Review: Nature Gastroenterology and Hepatology

Manuscript ID: NRGH-20-009V1B

Title:

Microbiome risk profiles as disease biomarkers for inflammatory and metabolic disorders

Authors

Amira Metwaly[1], Sandra Reitmeier[2], Dirk Haller[1,2*]

Authors' affiliation

[1] Technical University of Munich, Chair of Nutrition and Immunology, School of Life Sciences, 85354 Freising, Germany

[2] Technical University of Munich, ZIEL Institute for Food & Health, Germany

Correspondence

dirk.haller@tum.de

Abstract

The intestine harbours a complex array of microorganisms, collectively known as the gut microbiota. The past two decade witnessed an increasing interest in studying gut microbiota changes in relation to health and disease, driven by the vast advancement in the innovation and application of high-throughput multi-omics technologies. Microbial dysbiosis has been linked to many human pathologies, including metabolic disorders, such as type-2 diabetes, as well as inflammatory bowel diseases. Nevertheless, and even though the gut provides a common interface, a comprehensive understanding of microbiome contribution to disease causality remains limited, largely due to the heterogeneity in microbial community structure, individually diverse disease evolution and incomplete understanding for the mechanisms related to signal integration. Multiple factors might explain these inconsistencies, including methodological, environmental, therapeutic exposure factors, in addition to the inherent microbiome variations within human populations. To gain a mechanistic insight of how microbes impact intestinal health, we need to move from correlation to causation. Integrated analysis of multi-omics data, including metagenomics and metabolomics, with measurements of host response and cataloguing bacterial isolates identified bacteria and bacterial products linked to disease pathology. In this Review, we provide a broader insight into microbiome signatures for inflammatory and metabolic disorders, discuss the standing challenges and propose areas to improve the application of multi-omics towards an improved mechanistic understanding of underlying microbe-host interactions.

Key Points:

- Several commonalities exist between inflammatory bowel diseases (IBD) and type-2 diabetes, both recognized as multifactorial diseases with a rising global incidence following industrialization patterns.

- Altered gut bacterial composition and host processing of bacteria-derived metabolites have been implicated in IBD and T2D and stand as a common underlying mechanism of disease pathogenesis.

- A causal link between dysbiotic microbial communities and IBD or T2D has been established through gnotobiotic mouse experiments and through integrative multi-omics analyses of prospective longitudinal cohorts and large-scale population studies.

- The challenge in disease-specific biomarker discovery lies in the timing of changes (cause or consequence), the functional redundancy of changes (similar signal integration into disease mechanisms) and the gut microbiota heterogeneity (across geography and ethnicities).

- Big data refinement, testing and validation of specific bacterial strains, their encoded genes and metabolic by-products are necessary to identify disease biomarkers.

Introduction

The human digestive tract harbours a complex array of microorganisms, including bacteria, archaea, viruses, and fungi. Trillions of bacteria colonize the gastrointestinal tract in a spatially structured manner and their genomes reach more than 200 times the number of genes in the human body [1]. Colonization density of bacteria follows a gradient from the proximal to the distal part of the gut, reaching highest numbers in the colon [2,3]. In contrast to the small intestine, reduced motility in the colon provides prolonged retention of luminal content (20 – 50 hours) and builds a vast reservoir of biologically active metabolites. The term microbiome additionally includes the environment inhabited by the microbiota, or the niche shaped by the host. The incorporation of the host provides a broader view of the ecosystem where bi-directional microbe-host interactions influence the physio-chemical characteristics of the microbial environment "theatre of activity" [4] **(Box 1).** Since the digestive tract and its microbiome is considered as a central organ at the intersection of immune- and metabolic processes, we focus in this review on inflammatory bowel diseases (IBD) and type-2 diabetes (T2D) as examples of microbiota-associated disorders.

Several commonalities exist between IBD and T2D, both recognized as multifactorial diseases with a rising global incidence following industrialization patterns [5–7]. Their aetiology is associated with a complex interplay of genetic susceptibility, environmental triggers, and urban lifestyles. In this commonality, metabolic diseases, such as T2D, are characterized by chronic subclinical inflammation in liver, adipose tissue, muscles, pancreas, and gut. On the other hand, inflammatory gastrointestinal disorders, such as Crohn's diseases (CD) or Ulcerative colitis (UC) are associated with inflammation-driven metabolic alterations [8]. Genome-wide association studies (GWAS) identified genetic variants associated with increased susceptibility to developing T2D (143 loci) [9] and IBD (> 240 loci)[10]. Nevertheless, the heritability explained by these variants are rather limited (<10% for T2D, <15% for UC and <50% for CD) [11–16], supporting the relevance of environmental triggers, in particular the gut microbiome as a major contributor to disease aetiology.

Despite the great advancement in GWAS and multi-omics driven risk profiling, the identification of disease susceptible individuals is still difficult and validated diagnostic or prognostic markers are lacking. Analysis of multiple population studies and IBD or T2D patient cohorts identified microbiome signatures linked to disease phenotypes [17–20], the risk of relapse [21] or response to treatment [22]. Therapy for complex diseases, such as T2D and IBD, remain challenging, but recent controlled trials using faecal microbiota transplantation (FMT) show clinical efficacy in both diseases [23–26]. In this Review, we summarize the current knowledge on the involvement of the gut microbiome in IBD and T2D. We critically assess the status of the currently available disease-associated microbiome signatures and discuss the limitations facing their use in clinical applications. Finally, we discuss the use of multi-omics big data in an integrative framework to disentangle the complexity of disease pathology. In particular, we focus on the mechanistic interaction between bacterial strains and gut-derived metabolites on promoting processes involved in inflammatory and metabolic diseases such as IBD and T2D.

Box 1 – G*ut microbiome and microbe-host interactions: Terminology*

Microbiome signature: unique pattern of microbiome configuration that can stratify defined physiological and pathological conditions including risk prediction in patients for disease development or progression.

Microbiota refers to the microorganisms of a defined environment. It comprises bacteria, fungi, archaea, and viruses.

Microbiome comprises all the microorganisms, their genomes and the surrounding host-shaped environmental conditions of a given habitat. Characterization of gut microbiome can be achieved through the application of metagenomics, metabolomics, metatranscriptomics, and metaproteomics combined with clinical or environmental metadata [4,27].

Dysbiosis or Pathobiome describe "an altered microbial community composition, which has a consequential impact on the host immune response and leads to the emergence and outbreak of pathogens" [4,28,29].

Pathobiont versus opportunistic pathogen: Pathobionts are microorganisms linked to chronic inflammatory conditions. Opportunistic pathogens can cause acute infections. While pathobionts are harmless to the host under normal conditions, pathogens can drive disease in a healthy host[30].

Microbial and metabolic dysbiosis as common features of IBD and T2D

Both IBD and T2D show microbial alterations, characterized by reduced community richness, in addition to the reduction of beneficial microbes and expansion of pathobionts[31]. The challenge in understanding the role of microbial alterations in disease initiation and progression lies in the timing of changes (cause or consequence), the functional redundancy of changes (similar signal integration into disease mechanisms) and the fluctuations of changes during disease course (lack of longitudinal sampling). Despite the differences in pathology, IBD and T2D share common mechanistic features. T2D exhibits chronic low-grade inflammation and gut barrier disruption, and *vice versa*, recurrent inflammatory flares in IBD coevolve with metabolic alterations at the cellular and systemic level [32,33] .

Evidence for causal relationship between microbiota and inflammatory, immune or metabolic disorders was shown by FMT trials, in which the stool of a healthy donor is transferred to the patient [34]. FMT has been shown to be highly effective in treating approximately 90% of patients with *Clostridium difficile* infections [35] and has been assessed for the treatment of T2D[36], obesity, graft-versus-host diseases (GvHD) [37]andIBD[38], including UC and to a less extent CD. Based on the results of four randomized clinical trials, FMT induced clinical remission in 28% of UC patients [25,39–41]. Few studies have examined the clinical efficacy of FMT for CD and the results were rather heterogenous. In a clinical study including 174 CD patients treated with FMT, clinical remission was achieved in 20% and clinical response was achieved in 43% of patients [42]. A recent randomized controlled trial conducted by Sokol and colleagues showed no significant impact of FMT on CD clinical remission, but higher engraftment of donor microbiota was associated with maintenance of remission [24]. Conversely and despite a multitude of microbiota association studies, evidence for FMT for metabolic diseases is less established. Recent landmark studies demonstrated metabolic improvements together with changes in intestinal microbiome in patients with metabolic syndrome who received FMT from lean healthy donors [26]. These effects were however inconsistent and transient, explained by limited donor microbiota engraftment[26] or varying donor fecal microbial diversity at baseline[36]. Intriguingly, supplementation with low-fermentable fiber following oral FMT lead to improved insulin sensitivity, changed microbiota composition and prolonged donor stool engraftment in obese patients with metabolic syndrome, emphasizing the value of microbial modulation therapy in reversing metabolic dysfunction[43]. In line with these findings, FMT from metabolically compromised obese donors transiently worsened insulin sensitivity in recipients with metabolic syndrome, whereas FMT from healthy post-gastric bypass donors induced a minimal increase in insulin sensitivity in recipient patients, providing evidence for the transmissibility of donor metabolic profile by FMT [44].

Microbial dysbiosis in IBD

Several large cohort studies **(Table 1)** investigated gut microbiota alterations in IBD based on microbial profiling of luminal and mucosal microbial communities. Overall, an overabundance of certain bacterial groups such as *Enterobacteriaceae, Fusobacterium, Ruminococcus gnavus, Streptococcus anginosus, Enterococcus, Megasphaera, Campylobacter,* and sulfate-reducing Gamma- and Deltaproteobacteria have been implicated in patients with active disease. Conversely, the loss of beneficial taxa such as *Faecalibacterium prausnitzii, Christensenellaceae, Collinsella, Roseburia, Ruminococcus* and other butyrate-producing bacteria has been linked to disease [18,21,22,45–49]. Shotgun metagenomics of stool samples provided a more comprehensive view of functional dysbiosis and showed perturbations of metabolic pathways in IBD. An upregulation of metabolic pathways involved in sulfur-containing amino acids synthesis, riboflavin metabolism, glutathione transporters , oxidative stress and nutrient transport were shown for IBD [19,48,50–52]. Assessment of strain-level intra-species resolution revealed increased strain diversity of pathobionts and reduced strain diversity in beneficial microbes in stool samples from patients with IBD or irritable bowel syndrome (IBS) compared with healthy controls[52]. In-depth analysis showed 219 taxa (including 152 species) associated with CD and 102 taxa (including 93 species) associated with UC. CD was predominantly characterized with a decrease in taxa belonging to *Lachnospiraceae* and *Ruminococcaceae* and an increase in taxa belonging to *Enterobacteriaceae* family, whereas a decrease in taxa belonging to *Bacteroidaceae* and increase in taxa belonging to *Lachnospiraceae* was observed for UC. In concordance with this heterogeneity, only few species were identified to be shared across different IBD studies[53], suggesting individual differences within similar CD phenotypes and disease courses. One of the first clinical evidence for the key role of the intestinal microbiota in IBD pathogenesis originated from experiments showing that diversion of the fecal stream from an inflamed segment of the small intestine improved disease symptoms in CD patients. Restoration of fecal stream and postoperative exposure of the neo-terminal ileum to luminal contents induced inflammation, suggesting that the microbiota triggers postoperative recurrence of CD[54,55] . Furthermore,efficacy of antibiotic treatment in subsets of patients with active CD emphasizes the causal link of bacteria to IBD [56].

Mechanistic studies in mouse models of acute and chronic intestinal inflammation provided evidence for a causal relationship between microbial dysbiosis and IBD [57,58]. For example,the transfer of faecal microbiota from patients with IBD to germ-free recipient mice was sufficient to transfer disease phenotype [21,59] and genetically susceptible IBD mouse models develop no spontaneous inflammation under germ-free conditions[60].Additinaly, the transfer of dysbiotic microbial communities from inflamed mice could transfer disease phenotype in recipient germ-free mice[61]. Likewise, the transfer of IBD microbiota into germ-free mice induced imbalance in intestinal Th17 and RORgt+ regulatory T cells and commensal bacteria of the intestinal microbiota *Bacteroides fragilis* was shown to induce Foxp3+ regulatory T-cell development[62], suggesting microbiota-driven disease mechanisms in IBD.

Microbial dysbiosis in T2D

Like IBD, a widely variable change in the abundance of several bacterial taxa has been described in T2D **(Table 2).** For instance, previous data showed an increased relative abundance of *Escherichia coli, Veillonella, Blautia, Anaerostipes, Lactobacillus, Faecalibacterium,* Clostridiales amongst others in patients with T2D. On the contrary, reduced abundance of *Bacteroides, Bifidobacterium, Parabacteroides, Oscillospira* and the mucin-degrading gut bacteria, *Akkermansia muciniphila* is associated with improved metabolic health [20,63,64]. In a recent study by Zhong et al., metagenomic and metaproteomic analysis were performed on fecal samples from a Chinese cohort to characterize the gut microbiota compositional and functional alterations[65]. The cohort included 254 individuals including 77 treatment-naïve type 2 diabetics, 80 pre-diabetics and 97 individuals with normal glucose tolerance. T2D and pre-diabetics showed lower abundance of bacterial species belonging to Clostridiales and higher abundance *of Megasphaera elsdenii* compared to controls. Functional differences were observed in the microbiome of patients with T2D or pre-diabetics. Significant enrichment in pathways involved in sugar phosphotransferase systems (PTS), ATP-binding cassette transporters (ABC transporters) of amino acids, and bacterial secretion systems in the gut microbiota was observed in pre-diabeteics compared to control subjects, suggesting unique changes in the gut microbiome of pre-diabetics before transition to T2D. Differences in gut microbiota composition and gene clusters have been used to classify individuals with T2D[20,66]. However, confounding factors like geographic location, ethnicity, health status and medication history lead to inconsistency in identifying microbial alterations associated with T2D[64].

Recent studies provided evidence for a causal link of specific members of the intestinal microbiota to pathogenesis of T2D. For example, *Akkermansia muciniphila* is one of the key taxa shown to have a protective effect in metabolic disorders in human and in mouse studies[67–69]. Interestingly, prebiotic feeding normalized *Akkermansia muciniphila* abundance and improved metabolic health, where the administration of *Akkermansia muciniphila* reversed high-fat diet-induced fat-mass gain, metabolic endotoxemia, adipose tissue inflammation, and insulin resistance in mice [70]. Despite its high oxygen sensitivity and need for animal-derived compounds in the growth medium, *Akkermansia muciniphila* was shown to retain its protective effects in mice when grown on a synthetic medium compatible with human administration[71], opening avenues for therapeutic options to target human obesity and associated disorders. Further, the butyrate-producing bacterium *Anaerobutyricum soehngenii* (previously designated *Eubacterium hallii* strain L2-7) showed an increased abundance that correlated with improved peripheral insulin sensitivity in recipient of lean donor fecal microbiota transfer[26]. The administration of *Anaerobutyricum soehngenii* strain in patients improved peripheral insulin sensitivity after 4 weeks of treatment, together with an altered microbiota composition and changes in bile acid metabolism [72].

IBD and T2D: overlapping microbiome signatures

Curiously, specific bacterial taxa are overlapping between IBD and T2D, suggesting that immune-mediated and metabolic disease share common features that lead to similar adaptations of the microbiota. Examples include the decreased levels of *Clostridium spp.*, *Faecalibacterium, Ruminococcus, Akkermansia, Collinsella* and *Roseburia*, and increased representation by *Enterobacteriaceace*, *Escherichia coli* and *Fusobecaterium nucleatum* species, emphasizing the challenge in defining disease-specific markers **(Figure 2).** An example to illustrate this challenge is a recent study on 2,045 IBD patients, that aimed at finding a microbial signature for CD [18]. The authors identified a signature of eight taxa including unknown members of the family *Christensenellaceae* and the genus *Fusobacterium* to discriminate between patients with CD and healthy individuals. Nevertheless, the abundance of *Christensenellaceae* is known to be associated with low body mass index (BMI) and weight loss [73], a catabolic condition frequently observed in IBD patients. Similarly, the enrichment of *Fusobacterium* is a considered a prognostic marker for metastatic colorectal cancer (CRC) [74]. Given the fact that IBD patients are at higher risk of developing CRC, the proposed microbiome signature might be an associated phenomenon with no causal link to the underlying disease mechanisms.

Additional meta-omics approaches, including shotgun metagenomics and metabolomics, together with patient treatment history, demographics and environmental data enabled deeper characterization of the gut microbiome functional capacity. Findings from the second phase of the HMP immensely improved our understanding of microbe-metabolite interactions in T2D [75] and IBD [19]. Integrative network analysis of microbiome, metabolome and transcriptome datasets from 132 individuals identified key disease-associated network hubs connecting bacteria (*Faecalibacterium prausnitzii*, unclassified *Subdoligranulum, Alistipes, Escherichia coli* and *Roseburia)* to certain metabolites (SCFAs, octanoyl carnitine and several lipids). Interestingly, differences between subjects with and without IBD were most apparent in the fecal metabolome compared to the fecal metagenome, metatranscriptome, or proteome [19]. In the second study of the iHMP - the Integrated Personal Omics Profiling Study (iPOP), the authors showed a strong correlation between plasma metabolites and insulin resistance in longitudinal samples from 106 subjects, suggesting perturbation of the host metabolome and gut microbiome interactions in insulin resistant individuals.

Biomarkers of gut microbiome dysbiosis

According to the National Institute of Health (NIH) Biomarker Definition Working Group, a biomarker is defined as ‘‘a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes, or pharmacologic responses to a therapeutic intervention’’[76]. An ideal clinical biomarker should be rapid, quantitative, objective, reproducible, non-invasive and exhibit high accuracy in predicting disease state across several populations or ethnicities **(Box 2)** [77]**.** The identification of microbiome biomarkers and their use for classification of disease entities require extensive computational and statistical tools to determine networks of bacterial taxa that can accurately discriminate between different disease phenotypes (e.g., healthy vs. IBD or pre-T2D vs. T2D), as well as closely related disease entities (e.g., IBD and IBS). Profiles of microbial biomarkers require further validation in large population-based cohorts to verify their diagnostic or prognostic value. In the following sections, we review the advancements made towards the development of microbiome-based biomarkers for disease risk profiling. These biomarkers range from single indicator bacterial taxa to a dysbiotic complex communities, to multi-omics-based biomarkers **(Box 3)**.

From single indicator strains to complex signature networks. Multiple studies investigated the utility of microbial alterations as disease biomarker, particularly in patients with CD or UC. First efforts pursued to define single bacterial taxa as indicators for disease activity. For instance, *Faecalibacterium prausnitzii*, a butyrate producing Firmicutes is depleted in patients with CD [78]. Lower abundance of this bacterium in ileal mucosa from CD patients strongly correlated with the risk of endoscopic recurrence

after ileal resection. Conversely, an increased abundance of adherent invasive *Escherichia coli* correlated with ileal CD [79]. Mostly, 16S rRNA amplicon sequences for family-level or genus-level taxonomic classification but rarely species-level associations were applied. However, most bacterial species comprise individual strains with massively different gene content, making strain diversity of great functional importance, particularly in terms of pathogenicity. For instance, subspecies of *Ruminococcus gnavus* and *Escherichia coli* has each been associated with IBD severity [80,81]. Nevertheless, a particular subspecies of *Ruminococcus gnavus* was found to be more abundant in IBD fecal microbiota and was linked to changes in oxidative stress response, adhesion and iron and mucus utilization [82]. Similarly, strains belonging to the species *Bacteroides fragilis* showed functional divergence leading to differential IgA induction in IBD-related mouse models. These genetically distinct strains showed differential colitogenic and immunomodulatory effects when colonizing mice [83]. To define key dysbiotic taxa to use for monitoring disease severity, Lopez-Siles and colleagues tested whether the co-abundances of *Faecalibacterium prausnitzii* and *Escherichia coli* could be used to diagnose patients with IBD by computing the absolute abundances ration of these two bacteria, using quantitative PCR analysis (F-E index). While using the F-E index improved the classification of UC and IBS from those with CD and allowed a better discrimination of CRC from other gut disorders, it failed to discriminate between IBD subtypes [84,85], suggesting the limited utility of single taxa indicators for disease sub-classification. Gut dysbiosis indices were mostly based on the bacterial community, however, Sokol *et al.* [45] defined dysbiosis based on the differential abundance of two fungal phyla Basidiomycota and Ascomycota which robustly separated samples originating from healthy subjects and IBD patients with different disease phenotypes[45]. In addition, reduced fungal diversity was shown in pediatric CD together with increased Candida taxa [86]. Interestingly, recent work by Sarrabayrouse and colleagues showed significant difference in fungal and bacterial loads between healthy relatives and non-related healthy controls and between patients with different IBD subtypes, demonstrating that bacteria and fungi contribute to IBD gut dysbiosis .[87].

Dysbiosis score as a quantifiable deviation from a healthy baseline: Moving beyond the simplistic view conveying the abundance of a single bacterium as a marker for disease, several studies evaluated dysbiotic indices that describe more complex bacterial co-occurrence abundances for disease classification. For instance, Halfvarson and colleagues demonstrated that microbiota of IBD patients fluctuates more dramatically than healthy subjects, based on deviation from a baseline they identified and termed as "healthy plane". The distance to the "healthy plane" varied overtime in IBD patients. Nevertheless, this distance does not necessarily correlate with disease activity [88]. Gevers *et al.* demonstrated in a large early onset pediatric CD cohort (RISK study) that the microbiome dysbiosis (MD) index, which is the log of total abundance in organisms increased in IBD over total abundance of organisms decreased in IBD strongly correlated with disease severity and could be used in the stratification of patients [48]. Nevertheless, MD index was limited in classifying disease and showed an overlap between IBD and healthy individuals [48]. In a recent study from the second phase of the integrative Human Microbiome Project (iHMP), the authors identified samples from IBD patients with highly dysbiotic metagenomic microbial structure using a dysbiosis score based on Bray–Curtis dissimilarities to non-IBD metagenomes. In addition to the microbial response to inflammation, their dysbiotic score encompassed other host (transcriptomic regulation) and biochemical (serum metabolites and chemokines) interactions, providing a more comprehensive view of the systemic and microbe-host interactions in IBD [19]. Nevertheless, it is important to emphasize that dysbiosis is not a well-defined condition and hence dysbiotic indices differ according to the methodology, disease entity and among different cohorts or groups of individuals.

Large-scale marker profiling using machine learning algorithms. Several studies used machine learning (ML) algorithms to validate complex microbiome signatures on cross-sectional and longitudinal patient cohorts. For example, Pascal *et al.* in 2017 analyzed fecal samples from large cohort of IBD and non-IBD individuals and based on the 16S microbiota profiling, they proposed a microbial signature specific for CD based on the abundance of 8 bacterial taxa. Additionally, Ananthakrishnan *et al.* showed that early changes in gut microbiome composition at baseline could predict IBD patients' response to therapy, 14 weeks after anti-integrin initiation with an AUC of 0.87 compared to a model

based on clinical covariates (AUC =0.62) [22]. In two recent studies, we assessed the utility of microbiome signatures as biomarkers of IBD and T2D. In the first study, we examined disease activity and response to therapy in a unique cohort of 29 CD patients who undergone autologous hematopoietic stem cell transplantation (HSCT). Integration of microbiome and metabolome profiles from human donors and humanized mice improved the predictive modelling of disease outcome from an AUC of 0.79 to an AUC of 0.96 and identified a network of disease-associated bacterial and metabolite factors involved in sulfur metabolism [21].While these findings sound promising, it is important to acknowledge that microbiome risk profiles are based on prediction models that are derived from population or individual patient prospective cohort studies, and hence could be more accurate for groups of similar patients "populations or cohorts" than they are for any individual patient. Therefore, it is important to keep in mind that the predicted risks might not translate directly to individual patients, possibly due to limited generalizability in heterogeneous settings.

In the second study, we investigated metabolic health and the diurnal rhythmicity of gut microbiota in a German population cohort of 1,976 individuals. Fecal microbiota profiling identified a risk signature of 13 microbial taxa that showed disrupted diurnal rhythmicity in T2D. A predictive model based on this arrhythmic risk signature could successfully predict individuals at risk of developing T2D with an AUC of 0.78 when body-mass index (BMI) was included [20]. These examples, among others provide evidence for the applicability of microbiome signatures in biomarker discovery for diagnostic and therapeutic purposes, however it is important to note that dysbiotic indices are not standalone measurements and need to be incorporated with additional host-derived and clinical data. Proper standardization and validation require large-scale studies with longitudinal assessment of potential biomarkers and the consideration of possible confounders, such as diet, age, ethnicity, medical history and lastly time of defecation, which all have proven to be involved in microbiome alterations.

Box 2 – *Gut microbiome Biomarker Discovery: Definitions*

A biomarker is ‘‘a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes, or pharmacological responses to a therapeutic intervention.’’ **(NIH,** [76,89]

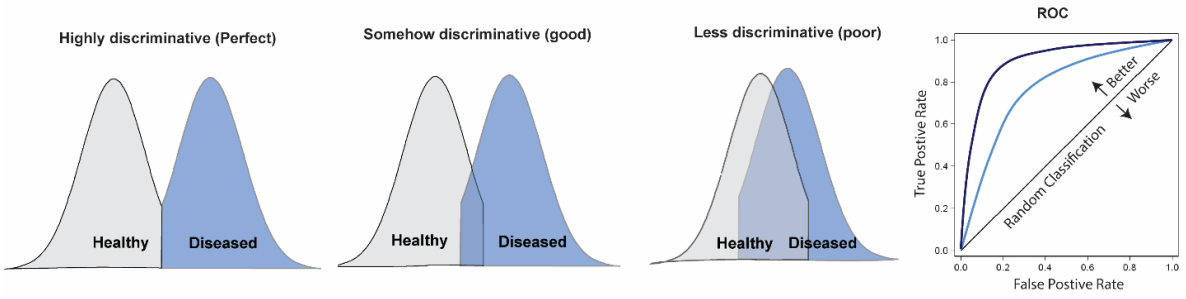
A ‘‘diagnostic’’ biomarker is a characteristic that is directly linked to the etiology of the disease, e.g., elevated blood glucose concentration for the diagnosis of T2D, a marker showing strong correlation with inflammation in IBD (e.g., Fecal calprotectin).

Sensitivity refers to the proportion of individuals who have the disease condition (reference standard positive) and give positive test results [90].

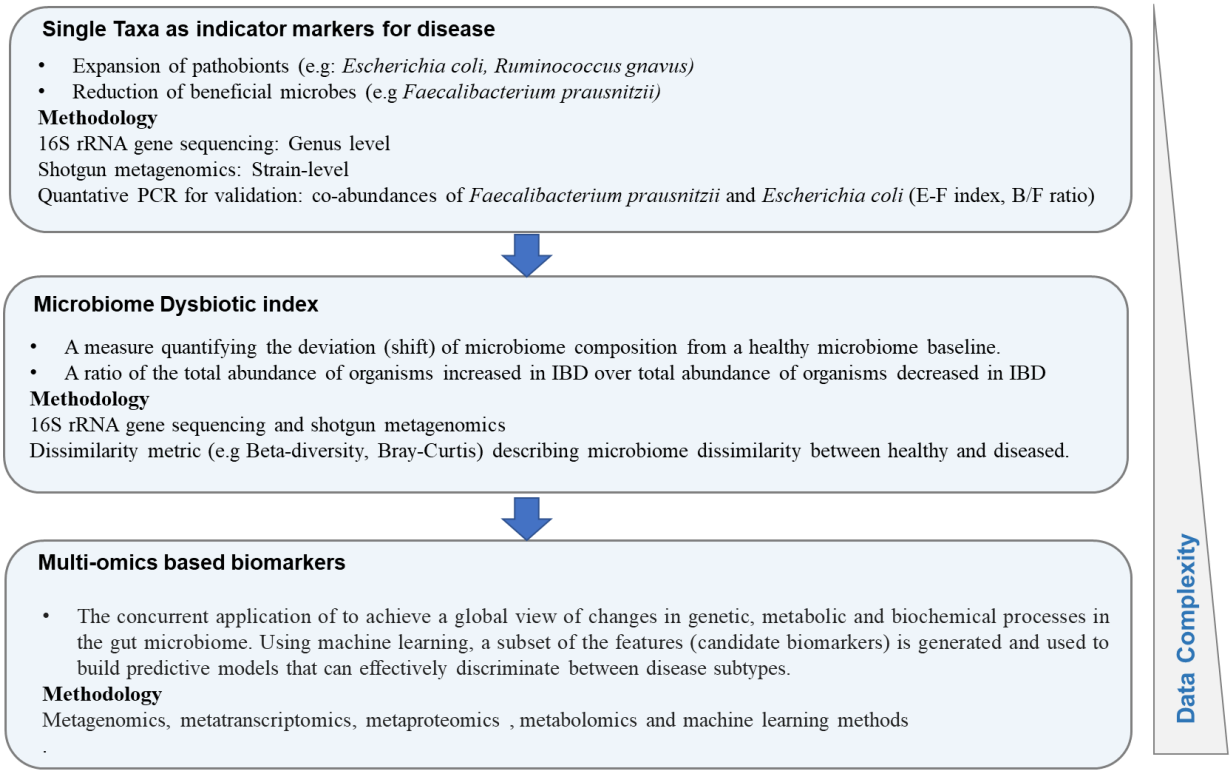
Specificity is the proportion of individuals without the disease condition and give negative test results [90].

Receiver’s Operating Characteristic (ROC) curve plots the specificity and sensitivity of a specific measurement to distinguish health status in a population of subjects.

The area under the curve (AUC) estimates the accuracy of a specific measurement to be a diagnostic tool. The higher the AUC, the better the model is. A perfect model has an AUC=1. This analysis is used to identify the appropriate thresholds to classify a certain population or to diagnose a patient with expected sensitivity and specificity. A ROC curve is generated, and the AUC is calculated.



Box 3 – *Gut microbiome Biomarker Discovery: from single taxa to complex networks*



Inconsistency of disease biomarker prediction across geography and ethnicities

Previous studies have compared gut microbiome profiles between western and rural or non-industrialized populations and identified dramatic differences in their gut microbiome characteristics. In 2015, Martínez *et al.* compared the fecal gut microbiota of individuals from two non-industrialized regions in Papua New Guinea (PNG) with that of United States (US) residents. Interestingly, gut microbiome profiles in PNG showed significantly higher microbial diversity and lower interindividual variations compared to US residents [91]. They also reported many shared bacterial species among PNG and the US with different abundance levels, explained by decreased bacterial dispersal rates in Western populations. In another pioneering study by Falony *et al.*, the authors aimed at identifying a global core microbiome in healthy populations. They showed a decrease in the number of core genera from 17 to 14 when they compared gut microbiome profiles from rural populations in Papua New Guinea, Peru, and Tanzania with that of western populations including Flemish and Dutch cohorts, in addition to UK and US populations [92]. This loss of resident microbes or the concept of "disappearing microbes", as coined by Blaser and Falkow [93] may explain the rising incidence of chronic diseases in the western world. In a more recent study and to robustly validate the generalizability of microbiota-based classifying models of metabolic health, He *at al*. characterized the gut microbiota of 7,009 individuals from 14 districts within 1 province in China and tested the effect of geographic location on the predictive power of the models they generated [64]. Interestingly, host location showed the strongest association with variations in gut microbiome. Further, in a large longitudinal intercontinental study on 531 patients with IBD from Ireland and Canada, geographic location was the major determinant of microbiome variations, yet most (90.3%) of the compositional variance remains unexplained [94].The importance of geography and related environmental exposure are well-illustrated with migration studies, where a strong association between microbiome functional strain diversity and migration was clearly demonstrated. In this context, Vangay *et al*. performed 16S and deep shotgun metagenomic sequencing

on stool samples collected from individuals living in Thailand and in the US, including first- and second-generation immigrants before and after immigration. Intriguingly, US immigration was found to be associated with significant alterations to the gut microbiome, including loss of diversity, loss of bacterial strains, functional loss of fiber degradation and a shift from the *Prevotella* to *Bacteroides* enterotype. Additionally, these perturbations were intensified by obesity and across generations [95].

To examine connections between geographical locations and gut microbial dysbiosis in IBD and T2D, we summarized the changes revealed in selected microbiome association studies from countries around the world. Due to the higher availability of 16S rRNA gene-based sequencing datasets, we focused on cohort studies with 16S microbial profiling, despite the variability in the sequenced 16S variable regions, or the sequencing platform in some cases **(Figure 1).** In case of the IBD stool-derived microbiome, Firmicutes, Proteobacteria and genera including *Fusobacterium*, *Escherichia coli*, *Ruminococcus gnavus* and *Streptococcus* showed consistent increased relative abundance correlating with disease manifestation. On the other hand, *Roseburia*, *Blautia* and *Faecalibacterium* consistently decreased cohorts. In T2D, a decrease in, *Akkermansia muciniphila*, *Clostridium*, *Roseburia* and *Faecalibacterium* was shown in most cases. However substantial divergences in the disease-associated profiles between individuals from different race and ethnicity remain. The relevance of these taxa in disease pathology has been validated in several clinical and translation gnotobiotic experiment, as discussed before[26,62,67,71,85,96,97]. Collectively, these data dictate the necessity for more global studies of human microbiota in different geographic locations across continents to rule out regional associated confounding factors and define specific and individualized microbiome signatures. Recently, studies within the integrative Human Microbiome Project (iHMP) aimed at exploring the link between gut microbiome alterations and the development of IBD or T2D in large cohort population studies [19,75]. Nevertheless, and up to date, most of these studies have predominantly focused on western populations, in most cases from US and Europe, representing at most 1/6th of the world's population. In the recent years, a few national and international projects have been initiated to characterize gut microbiome variations in diverse populations and ethnicities, including studies in Africa, Asia, South America and the Middle East [98]

.

Table 1 Gut microbiome signatures associated with IBD in selected regions across the globe

Region	Population	Biomarker	Risk signature (increased)	Risk signature (decreased)	Sequencing Technology	Statistical Analysis	Validation Cohort	Reference
Spain	Spanish IBD cohort (34 CD +33 UC, 111 HC)	Diagnosis confirmed by endoscopy and histology clinical remission for at least 3 months—defined by (CAI) for UC and (CDAI) for CD. HC were without previous history of chronic disease.	*Fusobacterium*, *Escherichia coli* loss of beneficial microorganisms is more associated with patients with CD than a gain of more pathogenic ones.	*Faecalibacterium*, *Peptostreptococcaceae*, *Anaerostipes*, *Methanobrevibacter*, *Christensenellaceae*, *Collinsella*	16S rRNA gene sequencing of V4 region (Illumina MiSeq)	MaAsLin 80% for CD, using the Spanish and Belgian cohorts, and a specificity of 94.3%, 94.4%%, 89.4% and 90.9% of CD detection versus HC, and patients with anorexia, IBS and UC, respectively.	Belgian CD cohort (n= 187), a Spanish IBS cohort (n=41 patients), a UK healthy twin cohort (n= 1016 samples) and a German anorexic cohort (158) and French cohort of IBD (146 CD and 86 UC) and the 38 HC.)	[18]
France	235 patients with IBD and 38 healthy subjects (HS)	A diagnosis of IBD was defined according to clinical, radiological, endoscopic, and histological criteria.	*Ruminococcus gnavus was increased in ileal CD.* *Streptococcus anginosus in IBD. Aggregatibacter segnis and Actinobacillus* (two members of the Pasteurellaceae family) in IBD flare compared with remission. disease-specific fungal microbiota dysbiosis	*Ruminococcus*, *Coprococcus*, *Blautia*, *Eubacterium* and *Dorea* (IBD), *Roseburia*, *Faecalibacterium*, *Dorea and Blautia* (IBD flare) *Anaerostipes* in IBD and particularly in flare and in ileal CD.	16S rRNA gene sequencing of V3-V5. PGM Ion Torrent	MaAsLin	NA	[45]
USA	85 patients with IBD (43 UC, 42 CD)	disease activity was assessed using the Harvey Bradshaw index for CD (Harvey and Bradshaw, 1980) and simple clinical colitis activity index for UC	***CD:*** *Bifidobacterium longum, Eggerthella, Ruminococcus gnavus, Roseburia inulinivorans, and Veillonella parvula* decreased in patients achieving remission. ***UC:*** *Streptococcus salivarium* increased in patients not achieving remission	relative abundance of *Roseburia inulinivorans* and Burkholderiales at baseline was predictive of week 14 remission.	Illumina based DNA shotgun sequencing	Neural network algorithms (vedoNet) to predict treatment response	External validation was performed in an independent cohort of 20 patients with moderate-to-severe CD or UC initiating therapy with an anti-TNF biologic therapy (infliximab or adalimumab)	[22]

China	72 CD patients, 51 UC patients, and 73 healthy	diagnoses based on standard endoscopic, radiographic, and histologic criteria.	***CD:*** *Streptococcus*, Proteobacteria, *Enterococcaceae* ***UC:*** Bacteroidia, and *Pseudomonadaceae*	*Roseburia*, *Coprococcus*, Clostridiales	16S rRNA gene sequencing of V4 region (Illumina MiSeq)	Random forest classification models were trained on features of the OTU data with 5 repeats of 10-fold cross-validation HC-CD, AUC=0.89 HC-UC, AUC=0.93	RISK and PRISM US cohorts	[75]
USA	447 treatment-naive patients with CD and 221 healthy subjects	newly diagnosed population of pediatric patients with CD	*Pasturellaceae, Veillonellaceae, Neisseriaceae, Fusobacteriaceae* species, and *Escherichia coli* .	*Bacteroides*, Clostridiales, *Faecalibacterium, Roseburia, Blautia, Ruminococcus, and Lachnospiraceae*	16S rRNA gene sequencing of V4 region (Illumina MiSeq)	MaAsLin	NA	[48]
Spain	29 CD patients under HSCT	diagnoses based on standard endoscopic, radiographic, and histologic criteria.	*sulfate-reducing* Gamma- *and* Deltaproteobacteria, *butyrate-producing* Clostridiales, *Enterococcus*, *Megasphaera*, *Campylobacter*, and *Fusobacterium.*	*Akkermansia*, *Barnesiella*, *Oscillibacter*, *Roseburia*, and *Odoribacter.*	16S rRNA gene sequencing of V3-V4 region (Illumina MiSeq)	Random forest classification models were trained on features of the OTU data with 5 repeats of 10-fold cross-validation OTU AUC=0.79 Bacteria-metabolite AUC=0.96	NA	[21]
Denmark	300 UC, 213 CD 30 healthy subjects	CD: HBI-score UC: SCCAI-score	Firmicutes and Proteobacteria	*Dorea*	16S rRNA gene sequencing of V3 region (Illumina MiSeq)		NA	[99]
Germany	62 individuals including twin CD, UC patients and healthy volunteers	CAI, Colitis Activity Index, endoscopic appearance and continuity of inflammation, histology and proven exclusive affection of the colon.	*Lachnospiraceae* and *Ruminococcaceae*	Actinobacteria and Proteobacteria	16S rRNA gene sequencing of V4-V5 region	NA		[100]

Functional Profiling Using Shotgun Metagenomics

Region	Population	Biomarker	Functional Risk signature (increased)	Functional Risk signature (decreased)	Sequencing Technology	Statistical Analysis	Validation Cohort	Reference
USA	N=4 CD N=7 Healthy	Endoscopic, pathologic, or radiographic findings	Modules involved in glycolysis and carbohydrate transport and metabolism (nutrient uptake) ***CD and UC:*** Metabolism of sulfur-containing AA Cysteine increased. Increase in Riboflavin metabolism, glutathione transporters.	Lower abundance of genes involved in lipid metabolism and catabolism Global decrease in nicotinamide, purine, and pyrimidine	Illumina MiSeq (2x150 bp, paired-end)	NA	NA	[101]

				nucleotide biosynthesis in IBD				
USA	N=366 children under 22 years old with CD	Pediatric Crohn's Disease Activity Index (PCDAI)	***CD:*** Top six pathways included sulfur relay systems, galactose metabolism, biosynthesis of siderophores, glycolipid metabolism, glutamine/glutamate metabolism and biosynthesis of Siderophores		Illumina HiSeq	Random Forest on samples at baseline=0.87 using gene pathway data Most predictive pathways: glycerophospholipid metabolism, aminobenzoate degradation, sulfur relay system and glutathione metabolism (increased) and selenocompound metabolism (decreased)	NA	[51]
USA	N=266 samples from N=12 controls and N=20 IBD	CD: CDAI or HBI-score UC: SCCAI-score	***IBD:*** Increase in facultative anaerobe abundance in IBD (*Streptococcus salivarius* and *Streptococcus parasanguinis*), *Ruminococcus gnavus* -Genes involved in Oxidative stress (NADH oxidase and peroxiredoxin-encoding genes) -Genes involved in biosynthesis of Cysteine, acquisition of iron 13 genes involved in sugar utilization and transport	***IBD:*** decrease in (*Blautia obeum* and *Alistipes putredinis*)	Illumina MiSeq (2x150 cycle runs)	NA	80 samples from (HMP) as healthy Data from [51] as IBD validation cohort	[82]
Denmark	N=1,792 participants: 33 IBD, 412 IBS and 1025 control Dutch cohorts (LifeLines DEEP, UMCG IBD, MIDS)	Endoscopic, pathologic, or radiographic findings	***CD:*** 219 taxa (including 152 species) associated with CD ***UC:*** 102 taxa (including 93 species) associated with UC. ***CD:*** an increase in taxa belonging to *Enterobacteriaceae* family. ***IBD***: Perturbations in multiple functional pathways, including pathways involved in amino acid synthesis, vitamins, neurotransmitters and SCFA synthesis	***CD:*** a decrease in taxa belonging to *Lachnospiraceae* and *Ruminococcaceae*. ***UC:*** a decrease in taxa belonging to *Bacteroidaceae* and increase in taxa belonging to *Lachnospiraceae*	Illumina HiSeq	10-fold cross validation To discriminate between IBD and IBS AUC (age+sex+BMI+calprotectin+ top taxa) = 0.90		[102]
USA	N=155 individuals N=68 CD, 53 UC, 34 non-IBD	Endoscopic, pathologic, or radiographic findings	***CD and UC:*** Unclassified *Roseburia* species were significantly elevated *Bifidobacterium breve* and *Clostridium symbiosum* were uniquely enriched in UC.	*Roseburia hominis*, *Doreaformicigenerans*, and *Ruminococcus obeum* were strongly	Illumina HiSeq 2500 platform (101 bp, paired end)	Random Forest 5-fold CV AUC (metabolites+ species) =0.92	Validation cohort from the Netherlands. LifeLines-DEEP (22 control), 43	[103]

			Twelve species were enriched in CD, including *Ruminococcus gnavus, Escherichia coli*, and *Clostridium clostridioforme.*	enriched in non-IBD controls.		Independent Validation AUC (metabolites+species) =0.89	IBD patients (UMCG cohort).	
USA	Longitudinal sampling of 132 patients with IBD, 1,638 stool samples	CD: HBI-score UC: SCCAI-score	*Faecalibacterium prausnitzii and Roseburia hominis* Nicotinuric acid found exclusively in IBD	*Escherichia coli, Ruminococcus torques, Ruminococcus gnavus, Clostridium hathewayi, Clostridium bolteae* Pantothenate and nicotinate (Vitamin B5 and B3) depleted in IBD	Shotgun metagenomics, HiSeq2000 or 2500 2x101 xx		NA	[19]

Table 2 Gut microbiome signatures associated with T2D in selected regions across the globe

Region	Population	Biomarker	Risk signature (increased)	Risk signature (decreased)	Sequencing Technology	Statistical Analysis	Validation Cohort	Reference
Mexico	CARE-In-DEEP Study (N = 427)	oral-glucose tolerance test	*Escherichia coli, Veillonella Blautia, Anaerostipes*		16S rRNA gene sequencing of V4 region (Illumina MiSeq)	Random forest; AUC = 0.69; all taxa Differential gene expression analysis (negative binomial distribution)	NA	[104]
Sweden	N=145 women with normal, impaired or diabetic glucose control.	HBA1C	*Lactobacillus gasseri, Lactobacillus salivarius*	*Desulfurispirillum indicum* *Clostridium beijerinckii* *Clostridium Eklund* *Clostridium botulinum* *Pyramidobacter* *Clostridium thermocellum*	Shotgun metagenomics	Random forest; AUC = [0.60; 0.71]; no. species = [1 ;952]	Chinese population ([105].), random forest; AUC = [0.60; 0.74]; no. species = [1 ;1152]	[66]
Israel	N = 800		*Bacteroides thetaiotaomicron, Alistipes putredinis*	*Eubacterium rectale* *Parabacteroides distasonis* *Roseburia inulinivorans* *Eubacterium eligens*	16S rRNA gene sequencing of V3-V4 region (Illumina MiSeq)	Stochastic gradient boosting regression; 4,000 estimators	Validation cohort of n = 100	[106]

				Bacteroides vulgatus				
Pakistan	N = 60	Fasting blood glucose	Bacteroidetes,Verrucomicrobia Proteobacteria,Elusimicrobia, Acidobacteria, Deferribacteres Gemmatimonadetes,*Porphyromonadaceae, Alistipes marseilloanorexic, Bacillus sporothermodurans, Staphylococcus, Prevotella*	Verrucomicrobia, Elusimicrobia, *Methanogenic archaeon*	16S rRNA gene sequencing of V3-V4 region (Illumina MiSeq)	Kruskal – Wallis test		[107]
China	N = 6,896	MetS =waist>90 cm (male) or waist>85 cm (female), FBG≥6.1 mmol/L (110 mg/dl) or diagnosis of diabetes mellitus, TG≥1.7 mmol/L (150 mg/dl), HDL<1.04 mmol/L (40 mg/dl) SBP/DBP≥130/85 mmHg or previous diagnosis of high blood pressure	Actinobacteria, *Fusobacterium Streptococcus, Lactobacillus*	*Akkermansia*, Synergistes *Methanobrevibacter, Oscillospira,Roseburia, Bifidobacterium*	16S rRNA gene sequencing of V4 region (Illumina MiSeq)	MaAsLin		[108]
China	N = 60	Fasting blood glucose	*Faecalibacterium*, Clostridiales, *Dorea, Clostridiaceae, Lachnospiraceae*	*Bifidobacterium, Parabacteroides, Oscillospira, Bacteroides*	16S rRNA gene sequencing of V3-V4 region (Illumina MiSeq)	Random forest; AUC = 0.90; 50 OTUs		[109][109]
China	Three Chinese cohort studies N = 1,832	Fasting blood glucose or HBA1C (ADA)	*Lactobacillaceae*	*Mogibacteriaceae, Clostridiaceae, Butyrivibrio, Roseburia, Megamonas, Clostridiaceae, Dorea*	16S rRNA gene sequencing of V1-V2 region (Illumina MiSeq)	LightGBM algorithm; AUC = 0.88; 21 features	Cohort 1 N = 203; AUC = 0.87, Cohort 2 N = 7,009; AUC = 0.83	[110]
Denmark	Inter99 study population N = 784	HBA1C	*Lactobacillus, Escherichia coli*	*Roseburia, Subdoligranulum, Intestinibacter*	Shotgun metagenomics			[63]

Africa	Africa America Diabetes Mellitus (AADM) study N = 291	ADA Oral glucose tolerance test	*Peptostreptococcus, Eubacterium, Prevotella, Desulfovibrio*	*Collinsella, Adlercreutzia Anaerostipes, Epulopiscium, Clostridium butyricum, Ruminococcus, Pediococcus*	16S rRNA gene sequencing of V4 region (Illumina MiSeq)	Differential gene expression analysis (negative binomial distribution) PERMANOVA Kruskal-Wallis rank		[111]
Germany	KORA cohort N = 1,976	WHO Oral glucose tolerance test	*Escherichia coli*	*Faecalibacterium prausnitzii, Bifidobacterium longum, Intestinales bartlettii, Coprococcus, Eubacterium rectale*	16S rRNA gene sequencing of V3-V4 region (Illumina MiSeq)	Random forest; AUC = 0.76; 14 OTUs; BMI Geralized linear model (AUC = 0.79; 13 OTUs; BMI)	FoCus cohort (N = 1,529); TwinsUK cohort (N = 1,399)	[20]
Germany	Popgen cohort (N= 436, and FoCus cohort (N = 844)	HOMA-IR > 5.0	*Bacteroides thetaiotaomicron*	*Clostridium sensu stricto, Escherichia coli, Romboutsia, Barnesiella, Pseudoflavonifractor, Veillonella, Roseburia*	16S rRNA gene sequencing of V1-V2 region (Illumina MiSeq)	MaAsLin	17 associations identified here, 15 were among the analysed taxa in the independent SHIP cohort (N = 800), and of these, 7 retained a significant association with obesity	[112]
United Kingdom	TwinUK cohort (N = 977)	Overweight BMI 25 – 30, obese BMI > 30		*Christensenellaceae*	16S rRNA gene sequencing of V4 region (Illumina MiSeq)	t- test. Benjamini-Hochberg, Wilcoxon rank sum one sided (lean higher)	NA	[73]

USA	CDC cohort N = 451, NYU study N = 239	Overweight (BMI 18 – 25) N = 246, Obese (> 30) N = 142	*Gemellaceae, Streptococcus, Blautia*	Parabacteroides, *Clostridiaceae, Lachnospiraceae, Ruminococcaceae*, Clostridiales, *Oscillospira*	16S rRNA gene sequencing V4 Illumina MiSeq with a 300-cycle (2 × 151 bp)	weighted UniFrac distance, PCoA and CAP, PERMANOVA, adjusting for age, sex, polyp status, and study, DESeq2, RF (1,825 OTU) repeated (20 times) 5-fold cross-validatio - optimal model included 49 OTUs and had an AUC of 0.81, testing sets was 0.72 and 0.68	Test Set: NYU study N = 239 Validation set = Baxter et al. 402 subjects	[113]

Functional Profiling Using Shotgun Metagenomics

Region	Population	Biomarker	Functional Risk signature (increased)	Functional Risk signature (decreased)	Sequencing Technology	Statistical Analysis	Validation Cohort	Reference
Sweden	N=46 T2D N=442 pre-T2D N=53 controls	Oral-glucose tolerance test (oGTT) Finnish diabetes risk score (FINDRISC)	- 118 metagenomic species significantly altered -Two component systems -Fructose and Mannose metabolism -Pentose phosphate pathway -BCAA synthesis -B group Vitamin biotin metabolism		Illumina HiSeq 4000 (150 bp; paired-end)	Random forest; AUC = 0.70 True Prediction, AUC=0.64 118 features selected	Swedish Cardiopulmonary Bioimage Study (SCAPIIS)	[114]
Swede	N=53 T2D N=49 IGT N=43 controls	Fasting glucose and HBA1C	Four members of *Lactobacillus* MGC model identified *Roseburia* and *Faecalibacterium prausnitzii* as highly discriminant for T2D - 7 of the T2D-enriched KEGG orthologues markers -starch and glucose metabolism -fructose and mannose metabolism -ABC transporters for amino acids -ions and simple sugars -cysteine and methionine metabolism	Five members of *Clostridium*	Illumina HiSeq 2000	RF and ten-fold cross-validation on microbial cluster AUC = 0.83 on species AUC= 0.71	Chinese population ((Qin et al., 2012).),	[66]

Netherlands	N= 1,179 LL-DEEP sample	oral glucose tolerance test (oGTT)	GABA degradation activity PWY-5022		Illumina HiSeq platform	Random Forest on log transformed data Microbial pathway involved in 4 aminobutanoate (GABA) degradation Aminobutanoate degradation V) on increased insulin secretion	Genotype and phenotype data from the UK Biobank, a study of 500,000 subjects from the United Kingdom aged 45–65 years of ag	115
China, Suzhou	N = 77 T2D N = 80 Pre N = 97 controls		The most discriminatory MLG for separating TN-T2D and NGT was *Akkermansia muciniphila*. *Faecalibacterium prausnitzii* and *Escherichia coli* both showed to be important in separating Pre samples from T2D and healthy samples TMAO producing enzyme Dimethylaniline monooxygenase Higher amylase (AMY1) levels Antimicrobial cathepsin G	higher levels of four AMPs in controls GTPase-activating-like protein (IQGAP1) and unconventional myosin-Ic (MYO1C) were uniquely identified in the healthy group lower levels of proteases (trypsin and chymotrypsin and their precursors) and lipases	BGISEQ-500 sequencing for metagenomics (single-end; read length of 100 bp)	Random Forest, five-fold cross validation AUC (T2D vs. Pre = 0.90) AUC (PRE vs. healthy = 0.88) AUC (T2D vs. Controls = 0.94)		65
China	N=71 T2D N=74 controls		*Bacteroides caccae,* *Clostridium hathewayi,* *Clostridium ramosum,* *Clostridium symbiosum,* *Eggerthella lenta,* *Escherichia coli* Membrane sugar transport BCAA transport methane metabolism xenobiotics degradation and metabolism sulphate reduction	*Clostridiales sp. SS3/4,* *Eubacterium rectale,* *Faecalibacterium prausnitzii, Roseburia intestinalis,* *Roseburia inulinivorans,* *Haemophilus parainfluenzae* -level of bacterial chemotaxis -flagellar assembly -butyrate biosynthesis --metabolism of cofactors and vitamins	HiSeq 2000	Random Forest AUC = 0.81 50 gut microbial gene marker		116

6

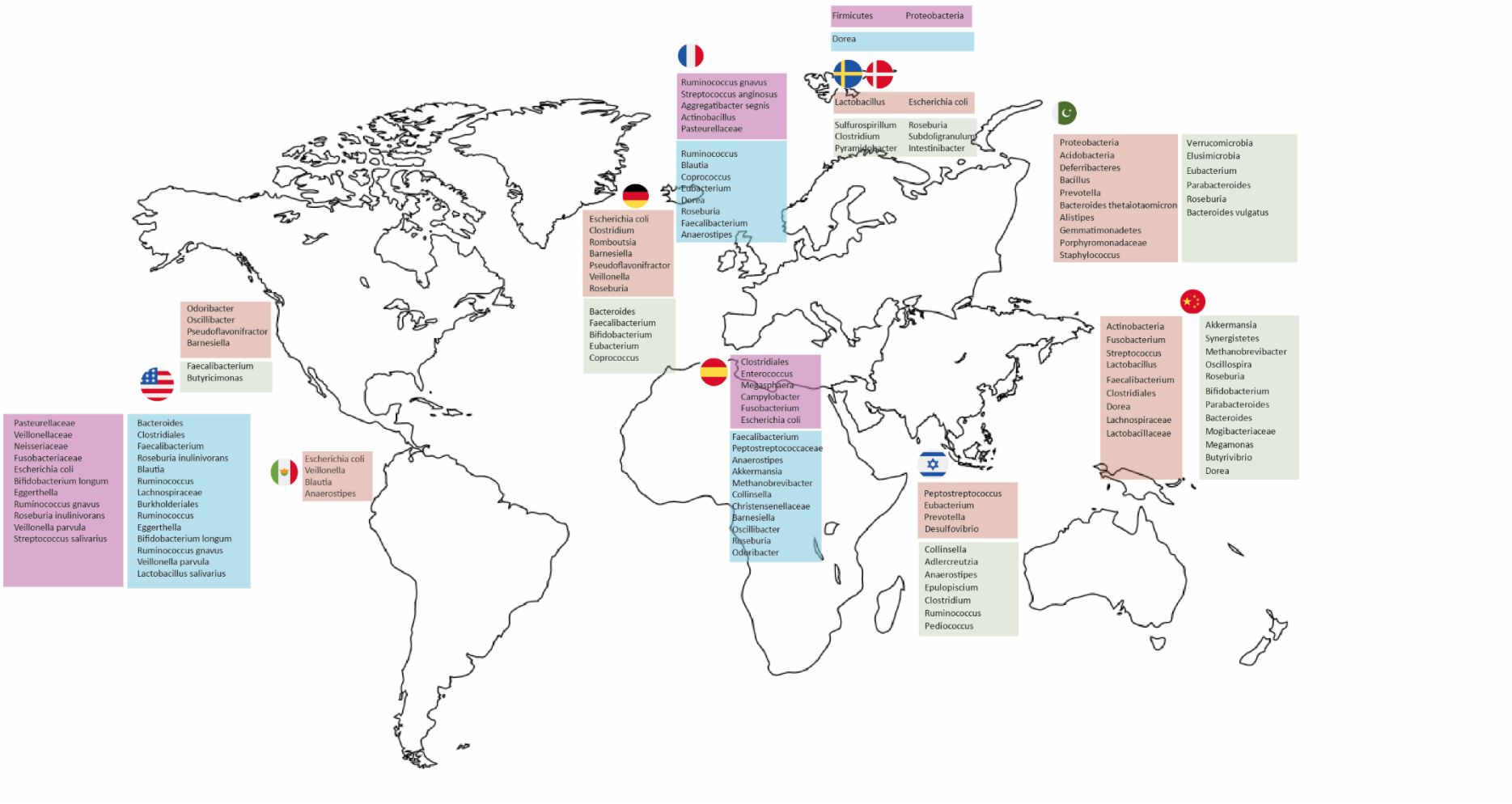


Figure 1. Bacterial risk signatures in gut microbiome of patients with IBD (enriched: magenta, decreased: blue) or T2D (enriched: orange, decreased: green) in diverse populations across the globe. Box color: enriched or decreased taxa (magenta, blue or orange, green) for IBD and T2D, respectively

In all these studies, the authors tried to identify disease-associated microbial risk profiles contributing to the development of IBD or T2D. However, it clearly remains challenging to reach a common consensus on disease-related bacterial taxa with a disease diagnostic value. Despite the heterogeneity, few bacterial taxa were found to be commonly implicated across different studies **(Figure 1).** For instance, *Akkermansia*, *Eubacterium rectale*, *Alistipes* and *Faecalibacterium prausnitzii* were found to be positively correlated with improved metabolic health in multiple reports [66,105,109,112,117]. Similarly, in IBD, an overabundance of *Escherichia coli, Enterococcus, Fusobacterium, Ruminococcus gnavus* and *Streptococcus* or a reduction in *Feacalibacterium prausnitzii*,*Roseburia*, members of *Ruminococcus* genus were described for IBD-associated dysbiosis [18,19,21,45,48,100,118]. Intriguingly, some taxa showed similar trends in IBD and T2D. For instance, an overabundance of family *Christensenellaceae* and *Escherichia coli* were linked to CD and T2D-associated dysbiosis [18,20,109], posing questions about the specificity of the available microbiome signatures for discriminating between different disease entities. Looking at overlapping IBD and T2D-ssociated core signatures showed similar or opposite trends. We focused on the 16S studies summarized in **Tables 1 and 2** and described bacterial signatures associated with each disease entity according to the highest reported taxonomic level (bacterial phylum, family, genus and species) **(Figure 2).**

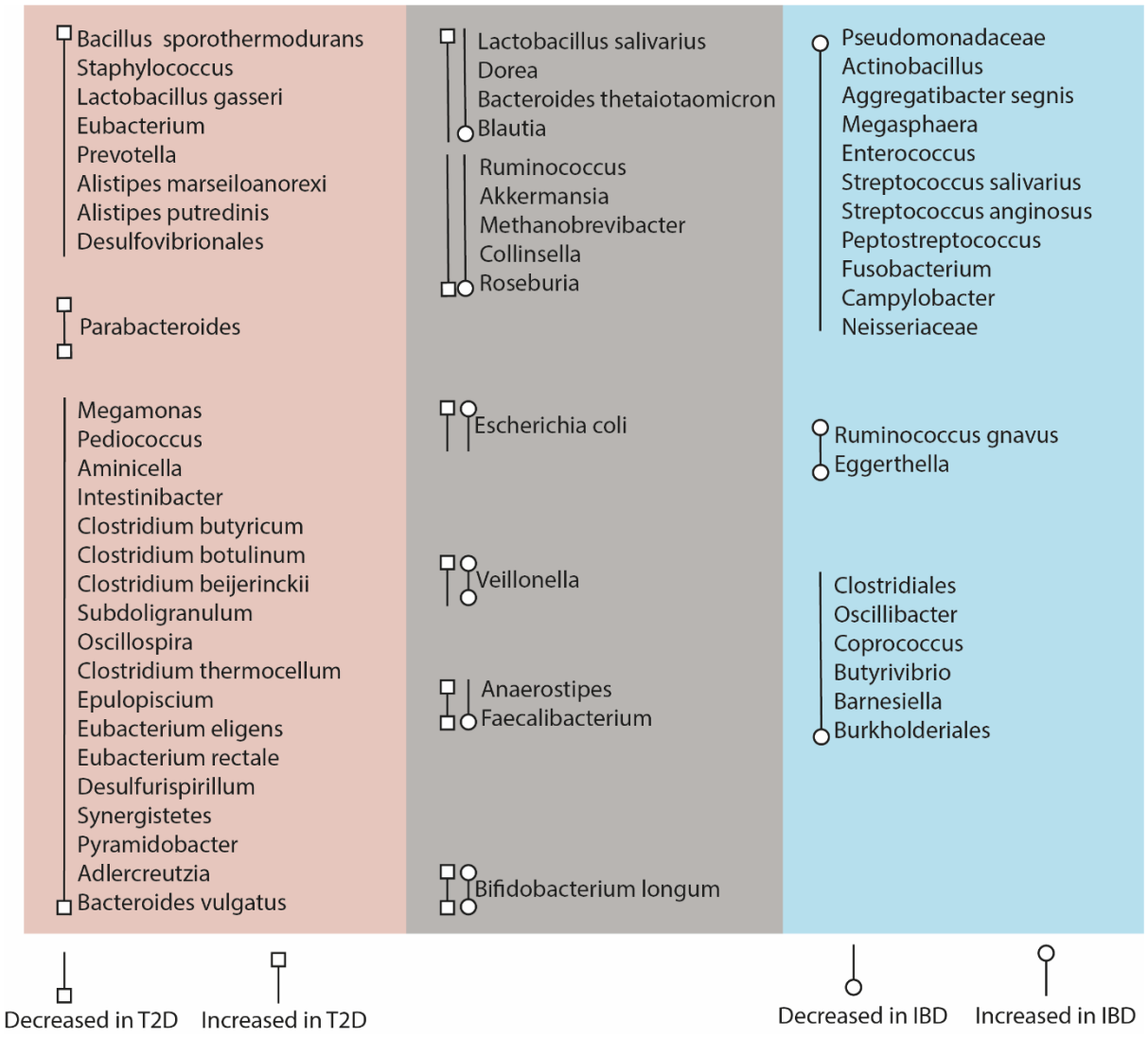


Figure 2. IBD and T2D overlapping global risk profile. Bacterial genera signatures associated with IBD, T2D and common shared genera based on IBD and T2D 16S rRNA gene sequencing microbiome association studies summarized in Table 1 and 2.

Metabolomics for IBD and T2D biomarker discovery

In the pursuit of identifying disease biomarkers, metabolites serve as the most proximal indicators of disease activity and are strongly linked to the underlying regulatory signals that modulate disease mechanisms. In fact, both the metabolome and microbiome fluctuate in relation to endogenous and exogenous factors such as diet, environment, aging, and health condition [119]. Numerous studies have reported substantial alterations in the gut metabolite profiles from patients with IBD [19,21,103,120,121] or T2D [75,122–124]. For instance, reduced levels of medium chain fatty acids, such as pentanoate and hexanoate [125] and reduced levels of vitamin B [126] were reported in fecal metabolome from IBD patients. Conversely, increased levels of amino acids, amines, and carnitines were reported in the feces and serum of adult and pediatric IBD patients, respectively [121,127] . A landmark study by Marchesi *et al*. showed that metabolite profiling could discriminate IBD patients from healthy individuals [128]. This was followed by numerous studies that consistently showed that the metabolite phenotype of IBD patients differ from healthy individuals [103,120,129,130]. Interestingly, metabolite profiling could also discriminate different disease subtypes, such as CD and UC [128,130], and further stratified CD to ileal or colonic inflammation [129]. Similarly, patients with T2D demonstrated altered metabolic activity [131,132] and serum levels of branched-chain and aromatic amino acids, such as leucine, isoleucine, valine, phenylalanine, tyrosine and tryptophan showed association with insulin resistance, obesity and the risk of T2D in multiple reports [133–135]. Metabolite profiling of T2D patients revealed significant associations between specific bacterial metabolites and disease onset [75,123,136,137].

As an example of promising metabolite biomarkers, tryptophan metabolism has attracted attention as a candidate biomarker due to its association with inflammatory and metabolic disease development in both human and mouse studies [138–141]. Tryptophan is an essential amino acid acquired from the diet, and is mainly absorbed in the small intestine, yet a small fraction is catabolized to indole metabolites in the colon [142]. Tryptophan metabolism and downstream cellular signaling have been reviewed by many [142–144] and will not be extensively discussed in this review. In a recent study, Chen *et al*. assessed the association of tryptophan with the risk of T2D development and they evaluated its performance as sole biomarker or in combination with existing amino acid biomarkers in a Chinese population [140]. In this study, they quantitively measured the baseline fasting serum tryptophan concentrations in 51 subjects who developed diabetes and 162 subjects who remained metabolically healthy 10 years later. Higher levels of tryptophan at baseline were associated with a higher risk of T2D development. Beyond associations, the predictive modelling of tryptophan as disease biomarker was comparable to the 5 existing amino acids in discriminating between T2D and non-T2D individuals. Noteworthy, prior reports showed that different amino acids could classify T2D patients from different populations with varying accuracy. For example, Phenylalanine and valine showed better performance in American populations [134], while tyrosine showed higher accuracy in South Asian populations [145], pointing to the importance of regional specific biomarker in achieving higher diagnostic accuracy.

Microbiome-based biomarkers versus clinical biomarkers

A valuable biomarker must contribute additional classification power to clinically relevant information. Fecal biomarkers provide a suitable target for mucosal disease diagnostics, given that the fecal stream is in direct contact with the intestinal mucosa. Fecal calprotectin (Fcal), a granulocyte-derived cytosolic protein detected in stool is the most utilized biomarker for inflammatory disorders. Schoepfer *et al*. showed a strong correlation between the severity of inflammation and Fcal levels [146]. Further, a number of reports confirmed the ability of Fcal to detect endoscopic inflammation with a sensitivity ranging between (70-100%) and a specificity of (44-100%), explained by the variations in the selected cut-off values applied in each study [147,148]. Nevertheless, elevated levels of Fcal are not specific for IBD but rather reflect inflammatory conditions also associated with other intestinal and metabolic pathologies (e.g., IBS, gastrointestinal malignancies, obesity and T2D). For instance, gut microbiota metagenomic profiling in 1792 individuals could distinguish IBD from IBS and machine learning algorithms showed improved IBD vs IBS prediction accuracy to AUC 0.91) compared to (AUC=0.80) based on Fcal [102].

Importantly, the model reached the highest prediction accuracy by combining Fcal with the top 20 selected taxa (AUC=0.93), suggesting that integrating clinical, and microbial biomarkers improves diagnostics accuracy. An example of a combined biomarker approach has recently been used to predict response to therapy in patients with IBD [22]. While baseline clinical data, including serological, endoscopic, and clinical markers, were insufficient in predicting remission (AUC=0.62), the use of taxonomic and metabolic profiles improved the diagnostic power to (AUC=0.72) and (AUC=0.74), respectively. Further, Dubinsky and colleagues showed that Fcal alone classified patients with Pouchitis or with normal Pouch with an AUC of 0.63. In contrast, the microbiome species model (with or without Fcal as an additional predictor) achieved an AUC of 0.78, confirming a superior diagnostic value of microbial profiles to Pouchitis classification [149].

In the diagnosis of T2D, serological biomarkers for impaired glucose metabolism in patients with T2D include fasting plasma glucose (FPG), 2-h plasma glucose (2-h PG) in a 75-g oral glucose tolerance (OGTT), or the presence of glycated haemoglobin (HBA1C) [150]. Combining biomarkers for predictive modelling of T2D has been shown in a recent study by Wu *et al.* using data from two Swedish cohorts [114]. Multivariate analyses demonstrated a strong correlation between insulin resistance and microbiome variations. Interestingly, using a microbiome-based machine learning model to distinguish between individuals with the lowest and the highest insulin resistance in the validation cohort yielded an AUC (0.78), suggesting that the gut microbiota is an important modifier of T2D progression. In fact, while a broad range of biomarkers have been proposed for T2D diagnosis, most of them fail to capture the disease complexity or to grasp both microbial and metabolic alterations. In this regard, metabolite biomarkers have been used in combination with established risk factors to significantly improve disease classification [151,152].

Mechanistic implications of microbiome signatures

The need to understand the functional role and specificity of single bacterial taxa (pathobiont) [153,154] or complex dysbiotic microbial communities (dysbiosis)[155–157] is essential to resolve mechanisms of microbe host interactions in the pathogenesis of IBD or T2D. In this context, functional alterations of the gut microbiome potentially represent the consequential changes of host adaptations. A causal link of gut microbiota to multiple diseases has been demonstrated in gnotobiotic mouse experiments [21,61,156,158–162]. Germ-free mouse models are selectively colonized with single bacterial strains, minimal bacterial consortia or defined complex gut microbial ecosystems from human stool or other donor material to study their impact on host phenotype. In IBD, mono-association of germ-free mouse models with a variety of commensal bacteria, including *Escherichia coli*, *Enterococcus faecalis*, *Bacteroides vulgatus*, and *Bilophila wadsworthia* allowed us to understand underlying mechanisms of disease initiation or protection[163–165]. Building complexity through the generation of well-characterized minimal bacterial consortia (e.g., SIHUMI[166] and Oligo MM[12 167]) provided the means to investigate complex mechanisms of host-microbe and microbe-microbe cross-talk under physiological and pathological conditions. In addition, colonization of germ-free mouse models with human fecal microbiota (also known as humanized mice or human microbiome-associated mice) has been used extensively as a translation tool to understand mechanisms of complex pathologies, including IBD, T2D, obesity[117,156], asthma, and malnutrition[21,59,117,156,158–160,168–170].

We recently showed that gut bacteria are required for driving inflammation in a colitis mouse model and is associated with the risk of relapse in CD patients [21]. Despite the known limitations of incomplete human bacterial transfer into germ-free mice [171], we captured key features of the disease-associated microbiome signatures. The transfer of gut microbiota into germ-free mice resulted in successful transfer of different disease states and revealed shared functional metabolic pathways implicated in inflammation. These shared patterns could serve as signatures for better classification of disease activity. Similarly, previous work showed that glucose tolerance [172] and insulin resistance [156,173] are influenced by gut microbiome composition, verified by a series of fecal microbiota transplantation trials in gnotobiotic mouse models[156,157]. At present, microbiome research is moving swiftly beyond the mere description of microbial community structure and disease association, towards a deeper understanding of the causative role of gut bacteria to the pathogenesis of complex chronic disorders **(Figure 3).** As

such, defining the functional capacity of a given microbial signature can be achieved by metagenomics and metabolomics interrogation. Strain-level shotgun metagenomic sequencing, can be used to identify strains, including those keystone species that are present in low abundance yet have an important role in disease development, and to infer bacterial metabolic pathways, microbial interactions, and microbial metabolites that affect host physiology. These collective efforts would enhance microbiome modelling and advance the development of microbiota-based signatures or risk profiles that can be utilized in clinical settings.

Conclusions

Over the past decades, evidence from human and mouse studies revealed a fundamental role of the intestinal microbiome in the pathogenesis of inflammatory and metabolic diseases, such as IBD and T2D. Changes in the structure and function of the gut microbial ecosystem (dysbiotic microbiome signatures) have been associated with disease activity, risk of relapse or response to therapy. Nevertheless, the multifactorial nature of most of these complex pathologies and the existence of a variety of confounding factors affecting human studies stand as a major challenge for the implementation of microbiome signatures for diagnosis, prognosis, or decision on therapy. In this context, the scientific community needs to move from correlation to causation. Beyond sequencing the identification, isolation and cultivation of functionally relevant bacterial strains and their metabolites is needed. To achieve this goal, the establishment and use of well-defined *in vivo* gnotobiotic mouse models provide fundamental information on the impact of microbial composition on host physiology and disease susceptibility. To address the heterogeneity and inter-individual variation in microbiome signatures identification, dense microbiome sampling and disease modelling across populations and ethnicities should be performed to improve predictive models' generalizability. Thus, the stratification of population and patient cohorts is necessary to improve individual disease risk assessment. To ensure reproducibility and comparability between microbiome studies, the specificity and sensitivity of microbiome signatures need to be assessed and validated in well-characterized multi-centered cohorts.

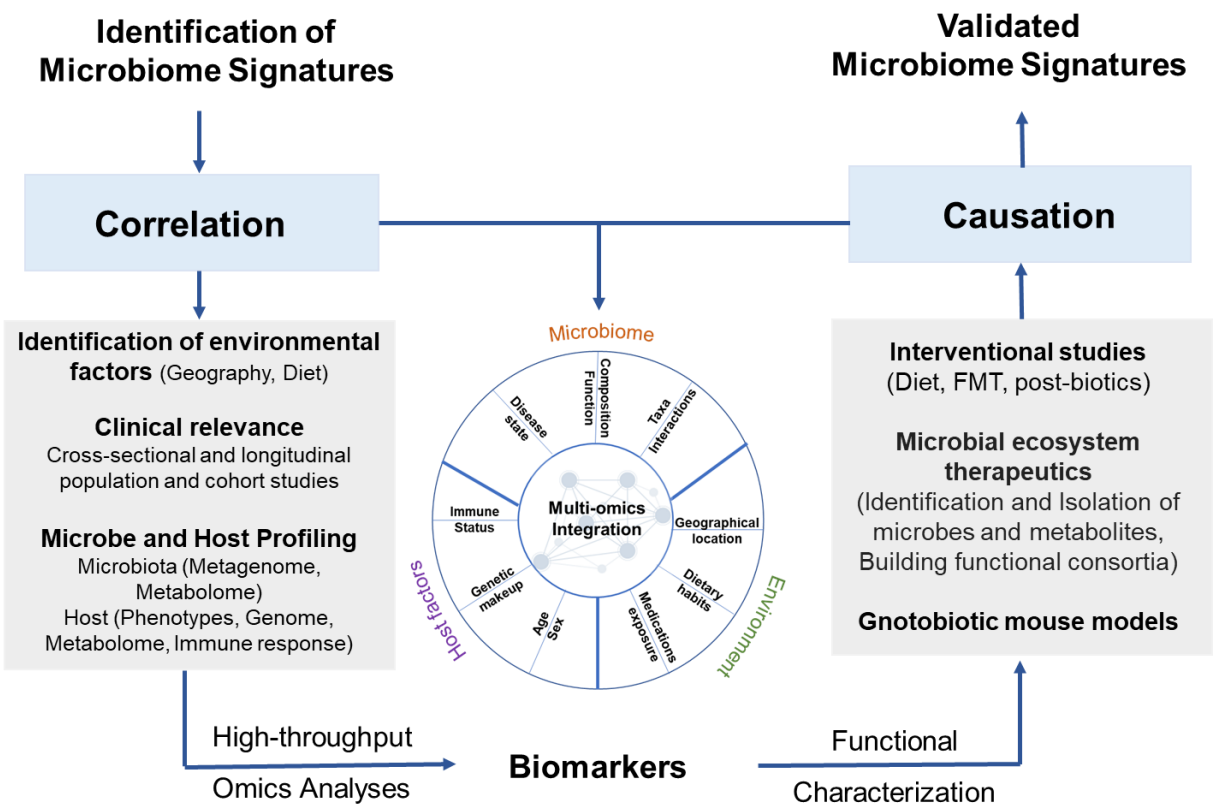


Figure 3 Moving beyond correlation to causation and microbiome signatures discovery. To investigate the causative role the gut microbiota in disease pathogenesis, cross-sectional or longitudinal population- based and patient cohort studies are carried out. Microbial, environmental, as well as host-related factors need to be considered when establishing correlations between disease entity and microbiome structure or function. Multi-omics data are generated based on samples collected from both host (phenotype, genotype, metabolome, transcriptome) and luminal or mucosal-associated microbiome (composition, metabolic and genetic functional pathways) using high-throughput omics technologies. Candidate microbiome-based biomarkers are identified using complex computational and machine-learning tools. Functional validation of candidate biomarkers is achieved through comprehensive *in vivo*, *in vitro* and pre-clinical studies. Isolation and identification of candidate bacterial taxa is performed for biobanking and for subsequent *in vivo* targeted mechanistic studies in gnotobiotic mice. Colonization of germ-free mice with single bacterial taxa or with synthetic minimal consortia could give insight into the causative role of specific microbes and the underlying microbe-host interactions. Intervention studies using FMT, or different dietary interventions help to validate the clinical relevance of the identified microbiome signatures.

References

1. Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400 (2018).
2. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host[ndash]bacterial mutualism in the human intestine. *Science (80-.).* **307**, 1915–1920 (2005).
3. Sender, R., Fuchs, S. & Milo, R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* **164**, 337–340 (2016).
4. Berg, G. *et al.* Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**, 103 (2020).
5. Kaplan, G. G. The global burden of IBD: from 2015 to 2025. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 720–727 (2015).
6. Saeedi, P. *et al.* Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res. Clin. Pract.* **157**, (2019).
7. Blüher, M. Obesity: global epidemiology and pathogenesis. *Nat. Rev. Endocrinol.* **15**, 288–298 (2019).
8. Chakaroun, R. M., Massier, L. & Kovacs, P. Gut Microbiome, Intestinal Permeability, and Tissue Bacteria in Metabolic Disease: Perpetrators or Bystanders? *Nutrients* **12**, 1082 (2020).
9. Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
10. de Lange, K. M., Moutsianas, L. & Lee, J. C. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
11. Billings, L. K. & Florez, J. C. The genetics of type 2 diabetes: what have we learned from GWAS? *Ann. N. Y. Acad. Sci.* **1212**, 59–77 (2010).
12. Yang, S. K. *et al.* Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. *Gut* **63**, 80–87 (2014).
13. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42**, 1118–1125 (2010).
14. Zhou, M. *et al.* Cause-specific mortality for 240 causes in China during 1990–2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. *Lancet* **387**, 251–272 (2016).
15. Momozawa, Y. *et al.* IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* **9**, 2427 (2018).
16. Qi, Q., Wang, X., Strizich, G. & Wang, T. Genetic Determinants of Type 2 Diabetes in Asians. *Int. J. Diabetol. Vasc. Dis. Res.* **2015**, 10.19070/2328-353X-SI01001 (2015).
17. Yilmaz, B. *et al.* Microbial network disturbances in relapsing refractory Crohn's disease. *Nat. Med.* **25**, 323–336 (2019).
18. Pascal, V. *et al.* A microbial signature for Crohn's disease. *Gut* **66**, 813–822 (2017).
19. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).
20. Reitmeier, S. *et al.* Arrhythmic Gut Microbiome Signatures Predict Risk of Type 2 Diabetes. *Cell Host Microbe* (2020). doi:10.1016/j.chom.2020.06.004
21. Metwaly, A. *et al.* Integrated microbiota and metabolite profiles link Crohn's disease to sulfur metabolism. *Nat. Commun.* **11**, 4322 (2020).
22. Ananthakrishnan, A. N. et al. Gut Microbiome Function Predicts Response to Anti-integrin Biologic Therapy in Inflammatory Bowel Diseases. *Cell Host Microbe* **21**, 603-610.e3 (2017).
23. Wang, H. *et al.* Promising Treatment for Type 2 Diabetes: Fecal Microbiota Transplantation Reverses Insulin Resistance and Impaired Islets. *Front. Cell. Infect. Microbiol.* **9**, 455 (2020).
24. Sokol, H. *et al.* Fecal microbiota transplantation to maintain remission in Crohn's disease: a pilot randomized controlled study. *Microbiome* **8**, 12 (2020).
25. Costello, S. P. *et al.* Systematic review with meta-analysis: faecal microbiota transplantation for the induction of remission for active ulcerative colitis. *Aliment. Pharmacol. Ther.* **46**, 213–224 (2017).
26. Vrieze, A. *et al.* Transfer of Intestinal Microbiota From Lean Donors Increases Insulin Sensitivity in Individuals With Metabolic Syndrome. *Gastroenterology* **143**, 913-916.e7 (2012).
27. Marchesi, J. R. & Ravel, J. The vocabulary of microbiome research: a proposal. *Microbiome* **3**, 31 (2015).
28. Tamboli, C. P. Dysbiosis in inflammatory bowel disease. *Gut* (2004). doi:10.1136/gut.53.1.1
29. Bello, M. G. D., Knight, R., Gilbert, J. A. & Blaser, M. J. Preserving microbial diversity. *Science* **362**, 33–34 (2018).
30. Chow, J., Tang, H. & Mazmanian, S. K. Pathobionts of the gastrointestinal microbiota and inflammatory disease. *Curr. Opin. Immunol.* **23**, 473–480 (2011).

31. Round, J. L. & Mazmanian, S. K. The gut microbiome shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9**, 313–323 (2009).
32. Jurjus, A. *et al.* Inflammatory bowel disease, colorectal cancer and type 2 diabetes mellitus: The links. *BBA Clin.* **5**, 16–24 (2015).
33. Verdugo-Meza, A., Ye, J., Dadlani, H., Ghosh, S. & Gibson, D. L. Connecting the Dots Between Inflammatory Bowel Disease and Metabolic Syndrome: A Focus on Gut-Derived Metabolites. *Nutrients* **12**, 1434 (2020).
34. Gupta, A. & Khanna, S. Fecal Microbiota Transplantation. *JAMA* **318**, 102 (2017).
35. Quraishi, M. N. *et al.* Systematic review with meta-analysis: the efficacy of faecal microbiota transplantation for the treatment of recurrent and refractory Clostridium difficile infection. *Aliment. Pharmacol. Ther.* **46**, 479–493 (2017).
36. Kootte, R. S. *et al.* Improvement of Insulin Sensitivity after Lean Donor Feces in Metabolic Syndrome Is Driven by Baseline Intestinal Microbiota Composition. *Cell Metab.* **26**, 611-619.e6 (2017).
37. van Lier, Y. F. *et al.* Donor fecal microbiota transplantation ameliorates intestinal graft-versus-host disease in allogeneic hematopoietic cell transplant recipients. *Sci. Transl. Med.* **12**, eaaz8926 (2020).
38. Paramsothy, S., Kamm, M. & Kaakoush, N. Multidonor intensive faecal microbiota transplantation for active ulcerative colitis: A randomised placebo-controlled trial. *Lancet* **389**, 1218–28 (2017).
39. Rossen, N. G. *et al.* Findings From a Randomized Controlled Trial of Fecal Transplantation for Patients With Ulcerative Colitis. *Gastroenterology* **149**, 110-118.e4 (2015).
40. Moayyedi, P. *et al.* Fecal Microbiota Transplantation Induces Remission in Patients With Active Ulcerative Colitis in a Randomized Controlled Trial. *Gastroenterology* **149**, 102-109.e6 (2015).
41. Sood, A. *et al.* Acceptability, tolerability, and safety of fecal microbiota transplantation in patients with active ulcerative colitis (AT&S Study). *J. Gastroenterol. Hepatol.* **35**, 418–424 (2020).
42. Xiang, L. *et al.* Efficacy of faecal microbiota transplantation in Crohn's disease: a new target treatment? *Microb. Biotechnol.* **13**, 760–769 (2020).
43. Mocanu, V. *et al.* Fecal microbial transplantation and fiber supplementation in patients with severe obesity and metabolic syndrome: a randomized double-blind, placebo-controlled phase 2 trial. *Nat. Med.* **27**, 1272–1279 (2021).
44. de Groot, P. *et al.* Donor metabolic characteristics drive effects of faecal microbiota transplantation on recipient insulin sensitivity, energy expenditure and intestinal transit time. *Gut* **69**, 502 LP – 512 (2020).
45. Sokol, H. *et al.* Fungal microbiota dysbiosis in IBD. *Gut* **66**, 1039–1048 (2017).
46. Sokol, H. *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci.* **105**, 16731–16736 (2008).
47. Schirmer, M. *et al.* Compositional and Temporal Changes in the Gut Microbiome of Pediatric Ulcerative Colitis Patients Are Linked to Disease Course. *Cell Host Microbe* **24**, 600-610.e4 (2018).
48. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**, 382–392 (2014).
49. Sokol, H. & Seksik, P. The intestinal microbiota in inflammatory bowel diseases: Time to connect with the host. *Curr. Opin. Gastroenterol.* **26**, 327–331 (2010).
50. Morgan, X. C. *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* **13**, R79–R79 (2012).
51. Lewis, J. D. *et al.* Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. *Cell Host Microbe* **18**, 489–500 (2015).
52. Vila, A. V. *et al.* Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. *Sci. Transl. Med.* **10**, (2019).
53. Schirmer, M., Garner, A., Vlamakis, H. & Xavier, R. J. Microbial genes and pathways in inflammatory bowel disease. *Nat. Rev. Microbiol.* **17**, 497–511 (2019).
54. D'Haens*, G. R. *et al.* Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded ileum. *Gastroenterology* **114**, 262–267 (1998).
55. Rutgeerts, P. *et al.* Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum. *Lancet (London, England)* **338**, 771–774 (1991).
56. Perencevich, M. & Burakoff, R. Use of antibiotics in the treatment of inflammatory bowel disease. *Inflamm. Bowel Dis.* **12**, 651–664 (2006).
57. Hörmannsperger, G., Schaubeck, M. & Haller, D. Intestinal Microbiota in Animal Models of Inflammatory Diseases. *ILAR J.* **56**, 179–191 (2015).
58. Ahmed, M., Metwaly, A. & Haller, D. Modeling microbe-host interaction in the pathogenesis of Crohn's disease. *Int. J. Med. Microbiol.* **311**, 151489 (2021).
59. Nagao-Kitamoto, H. *et al.* Functional Characterization of Inflammatory Bowel Disease-Associated Gut Dysbiosis in Gnotobiotic Mice. *Cell. Mol. Gastroenterol. Hepatol.* **2**, 468–481 (2016).
60. Sellon, R. K. *et al.* Resident enteric bacteria are necessary for development of spontaneous colitis and

immune system activation in interleukin-10-deficient mice. *Infect. Immun.* **66**, 5224–5231 (1998).
61. Schaubeck, M. *et al.* Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut* **65**, 225–237 (2016).
62. Round, J. L. & Mazmanian, S. K. Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12204–12209 (2010).
63. Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262–266 (2015).
64. He, Y. *et al.* Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* **24**, 1532–1535 (2018).
65. Zhong, H. *et al.* Distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naïve type 2 diabetics. *EBioMedicine* **47**, 373–383 (2019).
66. Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103 (2013).
67. Shin, N.-R. *et al.* An increase in the Akkermansia spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice. *Gut* **63**, 727–735 (2014).
68. Org, E. *et al.* Genetic and environmental control of host-gut microbiota interactions. *Genome Res.* **25**, 1558–1569 (2015).
69. Depommier, C. *et al.* Supplementation with Akkermansia muciniphila in overweight and obese human volunteers: a proof-of-concept exploratory study. *Nat. Med.* **25**, 1096–1103 (2019).
70. Everard, A. *et al.* Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9066–9071 (2013).
71. Plovier, H. *et al.* A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat. Med.* **23**, 107–113 (2017).
72. Gilijamse, P. W. *et al.* Treatment with Anaerobutyricum soehngenii: a pilot study of safety and dose–response effects on glucose metabolism in human subjects with metabolic syndrome. *npj Biofilms Microbiomes* **6**, 16 (2020).
73. Goodrich, J. K. *et al.* Human Genetics Shape the Gut Microbiome. *Cell* **159**, 789–799 (2014).
74. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* **10**, (2014).
75. Zhou, W. *et al.* Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature* **569**, 663–671 (2019).
76. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69**, 89–95 (2001).
77. Califf, R. M. Biomarker definitions and their applications. *Exp. Biol. Med. (Maywood).* **243**, 213–221 (2018).
78. Sokol, H. *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci.* **105**, 16731–16736 (2008).
79. Darfeuille-Michaud, A. *et al.* High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. *Gastroenterology* **127**, 412–421 (2004).
80. Joossens, M. *et al.* Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* **60**, 631–637 (2011).
81. Fang, X. *et al.* Metagenomics-Based, Strain-Level Analysis of Escherichia coli From a Time-Series of Microbiome Samples From a Crohn's Disease Patient. *Front. Microbiol.* **9**, 2559 (2018).
82. Hall, A. B. *et al.* A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Med.* **9**, 103 (2017).
83. Palm, N. W. *et al.* Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000–1010 (2014).
84. Lopez-Siles, M., Duncan, S. H., Garcia-Gil, L. J. & Martinez-Medina, M. Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics. *ISME J.* **11**, 841–852 (2017).
85. Lopez-Siles, M. *et al.* Mucosa-associated Faecalibacterium prausnitzii and Escherichia coli co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int. J. Med. Microbiol.* **304**, 464–475 (2014).
86. Chehoud, C. *et al.* Fungal Signature in the Gut Microbiota of Pediatric Patients With Inflammatory Bowel Disease. *Inflamm. Bowel Dis.* **21**, 1948–1956 (2015).
87. Sarrabayrouse, G. *et al.* Fungal and Bacterial Loads: Noninvasive Inflammatory Bowel Disease Biomarkers for the Clinical Setting. *mSystems* **6**, (2021).
88. Halfvarson, J. *et al.* Dynamics of the human gut microbiome in Inflammatory Bowel Disease. *Nat. Microbiol.* **2**, (2017).
89. Wagner, J. A. Strategic Approach to Fit-for-Purpose Biomarkers in Drug Development. *Annu. Rev. Pharmacol. Toxicol.* **48**, 631–651 (2008).

90. Florkowski, C. M. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin. Biochem. Rev.* **29 Suppl 1**, S83–S87 (2008).
91. Martínez, I. *et al.* The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes. *Cell Rep.* **11**, 527–538 (2015).
92. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science (80-.).* **352**, 560–564 (2016).
93. Blaser, M. J. & Falkow, S. What are the consequences of the disappearing human microbiota? *Nat. Rev. Microbiol.* **7**, 887–894 (2009).
94. Clooney, A. G. *et al.* Ranking microbiome variance in inflammatory bowel disease: a large longitudinal intercontinental study. *Gut* gutjnl-2020-321106 (2020). doi:10.1136/gutjnl-2020-321106
95. Vangay, P. *et al.* US Immigration Westernizes the Human Gut Microbiome. *Cell* **175**, 962-972.e10 (2018).
96. Lopez-siles, M. *et al.* Cultured Representatives of Two Major Phylogroups of Human Colonic Faecalibacterium prausnitzii Can Utilize Pectin , Uronic Acids , and Host-Derived Substrates for Growth. 420–428 (2012). doi:10.1128/AEM.06858-11
97. Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. Roseburia intestinalis sp . nov ., a novel saccharolytic , butyrate-producing bacterium from human faeces. *Int. J. Syst. evoluationary Microbiol.* 1615–1620 (2002).
98. Moossavi, S. The necessity for an Iranian gut microbiome initiative. *Middle East J. Dig. Dis.* **6**, 109–110 (2014).
99. Vester-Andersen, M. K. *et al.* Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. *Sci. Rep.* **9**, 13473 (2019).
100. Lepage, P. *et al.* Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* **141**, 227–236 (2011).
101. Morgan, X. C. *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* (2012). doi:10.1186/gb-2012-13-9-r79
102. Vich Vila, A. *et al.* Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. *Sci. Transl. Med.* **10**, (2018).
103. Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat. Microbiol.* **4**, 293–305 (2019).
104. Chávez-Carbajal, A. *et al.* Characterization of the Gut Microbiota of Individuals at Different T2D Stages Reveals a Complex Relationship with the Host. *Microorganisms* **8**, 94 (2020).
105. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
106. Zeevi, D. *et al.* Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163**, 1079–1094 (2015).
107. Ahmad, A. *et al.* Analysis of gut microbiota of obese individuals with type 2 diabetes and healthy individuals. *PLoS One* **14**, e0226372 (2020).
108. He, Y. *et al.* Linking gut microbiota, metabolic syndrome and economic status based on a population-level analysis. *Microbiome* **6**, 172 (2018).
109. Li, Q. *et al.* Implication of the gut microbiome composition of type 2 diabetic patients from northern China. *Sci. Rep.* **10**, 5450 (2020).
110. Zhang, F. *et al.* Response of gut microbiota in type 2 diabetes to hypoglycemic agents. *Endocrine* **66**, 485–493 (2019).
111. Doumatey, A. P. *et al.* Gut Microbiome Profiles Are Associated With Type 2 Diabetes in Urban Africans. *Front. Cell. Infect. Microbiol.* **10**, 63 (2020).
112. Thingholm, L. B. *et al.* Obese Individuals with and without Type 2 Diabetes Show Different Gut Microbial Functional Capacity and Composition. *Cell Host Microbe* **26**, 252-264.e10 (2019).
113. Peters, B. A. *et al.* A taxonomic signature of obesity in a large study of American adults. *Sci. Rep.* **8**, 9749 (2018).
114. Wu, H. *et al.* The Gut Microbiota in Prediabetes and Diabetes: A Population-Based Cross-Sectional Study. *Cell Metab.* **32**, 379-390.e3 (2020).
115. Sanna, S. *et al.* Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat. Genet.* **51**, 600–605 (2019).
116. Kishikawa, T. *et al.* A Metagenome-Wide Association Study of Gut Microbiome in Patients With Multiple Sclerosis Revealed Novel Disease Pathology . *Frontiers in Cellular and Infection Microbiology* **10**, 780 (2020).
117. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).

118. Yilmaz, B. *et al.* Microbial network disturbances in relapsing refractory Crohn's disease. *Nat. Med.* **25**, (2019).
119. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **13**, 260–270 (2012).
120. Jacobs, J. P. *et al.* A Disease-Associated Microbial and Metabolomics State in Relatives of Pediatric Inflammatory Bowel Disease Patients. *Cmgh* **2**, 750–766 (2016).
121. Kolho, K.-L., Pessia, A., Jaakkola, T., de Vos, W. M. & Velagapudi, V. Faecal and Serum Metabolomics in Paediatric Inflammatory Bowel Disease. *J. Crohns. Colitis* **11**, 321–334 (2017).
122. Guasch-Ferré, M. *et al.* Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. *Diabetes Care* **39**, 833–846 (2016).
123. Tam, Z. Y. *et al.* Metabolite profiling in identifying metabolic biomarkers in older people with late-onset type 2 diabetes mellitus. *Sci. Rep.* **7**, 4392 (2017).
124. Urpi-Sarda, M. *et al.* Metabolomics for Biomarkers of Type 2 Diabetes Mellitus: Advances and Nutritional Intervention Trends. *Curr. Cardiovasc. Risk Rep.* **9**, 12 (2015).
125. Preter, V. De *et al.* Faecal metabolite profiling identifies medium-chain fatty acids as discriminating compounds in IBD. *Gut* **64**, 447–458 (2015).
126. Santoru, M. L. *et al.* Cross sectional evaluation of the gut-microbiome metabolome axis in an Italian cohort of IBD patients. *Sci. Rep.* **7**, 9523 (2017).
127. Ni, J. *et al.* A role for bacterial urease in gut dysbiosis and Crohn's disease. *Sci. Transl. Med.* **9**, eaah6888 (2017).
128. Marchesi, J. R. *et al.* Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. *J. Proteome Res.* **6**, 546–551 (2007).
129. Jansson, J. *et al.* Metabolomics Reveals Metabolic Biomarkers of Crohn's Disease. *PLoS One* **4**, e6386 (2009).
130. Le Gall, G. *et al.* Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome. *J. Proteome Res.* **10**, 4208–4218 (2011).
131. Bain, J. R. *et al.* Metabolomics applied to diabetes research: moving from information to knowledge. *Diabetes* **58**, 2429–2443 (2009).
132. Menni, C. *et al.* Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes* **62**, 4270–4276 (2013).
133. Batch, B. C. *et al.* Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. *Metabolism.* **62**, 961–969 (2013).
134. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **17**, 448–453 (2011).
135. Würtz, P. *et al.* Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. *Diabetes Care* **36**, 648–655 (2013).
136. Salihovic, S. *et al.* Non-targeted urine metabolomics and associations with prevalent and incident type 2 diabetes. *Sci. Rep.* **10**, 16474 (2020).
137. Wang-Sattler, R. *et al.* Novel biomarkers for pre-diabetes identified by metabolomics. *Mol. Syst. Biol.* **8**, 615 (2012).
138. Lamas, B. *et al.* CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. *Nat. Med.* **22**, 598–605 (2016).
139. Nikolaus, S. *et al.* Increased Tryptophan Metabolism Is Associated With Activity of Inflammatory Bowel Diseases. *Gastroenterology* **153**, 1504-1516.e2 (2017).
140. Chen, T. *et al.* Tryptophan Predicts the Risk for Future Type 2 Diabetes. *PLoS One* **11**, e0162192–e0162192 (2016).
141. Hisamatsu, T. *et al.* Novel, objective, multivariate biomarkers composed of plasma amino acid profiles for the diagnosis and assessment of inflammatory bowel disease. *PLoS One* **7**, e31131 (2012).
142. Roager, H. M. & Licht, T. R. Microbial tryptophan catabolites in health and disease. *Nat. Commun.* **9**, 3294 (2018).
143. Lavelle, A. & Sokol, H. Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 223–237 (2020).
144. Sorgdrager, F. J. H., Naudé, P. J. W., Kema, I. P., Nollen, E. A. & Deyn, P. P. De. Tryptophan Metabolism in Inflammaging: From Biomarker to Therapeutic Target. *Front. Immunol.* **10**, 2565 (2019).
145. Tillin, T. *et al.* Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and diabetes in a South Asian and European cohort from the SABRE (Southall And Brent REvisited) Study. *Diabetologia* **58**, 968–979 (2015).
146. Schoepfer, A. M. *et al.* Fecal calprotectin correlates more closely with the Simple Endoscopic Score for Crohn's disease (SES-CD) than CRP, blood leukocytes, and the CDAI. *Am. J. Gastroenterol.* **105**, 162–169 (2010).
147. Lewis, J. D. The utility of biomarkers in the diagnosis and therapy of inflammatory bowel disease.

Gastroenterology **140**, 1817-1826.e2 (2011).
148. Sipponen, T. & Kolho, K.-L. Fecal calprotectin in diagnosis and clinical assessment of inflammatory bowel disease. *Scand. J. Gastroenterol.* **50**, 74–80 (2015).
149. Dubinsky, V. *et al.* Dysbiosis in metabolic genes of the gut microbiomes of patients with an ileo-anal pouch resembles that observed in Crohn's Disease. *medRxiv* 2020.09.23.20199315 (2020). doi:10.1101/2020.09.23.20199315
150. Association, A. D. Classification and diagnosis of diabetes: Standards of medical care in diabetes-2018. *Diabetes Care* **41**, S13–S27 (2018).
151. Floegel, A. *et al.* Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. *Diabetes* **62**, 639 LP – 648 (2013).
152. Lu, Y. *et al.* Metabolic signatures and risk of type 2 diabetes in a Chinese population: an untargeted metabolomics study using both LC-MS and GC-MS. *Diabetologia* **59**, 2349–2359 (2016).
153. Buttó, L. F., Schaubeck, M. & Haller, D. Mechanisms of microbe-host interaction in Crohn's disease: Dysbiosis vs. Pathobiont Selection. *Frontiers in Immunology* (2015). doi:10.3389/fimmu.2015.00555
154. Jochum, L. & Stecher, B. Label or Concept – What Is a Pathobiont? *Trends Microbiol.* (2020). doi:https://doi.org/10.1016/j.tim.2020.04.011
155. Buttó, L. F. & Haller, D. Dysbiosis in Crohn's disease - Joint action of stochastic injuries and focal inflammation in the gut. *Gut Microbes* **8**, 53–58 (2017).
156. Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science (80-.).* (2013). doi:10.1126/science.1241214
157. Turnbaugh, P. J. *et al.* A core gut microbiom in obese and lean twins. *Nature* **457**, 480–484 (2009).
158. Arrieta, M., Sadarangani, M., Brown, E. M., Russell, S. L. & Nimmo, M. A humanized microbiota mouse model of ovalbumin-induced lung inflammation. *Gut Microbes* **7**, 342–352 (2016).
159. Blanton, L. V. Gut bacteria that rescue growth impairments transmitted by immature microbiota from undernourished children. **351**, 1–18 (2016).
160. Kostic, A. D. *et al.* The Dynamics of the Human Infant Gut Microbiome in Development and in Progression towards Type 1 Diabetes. *Cell Host Microbe* **17**, 260–273 (2016).
161. Nagao-Kitamoto, H. *et al.* Functional Characterization of Inflammatory Bowel Disease-Associated Gut Dysbiosis in Gnotobiotic Mice. *CMGH* (2016). doi:10.1016/j.jcmgh.2016.02.003
162. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. **444**, 1027–1031 (2006).
163. Steck, N. *et al.* Enterococcus faecalis metalloprotease compromises epithelial barrier and contributes to intestinal inflammation. *Gastroenterology* **141**, 959–971 (2011).
164. Lengfelder, I. *et al.* Complex Bacterial Consortia Reprogram the Colitogenic Activity of Enterococcus faecalis in a Gnotobiotic Mouse Model of Chronic, Immune-Mediated Colitis. *Front. Immunol.* **10**, 1420 (2019).
165. Kim, S. C., Tonkonogy, S. L., Karrasch, T., Jobin, C. & Balfour Sartor, R. Dual-association of gnotobiotic IL-10-/- mice with 2 nonpathogenic commensal bacteria induces aggressive pancolitis. *Inflamm. Bowel Dis.* **13**, 1457–1466 (2007).
166. Eun, C. S. *et al.* Induction of bacterial antigen-specific colitis by a simplified human microbiota consortium in gnotobiotic interleukin-10-/- mice. *Infect. Immun.* **82**, 2239–2246 (2014).
167. Brugiroux, S. *et al.* Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. *Nat. Microbiol.* **2**, 16215 (2016).
168. Lepage P, Häsler R, Spehlmann ME, Rehman A, Zvirbliene A, Begun A, Ott S, Kupcinskas L, Doré J, Raedler A, S. S. Twin study indicates loss of interaction between microbiota and mucosa patients with ulcerative colitis. *Gastroenterology* **141**, 227–36 (2011).
169. Rehman, A. *et al.* Geographical patterns of the standing and active human gut microbiome in health and IBD. *Gut* **65**, 238—248 (2016).
170. Turnbaugh, P. J. & Gordon, J. I. The core gut microbiome, energy balance and obesity. *J. Physiol.* **587**, 4153–4158 (2009).
171. Walter, J., Armet, A. M., Finlay, B. B. & Shanahan, F. Establishing or Exaggerating Causality for the Gut Microbiome: Lessons from Human Microbiota-Associated Rodents. *Cell* **180**, 221–232 (2020).
172. Brown, K. *et al.* Prolonged antibiotic treatment induces a diabetogenic intestinal microbiome that accelerates diabetes in NOD mice. *ISME J.* **10**, 321–332 (2016).
173. Cani, P. D. *et al.* Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **56**, 1761–1772 (2007).

Competing interests

The authors declare no competing interests.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 395357507 (SFB 1371, Microbiome Signatures), and has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N°964590.