterial taxa [81–85].

Approaches to increase power in *trans*-QTL studies are also likely to be beneficial to mGWAS, because the two study designs face similar challenges. Several recent papers proposed increasing power in *trans*-QTL studies by finding a small set of latent variables that can approximately represent all the molecular phenotypes in the data, and then testing for association with these latent variables [86–88]. This reduced representation can help reduce the number of tested hypotheses and to account for hidden confounding factors. A recent paper proposed increasing power by searching for variants associated with multiple molecular phenotypes [89]. A similar strategy can also be employed in mGWAS, by finding variants associated with microbiome dissimilarity instead of individual taxa, as discussed below.

Despite their similarities, there are some aspects of mGWAS with no analogue in standard GWAS. These include the multiple taxonomic levels that can be investigated, the possibility to investigate the functional composition of the microbiome (i.e. bacterial genes) rather than its taxonomic composition, and statistical modeling of zero-inflation. Previous mGWAS handled these aspects in different ways [8••,27,28••,29••,30••,31••,32], but a systematic comparison of the various approaches is still lacking.

Global versus local approaches

A common strategy often employed in GWAS to address power limitations is 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 recent studies investigate the genetic architecture of common traits as a distinct entity. Such global approaches have arguably provided more insights into the underlying biology of common traits than direct genetic associations. Below, we provide a brief overview of global approaches in GWAS, and then explain how such approaches can be carried over to mGWAS.

Global approaches in GWAS

Global approaches in GWAS typically use polygenic models, which assume that all genetic variants exert a causal effect on the phenotype [90,91]. This line of research arguably began with the seminal heritability estimation study of Yang et al., which quantified the overall association between all genotyped genetic variants and height in a cohort of 3,925 individuals [92]. The underlying idea is that individuals who are more genetically similar are likely to have more similar phenotypes. Heritability estimation provided the first principled demonstration that virtually all complex traits are highly polygenic.

Recent studies extended heritability estimation to investigate the polygenic contribution to heritability originating from different functional annotations, such as coding or conserved variants [93–97]. Such studies can provide valuable insights into the genetic architecture of diseases and traits. For example, a recent study demonstrated that variants expressed in the central nervous system contribute disproportionately to the heritability of smoking behavior, thus providing evidence that nicotine processing is a heritable trait [93].

Another extension of heritability is genetic correlation, which quantifies genetic similarity between pairs of traits [98–100]. A large genetic correlation indicates a shared biological mechanism for two traits. Recent genetic correlation studies have uncovered surprising biological similarities between seemingly unrelated diseases, such as a positive genetic correlation between anorexia nervosa and schizophrenia [98].

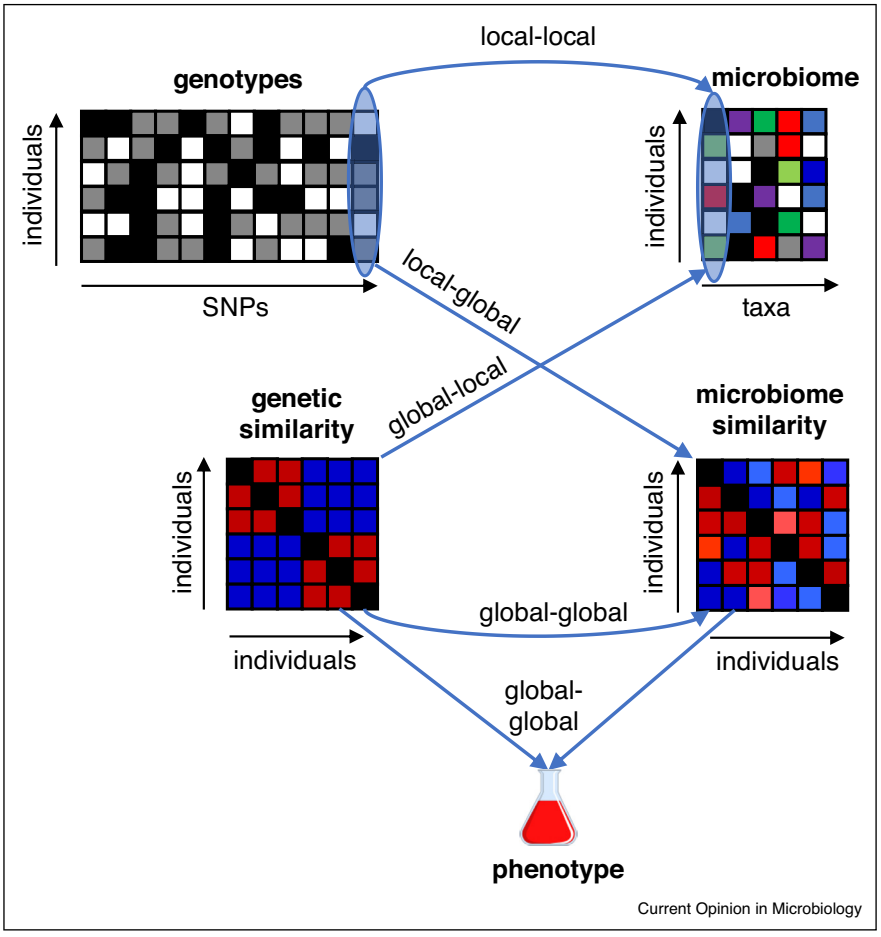
Finally, another global GWAS approach is construction of polygenic risk scores (PRS) [101–103]. PRS enable ranking individuals according to their predisposition to a certain disease, which can be used for early screening of individuals at high risk. In recent years PRS have also been used for finding unexpected genetic correlations. For example, a recent study demonstrated that PRS of schizophrenia and of bipolar disorder can predict creativity [104].

Global approaches in mGWAS

We encourage mGWAS to adopt global approaches, which treat the microbiome and the host genome as distinct entities, rather than local approaches which consider them as a collection of taxa and of variants (Figure 3 and Table 1). As discussed above, global approaches arguably led to a greater understanding of genetic diseases and traits than analysis of individual variants, and we expect that similar trends can hold for mGWAS. Global approaches are arguably more suitable for microbiome analysis 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 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.

To date, global association tests have been employed in mGWAS in a limited manner. One study tested if genetically similar individuals have similar abundances of an investigated taxon [40]. Several studies tested if individuals with similar microbiomes share similar alleles at one or several investigated variants, or if they share similar top

Figure 3



A demonstration of various types of mGWAS association tests. The top row shows matrices of genotypes (left) and of taxa abundances (right) for six individuals. The middle row shows the corresponding matrices of genetic similarity (left) and microbiome similarity (right). The bottom row shows a phenotype. The arrows represent types of tests; the text above each arrow states whether the test is global with respect to host genetics (first word) and to the microbiome (second word). Local genetic tests examine if a specific variant (a column of the genotypes matrix) is associated with some other factor, whereas global genetic tests examine if individuals who are genetically similar share some other property. Similarly, local microbiome tests examine if a specific taxon (a column of the microbiome matrix) is associated with some other factor, whereas global microbiome tests examine if individuals with similar microbiomes share some other property. There are two types of global–global tests: The first (top global–global arrow) tests if individuals who are genetically similar have similar micorbiomes; the second (bottom global–global arrow) tests if individuals with similar genomes and/or similar microbiomes are likely to share similar phenotypes. The genotypes matrix includes only three colors, corresponding to the number of SNP minor alleles that can be carried by an individual.

genetic PCs [8••,27,29••,30••,32]. Finally, several studies asked if genetically similar individuals have a similar microbiome diversity [8••,28••,30••,32,40] or similar microbiomes [8••]. However, existing mGWAS are still likely underpowered for such approaches due to limited sample sizes.

We expect global mGWAS analyses to be useful mainly for improved understanding of biological mechanisms involving both host genetics and the microbiome, such as fasting glucose levels, rather than identification of specific variant-taxa associations. This is in line with global GWAS analyses, which provide insights about the genetic architecture of traits rather than pinpointing specific associations. One example of a global analysis in mGWAS is the recently proposed microbiome-association index, which quantifies the association of a host phenotype with the entire genomes of the microbiome and of the host in a single analysis [8••]. We anticipate that this approach could be extended and applied in different study designs, such as microbiome–metabolome association studies [105].

Despite their appeal, global approaches are no panacea. The main limitation of global association tests is reduced interpretability and actionability, compared to local

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

		Microbiome	
		Local	Global
Host genome	**Local**	**SNP-taxon association test**: Associating a specific host SNP with a specific bacterial taxon. **Requires hundreds of millions of tests**	**SNP-microbiome association tests**: Associating a specific host SNP with microbiome β-diversity. **Requires one million tests**
	Global	**Taxon heritability test**: Estimating the fraction of variance of the abundance of a specific taxon inferred by the host genome **Requires hundreds of tests**	**Microbiome–host association tests** Measuring the correspondence between host-genome similarity and microbiome β-diversity (e.g. via a Mantel test); **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 **Requires a single test**

associations. Specifically, it may be difficult to gain biological insights from an association that involves the entire microbiome. The clinical utility of such an association may also be limited. However, these limitations may be an inherent biological property rather than a statistical limitation. Since the microbiome consists of multiple co-interacting taxa, there may simply not exist an individual taxon that can be acted upon for clinical applications. As an analogue, associations identified in disease GWAS are typically not directly actionable due to the complex genetic architecture of complex diseases. Despite these shortcomings, we believe that global microbiome associations are valuable for two reasons: such associations can shed light on the genetic architecture of microbiome-genetic associations; and the microbiome composition can be manipulated, which allows acting upon a global association by replacing the entire microbiome composition of an individual, without having to implicate individual taxa.

Global approaches in statistical genetics are often carried out using linear mixed models [92,106–108], whereas in statistical ecology they are typically performed via ordination methods [109], Mantel tests [109] or multivariate analysis of variance [110]. To date, relatively little work has been done on combining these two frameworks together. Notable examples include MiRKAT [111] and microbiome-association index estimation [8••]. Both methods test for microbiome association with a phenotype via a linear mixed model, whose covariance matrix is induced from a β-diversity dissimilarity measure. 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 last 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.

Conflict of interest

None.

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 and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as

- • of special interest
- •• of outstanding interest

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**:1027-1131.

2. Ivanov II, Frutos Rde 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.

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.

This study assembled a cohort of 1046 individuals with several distinct ancestries and a relatively common environment, and did not find significant associations between genetic ancestry or individual host SNPs and between gut microbiome composition or individual taxa. The study additionally demonstrated that household sharing and several environmental factors can infer a large fraction of the variance of microbiome β-diversity, whereas genetic variants cannot significantly infer this quantity. Finally, the authors proposed a measure called microbiome-association index for quantifying the association between the gut microbiome and a host phenotype after accounting for host genetics, and demonstrated that the gut microbiome can infer 22–36% of the variance of several anthropometric and metabolic related traits.

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.

A review of studies implicating host genetic variants with microbiome dysbiosis, which can increase risk for certain types of diseases.

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. Yatsunenko 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, Nakielny 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.

This study assembled two cohorts totaling 1561 healthy individuals with measured genotypes and 16S rRNA gene sequenced gut microbiomes, and identified several variants suggestively associated with specific taxa. This study additionally found significant over-sharing of almost one-third of bacterial taxa among close family members, but it is not clear if this is due to genetic or environmental inheritance.

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.

This study assembled two cohorts totaling 1812 healthy individuals with measured genotypes and 16S rRNA gene sequenced gut microbiomes, and identified several variants suggestively associated with specific taxa. The study additionally identified over 40 loci significantly associated with gut microbiome β-diversity, and argued that genetic and environmental factors can both infer approximately 10% of the variance of the gut microbiome β-diversity.

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.

This study, which follows up on a previous 2014 study, assembled a cohort of 1,126 twin pairs and demonstrated that at least 33 gut bacterial taxa are significantly heritable, with heritability levels typically ranging between 10% and 30%. The study also reported that the gut microbiome is suggestively associated with several genes involved with diet-sensing, metabolism and immune defense mechanisms.

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.

This study assembled a cohort of 1514 individuals with measured genotypes and metagenomic-sequenced gut microbiomes, and identified several variants suggestively associated with specific taxa, pathways or GO oncology terms, one of which has been replicated in additional studies.

32. Kolde R, Franzosa EA, Rahnavard G, Hall AB, Vlamakis H, Stevens C, Daly MJ, Xavier RJ, Huttenhower C: **Host genetic variation and its microbiome interactions within the Human Microbiome Project**. *Genome Med* 2018, **10**.

33. • 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.

This study assembled a cohort of 125 twin pairs with metagenome-sequenced gut microbiomes, and identified several significantly heritable bacterial taxa and functional modules. The study also found that the gut microbiomes of twins become increasingly dissimilar when living apart.

34. • 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.

This study assembled a cohort of 1106 individuals with 16S rRNA gene sequenced gut microbiomes, and identified several environmental factors significantly associated with the gut microbiome. Stool consistency was shown to be the factor most strongly associated with gut microbiome β-diversity, and medication explained the largest fraction of β-diversity variance out of all examined categories of environmental factors.

35. • 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.

This study assembled a cohort of 1135 metagenome-sequenced individuals and identified several environmental factors significantly associated with the gut microbiome, which jointly inferred over 18% of microbiome β-diversity.

36. • 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.

This study assembled a cohort of 56 genetically unrelated individuals, with 24 pairs of individuals sharing a household. The study found significant similarities between the microbiomes of cohabitating individuals, which could be identified even after intake of antibiotics.

37. • 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.

A review of the main discoveries made by twin and mGWAS studies, along with a discussion about the main challenges faced by present and future mGWAS.

38. • Kurilshikov A, Wijmenga C, Fu J, Zhernakova A: **Host genetics and gut microbiome: challenges and perspectives**. *Trends Immunol* 2017, **38**:633-647.

A thorough review of the history of mGWAS, its main findings and their implications.

39. 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.

40. 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.

41. 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.

42. Kendler KS, Neale MC: **"Familiality" or heritability**. *Arch Gen Psychiatry* 2009, **66**:452.

43. 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.

44. 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.

45. Tong M, McHardy I, Ruegger P, Goudarzi M, Kashyap PC, Haritunians T, Li X, Graeber TG, Schwager E, Huttenhower C *et al.*: **Reprograming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism**. *ISME J* 2014, **8**:2193-2206.

46. 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.

47. 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.

48. 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.

49. 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 U S A* 2013, **110**:9066-9071.

50. 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.

51. 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.

52. 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.

53. Visscher PM, Brown MA, McCarthy MI, Yang J: **Five years of GWAS discovery**. *Am J Hum Genet* 2012, **90**:7-24.

54. • 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.

An in-depth review of the history of GWAS and some examples of their implications.

55. 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**.

56. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L *et al.*: **The NHGRI GWAS Catalog, a curated resource of SNP-trait associations**. *Nucleic Acids Res* 2014, **42**:D1001-D1006.

57. Kraft P, Zeggini E, Ioannidis JPA: **replication in genome-wide association studies**. *Stat Sci* 2009, **24**:561-573.

58. 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.

59. Bush WS, Moore JH: **Chapter 11: genome-wide association studies**. *PLoS Comput Biol* 2012, **8**:e1002822.

60. 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**:559-575.

61. Shi H, Kichaev G, Pasaniuc B: **Contrasting the genetic architecture of 30 complex traits from summary association data**. *Am J Hum Genet* 2016, **99**:139-153.

62. •• Boyle EA, Li YI, Pritchard JK: **An expanded view of complex traits: from polygenic to omnigenic**. *Cell* 2017, **169**:1177-1186.

This paper makes a compelling argument that almost every genetic variant affects almost every trait, owing to properties of regulatory networks. Consequently, the authors advocate employing global approaches that consider the entire genome jointly, rather than associating specific variants with specific traits.

63. Benson AK: **The gut microbiome — an emerging complex trait**. *Nat Genet* 2016, **48**:1301-1302.

64. 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.

65. 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.

66. Pasaniuc B, Price AL: **Dissecting the genetics of complex traits using summary association statistics**. *Nat Rev Genet* 2017, **18**:117-127.

67. Nica AC, Dermitzakis ET: **Expression quantitative trait loci: present and future**. *Philos Trans R Soc B Biol Sci* 2013, **368**:20120362.

68. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, Bryois J, Giger T, Romano L, Planchon A *et al.*: **Passive and active DNA methylation and the interplay with genetic variation in gene regulation**. *eLife* 2013, **2**.

69. Williams RBH, Cotsapas CJ, Cowley MJ, Chan E, Nott DJ, Little PFR: **Normalization procedures and detection of linkage signal in genetical-genomics experiments**. *Nat Genet* 2006, **38**:855-856.

70. Fish AE, Capra JA, Bush WS: **Are interactions between *cis*-regulatory variants evidence for biological epistasis or statistical artifacts?** *Am J Hum Genet* 2016, **99**:817-830.

71. Small KS, Hedman ÅK, Grundberg E, Nica AC, Thorleifsson G, Kong A, Thorsteindottir U, Shin S-Y, Richards HB, Soranzo N *et al.*: **Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes**. *Nat Genet* 2011, **43**:561-564.

72. The Multiple Tissue Human Expression Resource (MuTHER) Consortium: Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, Bell JT, Yang T-P, Meduri E *et al.*: **Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins**. *Nat Genet* 2012, **44**:1084-1089.

73. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE *et al.*: **Systematic identification of *trans* eQTLs as putative drivers of known disease associations**. *Nat Genet* 2013, **45**:1238-1243.

74. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R *et al.*: **Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals**. *Genome Res* 2014, **24**:14-24.

75. Kirsten H, Al-Hasani H, Holdt L, Gross A, Beutner F, Krohn K, Horn K, Ahnert P, Burkhardt R, Reiche K *et al.*: **Dissecting the genetics of the human transcriptome identifies novel trait-related *trans*-eQTLs and corroborates the regulatory relevance of non-protein coding loci**. *Hum Mol Genet* 2015, **24**:4746-4763.

76. Joehanes R, Zhang X, Huan T, Yao C, Ying S, Nguyen QT, Demirkale CY, Feolo ML, Sharopova NR, Sturcke A *et al.*: **Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies**. *Genome Biol* 2017, **18**.

77. McRae A, Marioni RE, Shah S, Yang J, Powell JE, Harris SE, Gibson J, Henders AK, Bowdler L, Painter JN *et al.*: **Identification of 55,000 replicated DNA methylation QTL**. *bioRxiv* 2017 http://dx.doi.org/10.1101/166710.

78. Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, Feng J, Liu R, Diao L, Guo A-Y *et al.*: **PancanQTL: systematic identification of *cis*-eQTLs and *trans*-eQTLs in 33 cancer types**. *Nucleic Acids Res* 2018, **46**:D971-D976.

79. 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.

80. Sham PC, Purcell SM: **Statistical power and significance testing in large-scale genetic studies**. *Nat Rev Genet* 2014, **15**:335-346.

81. Hunter DJ: **Gene–environment interactions in human diseases**. *Nat Rev Genet* 2005, **6**:287-298.

82. Thomas D: **Gene–environment-wide association studies: emerging approaches**. *Nat Rev Genet* 2010, **11**:259-272.

83. Robinson MR, English G, Moser G, Lloyd-Jones LR, Triplett MA, Zhu Z, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H, Esko T *et al.*: **Genotype–covariate interaction effects and the heritability of adult body mass index**. *Nat Genet* 2017, **49**:1174-1181.

84. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å: **Gene–environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status**. *PLOS Genet* 2017, **13**: e1006977.

85. Moore R, Casale FP, Bonder MJ, Horta D, Franke L, Barroso I, Stegle O: **A linear mixed model approach to study multivariate gene–environment interactions**. *bioRxiv* 2018 http://dx.doi.org/10.1101/270611.

86. Stegle O, Parts L, Durbin R, Winn J: **A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL Studies**. *PLoS Comput Biol* 2010, **6**:e1000770.

87. Fusi N, Stegle O, Lawrence ND: **Joint modelling of confounding factors and prominent genetic regulators provides increased accuracy in genetical genomics studies**. *PLoS Comput Biol* 2012, **8**:e1002330.

88. Hore V, Viñuela A, Buil A, Knight J, McCarthy MI, Small K, Marchini J: **Tensor decomposition for multiple-tissue gene expression experiments**. *Nat Genet* 2016, **48**:1094-1100.

89. Brynedal B, Choi J, Raj T, Bjornson R, Stranger BE, Neale BM, Voight BF, Cotsapas C: **Large-scale *trans*-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation**. *Am J Hum Genet* 2017, **100**:581-591.

90. Dudbridge F: **Polygenic Epidemiology**. *Genet Epidemiol* 2016, **40**:268-272.

91. 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.

92. 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.

93. 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.

94. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjalmsson 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.

95. 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.

96. 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.

97. 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.

98. 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.

99. 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.

100. 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.

101. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh P-R, Bhatia G, Do R *et al.*: **modeling linkage disequilibrium increases accuracy of polygenic risk scores**. *Am J Hum Genet* 2015, **97**:576-592.

102. Abraham G, Inouye M: **Genomic risk prediction of complex human disease and its clinical application**. *Curr Opin Genet Dev* 2015, **33**:10-16.

103. Chatterjee N, Shi J, García-Closas M: **Developing and evaluating polygenic risk prediction models for stratified disease prevention**. *Nat Rev Genet* 2016, **17**:392-406.

104. Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM, Johannesson M, Galesloot TE, Hottenga JJ, Willemsen G *et al.*: **Polygenic risk scores for schizophrenia and bipolar disorder predict creativity**. *Nat Neurosci* 2015, **18**:953-955.

105. Pearson H: **Meet the human metabolome**. *Nature* 2007, **446**:8.

106. 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.

107. Golan D, Rosset S: **Effective genetic-risk prediction using mixed models**. *Am J Hum Genet* 2014, **95**:383-393.

108. Weissbrod O, Geiger D, Rosset S: **Multikernel linear mixed models for complex phenotype prediction**. *Genome Res* 2016, **26**:969-979.

109. Legendre P, Legendre L: *Numerical Ecology*. Elsevier; 2012.

110. Anderson MJ: **A new method for non-parametric multivariate analysis of variance**. *Austral Ecol* 2001, **26**:32-46.

111. Zhao N, Chen J, Carroll IM, Ringel-Kulka T, Epstein MP, Zhou H, Zhou JJ, Ringel Y, Li H, Wu MC: **Testing in microbiome-profiling studies with MiRKAT, the Microbiome Regression-Based Kernel Association Test**. *Am J Hum Genet* 2015, **96**:797-807.