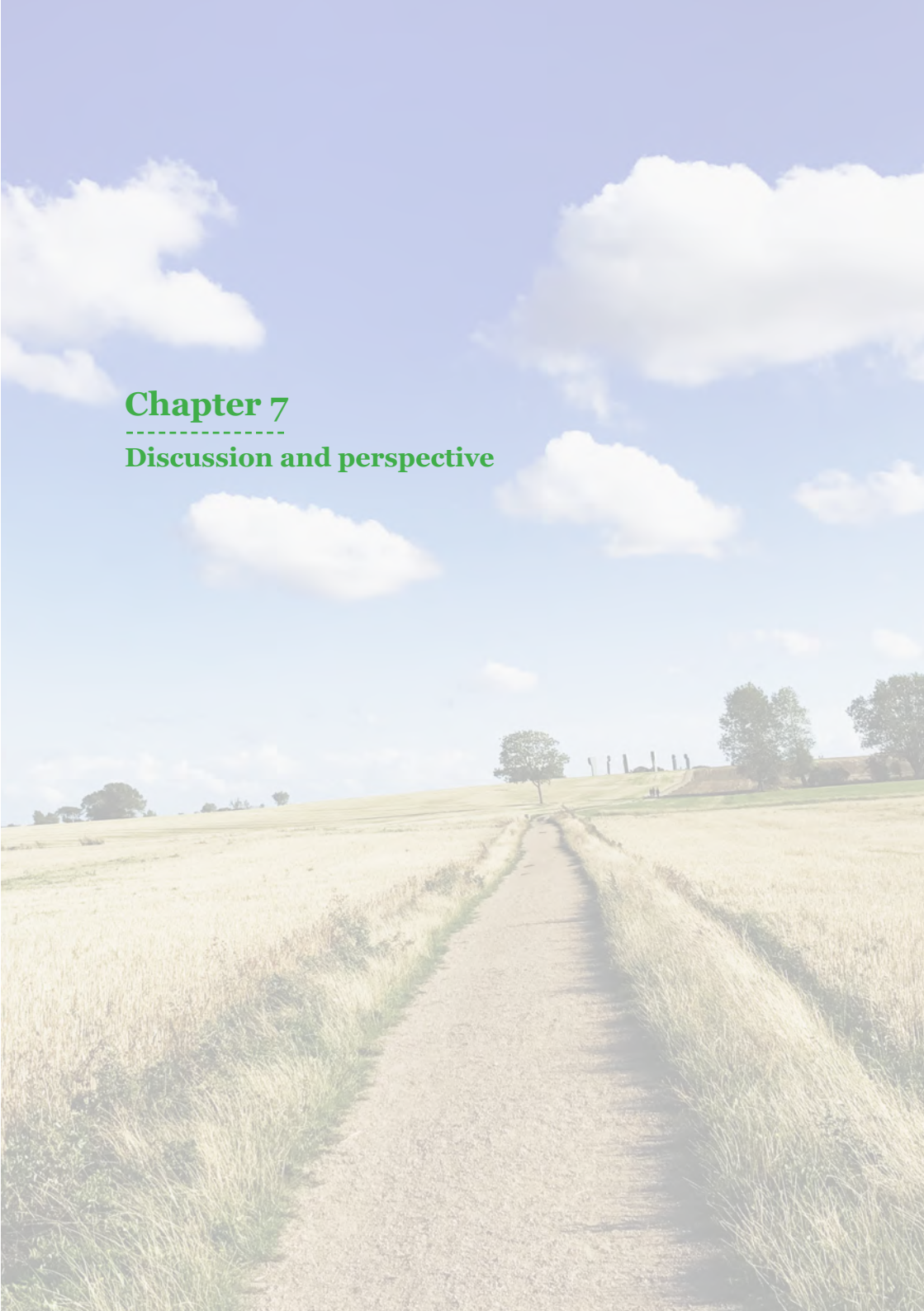
Environment-host-microbe interactions shape human metabolism

Chen, L; Fu, Jingyuan

DOI:
10.33612/diss.171839167



Chapter 7

Discussion and perspective

Discussion

The aetiologies of complex diseases such as cardiovascular disease (CVD), type 2 diabetes (T2D) and inflammatory bowel disease (IBD) are still not entirely resolved, but their initiation and development have traditionally been recognized as the consequence of a combination of genetics, environment and their interactions [1]. To date, numerous genome-wide association studies (GWAS) have linked genetic variants to the susceptibility of these diseases [2]. However, it has become increasingly clear that genetics can only explain a limited proportion of an individual's risk of developing a complex disease [3]. Thus, it is crucial to identify other risk factors and to understand their interactions with genetics. At the time of starting my PhD, gut microbiome research had just developed into a booming field of research and several large cohort-based studies had generated data that linked gut microbial composition to various complex diseases [4]. As a consequence, significant associations between gut microbial taxonomies and complex diseases were well-characterized among different cohorts [4]. Unlike the human genome, modification of gut microbial communities is feasible and ethical, making gut microbes promising therapeutic targets for disease prevention and treatment. But, before using microbiome adaptations for clinical application, we have to recognize that our knowledge of gut microbial functionality, and of its interactions with host genetics and environmental factors, is limited, which may hamper clinical translation of microbiome research.

In this thesis, I hypothesized that microbial dysbiosis, through its interactions with host genetics and environment, can dysregulate human metabolism and contribute to an individual's risk of developing complex disease. To test this hypothesis, I made use of various layers of "omics" datasets (metagenomics, metabolomics and genetics) in combination with the extensive phenotypic information that has been generated for a unique series of population-based prospective cohorts and patient cohorts, including Lifelines-DEEP, 500FG, 300OB and 1000IBD. Firstly, I linked the gut microbiome to host plasma metabolites generated by various platforms and showed that the gut microbiome can explain a substantial proportion of inter-individual metabolite variations. Secondly, we assessed genetics-microbiome-diet interactions in the control of plasma metabolite concentrations to reveal their role in metabolic dysregulation and their relevance to human health and disease. Furthermore, I applied statistical models to infer causal relationships between the gut microbiome and plasma metabolite concentrations. I was able to show that the gut microbiome may causally contribute to host phenotypes via the regulation of plasma metabolites. Additionally, I inferred microbial interactions through co-abundance analysis and characterized many IBD- and obesity-specific microbial interactions that pinpoint key microbial species and pathways in these diseases. Overall, the findings reported in this thesis have extended our understanding of the

role of environment-genetics-microbiome interactions in the development of complex disease, and this knowledge will ultimately contribute to better therapeutic treatment options for these complex diseases.

Understanding the functional role of the gut microbiome through metabolites

Over the last decade, microbiome-wide association studies (MWAS) have established numerous associations between gut microbiome composition and human complex diseases [5-7] such as obesity, CVD, IBD and T2D. MWAS typically generate a long list of commensals that are implicated as biomarkers or risk factors for disease but provide no clear explanation about the functional role of these gut microbes. Gut microbes perform a diverse set of metabolic functions, such as extracting energy from a wide array of indigestible carbohydrates, i.e. dietary fibres, to produce short-chain fatty acids (SCFAs) that are crucial for host energy and immune homeostasis. Dysbiosis of the gut microbiome may thus contribute to human health or to disease-onset through unfavourably altered metabolic activities [8]. Generating various types of metabolomic datasets and linking them to the gut microbiome may thus help us to extend our understanding of the importance of gut microbes in complex diseases, moving us from taxonomical association to potential functionality.

As a part of the initial efforts to understand the functional potential of the gut microbiome in human disease, we generated lipidomic profiles (231 metabolites) in the population-based Lifelines-DEEP (LLD) cohort and in an obesity cohort with 300 participants (300OB) and associated these metabolites to microbial species and pathways **(Chapter 2)**. In general, the gut microbiome could explain up to 11.1% of the variation in plasma lipids in LLD and 16.4% of the variation in 300OB, after correcting for age, gender and BMI. Microbial associations were detected for 210 metabolites in LLD at FDR<0.05, with 64 associations to 12 species and 4,135 associations to 308 pathways. In the 300OB cohort, microbial associations were detected for 42 metabolites, with one association to species and 105 associations to 19 pathways. The smaller number of associations in 300OB was likely due to its smaller sample size, but our data also showed that the microbial factors identified in LLD generally had lower predictive value for metabolic variation in 300OB. This indicates that there are some genuine differences in microbial associations between the population-based and obese cohorts. By comparing association strength and directions for the top associations from both cohorts, we observed some obesity-specific associations, particularly for the relative composition of lipoprotein subclasses. For example, we found significant associations of *Ruminococcus* species *sp_5_1_39BFAA* to the phospholipid content of very large VLDL particles (XXL-VLDL_PL_percent)

in 300OB that were completely absent in LLD, even with its much larger sample size. Interestingly, microbial factors that were found to be associated with plasma metabolites were also related to end-point phenotypes. For instance, *Ruminococcus sp_5_1_39BFAA* was positively associated to hepatic fat content, indicating a relevance for non-alcoholic fatty liver disease (NAFLD).

To further understand how gut microbes regulate host lipid metabolism, we profiled plasma and faecal bile acids (BAs) in the 300OB cohort. BAs are a category of steroids that are implicated in the aetiology of CVD-related conditions such as dyslipidaemia by affecting lipid absorption and activating the nuclear Farnesoid X Receptor (*FXR*) or the membrane-bound Takeda G protein-coupled bile acid receptor 1 (*TGR5*) [9]. Primary BAs are synthesized in the liver from cholesterol and then further modified by the gut microbiome to form secondary BAs. These differently structured primary and secondary BAs have a highly variable potency to activate *FXR* and *TGR5* and thus variable functionalities in control of lipid metabolism [10]. Thus, if we wish to understand the contribution of endogenous BAs to obesity-related disease development, rationalize their potential use as biomarkers and personalize future therapeutic applications of *FXR* and *TGR5* modulators, it is important to have insight into the determinants of human BA pool size, composition and turnover. Furthermore, the rate at which BAs are formed in the liver, as reflected by plasma C4 concentrations, is the major determinant of cholesterol turnover in the human body and contributes to control of plasma lipoprotein levels. However, it is still not clear which gut microbial species actually determine BA metabolism in humans. By checking associations between microbial factors and BA entities, we observed 439 BA-microbial associations for 45 BA entities at FDR < 0.05 **(Chapter 3)**. In particular, 44 BA entities associated with 61 bacterial species and 30 BA entities associated with 112 bacterial pathways. The microbial associations with BA entities that we identified not only confirmed previous microbial BA determinants [11-13], they also included novel associations, such as those with *Ruminococcus sp_5_1_39BFAA* and *Faecalibacterium prausnitzii*, that may reflect conversion of BAs by the bacterium or, conversely, modulation of the bacterium by the actions of (bacteriostatic) BAs. Intriguingly, we observed that *Ruminococcus sp_5_1_39BFAA*, a species expressing the gene encoding choloylglycine hydrolase that catalyses BA deconjugation, not only associated with bacterial BA entities but also with hepatic fat content [14]. These results support the hypothesis that *Ruminococcus sp_5_1_39BFAA* might regulate lipid metabolism through intestinal BAs and thereby contribute to NAFLD development. However, our study only focused on the most prominent BA species. It would be interesting to also assess the determinants of much less abundant secondary BA species, such as the LCA metabolites 3-oxo-LCA and isoallo-LCA that were recently shown to control T cell differentiation in the

colonic lamina propria, and hence modulate adaptive immunity [15].

In addition to microbial impacts on lipid and BA metabolism, I have also examined how the gut microbiome influences broad categories of plasma metabolites (1183 in total) generated by an un-targeted metabolomics approach employing LC-MS-MS on plasma samples from LLD. Using the association-based approach, we explored the associations between metabolites and microbial composition, metabolic pathways and structural variants (SVs) **(Chapter 4)**. Interestingly, many metabolites that associated with microbial species and pathways were already known to be gut microbe-related based on their annotations [16]. For instance, I observed 919 associations with 25 uremic toxins, 142 associations with thiamine (vitamin B1) and 117 associations with 5 phytoestrogens. Uremic toxins and thiamine derived from gut microbes have been shown to be related to various diseases, including chronic kidney disease and CVD [17]. Phytoestrogens are a class of plant-derived polyphenolic compounds that can be transformed by gut microbiota into metabolites that promote the host's metabolism and immune system [18]. More importantly, established associations also connect the genetically encoded function of microbes with metabolites to provide putative mechanistic information underlying the functional output of the gut microbiome. For instance, we observed that microbial uremic toxin biosynthesis pathways, including the glycine-cleavage pathway (in species *Olsenella sp* and *Clostridium sp*) and the hydroxybenzoate to phenol pathway (in *Clostridium sp*) are responsible for hippuric acid and phenol sulphate biosynthesis, are associated with hippuric acid and phenol sulphate levels measured in plasma, respectively. Importantly, the associations we observed between unannotated SVs and metabolites may potentially reveal the functional capacity of microbial genetic makeup. Importantly, these cross-sectional associations between microbiome and metabolites could be further confirmed by making use of the longitudinal study design of the Lifelines cohort, as we see in the consistent associations we observed between microbial composition changes and changes in metabolite levels in samples from the same individuals taken 4-years apart **(Chapter 5)**. This indicates that the within-individual gut microbial changes are relevant for metabolic changes that relate to host metabolic health status.

The established microbial associations to plasma metabolite concentrations have greatly improved our understanding of the functional potential of gut microbes. However, challenges remain on both the metabolomics and metagenomics sides of the equation that prohibit us from further exploring the functionality of gut microbes. While "traditional targeted metabolomics" can provide accurate identification and quantification for individual (classes of) metabolites, its low throughput and relatively high costs make it less suitable for application in large cohort studies. Untargeted metabolomics by innovative LC-MS-MS approaches can profile thousands of metabolites

after a single injection, but metabolite annotation and quantification remain a major bottleneck in untargeted metabolomics. Although community guidelines for metabolite identification were published over a decade ago, adoption of the recommended standards has been limited [19]. Developing targeted extraction/identification protocols for particular metabolite classes might be a promising approach to resolve these issues. In addition, the gap in unknown metabolites might be closing as more enzymatic functions become understood. Integrating the knowledge of metabolic reactions should promote the development of more powerful identification tools.

For metagenomics, similar issues also remain, as only around 60% of microbiome genomes can be annotated [20]. As a consequence, nearly half of the microbial functionalities are still a mystery. The central principle in the current microbial genomic annotation pipelines is based on the sequence similarity with existing databases, such as UniProt [21], Pfam [22] and SMART [23]. However, we have to bear in mind that similarity in sequence does not necessarily mean similarity in functionality. A prevailing belief across modern molecular biology is that a gene sequence will define the structure of the gene product and that this structure, in turn, will designate a unique function [24]. In other words, even with 99% similarity between the sequences of two genes, their functionalities may still be totally different due to the structure differences caused by the remaining 1% difference. Thus, predicting the functionality of microbial genes based on the end product structure, e.g. protein structure, can be a promising approach. Additionally, orthologs, paralogs and xenologs have always been ignored in sequence similarity-based annotation, as was recently intensively discussed [25]. For example, in most members of the *Crenarchaeota*, the family B DNA polymerases are represented by several paralogs that form distinct orthologous families within this archaeal phylum. In contrast, most bacteria that possess the *polB* gene have a single copy, which is co-orthologous to all archaeal *polB* genes. Thus, archaea and bacteria share only one orthologous family of *polB* [25]. Such complex relationships among homologous genes confound the analysis of clusters of orthologous gene because the definition of orthology becomes mutually dependent with the phyletic patterns. Apart from the challenges in microbial genomic annotation described above, another challenge relates to gene single-nucleotide variations (SNVs). Genetic polymorphisms rapidly arise through *de novo* mutations, and SNVs can regulate the expression of genes. As a consequence, the functionality of a given microbial gene may vary between individuals due to different SNVs in the gene. Thus, profiling microbial SNVs and further delineating their roles in regulating the downstream activities is also essential for understanding microbial functionality.

To sum up, understanding the functionality of gut microbes is emerging as a key point to reveal the microbial contributions to human health and disease.

While the study of microbiome-metabolite associations had pinpointed potential functionalities of microbes, in my opinion, the in-depth evaluation of microbial genomes, genes and SNVs will be the inflection point in the field of human metabolic health.

Genetics-microbiome interactions in metabolism

Despite the fact that the gut microbiome acts as determinant of various disease-relevant metabolites, host genetics are also involved in the regulation of some of these metabolites, such as BAs and amino acids. Thus, understanding the impact of genetics-microbiome interactions in various categories of circulating plasma metabolites will provide deeper insights into host-microbe interactions in health and disease.

For the 1183 plasma metabolites that were measured in LLD, we observed that microbial composition contributed 0.6% to 26% of the variance in 198 metabolites, while genetic variants contributed 3% to 28% of the variance in 62 metabolites **(Chapter 4)**. Inter-individual variations in 85 metabolites were dominantly explained by the gut microbiome and those of 56 metabolites by genetics. For instance, out of the 85 metabolites whose variation was dominantly explained by the gut microbiome, 27 were defined to be microbiome-related metabolites. Out of the 56 metabolites whose inter-individual variations were dominantly driven by genetic variants, 18 were lipids and 12 were amino acids. In addition, there were 7 metabolites that were associated with both genetics and microbiome. For example, genetics and microbiome explained 6% and 5%, respectively, of the variation in plasma 5'-carboxy-gamma-chromanol, a dehydrogenated carboxylate product of 5'-hydroxy-r-tocopherol [26] that has been related to cancer and cardiovascular risk [27]. We further investigated the association between metabolite loci and the gut microbiome and observed that genetics and microbiome likely play additive roles in shaping plasma metabolite patterns **(Chapter 4)**. We also observed a similar phenomenon for BAs **(Chapter 3)**, and a recent study assessed the prediction power of genetics and gut microbiome for the plasma metabolome [28]. These observations indicate that genetics-microbiome interactions are largely acting in the form of the "additive model" proposed in the Introduction **(Chapter 1)**. In the additive model, the gut microbiome exerts an additive effect on the regulation of metabolites, i.e. acting in addition to the known genetic and environmental contributions. The gut microbiome can explain extra inter-individual variation of a trait, suggesting that microbiome-targeting approaches may have a better control on a complex trait on top of other approaches of modulating genetic and environment factors.

On the other hand, it has been generally recognized that genetics-

microbiome interactions established by the association-based approach are always weak, rarely surpassing the statistical threshold, and are poorly replicated in different cohorts [29]. Increasing the study sample size and standardizing bioinformatic pipelines have been proposed as promising approaches to reveal more *in silico* genetics-microbiome interactions for biological hypothesis generation. However, recent studies indicate that even with 10 times larger samples sizes (up to 20,000 subjects) and standardized protocols, the signals observed remain limited [30-32]. It is likely that the effects of exogenous and environmental factors on the gut microbiome may have masked the potential effects of genetic variants. It has been estimated that established environmental factors explain 10-20% of microbiome variance, whereas the effect of genetics is identified as approximately 10% [33]. It is not yet clear what explains the remaining microbiome variance. The effect of exogenous and environmental factors is likely to be underestimated because not all potential confounding factors have been investigated, even though control of such factors may substantially increase detection power of genetics-microbiome interactions.

Although there are other challenges that remain in decoding the role of genetics-microbiome interactions, I anticipate that intensive collection of data on environmental factors in large cohort studies in the near future will bring new insights into the roles of genetics-microbiome interactions in control of metabolism. The biological significance of these *in silico* interactions in regulating host metabolism will, however, still need functional validation, and setting up animal models or other state-of-the-art systems with well-controlled environments will help to reveal the actual modes of action.

Diet-microbiome interactions in metabolism

Complex diseases such as CVD, obesity and T2D are intimately linked to diet, and increasing evidence now points to the gut microbiome as a mediator of the dietary impact on host metabolism [34]. Numerous studies have shown that diet has a pronounced effect on the composition and function of gut microbes. In one example of this effect, *Ruminococcus bromii* species bloomed in a majority of obese men in response to a resistant-starch intervention, while the lack of response in the other individuals might reflect an absence of this species [35].

Gut microbes interact with dietary components and consequently interfere with the regulation of plasma metabolite levels in the host, which may contribute to development of metabolic diseases. For instance, cheese, seafood, eggs and red meat are abundant sources of L-carnitine and phosphatidylcholine. Microbes have choline-TMA lyase (*CutC*), carnitine

oxygenase (*CntA*) and betaine reductase (*GrdH*) in the gut can convert these compounds into trimethylamine (TMA). Once it has been absorbed from the gut into the bloodstream, TMA is transported to the liver, where it is enzymatically oxidized to TMA N-oxide (TMAO), a compound that has been associated with poor cardiovascular outcomes in humans [36]. Other well-studied metabolites with relevance for human diseases include SCFAs and BAs.

Associations between dietary habits and gut microbial composition were established in a previous study from our laboratory [37]. By further linking thousands of metabolites that cover a wide range of lipids, organic acids, phenylpropanoids and benzenoids to dietary factors collected from questionnaire in LLD and to the gut microbiome **(Chapter 4)**, we observed that dietary habits contributed from 0.4% to 34% of the variance in 690 metabolites. Further analysis showed that inter-individual variations in 156 these metabolites could be significantly explained by the gut microbiome, suggesting the importance of diet-microbiome interactions in regulating these metabolites. For example, diet and microbiome together could explain 12% of the variation in hippuric acid, a CVD- and kidney disease-related uremic toxin that originates from bacterial conversion of dietary proteins [38]. Other bioactive metabolites that underlie diet-microbiome interactions include anthocyanins, phytoestrogens, vitamins and amino acids.

We further applied bi-directional mediation analysis to evaluate mediation effects of microbiome and metabolites for diet. In total, we established 195 mediation linkages: 185 for dietary impact on microbiome through metabolites and 10 for dietary impact on metabolites through microbiome. Most of these linkages were related to the impact of coffee and alcohol consumption on microbial metabolic functionalities. Coffee contains various hydroxycinnamic acids, including ferulic acid, which can be catabolized by gut microbes and may play a beneficial role in alleviating features of diabetes [39,40]. For example, we observed that ferulic acid can mediate the impact of coffee on a variable structural variant (vSV) of *Ruminococcus sp* that encodes an ATPase component that can be activated by ferulic acid [41]. We also observed that hulupinic acid, which is commonly detected in alcoholic drinks, can mediate the impact of beer consumption on the *Clostridium methylpentosum* ferredoxin-NAD:oxidoreductase (Rnf) complex, an important membrane protein in driving ATP synthesis that is essential for all bacterial metabolic activities [42]. With respect to dietary impacts on metabolites through microbiomes, a *Eubacterium hallii* vSV provides an interesting example. This *Eubacterium hallii* vSV encodes an ATPase that is responsible for transmembrane transport of various substrates [43] that mediate the effect of white wine consumption on plasma levels of pipecolic acid. Pipecolic acid present in human plasma is mainly derived from the catabolism of dietary lysine by intestinal bacteria, rather than by direct

intake from diet [44], and has been evaluated in chronic liver disease [45]. Notably, lysine is one of the most abundant amino acids in white wine [46]. Taken together, our data provide functional insights into diet-microbiome and diet-metabolite interactions, and all these examples could be harnessed for personalized nutrition strategies to improve human metabolic health through gut microbes.

However, before utilizing diet-microbiome interactions to improve human metabolic health, we have to note that established interactions in cohort-based studies may be biased by the facts that people are poor at adhering to dietary regimes and that it is difficult to accurately measure the extent of their adherence because the self-assessment of food intake can be confounded by numerous factors. Thus, it is hard to distinguish which dietary component(s) was responsible for the variation in the microbiota. A further complication is that many of the diets also have the potential to directly influence host metabolism in a microbiome-independent way.

To summarize, the gut microbiome is increasingly recognized as an important factor in the field of personalized nutrition. The capacity of microbes to metabolize compounds present in food has an impact on host metabolic health. While the study of compositional signatures of the gut microbiome has identified interesting links between microbes and diet, in my opinion, the in-depth evaluation of microbial metabolic capacities will be key for the development of personalized nutrition.

Causal role of the gut microbiome in host metabolic health

Dysbiosis is associated with metabolic diseases in humans that range from obesity to T2D and CVD. Causality has been demonstrated in animal models by applying faecal microbiota transplantation (FMT) strategies. Moving forward, it is essential to investigate whether the human gut microbiome can causally contribute to host metabolic health. This requires large longitudinal cohorts with sufficient numbers of participants and sufficient duration to actually change their phenotypes and high-resolution monitoring of host and microbial parameters to determine the progression of derangements.

I have examined whether alterations of the gut microbiome are related to changes in host health status in the 338 participants of LLD cohort for whom 4-year follow-up data became available **(Chapter 5)**. We identified 190 associations involving 169 microbial features and 34 phenotypes. These included 84 associations with species and pathway abundances and 106 associations with microbial SVs. BMI, blood pressure, glycated haemoglobin (HbA1c) and depression were the phenotypic factors with the most associations. For instance, we observed a positive association between systolic blood pressure and the abundance of *Lachnospiraceae bacterium* and

a negative association between HbA1c and the flavin biosynthesis pathway. In addition, temporal changes in microbial SVs were associated with host immune phenotypes. For instance, a vSV in *Blautia obeum* that contains the virulence protein E and chloramphenicol resistance genes was negatively associated with a change in blood lymphocyte cell counts. For disease onset, we detected 22 associations, namely 15 associations for depression, 3 associations for irritable bowel syndrome and 3 for asthma. The top association for depression onset was found for a deletion structural variant (dSV) region in *Collinsella sp* that encodes the histidine kinase A.

To further evaluate whether metabolites can mediate the microbial impact on host phenotypes, we applied bi-directional mediation analysis, which revealed 21 mediation linkages **(Chapter 5)**. Most of these linkages were related to microbial impact on blood pressure via thiamine (9 linkages) and acetyl-N-formyl-5-methoxykynurenamine (AFMK, 9 linkages). The impact of thiamine on cardiometabolic health has been well-documented and was confirmed in a randomized controlled trial that showed that thiamine can reduce diastolic blood pressure [47]. AFMK is a degradation product of melatonin that contributes to blood pressure reduction by inhibiting the synthesis of prostaglandins [48,49]. Our mediation analysis suggested that various bacterial pathways may contribute to these effects. For instance, the microbial sulphate reduction pathway may contribute to a decrease in diastolic blood pressure by increasing plasma thiamine levels and bacterial lipopolysaccharide biosynthesis may contribute to a decrease in systolic blood pressure by affecting plasma levels of AFMK. Metabolic products of the bacterial sulphate reduction pathway, such as cysteine, are essential for bacterial thiamine (vitamin B1) biosynthesis [50], and lipopolysaccharides can activate melatonin oxidation to produce AFMK [51].

We also inferred causal relationships between gut microbiome and disease-associated metabolites using bi-directional Mendelian randomization (MR) analyses **(Chapter 4)**. For instance, we observed that increased *E. rectale* abundance may causally contribute to decreased levels of plasma p-cresol and p-cresol sulphate. As uremic toxins, p-cresol and its sulphate are involved in the aetiologies of chronic kidney disease and cardiometabolic diseases [17]. Importantly, by further annotating the genome of *E. rectale*, we characterized genes the encode a sulfatase and 4-hydroxybenzoyl-CoA reductase that are responsible for consecutive solvolysis of p-cresol sulphate into p-cresol and the subsequent reduction of p-cresol into benzoyl-CoA. Thus, our results revealed a potential new beneficial effect of *E. rectale* through degradation of uremic toxins.

Altogether, our longitudinal microbial associations and our mediation and MR analyses on host phenotypes and plasma metabolites have revealed novel functional insights and putative causality for the role of the gut microbiome

in human health and disease onset. However, longitudinal associations are not proof of causation, although we did carry out causal mediation analysis to infer *in silico* causality. We primarily focused on biologically plausible mechanisms by integrating the longitudinal metabolism dataset in order to provide mechanistic hypotheses that pinpoint specific microbial genetics and function but also to demonstrate which metabolites are likely to mediate the impact of the gut microbiome on the host phenotype. However, experimental validation is still needed to further substantiate these observations.

Microbial interactions show disease-specificity

Despite the complex interactions between gut microbiome, genetics and environment in metabolic regulation, the diverse microbial communities in our gut make up a complicated ecosystem in which microbes can exchange or compete for nutrients, signalling molecules, or immune-evasion mechanisms through ecological interactions that are far from fully understood [52]. Enthusiasm has thus been rising to decipher these microbial interactions in order to detect key microbes in health and disease. It is not currently possible to construct a human gut microbiota to study microbial interactions in health and disease because many microbes are not yet culturable. However, the gut microbial composition profile generated by shotgun metagenomics sequencing of large human cohorts does allow us to make inferences about microbial interactions *in silico* by assessing their co-abundance relationship and further comparing their differences between health and disease.

To look at these kinds of relationships, we compared the gut microbiome of 2,379 participants from LLD, 300OB, 500FG and 1000IBD. By constructing and comparing microbial co-abundance networks, we observed that the strengths of 39% of microbial species co-abundances and 64% of microbial pathway co-abundances varied significantly between the cohorts. In addition, hundreds of microbial co-abundance relationships showed IBD- and obesity-specific effects that could be replicated in independent cohorts. Moreover, we identified key microbes that potentially dominate the diseased gut microbial ecosystem, e.g. *Escherichia coli, Oxalobacter formigenes* and *Actinomyces graevenitzii* for IBD.

In summary, we show that microbial dysbiosis in disease may not only be driven by differences in microbial abundance level, but also by shifts in microbial interactions that are mirrored in co-abundance analyses. This extends the current knowledge about the role of the microbiome in disease. In particular, the disease-specific microbial interactions that we identified provide further insight into functional dysbiosis in IBD and obesity. However, the mechanism behind these disease-specific microbial interactions is still not clear. In-depth sequencing of the microbiome to

compare the genomic functional content difference between species/strains, e.g. via horizontal gene transfer, might be helpful to decode such disease-specific microbial interactions.

Perspective

Despite several illustrative examples regarding the role of environment-genetics-microbiome interactions in human metabolic health presented in this thesis, the underlying causal inference and mechanisms of environment-genetics-microbiome interactions in the development of complex diseases still remain largely unknown. Recently, we highlighted the importance of cohort studies in studying the aetiology of complex diseases in the post-GWAS era [53]. Below, I would like to further emphasize how integration of an approach using omics data, systems biology and genetics in combination with other cutting-edge bacterial culture-omics and organ-on-chip technologies can accelerate our understanding of the causality and mechanisms involved in host-microbe interactions (Figure 1).

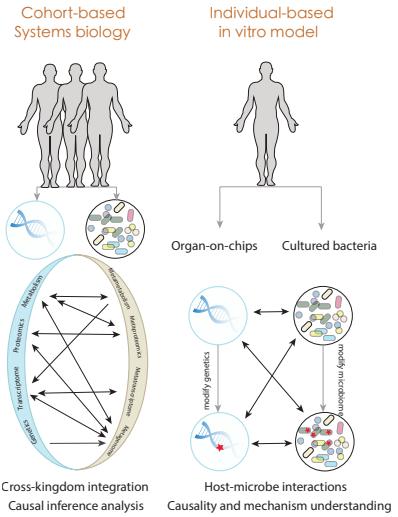


Figure 1. Analysis frame that combines a cohort-based systems biology approach with individual-based in vitro models to study host-microbe interactions.

Moving from metagenome to meta-omics

In addition to omics data on the human genome, omics datasets have been emerging from the metagenome, including metatranscriptomics, metaproteomics and metametabolomics. Several decades of DNA sequencing applications to determine the differences in microbial composition between healthy subjects and patients with specific phenotypes has produced evidence for dysregulation of microbial composition in human diseases. In one study, a longitudinal analysis of both metagenome and metatranscriptomics in IBD patients showed that certain species pathways exhibit different changes at the

transcriptional level compared to the DNA level [54]. Similarly, a comparison between 372 human faecal metatranscriptomes and 929 metagenomes identified both a "house-keeping" core of metatranscriptomes that is universally expressed over time and highly variable metatranscriptional activity that may reflect dynamic regulation of microbial composition in response to environmental perturbations [55]. Metaproteomics and metametabolomics have been proposed as complementary approaches to study the functional properties of the gut microbiome, and these methods combined have revealed species-specific metabolic shifts and variability in the gut microbiome of preterm infants and during the early years of development [56]. Studies using mouse and other animal models have reported that early life exposures, host genetics and diet can affect gut microbiome and metabolome [57], that diet can impact lipid metabolism in the gut [58] and that microbial metabolites can further affect host development, hormonal signalling, behaviour and gut physicochemical conditions [59,60]. Despite some technical challenges in data profiling, e.g. problems with the stability and reproducibility of microbial transcriptomics profiling and with the identification of proteomics and metabolomics based on mass-spectrometry data, meta-omics data has the potential to deliver a direct functional readout of the metagenome.

Cross-kingdom integration

Simultaneously profiling diverse omics data from the host genome and the metagenome, and incorporating these omics data, will substantially aid our understanding of cross-kingdom regulation and interaction at the molecular level. One of the clearest technological challenges here is the statistical complexity of integrating heterogeneous omics datasets [61]. Commonly used statistical approaches may be applied to integrate host omics and meta-omics data, including co-abundance network analysis, Bayesian network analysis, mediation analysis and causal inference analysis. However, it has been noted that the distribution of bacterial data often diverges from the normal distribution that many of these statistical methods assume. For instance, some bacterial species can be very abundant in some individuals but completely absent in others. Dealing with those 'zeros' may require a two-part model that deals with presence/absence and abundance levels separately [62] or the use of zero-inflated models [63,64]. This complexity in distribution greatly increases the complexity of data analysis, particularly in complex multi-omics models. Secondly, given that the gut microbiome harbours 100 times more bacterial genes than the human genome, the number of factors under study exponentially increases when integrating meta-omics data with host omics data. The power issue thus becomes a major burden in minimizing the false discovery rate. Although we can increase sample size

and conduct meta-analyses to combine association signals across multiple cohorts, for the time being the number of samples will continue to be far lower than the number of factors under study.

Furthermore, additional microbiome features will soon add to the complexity of cross-kingdom interaction studies. The gut virome and phageome compositions, for example, are important regulators of bacterial abundance and function that have not yet properly investigated in the majority of microbiome studies [65]. It is therefore essential to develop more advanced statistical algorithms and to take advantage of newly developed machine-learning algorithms and artificial intelligence methods to build models that can dissect the complexity of big data.

Even with these challenges, integrating diverse omics data from the genome and metagenome offers a great opportunity to study the underlying causality. For example, each kind of meta-omics data can be treated as a complex trait and subjected to genetic analysis. We can then use genetic variants as the instrumental variables and apply a MR approach to investigate the causal relationship between the genome and metagenome.

Innovative in vitro models for host-microbe interactions

For obvious ethical and practical reasons, human omics data is largely based on blood samples. This has, however, greatly limited our mechanistic understanding of the interaction of the gut microbiome with other human organs, including the intestine, liver and brain. Although mouse and other animal models have been used for these tasks, these animal models cannot fully mimic biological processes in humans. Nor can they take human genetic make-up into account, and thus cannot be used to study human genetics-microbe interactions.

Over the past few years, organ-on-chip technology has emerged as a next-generation disease and drug model [66,67]. In this new technology, human induced pluripotent stem cells (e.g. those developed from urine renal tubular cells that can be collected non-invasively from urine) can be further differentiated to different tissue cell types that can be used to construct organs-on-chips. One can now imagine an analysis frame that combines (1) cohort-based systems biology studies using well-characterized human cohorts with (2) individual-based studies using innovative in vitro model systems to investigate host-microbe interactions in health and disease (Figure 1). Two types of organs-on-chips, in particular, would be very interesting in this respect due to their direct interactions with the gut microbiome: gut-on-a-chip to study microbe-intestine interactions and liver-on-a-chip to study host-microbe metabolic interactions. With recent advances in bacterial culturomics, around 80% of gut microbes can be cultured [68], enabling

functional studies at both whole composition level and single strain level. This will allow for the whole gut microbiome, a specific strain, or their metabolic products to be applied to the organs-on-chips in order to assess the immune or metabolic response of human cells. Moreover, in such systems, the genetic background (e.g. via CRISPR-Cas9) and/or gut microbiome (e.g. specific strains) can be modified to test causality and genetics-microbe interactions.

In conclusion, the past several decades have witnessed an increased awareness and understanding of the role of the gut microbiome in human health and of its interactions with host genome and environmental factors. The gut microbiome is now emerging as an important player in personalized medicine. With the aid of well-characterized human cohorts and cutting-edge technologies, we are now on the verge of a major breakthrough in our understanding of host-microbe interactions that will lay the foundation for the development of the next phase of personalized medicine, one that coordinates and encompasses both the human genome and metagenome.

References

1 Hunter, D. J. Gene-environment interactions in human diseases. Nat Rev Genet 6, 287-298, doi:10.1038/nrg1578 (2005).

2 McCarthy, M. I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9, 356-369, doi:10.1038/nrg2344 (2008).

3 Manolio, T. A. et al. Finding the missing heritability of complex diseases. Nature 461, 747-753, doi:10.1038/nature08494 (2009).

4 Wang, J. & Jia, H. Metagenome-wide association studies: fine-mining the microbiome. Nat Rev Microbiol 14, 508-522, doi:10.1038/nrmicro.2016.83 (2016).

5 Falony, G. et al. Population-level analysis of gut microbiome variation. Science 352, 560-564, doi:10.1126/science.aad3503 (2016).

6 Lynch, S. V. & Pedersen, O. The Human Intestinal Microbiome in Health and Disease. N Engl J Med 375, 2369-2379, doi:10.1056/NEJMra1600266 (2016).

7 Zhernakova, A. et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 352, 565-569, doi:10.1126/science.aad3369 (2016).

8 Koh, A. & Backhed, F. From Association to Causality: the Role of the Gut Microbiota and Its Functional Products on Host Metabolism. Mol Cell 78, 584-596, doi:10.1016/j.molcel.2020.03.005 (2020).

9 Chavez-Talavera, O., Haas, J., Grzych, G., Tailleux, A. & Staels, B. Bile acid alterations in nonalcoholic fatty liver disease, obesity, insulin resistance and type 2 diabetes: what do the human studies tell? Curr Opin Lipidol 30, 244-254, doi:10.1097/MOL.0000000000000597 (2019).

10 Fu, J. & Kuipers, F. Systems genetics approach reveals cross-talk between bile acids and intestinal microbes. PLoS Genet 15, e1008307, doi:10.1371/journal.pgen.1008307 (2019).

11 Udayappan, S. et al. Oral treatment with Eubacterium hallii improves insulin sensitivity in db/db mice. NPJ Biofilms Microbiomes 2, 16009, doi:10.1038/npjbiofilms.2016.9 (2016).

12 Tonin, F. & Arends, I. Latest development in the synthesis of ursodeoxycholic acid (UDCA): a critical review. Beilstein J Org Chem 14, 470-483, doi:10.3762/bjoc.14.33 (2018).
13 Liu, H., Hu, C., Zhang, X. & Jia, W. Role of gut microbiota, bile acids and their cross-talk in the effects of bariatric surgery on obesity and type 2 diabetes. Journal of diabetes investigation 9, 13-20 (2018).
14 Kurilshikov, A. et al. Gut Microbial Associations to Plasma Metabolites Linked to Cardiovascular Phenotypes and Risk: A Cross-Sectional Study. Circ Res 124, doi:10.1161/CIRCRESAHA.118.314642 (2019).
15 Hang, S. et al. Bile acid metabolites control TH17 and Treg cell differentiation. Nature 576, 143-148, doi:10.1038/s41586-019-1785-z (2019).
16 Wishart, D. S. et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 46, D608-D617, doi:10.1093/nar/gkx1089 (2018).
17 Wang, Z. & Zhao, Y. Gut microbiota derived metabolites in cardiovascular health and disease. Protein Cell 9, 416-431, doi:10.1007/s13238-018-0549-0 (2018).
18 Seyed Hameed, A. S., Rawat, P. S., Meng, X. & Liu, W. Biotransformation of dietary phytoestrogens by gut microbes: A review on bidirectional interaction between phytoestrogen metabolism and gut microbiota. Biotechnol Adv 43, 107576, doi:10.1016/j.biotechadv.2020.107576 (2020).
19 Chaleckis, R., Meister, I., Zhang, P. & Wheelock, C. E. Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics. Curr Opin Biotechnol 55, 44-50, doi:10.1016/j.copbio.2018.07.010 (2019).
20 Almeida, A. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol 39, 105-114, doi:10.1038/s41587-020-0603-3 (2021).
21 UniProt, C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49, D480-D489, doi:10.1093/nar/gkaa1100 (2021).
22 Mistry, J. et al. Pfam: The protein families database in 2021. Nucleic Acids Res 49, D412-D419, doi:10.1093/nar/gkaa913 (2021).
23 Letunic, I., Khedkar, S. & Bork, P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res 49, D458-D460, doi:10.1093/nar/gkaa937 (2021).
24 Danchin, A., Ouzounis, C., Tokuyasu, T. & Zucker, J. D. No wisdom in the crowd: genome annotation in the era of big data - current status and future prospects. Microb Biotechnol 11, 588-605, doi:10.1111/1751-7915.13284 (2018).
25 Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Microbial genome analysis: the COG approach. Brief Bioinform 20, 1063-1070, doi:10.1093/bib/bbx117 (2019).
26 Zhao, Y. et al. Analysis of multiple metabolites of tocopherols and tocotrienols in mice and humans. J Agric Food Chem 58, 4844-4852, doi:10.1021/jf904464u (2010).
27 Jiang, Q., Christen, S., Shigenaga, M. K. & Ames, B. N. gamma-tocopherol, the major form of vitamin E in the US diet, deserves more attention. Am J Clin Nutr 74, 714-722, doi:10.1093/ajcn/74.6.714 (2001).
28 Bar, N. et al. A reference map of potential determinants for the human serum metabolome. Nature 588, 135-140, doi:10.1038/s41586-020-2896-2 (2020).
29 Kurilshikov, A., Wijmenga, C., Fu, J. & Zhernakova, A. Host Genetics and Gut Microbiome: Challenges and Perspectives. Trends Immunol 38, 633-647, doi:10.1016/j.it.2017.06.003 (2017).
30 Kurilshikov, A. et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet, doi:10.1038/s41588-020-00763-1 (2021).
31 Ruhlemann, M. C. et al. Genome-wide association study in 8,956 German individuals

identifies influence of ABO histo-blood groups on gut microbiome. Nat Genet 53, 147-155, doi:10.1038/s41588-020-00747-1 (2021).
32 Lopera-Maya, E. A. et al. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. bioRxiv, 2020.2012.2009.417642, doi:10.1101/2020.12.09.417642 (2020).
33 Wang, J. et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat Genet 48, 1396-1406, doi:10.1038/ng.3695 (2016).
34 Sonnenburg, J. L. & Backhed, F. Diet-microbiota interactions as moderators of human metabolism. Nature 535, 56-64, doi:10.1038/nature18846 (2016).
35 Walker, A. W. et al. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. The ISME journal 5, 220-230 (2011).
36 Zhu, W. et al. Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk. Cell 165, 111-124, doi:10.1016/j.cell.2016.02.011 (2016).
37 Zhernakova, A. et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 352, 565-569 (2016).
38 Pallister, T. et al. Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome. Sci Rep 7, 13670, doi:10.1038/s41598-017-13722-4 (2017).
39 Natella, F., Nardini, M., Belelli, F. & Scaccini, C. Coffee drinking induces incorporation of phenolic acids into LDL and increases the resistance of LDL to ex vivo oxidation in humans. Am J Clin Nutr 86, 604-609, doi:10.1093/ajcn/86.3.604 (2007).
40 Song, Y. et al. Feruloylated oligosaccharides and ferulic acid alter gut microbiome to alleviate diabetic syndrome. Food Res Int 137, 109410, doi:10.1016/j.foodres.2020.109410 (2020).
41 Salau, V. F. et al. Ferulic Acid Modulates Dysfunctional Metabolic Pathways and Purinergic Activities, While Stalling Redox Imbalance and Cholinergic Activities in Oxidative Brain Injury. Neurotox Res 37, 944-955, doi:10.1007/s12640-019-00099-7 (2020).
42 Biegel, E. & Muller, V. Bacterial Na+-translocating ferredoxin:NAD+ oxidoreductase. Proc Natl Acad Sci U S A 107, 18138-18142, doi:10.1073/pnas.1010318107 (2010).
43 Holland, I. B. & Blight, M. A. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. J Mol Biol 293, 381-399, doi:10.1006/jmbi.1999.2993 (1999).
44 Fujita, T., Fujita, M., Kodama, T., Hada, T. & Higashino, K. Determination of D- and L-pipecolic acid in food samples including processed foods. Ann Nutr Metab 47, 165-169, doi:10.1159/000070040 (2003).
45 Kawasaki, H., Hori, T., Nakajima, M. & Takeshita, K. Plasma levels of pipecolic acid in patients with chronic liver disease. Hepatology 8, 286-289, doi:10.1002/hep.1840080216 (1988).
46 Miras-Avalos, J. M., Bouzas-Cid, Y., Trigo-Cordoba, E., Orriols, I. & Falque, E. Amino Acid Profiles to Differentiate White Wines from Three Autochtonous Galician Varieties. Foods 9, doi:10.3390/foods9020114 (2020).
47 Alaei-Shahmiri, F., Soares, M. J., Zhao, Y. & Sherriff, J. The impact of thiamine supplementation on blood pressure, serum lipids and C-reactive protein in individuals with hyperglycemia: a randomised, double-blind cross-over trial. Diabetes Metab Syndr 9, 213-217, doi:10.1016/j.dsx.2015.04.014 (2015).
48 Mayo, J. C. et al. Anti-inflammatory actions of melatonin and its metabolites, N1-acetyl-N2-formyl-5-methoxykynuramine (AFMK) and N1-acetyl-5-methoxykynuramine (AMK), in

macrophages. J Neuroimmunol 165, 139-149, doi:10.1016/j.jneuroim.2005.05.002 (2005).
49 Rezzani, R. et al. Effects of melatonin and Pycnogenol on small artery structure and function in spontaneously hypertensive rats. Hypertension 55, 1373-1380, doi:10.1161/HYPERTENSIONAHA.109.148254 (2010).
50 Begley, T. P. The biosynthesis and degradation of thiamin (vitamin B1). Nat Prod Rep 13, 177-185, doi:10.1039/np9961300177 (1996).
51 Silva, S. O. et al. Oxidation of melatonin and its catabolites, N1-acetyl-N2 -formyl-5-methoxykynuramine and N1-acetyl-5-methoxykynuramine, by activated leukocytes. J Pineal Res 37, 171-175, doi:10.1111/j.1600-079X.2004.00149.x (2004).
52 Banerjee, S., Schlaeppi, K. & van der Heijden, M. G. A. Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol 16, 567-576, doi:10.1038/s41579-018-0024-1 (2018).
53 Wijmenga, C. & Zhernakova, A. The importance of cohort studies in the post-GWAS era. Nat Genet, 322-328, doi:10.1038/s41588-018-0066-3 (2018).
54 Schirmer, M. et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat Microbiol, doi:10.1038/s41564-017-0089-z (2018).
55 Abu-Ali, G. S. et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. Nat Microbiol, doi:10.1038/s41564-017-0084-4 (2018).
56 Xiong, W., Brown, C. T., Morowitz, M. J., Banfield, J. F. & Hettich, R. L. Genome-resolved metaproteomic characterization of preterm infant gut microbiota development reveals species-specific metabolic shifts and variabilities during early life. Microbiome 5, 72, doi:10.1186/s40168-017-0290-6 (2017).
57 Snijders, A. M. et al. Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. Nat Microbiol 2, 16221, doi:10.1038/nmicrobiol.2016.221 (2016).
58 Maier, T. V. et al. Impact of Dietary Resistant Starch on the Human Gut Microbiome, Metaproteome, and Metabolome. MBio 8, doi:10.1128/mBio.01343-17 (2017).
59 Zheng, H., Powell, J. E., Steele, M. I., Dietrich, C. & Moran, N. A. Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. Proc Natl Acad Sci U S A 114, 4775-4780, doi:10.1073/pnas.1701819114 (2017).
60 Kesnerova, L. et al. Disentangling metabolic functions of bacteria in the honey bee gut. PLoS Biol 15, e2003467, doi:10.1371/journal.pbio.2003467 (2017).
61 Schmidt, T. S. B., Raes, J. & Bork, P. The Human Gut Microbiome: From Association to Modulation. Cell 172, 1198-1215, doi:10.1016/j.cell.2018.02.044 (2018).
62 Fu, J. et al. The Gut Microbiome Contributes to a Substantial Proportion of the Variation in Blood Lipids. Circ Res 117, 817-824, doi:10.1161/CIRCRESAHA.115.306807 (2015).
63 Liu, F., Wang, C., Wu, Z., Zhang, Q. & Liu, P. A zero-inflated Poisson model for insertion tolerance analysis of genes based on Tn-seq data. Bioinformatics 32, 1701-1708, doi:10.1093/bioinformatics/btw061 (2016).
64 Chen, E. Z. & Li, H. A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. Bioinformatics 32, 2611-2617, doi:10.1093/bioinformatics/btw308 (2016).
65 Norman, J. M. et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell 160, 447-460, doi:10.1016/j.cell.2015.01.002 (2015).
66 Huh, D. et al. Reconstituting organ-level lung functions on a chip. Science 328, 1662-1668, doi:10.1126/science.1188302 (2010).
67 Zhang, B. et al. Biodegradable scaffold with built-in vasculature for organ-on-a-chip

engineering and direct surgical anastomosis. Nat Mater 15, 669-678, doi:10.1038/nmat4570 (2016).
68 Lagier, J. C. et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol 1, 16203, doi:10.1038/nmicrobiol.2016.203 (2016).