Facultad de Ciencias Biológicas

Unidad Mixta en Genómica y Salud de la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO) y el Instituto Cavanilles de Biodiversidad y Biología Evolutiva de la Universitat de València

Study of the intestinal microbiota in the HIV infection and the effect of a nutritional intervention

Memoria presentada por Jorge Francisco Vázquez Castellanos para optar al grado de Doctor por la Universitat de València.

DIRECTORES:
Prof. Andrés Moya Simarro.
Profa. Amparo Latorre Castillo.
Dra. María José Gosalbes Soler.

Valencia, mayo 2017

MARÍA JOSÉ GOSALBES SOLER, Investigadora del Centro de Investigación Biomédica en Red,

AMPARO LATORRE CASTILLO, Catedrática del Departamento de Genética de la Universidad de Valencia,

ANDRÉS MOYA SIMARRO, Catedrático del Departamento de Genética de la Universidad de Valencia,

CERTIFICAN que el trabajo para optar al grado de Doctor en Biotecnología, y que lleva por título "Study of the intestinal microbiota in the HIV infection and the effect of a nutritional intervention", ha estado realizado bajo su dirección en la Unidad Mixta en Genómica y Salud de la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO) y el Instituto Cavanilles de Biodiversidad y Biología Evolutiva de la Universitat de València por JORGE FRANCISCO VÁZQUEZ CASTELLANOS.

Y para que así conste, en el cumplimiento de la legislación vigente, firmo el presente certificado en: Valencia, a de 2017.

MARÍA JOSÉ GOSALBES SOLER

ANDRÉS MOYA SIMARRO

AMPARO LATORRE CASTILLO

“All disease begins in the gut”

Hippocrates

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Primero que nada, me gustaría agradecer a mi querida familia por el apoyo que me han dado a lo largo de estos años. Particularmente a mis padres, Luz y Jorge, a mis hermanos, Luz y Pablo, así como a mis tías, Miroslava y Asumari, y a mis primos Daniel y Lucía, ellos siempre estuvieron al pendiente de mí y me apoyaron en las situaciones más difíciles. De la misma forma también me gustaría agradecer a Bea por el cariño, apoyo y los buenos momentos que compartimos a lo largo de estos casi siete años.

Agradezco a mis tutores Andrés, Amparo y Pepa por permitirme realizar el doctorado en su grupo de investigación, por haberme formado como académico, por sus buenos consejos y por el gran apoyo que me dieron durante mi estadía en Valencia. Así mismo me gustaría agradecerle a Thomas Rattei, por aceptarme en su grupo de investigación durante mi estancia, sus buenos consejos y por su generosidad al permitirme utilizar sus recursos computacionales.

Me gustaría agradecer a mis compañeros de doctorado (de ambas salas) especialmente a Mària y a Rodrigo, porque aparte de ser grandes compañeros de trabajo me brindaron su valiosa amistad.

Me gustaría agradecer las personas del grupo de investigación que me ayudaron a sacar la tesis adelante como Nuria, Loreto, Artacho, Sergio y Manolo, como a las personas que creyeron que podría llevar a cabo una tesis doctoral, al Dr. Federico Sanchez y al Dr. Víctor Gonzales.

A los muchos amigos que tuve a lo largo de estos años en particular a Lina, Letty, Vicente, Mariloli, Danny, Orly, Bastien, Gonzalo y Benja entre varios otros, sin ellos las cosas no hubiesen sido igual de divertidas.

Al CONACYT y al Instituto de Salud Carlos III por haberme financiado durante estos casi 7 años.

En fin, agradezco a todas las personas que hicieron de mi estancia en Valencia una etapa de mi vida en la que puedo decir que fui feliz.

THESIS ABSTRACT

The use of ART in HIV+ subjects has increased considerably the life expectancy restoring the CD4+ T-cell counts and maintaining at low levels the viral load. However, their life expectancy is 10 years lower than the average population. This reduction is given due to unrelated AIDS illness, such as cardiovascular diseases and atherosclerosis that are caused by persistent immune activation and chronic inflammation. A possible explanation for this phenomenon is a constant bacterial translocation from the intestinal lumen to the systemic circulation given a prior disruption of the GALT. Moreover, the loss of the lymphoid tissue leads to a microbial imbalance that could be related to the systemic immune activation.

In the present thesis, we describe in a holistic way the fundamental role of the microbiome in the pathogenesis of HIV infection. Here we present the results of a cross-sectional study of a cohort of three different HIV-infected groups of subjects (with a different response to the ART) and controls target to understand the alterations of the gut-microbiome given the HIV infection. The microbiome was characterized implementing different "omic" technologies and the impact on the host health was determined based on measuring clinical data related to the immune response and the bacterial translocation. Finally, a pilot study based on dietary supplementation with prebiotics and glutamine was carried out with the aim of ameliorating the HIV-associated dysbiosis.

The HIV infection causes a disruption of the GALT leading the dysbiosis of the microbial community that cannot be restored by the ART. Moreover, the infection time would affect the diversity of the microbiota and the ecosystem stability. This dysbiotic community is enriched in Gram-negative species which are adapted to the inflammatory environment of the gut produced by HIV infection and produces pro-inflammatory metabolites which trigger the systemic immune activation and inflammation. Moreover, the HIV-dysbiosis is depleted for SCFA producer species and in the expression of genes related to anti-inflammatory metabolic pathways such as butanoate metabolism, propanoate metabolism or fatty acid metabolism.

The prebiotic has an effect on a community whose original configuration is receptive to the nutritional intervention; this is related to the time exposure to HIV infection. The prebiotic intervention increases the butyrate levels by means of the increase of SCFA-producer species such as *Faecalibacterium sp*. The increment of the levels of the butyrate is related to the decrease of the bacterial translocation and systemic inflammation.

Finally, we show that the dysbiotic-community is able to establish a stable-community which is associated with the deterioration of the patient's health. More importantly, we suggest that the microbiota may be a new target for clinical interventions in patients infected with HIV and proposed putative candidates for been viable targets for such interventions.

1. GENERAL INTRODUCTION

1.1. Human gut microbiome

The term “microbiome” was first suggested by Joshua Lederberg and was further coined to signify “the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space and have been all but ignored as determinants of health and disease’’ [1,2]. The gut microbiome has the largest number of bacteria and the greatest number of species compared to other areas of the body. The gut microbiota is a complex community of bacteria that coexist in the gastrointestinal tract (GIT), particularly in the large intestine. Its composition depends on age, diet, genotype, and health status. Due to its co-evolution with the host, it is involved in important mutualistic functions such as the correct development of the immune system, the fermentation of dietary fiber, the synthesis of essential amino acids and vitamins and the protection against the pathogens.

In the last decade, due to the development of the next generation sequence (NGS) technologies, the scientific community turned their efforts in the characterization of the uncultivable bacteria that reside in the gut microbiome. Nowadays, the number of available metagenomes retrieved from the gut and other body sites exceed 2000 from individuals spanning across five continents [3].

1.1.1. Overview of the human gut associated microbiome

The GIT is an organ which takes in food, digests it to extract and absorb energy and nutrients, and expels the remaining waste as feces and urine [4,5]. The GIT includes the pharynx, esophagus, stomach, the small intestine (composed by the duodenum, jejunum, and ileum) and large intestine (composed by the cecum, ascending colon, transverse colon, descending colon, sigmoid colon and the anus) [4,5] (**Figure I.1a**). The gastrointestinal wall that surrounds the lumen of the GIT is composed of four concentric layers (**Figure I.1b**): the mucosa, the submucosa, the muscular layer

and the serosa/adventitia layer. The mucosa is the innermost layer of the GIT and it is in direct contact with the digested food and the bacterial population. The submucosa, which consists of a dense and irregular layer of connective tissue, is composed of blood vessels, lymphatics vessels and nerves branching the mucosa and the muscular layer. The muscular layer, made of the inner and outer muscle layers, controls the peristalsis. Finally, the serosa/adventitia layer is the outermost layer of the GIT and consists of different sublayers of connective tissue, which depending on the part of the GIT, could work for lubrication (the serosa layer) or to binds tissues layers (adventitia layer) [4] (**Figure I.1b**).

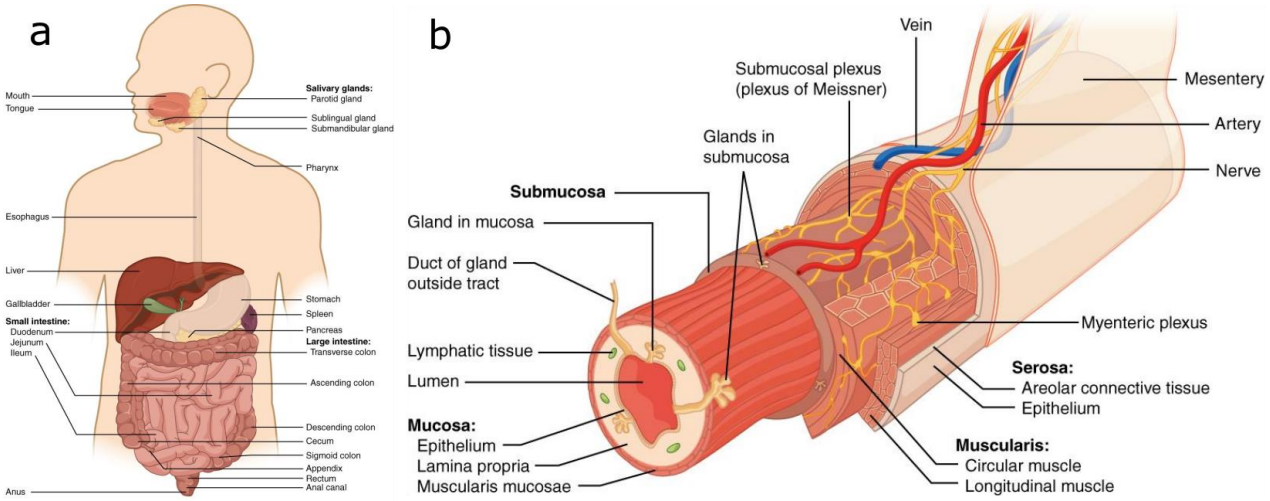


Figure I.1 Human intestinal tract. Panel **a**: Components of the digestive system. Panel **b**: Layers of the alimentary canal. Reproduced from: OpenStax College with permission of the Rice University under a Creative Commons Attribution License (CC-BY 3.0), and is an Open Educational Resource.

The GIT contains an assembly of distinct ecological habitats that harbor different bacterial populations. For instance, the stomach acidic environment harbors a reduced bacteria community dominated by species of the phyla Proteobacteria, Firmicutes, Bacteroidetes, Actinobacteria, and Fusobacteria (**Figure I.2**). The small intestine shows a bigger and more diverse bacterial community than the one observed in the stomach (**Figure I.2**), but the peristalsis together with the bile and the pancreatic secretions maintain small bacterial population comparing to those observed in the colon [6]. The small-intestine-associated microbiome is mainly composed of facultative species from the genera *Streptococcus*, *Lactobacillus* but also strict anaerobes such as *Clostridium* and *Veillonella* [7].

The human colonic-associated microbiome possesses more than 1,000 'species-level' or phylotypes [8] and contains around 10^{10} to 10^{13} CFU/g of intestinal contents (**Figure I.2**), being 99% of them anaerobic species [9]. This bacterial community is the most studied from all the GIT, being the most common method for its characterization the fecal sampling. However, although the microbial composition retrieved from the fecal samples does not mimic the observed in the mucosal samples it has been reported that luminal microbial contents of the colon correlate with feces in terms of species diversity and bacterial abundance [10–12]. For all these reasons in the present thesis we use the fecal samples as a proxy of the colon-associated microbiome and will be referred as the gut-microbiome.

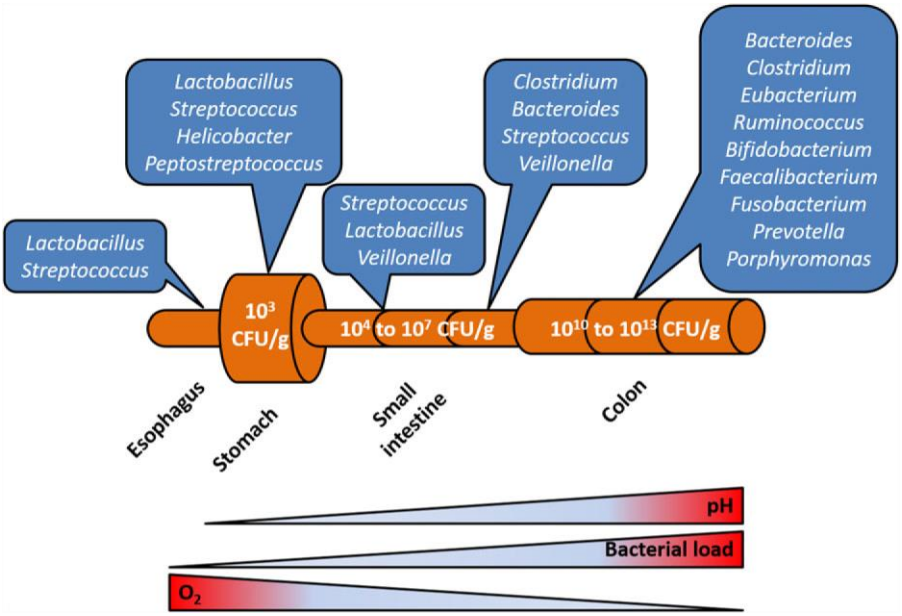


Figure I.2 Microbiota composition and physiological conditions along the digestive tract. The figure illustrates the composition of the most abundant genera found in the digestive tract. It also highlights how the pH, the oxygen concentration, and the bacterial load vary along the gut**.** Reproduced from: Mondot and Lepage (2016)[6] Wiley Online Library Copyright © 1999-2017 John Wiley & Sons, Inc. All Rights Reserved.

In adult populations, most of the gut associated microbiota mainly belong to the phyla Bacteroidetes and Firmicutes, whereas Actinobacteria, Proteobacteria, and Verrucomicrobia, although found in most of the human microbiomes, are generally minor constituents [13–18] (**Figure I.3a**). Methanogenic archaea (mainly *Methanobrevibacter smithii*), eukaryotes (mainly yeasts) and viruses (mainly bacteriophage) are also present as members of the gut microbiome [19]. Despite the consistency of these main phyla, the relative proportions of the gut microbiome species vary

markedly across individuals being *Faecalibacterium prausnitzii*, *Roseburia intestinalis* and *Bacteroides uniformis* the most abundant [16]; however, in some individuals, even these can represent less than 0.5% of the microbes present [18]. In 2011, the MetaHit consortia published their results about the diversity in gut associated-metagenomes in European western populations [14]. They showed that the western individuals could be classified, based on the differences of their microbial composition in three clusters denominated enterotypes. Each of the enterotypes was enriched in different species from a specific genus: Enterotype 1 (*Bacteroides* and *Parabacteroides* species), enterotype 2 (*Prevotella* and *Desulfovibrio*) and the most frequent of the three, the enterotype 3 (*Ruminococcus* and *Akkermansia*) (**Figure I.3b**). However, those enterotypes could not be found in higher size cohort and in non-European population, as shown in the results of the Human Microbiome Project [15]. This study found that the US individuals' microbiomes were enriched in *Bacteroides* species while the Malawin-Amerindian population was in *Prevotella* species, being the differences attributed to the lifestyle (**Figure I.3c**).

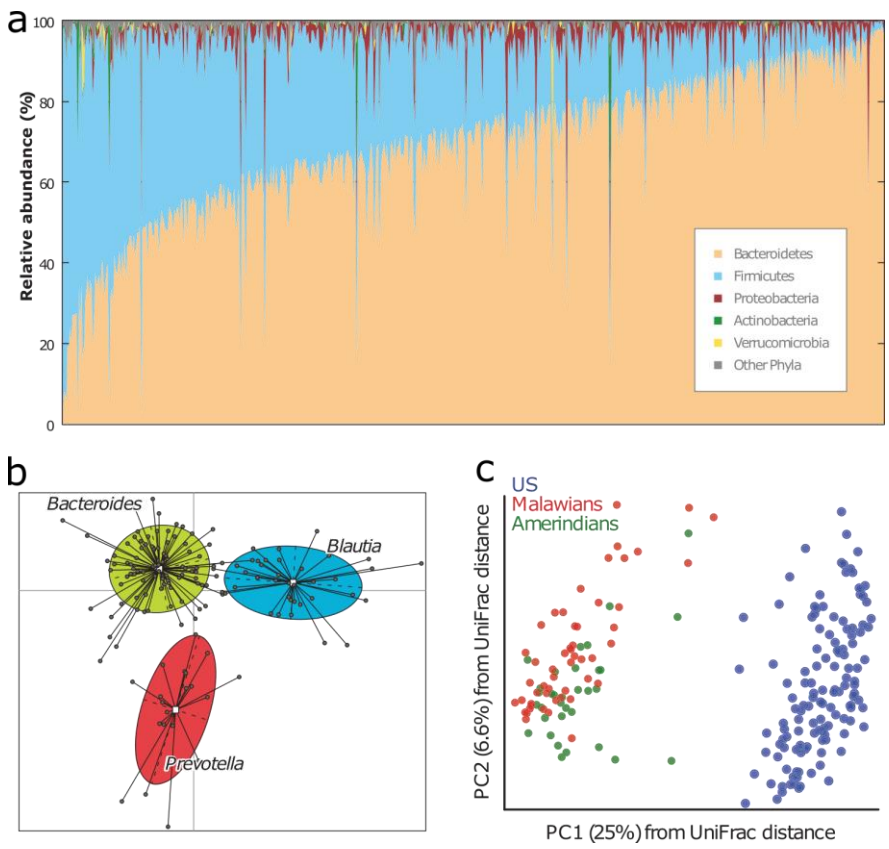


Figure I.3 Phylogenetic composition of the gut-associated human microbiome and differences between healthy individuals. Panel **a:** Phylogenetic composition, at the phylum level, from 648 samples collected as part of the NIH Human Microbiome Project (HMP). Panel **b**: Principal component analysis (PCA) of the distribution of the healthy samples from the Metahit project representing the three enterotypes clusters retrieved from 154 pyrosequencing-based on 16S rRNA gene sequences. Panel **c**: Principal coordinates analysis (PCoA) of the unweighted UniFrac distances for the fecal microbiota from 181 adults from the Malawian and Amerindian agrarian-communities and US subjects. Adapted from: Eloe-Fadrosh and Rasko (2013) [20], Arumugam et al. 2011 [14] and Yatsunenko et al. (2012) [15] with permission of Annual Review of Medicine® and Nature Publishing Group

1.1.2. Composition, diversity, and sources of variation of the gut microbial communities

The definition of a "healthy microbiome" is not an easy task given the vast microbial diversity that exists over the time and across individuals [8,21,22]. Several factors influence the microbial composition being in addition to host genotype, the diet and the age the principal drivers of variation in the healthy microbiome. However, there are other sources of variation such as disease and the use of antibiotics or medication [23] that produce altered microbial communities defined as dysbiotic in the setting of the disease.

1.1.2.1. Host genotype

Recently, several works have shown the microbiome heritability and its association with several loci of the host genome, as reviewed in Goodrich et al. (2014) [24]. For instance, it has been identified relationships such as members of the phylum Firmicutes and variants of the host-genes involved in the Toll-like receptor and T-cell receptor [25,26]; species from the Rikenellaceae family with a region encompassing signaling Toll-like receptors kinase *IRAK4* [27] and Prevotellaceae sp. with the TGFB3, a cytokine that modulates barrier function of the intestine [27].

Although most of the relationships are with genes related to the immune system [28], it has also been observed associations between species related to diet and several loci related to the food intake. For instance, the associations between species from the genus *Bifidobacterium* and the gene that encodes for the lactase enzyme [29]; another example is shown in *Akkermansia*, a mucin-dwelling and degrading genus, that has been associated with several loci related to lipids [28,30] and with the *SIGLEC15*, a sialic acid binding lectin (sugar epitopes of the mucin) [31]. Some species have been associated to cancer-related-genes. In humans and murine models, species from the *Turicibacter* genus has been associated with tissue-specific expression QTLs [30] inflammation and cancer [32,33]. In murine models, such species are associated with a QTL on MMU7 that overlaps with the HCS1 QTL for susceptibility to murine hepatocellular carcinomas [33].

1.1.2.2. Host age

The gut microbial community compositions vary over human lifetimes [15] (**Figure I.4**). The microbial colonization may begin in the uterus and it is established at three years after birth [34,35]. At birth, neonates are already colonized by species from the Firmicutes, Proteobacteria, and Actinobacteria with lower levels of Bacteroidetes [36–38].

The infant's microbiome is considerably less diverse than those observed in adults [15,39] and varies depending on if the birth was given by cesarean or vaginally delivered [39–44] and by the feeding habits: breast feeding versus formula feeding [41,45–47]. The infant gut microbiota becomes more diverse over time mainly with the introduction of solid food. The microbial composition in the childhood has shown an increase of the species related to *Roseburia, Faecalibacterium* and *Ruminococcus* genera and the species *Bacteroides vulgatus* and *Bacteroides xylanisolvens* when comparing with the adults´ gut-microbiota [48].

The adult microbiome configuration is reached approximately at 1-3 years of life and it's vary depending on the diet and geographical location, [15] remaining stable (except perturbations) until the age 65 years [49]. In the old age, the microbiota is different from the one observed in adults [50]. The elderly´s microbiome exhibits a higher Firmicutes to Bacteroidetes ratio [50] and a reduction of species from the *Bifidobacteria* genus [50] and *F. prausnitzii* [50,51]. The factors that shape the elder´s microbiome are probably the overall increased use of medication, dietary deficiency, as well as changing hormonal levels [52]. The GIT microbiome is thought to influence the overall health of the elderly, as changes in its composition have been associated with an increase of species from the Enterobacteriaceae family [53] and a decline in the health [54].

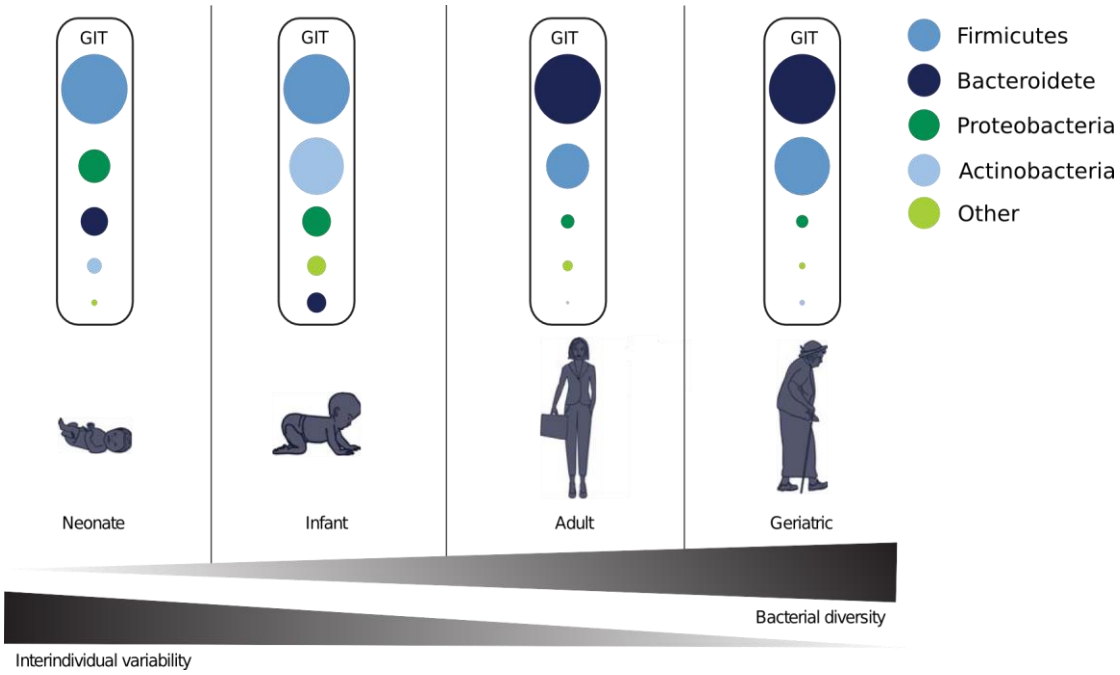


Figure I.4 Development of the gut-microbiome over time. Adapted from: Greenhalgh et al. (2016) [55] with permission of John Wiley & Sons, Inc.

1.1.2.3. Host diet

The diet is an essential factor for the establishment of the gut-microbiome assembly in the earlier steps and remains as an important factor in the adulthood [12]. In fact, it is considered that the diet has a dominant role in comparison with all the other causes of variation in the healthy gut-microbiome [56]. For example, when comparing US protein-rich diet versus the Malawi/Amerindian starch-rich diet [15] several genes related to the degradation of starch (glutamate synthase and alpha-amylase enzymes) were increased in the agricultural populations. On other hand, the US microbiomes had significant over representation of enzymes related to metabolic pathways of vitamin biosynthesis, the lipoic acid, the xenobiotics metabolism and the degradation of amino acids. Additionally, the taxonomic composition varies depending on the population and diet; a trade-off between *Prevotella* and *Bacteroides* has been observed in agrarian and western diet populations [15,56]. The *Prevotella*-rich-microbiomes have been associated with high-fiber diets, such as the diet that follows the agrarian communities (Amerindian, Malawian, and Burkina Faso villagers). *Prevotella* species are known to be involved in the fermentation of xylan and cellulose through carbohydrate-active enzymes such as xylanase, carboxymethylcellulase, and endoglucanase

[56]. Contrary to this, the *Bacteroides*-rich-microbiomes have been linked to western diet [15]. The western diet is associated with the consumption of sugar, animal fat and calorie-dense foods [15]. Microbiomes enriched in species of the *Bacteroides* genus also present a higher number of enzymes related to the degradation of amino acids, the catabolism of simple sugars and the vitamin metabolism [15].

The adaptive effect is more evident when comparing the microbiota of animals with a different type of diets. This was remarkable for the carnivores and herbivores by clearly separating the samples into two groups [57], indeed the effect of the microbial-diet adaptation is so high that the clustering retrieved from the gut-associated microbial communities does not mirror the clustering retrieved by the mammalian phylogeny [57]. The microbiome from herbivores was enriched in enzymes involved in the amino acid biosynthesis, meanwhile, the carnivores' metagenomes present an increase of enzymes related to the degradation of amino acids and not a single gene was related to amino acid biosynthesis. The results of Muegge et al. study [57] established that the carnivorous microbiomes have specialized to degrade proteins as an energy source, whereas herbivorous communities have specialized to synthesize amino acid building blocks.

1.1.2.4. Dysbiosis and host health

It has been suggested that the human microbiome is ecological stable [21], this refers to the ability of a community to resist changes in the setting of an ecologic stress or to return to an equilibrium state following a stress-related perturbation [58] (**Figure I.5**). The healthy microbial state could be seen as an attractor that represents a group of species towards which the microbiome tends to evolve independently from its initial condition [58] (**Figure I.5**). An important feature of those attractors is their high diversity, which brings a high tolerance against the pathogenic invasion [15,22,59] and environmental changes. This feature is commonly referred as the resilience and it is defined as the amount of stress or perturbation that a system can tolerate before its changes in a different equilibrium state [60] (**Figure I.5**). However, when the environmental perturbation is high enough the microbiome moves to a disease state which is known as dysbiosis. The term dysbiosis refers to an imbalance of the microbiome commensal species and its replacement with a pathogen or opportunistic bacteria [61]. The gut microbiome dysbiosis has been related to numerous diseases

such as cancer, allergies, HIV, inflammatory bowel disease (IBD), diabetes (types 1 and 2), obesity and asthma. Here, I am going to review the most important of this microbial imbalance states.

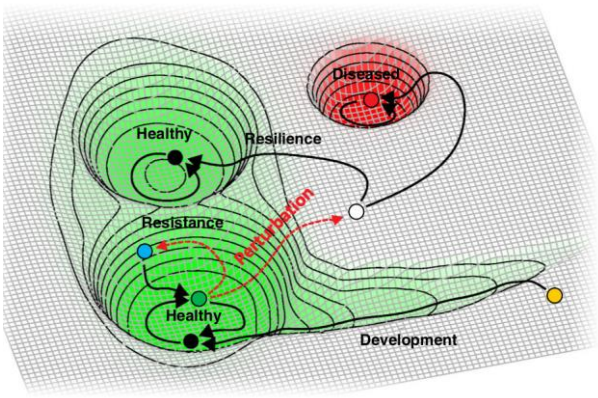


Figure I.5 Energy landscape. The microbiome development can be characterized in terms of its dynamics and represented as a conceptual energy landscape. The infant microbiome (yellow point) evolves into an adult microbiome (black point) which often falls into one healthy attractor (green depression). The attractor normally supports alterations (red lines) which return to the microbiome to its original attractor or its moves to another healthy attractor (black lines). However, if the perturbation is high enough the microbiome moves to a disease state (red depression) and states a dysbiotic microbiome (red point). Adapted from: Lloyd-Price, Abu-Ali, and Huttenhower (2016) [21] with permission of BioMed Central

Antibiotic and dysbiosis

Nowadays, most of the administered antibiotics have broad-spectrum activity and are used to treat many infections [62]. Thus, the antibiotic therapies affected not only the specific pathogens but also related members of the human microbiota. The overuse of the antibiotics caused a reduction in the microbiome diversity [63–66] associated with a transiently or permanently dysbiosis state.

The antibiotic effects depend on the initial microbial diversity, the drug concentration and the different microbial growth stages [67]. For instance, oral amoxicillin exposure caused microbial shifts that long on average 30 days but the effects have been observed until 60 days after the treatment [68]. The amoxicillin treatment has a strong effect on specific groups of the gut environment, among aerobic and anaerobic species, leading a decrease in *Clostridium* and *Eubacterium* and an increase in Enterobacteriaceae, *Bacteroides* and *Prevotella* taxa [69,70]. Other antibiotics, such as the clindamycin or tigecycline, are particularly active against the anaerobic bacteria of the gut-microbiome [71,72] and increases susceptibility to *Clostridium difficile* infection (CDI) [73,74]; particularly these antibiotics deplete the bile acid-hydroxylating activity of *Clostridium scindens*, which is required for protection against CDI [75]. Similarly, the use of the streptomycin and vancomycin has shown to cause an increased susceptibility to *Salmonella typhimurium* infection [76]. Finally, it has been observed that the triple antibiotic therapy (clarithromycin, metronidazole, and omeprazole) against the *H. pylori* gastritis decrease many beneficial species from the Clostridia and Bifidobacteria classes [77,78].

The use of antibiotics interferes with the immune-system-microbiome interactions [79,80], affects the mitochondrial gene expression [81] and increases the pathogen susceptibility [73,76,82]. In fact, high concentrations of antibiotics reduce or eliminate most of the beneficial bacterial metabolites such as short-chain fatty acids (SCFA) [83,84] and secondary bile acids (reviewed in [85]). Those effects seem to be independent of the antibiotic class and be more dependent on the dose [86].
A common side effect of the antibiotics is the appearance of diarrhea, given by the drug itself or by the loss of the fiber-degrading bacteria due to the antibiotic. This is because the normal digestion of these compounds prevents episodes of osmotic diarrhea [87]. However, opportunistic bacteria such as *Clostridium perfringens, Klebsiella oxytoca, Candida albicans or C. difficile* [88] can also be the factors that trigger the disease.

Infection and dysbiosis

One of the most important functions of the gut-microbiome is to protect against the pathogen colonization [89]. For instance, the commensal microbiota produces bacteriocins and proteinaceous toxins that specifically inhibit pathogenic bacterial species [90]. It also is known that the microbiome alters the gut environment conditions (e.g., pH or the oxygen concentrations) in order to avoid pathogen colonization [90,91]. Moreover, an alternative strategy utilized by the indigenous microbial community is the preferential consumption of nutrients, which in consequence outsource the growth of competing pathogens.

Commensal bacteria, through the production of specific metabolites, can also affect the expression of pathogenic virulence factors or totally inhibit the pathogen growth. It has been shown that the bacteria-butyrate production downregulates the expression of several virulence genes including those encoding the type 3 secretion system (T3SS) in *Salmonella enterica* Serovar *Enteritidis* and *Typhimurium* [92]. Similarly, *R. obeum* can restrict the *V. cholerae* colonization by the expression of the *lux*S (autoinducer-2 (AI-2) synthase) and AI-2 production, a furanosyl-borate-diester member of a family of signaling molecules used in quorum sensing.

Commensal bacteria also prevent pathogen colonization and infection indirectly by enhancing host defense mechanisms such as functionally promoting mucosal barrier function and enhancing either innate immune response. This section will be reviewed in the section "1.2.2. The role of the microbiota in the adaptation and maturation of the immune system".

When the pathogen evades all the defense lines, the bacterial infections could disturb the microbial balance in the gut-microbiome. For instance, some pathogens such as *V. cholerae*, cause an acute infection that disrupts the microbial community. The microbial composition of the subjects infected is dominated on average of a 56% of *V. cholerae* bacteria [93], however after the patient gets stabilized, the bacteria composition shows an increase of unidentified *Streptococcus* or *Fusobacterium* species [93]. Similarly, in a salmonellosis infection, species from the *Salmonella* genus used the hydrogen produced by the gastrointestinal microbiota and triggers potent inflammatory responses which generate an alternative terminal electron acceptors (reviewed in [94]). *Salmonella* species exclusively utilize these electron acceptors for anaerobic respiration, permitting metabolic access to abundant substrates such as ethanolamine to power growth blooms. Thus, the potent inflammatory response and the *Salmonella* resource availability cause a reduction of the bacterial diversity and an overgrowth of the pathogen species [94].

Inflammatory bowel disease dysbiosis

The IBD is a multi-factorial disorder characterized by chronic and relapsing intestinal inflammation and is mainly defined as either ulcerative colitis (UC) or Crohn's disease (CD) [95]. Although the factors that cause IBD are unknown, the most accepted hypothesis is that an exacerbate reaction of the immune response against the gut microbiota is triggered by environmental factors in a genetically susceptible host [95]. There are around 160 IBD susceptibility genes [96] that are mainly related to the acquired immunity, bacterial recognition, autophagy and mucosal barrier [95].

The IBD is characterized by a reduced bacterial diversity [97–100], a decrease of Firmicutes, and an increase of Proteobacteria [98,99,101,102]. Among Firmicutes, a decrease in the *Clostridium leptum* groups, especially *F. prausnitzii*, has been reported in many studies [103,104].

Metagenomic studies in IBD patients have shown a decrease in genes responsible for carbohydrate and amino acid metabolism and an increase in those in the oxidative stress pathway [105], raising the possibility that oxidative stress from the gut microbiota causes intestinal inflammation in IBD patients. The inflammation could also be given by the reduction of the butyrate-producer species such as *F. prausnitzii*.

Rheumatoid arthritis dysbiosis

The rheumatoid arthritis is a long-term autoimmune disorder that primarily affects joints caused by a combination of genetic and environmental factors. Studies of the gut microbiota in murine rheumatoid-arthritis models have shown an increase of the bacteria *P. copri* which strongly correlated with disease in new-onset untreated rheumatoid arthritis patients [106]. The *P. copri* dominate the intestinal microbiota and resulted in an increased sensitivity to chemically induced colitis [106].

Obesity dysbiosis

The microbial gut composition of obese subjects has been associated with changes in the relative abundance of the two dominant phyla, Bacteroidetes and Firmicutes, a reduced bacterial diversity, and an altered representation of bacterial genes and metabolic pathways [18,107]. The obese microbiome has an increased capacity to harvest energy from the diet. Furthermore, this obese-associated microbial community is transmissible as it has been observed in mouse experiments where the inoculation of the human-gut microbiome in the lean mice results in a significantly greater increase in total body fat [107]. These results identified the gut microbiota as an additional contributing factor to the pathophysiology of obesity [107] and that deviations from this "microbial-core" were associated with different physiological states of the obesity [18].

Cancer dysbiosis

Studies on colorectal cancer in murine models revealed an association between colonic microbiota and the cancer development [108]. The mice that presented colitis had lower diversity in their gut microbiota than wild-type mice and an increased number of *Escherichia coli* bacteria. The increase in the abundance of *E. coli* is related to the increase in the production of the genotoxic colibactin, which is an important toxin that induces DNA damage and it is critical to triggers cancer. Importantly, this toxin was detected in clinical isolates from patients with IBD (14 of 35; 40.0%) or colorectal cancer (14 of 21; 66.7%), demonstrating that genotoxic *E. coli* is associated with chronic intestinal inflammation and colorectal cancer in humans [108]. Interestingly, the mice that were inoculated with the human-commensal *Enterococcus faecalis* rarely developed tumors, despite similar levels of intestinal inflammation.

1.1.3. Nutritional functions of the human gut microbiota

The gut microbiome, known as "our forgotten organ", encodes about 10 million genes, which perform many of the functions required for host physiology. Thus, the gut microbiome plays an important role in the maintenance of a healthy state in adulthood. In this section, I am going to review the most important aspects of these functions.

Protein metabolism and catabolism

Amino acids biosynthesis, such as lysine and threonine, mainly occurs from the urea degradation by *Bacteroides, Roseburia,* and *Streptococcus* in the distal colon [15]. This function of the microbiota is particularly relevant when diets are deficient in protein [109]. The microbiota can also degrade the amino acids as an energy source. The amino acid fermentation yields both beneficial short chain fatty acids (SCFAs), branched-chain fatty acids (BCFA) [110] and other potentially toxic metabolites such as ammonia, amines, N-nitroso compounds, phenolic compounds and sulfides; however, the products of the fermentation depend on the type of the side chain molecule in amino acids [111].

The protein fermentation results in the formation of hydrogen-sulfide (H_2S) by sulfate-reducing bacteria that utilize hydrogen and sulfate, which can be diet-derived or released from sulfated mucins [112]. H_2S is highly toxic to humans as it increases mucosal apoptosis, goblet cell depletion, superficial ulceration, and causes genomic DNA damage [113].

Metabolic transformations of bile acids by gut microbiota

Ingested lipids can be transformed in bio-active metabolites by the intestinal bacteria, for instance, the bile acids that are not absorbed are normally used by the colonic microbiota [114]. Although it has been discovered that the microbiota has the potential to de-conjugate bile acids [115], mostly species of *Bifidobacterium* and *Lactobacillus* genera are the most likely producers [116]. The metabolic transformation of bile acids might also be beneficial for health, as deconjugation of bile acids by the intestinal microbes is a mechanism by which the colonic epithelium is protected from these, otherwise, genotoxic agents [117].

Vitamins metabolism and synthesis

In humans, it has been shown that gut-microbial species such as lactic acid bacteria and *Bifidobacteria* can *de novo* synthesize and supply the vitamins K and the B complex (biotin, cobalamin, folates, nicotinic acid, pathogenic acid, pyridoxine, riboflavin, and thiamine) [118–120]. Indeed, it has been observed that vitamin metabolism pathways were presented in each of the human-gut enterotypes [14]. As an example, folate is metabolized by species of the genera *Bifidobacterium* [120] and *Lactobacilli* (reviewed in [121]). The vitamin B-2 (riboflavin) biosynthesis starts from the precursor guanosine triphosphate and D-ribulose 5-phosphate and it is mostly synthesized by *Bacillus subtilis* [122] and *E. coli* [123]. The vitamin B12 (cobalamin) is exclusively produced by microorganisms, particularly by anaerobes [124–126] such as *Lactobacillus reuteri* [127,128], meanwhile, the production of the vitamin B complex (niacin and pyridoxine) has been reported for certain lactic acid bacteria such as *Lactobacillus helveticus* or *Bifidobacterium longum* [129]. Finally, vitamin K daily requirement is fulfilled by dietary phylloquinone and, to an undetermined extent, by bacterial metabolism [130,131].

Fiber degradation and production of SCFA

Non-digestible carbohydrates, as dietary fiber, are fermented in the proximal colon by saccharolytic bacteria, being mainly primary fermenters Bacteroidetes and Clostridia species. Major bacterial metabolic routes are the glycolysis and the pentose-phosphate pathway, which convert monosaccharide in phosphoenolpyruvate (PEP). Subsequently, PEP is converted into fermentation products such as the SCFA or alcohols. The acetate ($C_2H_3O_2^-$), propionate ($C_3H_6O_2$), and butyrate ($C_4H_7O_2^-$) represented 90–95% of the SCFA present in the colon [132].

In the colon and feces, the approximate molar ratio of the acetate, propionate, and butyrate are of 60:20:20 [133–135], but depending on the diet, the total concentration of SCFAs decreases from 70 to 140 mM in the proximal colon to 20 to 70 mM in the distal colon [136]. For example, a high-fiber diet with low-fat-meat consumption is characterized by the presence of higher amount of fecal SCFA than diets with reduced fiber intake [56,137,138].

The acetate is the most abundant SCFA in the colon [139] and it is mainly produced by species from the Bacteroidetes phylum [140]. Two different metabolic routes can synthesize it: Via the acetyl-CoA pathway or via the hydrolysis of acetyl-CoA (**Figure I.6a**). Most of the acetate production will be mostly introduced in the lipid biosynthesis. The acetate has been found to be a

key player in the ability of Bifidobacteria to inhibit enteropathogens [141] and interact with the central nervous system [142]. The propionate is also mainly produced by species from the Bacteroidetes phylum via the succinate pathway [143] (**Figure I.6b**). However, this is not the only metabolic route for its production, species from the families Veillonellaceae and Lachnospiraceae can metabolize it via the acrylate pathway [144] and it has been observed that in phylogenetically distant bacteria, including members of families Proteobacteria and Lachnospiraceae, the propionate can be obtained by means of the propanediol pathway [145,146].

The butyric acid is produced by members of the Firmicutes phylum [140]. The acetate CoA-transferase pathway is the most common way to produce butyrate (**Figure I.6c**) and it is carried out by some of the most abundant genera of the intestinal microbiota, such as *Faecalibacterium*, *Eubacterium*, and *Roseburia*. Interestingly, the production of butyrate and propionate by the same bacterium is not common and only a few anaerobes, such as *Roseburia inulinivorans* and *Coprococcus catus*, are able to produce both [145]. An alternative route for the butyrate synthesis is via the butyrate kinase pathway, which employs the phosphotransbutyrylase and butyrate kinase enzymes to convert butyryl-CoA in butyrate [147]. However, this route is not very common in bacteria and is only found in members of the *Coprococcus* genus [144].

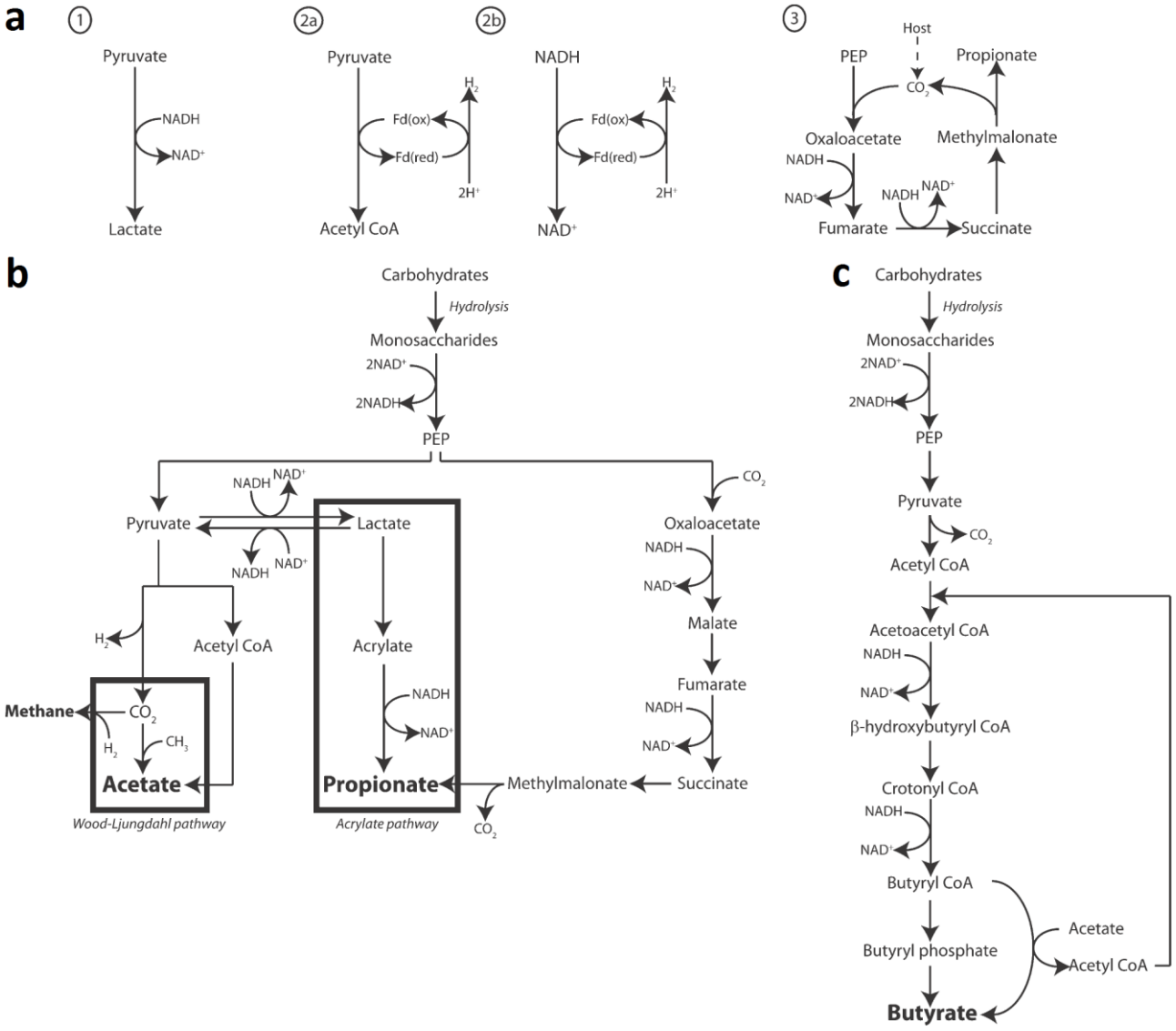


Figure I.6 Schematic overview of the short-chain fatty synthesis. Panel **a**: The pyruvate is reduced to lactate (sub panel 1) or to Acetyl CoA by the ferredoxin oxidoreductase (Fd(ox)) and hydrogenase (Fd(red)) (sub panel 2a). Similarly, the Fd(ox) and Fd(red) can reduce the NADH to molecular H_2 (2b). Finally, PEP can be reduced using the primitive anaerobic electron transport chain for reducing NADH (3). Panel **b**: The acetate production is given by the acetyl CoA or via the Wood-Ljungdahl pathway using formate. The propionate is obtained through the acrylate pathway in which lactate is reduced to acrylate and then reduced to propionate or by the succinate de-carboxylation pathway. Panel **c**: The butyrate is metabolized by the condensation of two molecules of acetyl CoA, by the enzyme butyrate-kinase, or by utilizing acetate through the enzyme butyryl- CoA:acetate-CoA-transferase. Reproduced from den Besten et al. (2013) [148] with permission of the American Society for Biochemistry and Molecular Biology.

1.1.3.1. Biological functions of the short chain fatty acid

Nowadays several molecular mechanisms of action have been ascribed to acetate, propionate, and butyrate that are relevant to their therapeutic potential to promote intestinal health, reduce

inflammation and inhibit cancer (reviewed in [149]). All three SCFAs are used as energy substrates, with propionate serving as a substrate for gluconeogenesis, while acetate and butyrate serve as substrates for fatty acid synthesis (reviewed in [149]). The production of SCFA reduced the luminal pH, which by itself inhibits pathogenic microorganisms and increases the absorption of ions [150]. Also, the SCFA have the capacity to promote mucus production by intestinal goblet cells, induce secretory IgA and regulates the inflammasome (a multiprotein oligomer which is responsible for activation of inflammatory processes) and has been shown to induce cell pyroptosis (a process of programmed cell death distinct from apoptosis) [151,152]. A decrease in luminal SCFAs is associated with ulcerative colitis and intestinal inflammation, which can be ameliorated with dietary fiber or administration [153–155]. For these reasons, several studies have investigated the therapeutic potential of SCFAs and fermentable fibers (reviewed in [149]).

The effect of the SCFAs extends beyond the gut. SCFAs have shown to confer anti-inflammatory effects in the lung [156,157] and brings protection against inflammatory arthritis [158]. Additionally, they are implicated in the improving of the kidney function [159] and are essential in the arterial response to injury, vessel development, and atherogenesis by inhibiting the vascular smooth muscle cell proliferation and migration [160]. The mechanisms underlying these effects appear to center largely on the histone deacetylases (HDAC) inhibition and protein-coupled receptor 43, 41 and 109A (GPR43, GPR41, and GPR109A) [161–163]. These three receptors are present throughout the gastrointestinal tract, as well as on immune cells and adipose tissues, and have been implicated in the regulation of inflammation and cancer.

Butyrate

Butyrate is the main source of energy of the colonocytes through the fatty acid b-oxidation and tricarboxylic acid cycle pathways [146]. The butyrate oxidation by epithelial cells isolated from normal human colonic mucosa produced more ATP than acetate, propionate, and glucose [164]. The butyrate that is not absorbed by colonocytes is transported by the hepatic vein and goes directly to the liver [148].

The butyric acid modulates the transcription of numerous genes through its ability to inhibit the histone HDAC activity [165]. The HDAC inhibition induces cell proliferation and differentiation of the gut epithelial cells [166]; improves tight-junctions integrity [167] and increases mucin

production, by the induction of the *MUC2* gene [168], which is related to the bacterial adhesion [169]. Additionally, it inhibits the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κ B) pathway, a prototypical proinflammatory signaling pathway that expresses genes for cytokines, chemokines and adhesion molecules, [170–174]. Similarly, it is important to induce the intestinal immune tolerance through the regulation of the intestinal macrophages [175], the development of the regulatory T (Treg) cells [176,177] and the regulation of their cytokine production [178]. A reduction in the butyrate concentrations has related to UC [179], obesity and insulin resistance syndrome [180].

In acute inflammations, such as the ones from the IBD patients, it has been observed a butyrate-oxidation deficiency given to a decrease of the expression of the colonic butyrate transporter MCT1 (reviewed in [181]). This leads to a switch from butyrate to glucose as the main source of energy. This shift affects the balance between cell death and proliferation and hence to increase the risk of colorectal cancer [182]. Indeed, it has shown to have a significant role in cancer prevention by promoting colon motility, reducing inflammation, increasing visceral irrigation, inducing apoptosis, and inhibiting tumor cell progression [145,183,184].

Propionate

The propionic acid induces the production of gut hormones, thus reducing food intake [185] and induce the differentiation of Treg cells. Additionally, it lowers fatty acids content in liver and plasma, exerts immunosuppressive actions and probably improves tissue insulin sensitivity. Therefore, the propionate is considered beneficial in the context of prevention of obesity and diabetes type 2 [186]. Additionally, it possesses anti-fungal [187] and antimicrobial activity against the colonization of the gastrointestinal tract by pathogenic bacteria such as *Salmonella* [188], via the inhibition of the expression of the invasion genes in *Salmonella typhimurium* [189].

The propionic acid via the GPCR43 activation has shown to control inflammatory diseases such as colitis, arthritis, and asthma (reviewed in [186]). Additionally, it has a moderate inhibitory activity on cyclooxygenase [190], a major enzyme in the production of pro-inflammatory eicosanoids. Studies in mice have shown that a prebiotic diet [191] and cyclooxygenase inhibition [192] are associated with reduced incidence of colorectal cancer.

Acetate

The acetate is probably the most debated in terms of its beneficial or detrimental metabolic effects. Some studies in rats confirm a direct link between the acetate and obesity [193] while others point out its beneficial effects [142,194]. For instance, a high fiber diet, which increases the acetate concentrations, was found to suppressed allergic airway disease by HDAC inhibition and increased the forkhead box P3 (FOXP3) acetylation in adult mice. This effect was conferrable to fetal mice, in which a high-fiber or acetate maternal diet was able to suppress the expression of certain genes related to asthma [157]. Similarly, "The Canadian Healthy Infant Longitudinal Development study" [195] found that infants at risk for asthma showed transient alterations in the composition of their gut microbiota compared to low-risk infants during the first 100 days of life [196]. These at-risk infants had reduced levels of microbial taxa involved in SCFA formation (specifically *Lachnospira, Veillonella, Faecalibacterium* and *Rothia*) and reduced fecal acetate.

Contrary to this observation, Perry et al. (2016) [193] showed that the microbiota-derived acetate may lead to obesity, insulin resistance, and metabolic syndrome in rats. In this work, they established that the increase of acetate produced by an altered gut microbiota in rodents leads to activation of the parasympathetic nervous system, which in turn promotes increased glucose-stimulated insulin secretion, increased ghrelin secretion, hyperphagia, obesity and related sequelae. This implies that the increased of acetate production is a driver of metabolic syndrome and that the microbial products can directly modify the host´s hormonal response.

1.2. Gut mucosa

The gut mucosa is the largest and most dynamic immunological environment of the body. It is often the first point of pathogen exposure and many microbes use it as a beachhead in the rest of the body. The gut mucosa is the innermost layer of the gastrointestinal tract, which surrounds the lumen. It evolves as a protective epithelial three-layered barrier which regulates the microbial-host interactions [197]. Those layers are:

- The epithelium layer: It consists of a single layer of columnar epithelial cells associated with several specialized secretory cells. It is responsible for the absorptive and secretory processes and is the principal component of the intestinal mucosal barrier (**Figure I.1b**).
- The lamina propia layer: a layer of connective tissue that along with the desmosomes, provide the adhesive bonds that maintain the integrity of the tissue, promote intracellular communication and are the principal determinant of mucosa permeability [198] (**Figure I.1b**).
- The muscularis mucosae, a thin layer of smooth muscle that aids the passing of material and enhances the interaction between the epithelial layers (**Figure I.1b**).

These layers work as a scaffold for the different immune cells, which in collaboration with the epithelial cells protect the host for pathogen invasion. Therefore, the mucosa main functions are: to segregate the luminal microbiota from the intestinal environment; to limit bacterial invasion in host tissues; to attenuate the immune response to beneficial-commensal bacteria and to guarantee the vital functions of the intestinal mucosa and the gut microbiota [197].

1.2.1. The gut mucosa lymphoid tissue

The gastrointestinal tract is a lymphoid organ, and the lymphoid tissue within it is collectively referred to as the gut-associated lymphoid tissue (GALT). Its importance lies in the regulation of the innate and adaptive immune system [198]. The GALT lays throughout the intestine, covering an area of approximately 260–300 m^2 [199] and it is the largest collection of lymphoid tissues in the body. Although its cell composition is homogeneous throughout the GI, it exists differences depending on the intestine zone [200]. In the large intestine, Paneth cells are absent and the

absorptive cells are termed colonocytes. Furthermore, due to the greater number of bacteria that exist in the colon, it is cover with a thick mucus layer which segregates microbiota from the intestinal epithelium [200] **(Figure I.7**). In contrast, in the small intestine, absorptive requirements for enterocytes results in discontinuous mucus layer [200] **(Figure I.7**).

The GALT is formed by different cells types which will be briefly described (**Figure I.7**):

-*Paneth cells:* They are the principal cell types of the epithelium of the small intestine and provide host defense against microbes in the small intestine. Paneth cells secrete a number of antimicrobial molecules into the lumen such as defensins [201].

-*Colonocytes/enterocytes*: columnar epithelial cells, which are predominant in the mucosa's villus. The colonocytes or enterocytes are involved in nutrient absorption such as water, ions, sugar, peptide and amino acid, lipids or the vitamin B12. Additionally, they are related to the secretion of immunoglobulins.

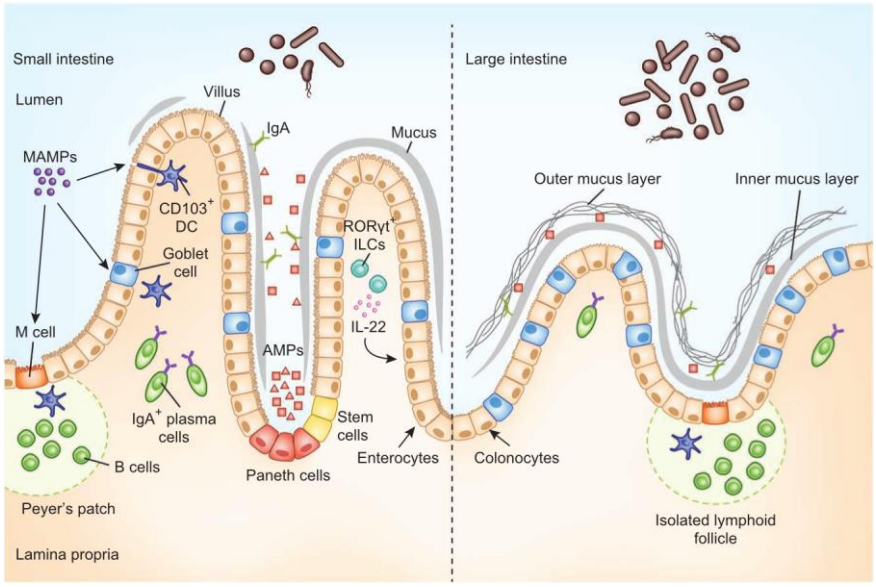


Figure I.7 Anatomical containment of the microbiota along the intestine. The intestinal epithelium comprises a single layer of enterocytes or colonocytes, and it is the role of the immune system to protect the integrity of this barrier. In the small intestine, absorptive requirements for enterocytes results in a discontinuous mucus layer, with fewer goblet cells. Here Paneth cells are enriched in the crypts, secreting antimicrobial peptides (AMPs), which can cross-link with the mucus layer. Through this barrier, a sampling of microbe-associated molecular patterns (MAMPs) can be mediated through antigen uptake by M cells and goblet cells to dendritic cells (DCs), along with a direct trans-epithelial luminal sampling from DCs. The RORγt innate lymphoid cells (ILCs) can sense microbial signals and produce Interleukin 22 (IL-22) to aid in intestinal epithelial cells (IEC) barrier function. Commensal-specific IgA is produced by plasma cells in the lamina propria, mediated by DCs in a T-cell–independent mechanism. The large intestine uses a thick, continuous mucus layer to compartmentalize the microbiota, with the immunoglobulin A (IgA) and AMPs having a secondary role. Reproduced from: Brown et al. (2013) [200] with permission of the Nature Publishing Group.

-*M cells:* these groups of cells initiate the mucosa immunity [202] by taking up antigen from the lumen, by means of the endocytosis, phagocytosis, or transcytosis, and crossing them to the immunity cells [202].

-*Goblet cells:* glandular epithelial cells, which are found sparse among other cells in the gut epithelium. Their main function is to secrete mucus. The mucus is composed of the mucin proteins, a heavily glycosylated family of proteins that form a viscous fluid. The mucus creates a barrier that prevents large particles, including bacteria, to contact directly the epithelial cell layer [203] as well as to maintain the intestinal hydration [204].

-*Enteroendocrine cell*: specialized cells of the gastrointestinal tract with endocrine function [205].

-*Intestinal villi:* villi are small protuberances (0.5–1.6 mm in length) that extend into the lumen of the intestine to increase the absorption area the nutrient intake.

-*Peyer's patches:* the Peyer's patches are an aggregated of lymphoid nodules which normally are found in the lowest portion of the small intestine, mainly in the distal jejunum and the ileum, but also could be detected in the duodenum [206]. The Lymphocytes B are often located in the follicles' germinal center whereas T lymphocytes are found in the zones between follicles. Its main role is to monitor the intestinal bacteria population and to prevent the growth of pathogenic bacteria. Also, the Payer's patches induce the antigen-specific immunoglobulin A (IgA) responses in the gut [207] that plays a crucial role in the immune function of the mucosa membranes.

-*T helper cells (Th cells)*: are a type of T-cell produced in the Thymus that play an important role in the adaptive immune system by suppressing or regulating immune responses by releasing cytokines. They activate T-killer cells (cytotoxic T cells) and B cells to kill infecting bacteria or cells that are infected by viruses. Mature Th cells express the surface protein CD4 and are referred to as CD4+ T-cells which are generally treated as having a pre-defined role as helper T cells within the immune system. These cells are preferentially infected by the HIV [208].

-*T helper 17 cell (Th17):* the Th17 cells are a subset of the proinflammatory T helper cells defined by the production of interleukin 17 (IL-17). The Th17 cells are very important to maintain the homeostasis of the GALT, indeed there is a relation of the Th17 population decay with the increase of the bacterial-metabolites translocation. The Th17 cells also promote neutrophil recruitment, the production of antimicrobial peptides (defensins) (reviewed in [209]) and produce IL-22 which enhances epithelial regeneration [210].

-*Regulatory T-cells (Treg):* Treg cells are a lineage of CD4+ T-cells that express the transcription factor Foxp3. They play a central role in the effective control of self-tolerance and maintenance of immune homeostasis [211]. The absence of the Treg cells is related to inflammatory diseases, autoimmune diseases or lymphoproliferative syndrome including inflammatory IBS, arthritis rheumatoid and systemic lupus erythematosus in humans and animals [212–215]. The secondary functions of this immunosuppressors cells are related to the regulation of the fucosylation (a type of glycosylation) of epithelial cells [216]. Glycosylation of the intestinal epithelium is a key component of the maintenance of the microbial-host interactions providing the first line of defense against pathogens [217]. The Treg cells also act as a central component in regulating the immune response T helpers 1 (Th1) cells [218], the humoral response cells T helpers 2 (Th2) [219] and the Th17 cells [218].

-*Macrophages*: a type of white blood cells that phagocyte debris, microorganisms and pathogens and cancer cells which were not recognized by the immune system. The macrophages also produce lysozyme, which is used as an antibacterial protein [220] and the IL-36α protein, which is a central mediator of cross-talk among intestinal epithelial and mesenchymal cells and macrophages for mucosal healing [221].

-*Dendritic cells*: Antigen presenting cells, which serve as the link between the innate and the adaptive immune system presenting antigens to the T-cell populations. They are in contact with the external environment and serve as agents of active transcytosis or uptake of luminal microorganisms [222,223]. Additionally, dendritic cells possess the capacity to rapidly produce interleukins to a rapid response of the immune system. For instance, the production of the IL-12 is a signal that helps in the differentiation of naive T-cells in Th1 cells [224] and stimulates the production of interferon-gamma (IFN-γ) [225].

-*B-lymphocytes*: specific types of white blood cells of the adaptive immune system that are involved in the humoral immunity by the antibody production. In the GALT, the B- lymphocytes are located in the Peyer's patches.

1.2.2. The role of the microbiota in the adaptation and maturation of the immune system

The GALT and the microbiota are closely related in a way in which the immune system maintains the bacterial population under control and the microbiota stimulates various cell populations [226]. The gut microbiome regulates the production of antibodies by the immune system [227,228] such as IgA that binds to microbes at mucosal surfaces, neutralizes toxins and contributes to microbial tolerance [226]. Additionally, the microbiota has a role in the maintenance of the intestinal epithelial cell by triggering signaling cascades that promote colonocyte regeneration, lymphocyte differentiation and the proliferation of the thigh junctions [229]. Furthermore, the microbiota contributes to the inflammasome induction by promoting the transcription of cytokines [230].

One of the most important tasks of the GALT is to distinguish innocuous antigens from pathogenic microorganisms in order to avoid an excessive immune response [231]. For instance, it has been demonstrated that the polysaccharide A from *Bacteroides fragilis* stimulates the generation of Treg cells via the Toll-like receptor 2 (TLR2) activation [232]. Similarly, *F. prausnitzii* induces the differentiation of the immunosuppressors CD4CD8αα T lymphocytes, which secretes the IL-10. Another important group of bacteria that regulate the immune response segmented filamentous bacteria (SFB) that are members of the Gram-positive Clostridiaceae family [233,234] and are potent promoters of T helper Th17 cells in the intestine [235] and inducer of the production of the IL-22 production [232].

A reduction in the gut microbial diversity in infants is associated with an allergic risk in school age children [236] and the development of food allergy [237]. Indeed, low abundance of species from the genus *Bifidobacterium, Akkermansia* and *Faecalibacterium,* and the high abundance of particular fungi including *Candida* and *Rhodotorula* in the newborn are associated to allergy susceptibility by influencing T-cell differentiation [238]. Moreover, it is well known that the Lipid A of the LPS is responsible for much of the toxicity of some pathogenic Gram-negative bacteria, given its capacity to produce a systemic inflammation. The lipid A activates cells via Toll-like receptor 4 (TLR4) in association with the lymphocyte antigen 96 (also known as MD-2) and the cluster of differentiation 14 (CD14), which in consequence triggers the inflammation response [239–241].

1.3. Human Immunodeficiency Virus infection

Although the clinical community has put their attention in the blood, it is important to note that two-thirds of the CD4+ T-cells are in the GALT, converting the Human Immunodeficiency Virus (HIV) infection in a gut disease. Indeed, the HIV infection is characterized by profound disruption of the GALT and a chronic inflammatory state that persists even after restoration of circulating CD4+ T-cells by successful antiretroviral therapy (ART). Additionally, the infection and the mucosal breakdown lead to compositional and functional dysbiosis of the gut-microbiome.

As the main objective of the present thesis is the study of the role of the microbiota in HIV infection, here I am going to review the most important aspect of the HIV infection, its effects on the gut mucosa and the HIV gut-associated dysbiosis.

1.3.1. Main features of HIV and AIDS

The HIV is a *Lentivirus* that causes the HIV infection and later on the acquired immunodeficiency syndrome (AIDS). The virus is roughly spherical with a diameter of about 120 nm [242]. It is composed of two copies of a single-stranded RNA genome of 9.2kb [243,244] that contain nine genes enclosed by a conical capsid composed of 2,000 copies of the viral capsid-protein p24. The nine viral genes encode for the nucleocapsid protein p7 and enzymes needed for the development of the virion such as reverse transcriptase, proteases, ribonuclease, and integrase. The HIV includes a diverse group of viruses as the HIV type 1 (HIV-1) and HIV-2 [245]. The HIV-1 type is more prevalent and pathogenic than the HIV-2. Homology-based analyses suggest that both HIV-1 and HIV-2 are the result of cross-species transmissions of simian immunodeficiency virus (SIV) from chimpanzees and sooty mangabeys, respectively. The origin of the HIV-pandemic infection occurs by the spread of the virus from non-human primates to humans sporadically throughout the 1900s [246,247]. Infection with HIV could occur by the transfer of blood, semen, vaginal fluid, pre-ejaculate, or breast milk. Within these body fluids, HIV is present as both free virus particles and virus within infected immune cells. By 1980s the virus comes to the world's attention when homosexual men in urban centers presented an advanced and unknown immunodeficiency [248].

UNAIDS organization [249] estimates that 75 million people worldwide have been infected, being approximately 36,7 million located in sub-Saharan Africa. The AIDS is defined in terms of either a

CD4+ T-cell count below 200 cells per µL, or the occurrence of specific diseases in association with AIDS such as the manifestation of infections from opportunistic microbes, as *Pneumocystis pneumonia* as well as viral-induced cancers including the Kaposi's sarcoma, Burkitt's lymphoma, primary central nervous system lymphoma and cervical cancer [250,251].

1.3.1.1. HIV replication and infection

The HIV enters the CD4+ T helper cell [252–254], particularly the Th17 type by the adsorption of glycoproteins on its surface to receptors on the target cells (CC-chemokine receptor 5 (CCR5)) or primers CXC-chemokine receptor 4 (CXCR4) followed by fusion of the viral envelope with the cell membrane and the release of the HIV capsid in the cell. After the fusion, the virus releases its RNA genome and by means of the reverse transcriptase, it is copied in DNA. The per-integration complex is imported to the lymphocyte nucleus and integrated into the host genome. Then, the host enzymes transcribe the viral genes which later on are exported to the cytoplasm where translation occurs to make viral proteins and eventually mature virions [208] (**Figure I.8**). Although the preferred targets for infection are activated T lymphocytes, other blood cells such as dendritic cells and macrophages are infected by the virus, being the infection of these considerably smaller.

During HIV infection, the GALT is disrupted given the loss of the Th17 cells. The depletion of the Th17 cells is greater in the gastrointestinal tract (reviewed in [209]) maybe because the Th17 lymphocytes express the HIV co-receptors CCR5 and CCR6, which are the receptors that the virus recognizes and uses to infect the cell [255]. The loss of the Th17 in the GALT leads to villous atrophy [256], a decrease in the lumen IgA level (reviewed [209]), B-cell dysfunction, damage of the colonocytes (reviewed in [209] and [209]), abnormal enterocyte differentiation [257] and enterocyte apoptosis [258] (**Figure I.9**). When the enterocyte harassment is severe, the tight junction also starts decaying only 14 days post-infection. All this together leads to an increase of the epithelial barrier permeability that facilitates the translocation of microbial products [259] (**Figure I.9**). Additionally, the loss of the Th17 reduces the capacity to control the microbiota and to regenerate the gut epithelium by the recruitment of neutrophils and macrophages.

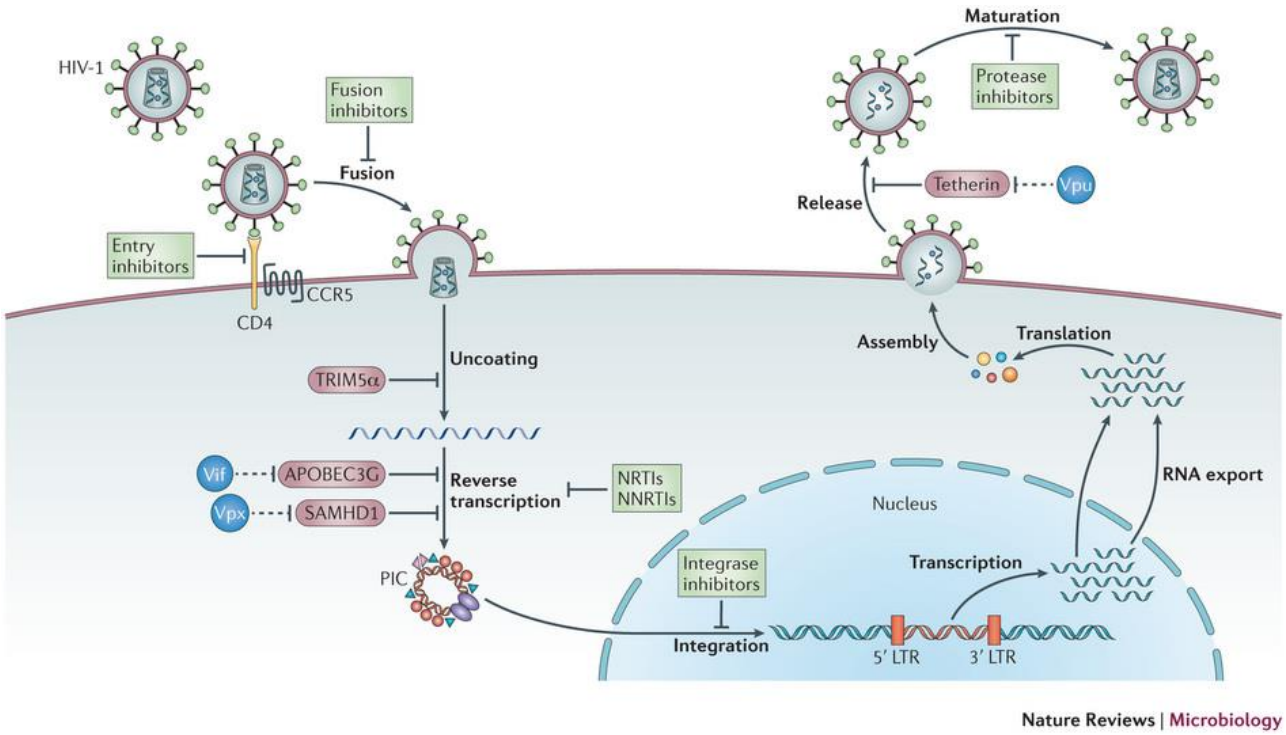


Figure I.8 Schematic overview of the HIV-1 replication cycle. The virus binds to the CD4 receptor and co-receptors (CCR5) to be fused with the lymphocyte membrane. The virus is uncoated in the host cytosol where the HIV-RNA genome and the viral proteins are released into the cytoplasm. Then, its genome is reverse-transcribed to DNA to form the pre-integration complex (PIC) which is translocated into the nucleus. The DNA-virus genome is integrated into the host DNA and subsequently transcribed and translated to a new viral RNA genome and new viral proteins. These elements are translocated to the cell surface to assemble a new immature virus. Then, the virus realized from the host-cell and, by means of the virus-protease, is mature by the cleavages of the structural polyprotein to form mature Gag proteins, resulting in the production of new infectious virions. In green are represented the major families of antiretroviral drugs indicating the life cycle that they block. In red is shown the key HIV restriction factors (tripartite motif-containing 5α (TRIM5α), APOBEC3G, SAMHD1 and tetherin) and in blue their corresponding viral antagonist (Vif, Vpx, and Vpu). CCR5, CC-chemokine receptor 5; LTR, long terminal repeat; NRTIs, nucleoside reverse transcriptase inhibitors; NNRTIs, non-nucleoside reverse transcriptase inhibitors. Image and text adapted from Barré-Sinoussi et al. (2013) [208] with permission of the Nature Publishing Group.

1.3.1.2. HIV treatment: the standard antiretroviral therapy

The ART consists of the combination of antiretroviral drugs to maximally suppress the HIV and to stop the progression of HIV disease [260]. The ART implements a specific drug for each step of the virus replication cycle (**Figure I.8**) such as nonnucleoside reverse transcriptase inhibitors, a protease inhibitor, entry inhibitors and integrase inhibitors. Nowadays several studies have shown that starting ART early after the infection has beneficial effects, indeed the World Health Organization recommends treatment initiation at 500 CD4 + cells/mm^3 per blood or less [261]. The

success of life-long ART is now clear, which markedly reduced morbidity and mortality [262–264] of the HIV infected subjects.

Although ART has markedly improved survival in HIV-infected individuals, the profound CD4+ T-cell depletion in the GALT is incompletely reversed by ART and microbial translocation continues long after peripheral CD4+ T-cell restoration [265,266]. This breakdown of the mucosal barrier results in a chronic exposure to microbial antigens and subsequently, to chronic activation of the innate and adaptive immune system contributing to an excess of morbidity and mortality during treated HIV infection [267,268]. However, this will be deeply reviewed in the next section.

1.3.1.3. Bacterial translocation and inflammation

HIV-infected individuals present a chronic systemic immune activation and inflammation [269–271] that have been associated with the metabolite bacterial translocation; this occurred independently of the response to the ART (reviews in [209]). In fact, among the subjects enrolled in the Strategies for Management of Antiretroviral Therapy (SMART), the level of the indirect marker of the bacterial translocation, the soluble cluster of differentiation 14 (sCD14), was the one that best-predicted mortality. Remarkably, less than 10% of the deaths were caused by opportunistic infections while the remaining 90% were mostly related to cardiovascular events [209]. The microbial translocation affects locally the GALT by driving a local immune activation that causes lymphatic tissue fibrosis and thereby interfere with the correct Th17 cell recovery [272]. Additionally, this microbial translocation triggers the systemic immune activation and chronic inflammation which is related to atherosclerosis and cardiovascular disease.

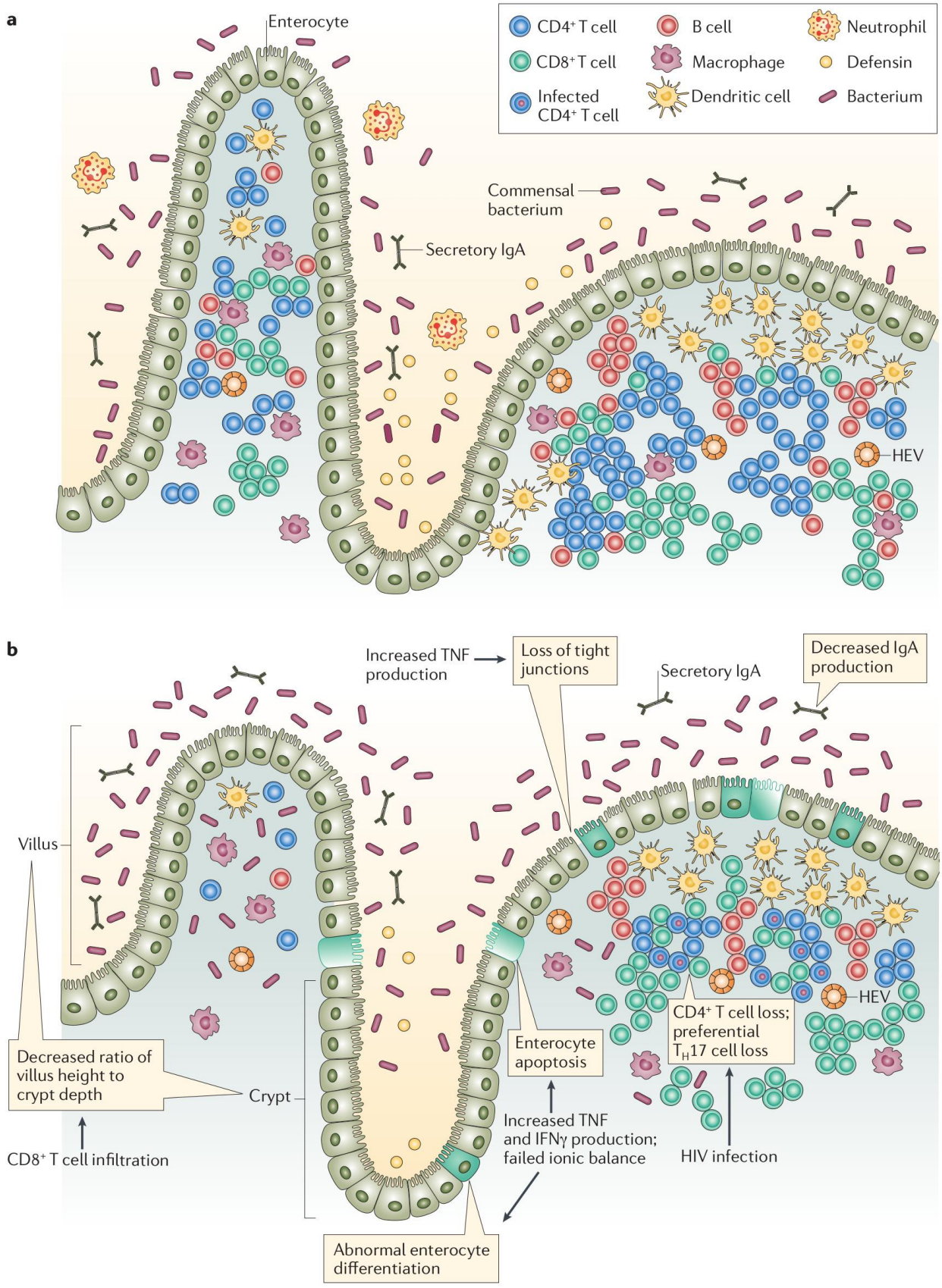


Figure I.9 The GALT in healthy and during HIV infection. Panel **a**: The intestinal epithelium in a healthy individual, which consists of a continuous layer of enterocyte with intact tight junctions that prevents the bacterial translocation. Neutrophils (recruited

by T helper 17 (Th 17) cells in GALT, defensins (produced by Th 17 cells) and secretory immunoglobulin A (IgA) maintain control over the growth of commensal bacteria, further impeding microbial translocation. Panel **b**: The intestinal epithelium in an HIV-infected individual is drastically modified by the loss of the Th17 cell. The decreased villus height/crypt depth ratio may be the result of abnormal enterocyte differentiation and enterocyte apoptosis, which may be caused by the failure of the cells to maintain ionic balance and by increased production of interferon-γ (IFNγ) and tumor necrosis factor (TNF). The loss of the enterocytes and the increase of the TNF production could lead to the destruction of the tight junctions. B cell dysfunction and the absence of Th17 cells contribute to decreased luminal IgA concentrations and to the increased microbial translocation. Reproduced from Sandler and Douek (2012) [209] with permission from © 2017 Macmillan Publishers Limited, part of Springer Nature. All Rights Reserved.

The cardiovascular disease occurs as a consequence of increased levels of microbial products such as LPS and flagellin, which induce the expression of tissue factors that initiate the coagulation cascade on the surface of monocytes [273]. The coagulation cascade leads to the production of fibrin, which is degraded into D-dimers (fibrin degradation product that serves as a proxy for sensing blood coagulation) [274,275], and to the cardiovascular disease. In addition, other diseases have been related to the bacterial translocation as the neurocognitive impairment and the brain atrophy. Individuals with HIV-associated dementia have higher lipopolysaccharide (LPS) and sCD14 levels and lower endotoxin core antibody levels than HIV-infected subjects without detectable cognitive disorders [276,277].

Finally, it has been observed a significant decrease in the Th17/Treg ratio in the HIV-infection. This decay is related to the enzyme indoleamine 2,3-dioxygenase 1 (IDO1), which is an intracellular heme-containing enzyme that is the first step of the kynurenine pathway, the O_2-dependent oxidation of L-tryptophan to N-formyl kynurenine. An important downstream metabolite is the 3-hydroxy anthranilic acid that has been related to the decrease of IL-17 and IL-22 production [278]. These cytokines are involved in the control of extracellular bacteria and they promote mucosal repair. The IDO1 enzyme can also be upregulated by TLR agonists such as LPS and bacterial and viral DNA, which is congruent with it over expression in the HIV infection. The expression of the enzyme correlates with the increased CD8 + T-cell activation and with a reduction in CD4 + T-cell counts [279]. Thus, the bacterial translocation and systemic immune activation are perpetuated by inducing the production of the 3-hydroxyanthranilic acid, which in turn increases the Treg cell frequency and reduce the Th17 and NK (Natural killer cells) levels with a reduction in the secretion of the IL-17 and IL-22 cytokines (reviewed in [209]).

1.3.2. HIV gut-associated dysbiosis

The Th17 cell depletion and the consequent mucosa breakdown lead to a shift in the gut-associated microbiome. Since 2011, thirteen publications have described the HIV gut-associated dysbiosis (**Table I.1**) using different methodological approaches, including cohort size, taxonomic profiling method, sequence technology, type of biological sample (tissue/feces) or the nationality population. Although there is no a full consensus between all the studies, consistent findings have been made in the field.

1.3.2.1. HIV gut-associated microbiota

The HIV-gut-associated dysbiosis is characterized by an increase in Gram-negative bacteria [280] particularly those from *Prevotella* genus [281–284] in both fecal samples [281–283] and colon biopsies [284], being called "HIV *Prevotella* enterotype". Indeed, the whole HIV+ microbiota composition was closely related to those from agrarian communities enriched in *Prevotella* species [281,282]. Other species enriched in the HIV-condition were species related to Negativicutes class (Veillonellaceae, Selenomonadaceae, and Acidaminococcaceae). Desulfovibrionaceae and Erysipelotrichaceae families and the *Clostridium* cluster XIII. Members of the Enterobacteriaceae family [280,285] were found in anal swabs and duodenal biopsies. Contrary, the HIV infection causes a decrease in the abundance of species from the genera *Bacteroides, Alistipes, Parabacteroides* [281–285] and the family Lachnospiraceae [284].

The current evidence indicates that the ART does not restore the HIV gut-associated dysbiosis [281–283,286,287] neither in the short-term treatment [286] nor in the long-term (more than a year) [258,259]. However, Santiago et., al (2013) [287] found that, after ART, less microbial translocation, less systemic immune activation, and less gut T lymphocyte proliferation occurred. Additionally, it has been observed differences in the fecal microbiota from HIV-negative, HIV-positive, and ART-treated infected individuals. Nowak et al. (2015) [286] suggested that the distinct microbial composition within HIV+ individuals may stem from differences in HIV viral load. Then, the HIV+ individuals with controlled viral load have a microbial composition that is distinct from that of viremic patients, and overall more similar to healthy controls.

It is important to state that there are microbial differences given the sampling method, and thus more work is in need to clarify this point. Dillon et al. (2014) [284] characterized the HIV-gut-associated dysbiosis in colonic biopsies, mucosa samples, fecal aspirates and rectal swabs. They discovered that the abundance of the bacteria species depended on the sampling zones. For instance, the increase of Proteobacteria (Comamonadaceae, Campylobacteraceae, Helicobacteraceae, and Moraxellacceae) was detected in the mucosa samples but was undetectable in the fecal aspirates or rectal swabs of the HIV+ subjects. The decrease of Lachnospiraceae and Ruminococcaceae families in HIV+ individuals was observed in fecal aspirates, but not in stool samples. Similarly, exploring other regions of the gut, Mutlu et al. (2014) [288] found that the different compositional patterns among the HIV+ patients that were under the ART and HIV- individuals were also present in samples obtained from the ileum and colon. Specifically, the HIV+ terminal ileum and the colon presented a reduced species richness (i.e., alpha diversity); while the luminal microbiota featured less pronounced differences. Those results were contrary to the ones found by Lozupone et al. (2013) [281] in fecal samples, where the HIV+ group of patients (a mixed cohort of HIV+ under the ART and naïve subjects) showed an increase of diversity. However, Noguera-Julian et al. (2016) [289] in two independent well-sampled cohorts showed that independently of HIV-1 status, the sexual preference markedly leads the shift in the microbiota species composition. Men who have sex with men (MSM) predominantly belonged to the *Prevotella*-rich enterotype whereas most non-MSM subjects were enriched in *Bacteroides*. Additionally, MSM had a significantly richer and more diverse fecal microbiota than non-MSM individuals. However, the HIV infection remained consistently associated with reduced bacterial richness.

1.3.2.2. Effects of the bacteria population on the immune response

It has been observed that HIV gut-associated dysbiosis correlates with an increased mucosal cellular immune activation, microbial translocation and blood T-cell activation [284]. Moreover, the CD4+ T-cells decrease was also related to the reduction of the bacterial diversity and richness [290]. Particularly, Enterobacteria species are prone to trigger the immune response. Species from this family have been related to the duodenal CD4+ T-cell depletion [280]; the increase of the bacterial translocation marker sCD14 [291]; the increase of the cytokine systemic inflammation markers

interleukin 1β [291], and the IL-6 [285] and the activation of the immunostimulatory and immunomodulatory cytokine, the interferon γ [291].

Species from the Bacteroidales order have also shown to be important factors related to the innate and immune activation. The *Prevotella* genus has shown positive correlations between the CD8+ and CD4+ T-cell activation [280,284] and the levels of the myeloid dendritic cells [284]; similarly, species from the *Barnesiella* genus correlates positively with systemic inflammation marker TNF [291]. Nevertheless, no all the species from this order trigger the immune response, the *Bacteroides* genus has shown negative correlations with the systemic inflammation marker IL-6 [288].

The TNF, indicator of systemic inflammation, has shown positive correlations with members of the Firmicutes phylum [288], such as the species from the Erysipelotrichaceae family [291]. However, it is important to note that several species related to this phylum seems to be related to the immunosuppression. The genus *Faecalibacterium*, abundant in colonic biopsies, correlates negatively with the bacterial translocation marker sCD14 [288] and the levels of the lipoteichoic acid, the major constituent of the cell wall of gram-positive bacteria. Moreover, species from the *Ruminococcus* genus show negative correlations with the systemic inflammation marker IL-6 [288]. Some other species depleted in the HIV-gut-associated dysbiosis such as the ones from the genus *Lactobacillus* were also associated with lower markers of microbial translocation (sCD14), higher CD4+ T-cells counts and lower viral load [287].

Table I.1 Studies related to the HIV gut-associated dysbiosis.

Study	Cohort	Sequencing Technique	Sampling method	Main findings
Ellis et., al (2011)	12 Antiretroviral (ART)-naive 5 seronegative	qPCR 16rDNA	Stool samples and duodenal biopsy	• A greater proportion of gram-negative bacteria, Order *Enterobacteriales* was seen in HIV positive individuals compared with seronegative controls. • The proportions of *Enterobacteriales* and *Bacteroidales* correlated with duodenal CD4+ T-cell depletion and peripheral CD8+ T-cell activation, respectively.
Santiago et., al (2013)	13 HIV positive individuals (before and during ART)	Pyrosequencing of the V6 region of 16S rRNA gene	Anal swabs	• Enrichment of *Lactobacillales* in HIV infected individuals before ART was associated with lower viral loads, higher CD4 T-cell concentrations and lower markers of microbial translocation. • Enrichment of *Lactobacillales* in HIV infected individuals after ART initiation was associated with lower translocation, lower systemic immune activation and higher CD4 T-cell concentrations
Mutlu et., al (2013)	21 subjects with HIV under the ART and 22 control subjects	Pyrosequencing of the V1-V3 region of 16S rRNA gene	Colonoscopy from terminal ileum, right colon, left colon and stool samples	• Reduced alpha diversity in the terminal ileum and colon were observed in HIV infection. • Bacteria that increased in the HIV positive group are potentially pathogenic in other disease states. • Higher diversity between microbiota samples in the HIV positive group compared with the seronegative group. • A significant increase in *Brachyspira, Campylobacter, Catenibacterium, Escherichia, Mogibacterium, Prevotella*, and *Ralstonia* was observed in the HIV positive group. • An increase in *Akkermansia, Bacteroides, Blautia, Coprococcus, Dialister, Dorea, Faecalibacterium, Lachnospira, Roseburia, Ruminococcus, Odoribacter, Oscillospira* was observed in the seronegative individuals group.
Lozupone et., al (2013)	22 chronic HIV infected individuals (with or without ART), 3 recently infected HIV positive individuals, 13 HIV seronegative individuals	Illumina sequencing of the V4 region of 16S rRNA gene	Stool samples	• Recently infected individuals have a microbiota that differs only slightly from the microbiota of uninfected individuals. • Short-term ART did not restore the microbiota to its uninfected composition. • Increase abundances of *Prevotellaceae, Erysipelotrichaceae, Veillonellaceae, Clostridium cluster XIII* and the genus *Desulfovibrio* in chronically untreated individuals compared to HIV seronegative individuals. • HIV seronegative individuals had increased abundance of *Bacteroidaceae, Rikenellaceae*, and *Porphyromonadaceae*.
McHardy et., al (2013)	20 HIV seronegative individuals, 20 HIV positive individuals on ART, 20 HIV positive individuals not on ART.	Illumina sequencing of the V4 region of 16S rRNA gene	Rectal mucosa secretions	• Depletion of specific *Lachnospira* and *Eubacterium* and enrichment of *Porphyromonas* and *Anaerococcus* in HIV positive individuals not on ART. • HIV positive individuals on ART showed similar trends but to a lesser extent. • Differences in attributed functionality were found between HIV positive individuals not receiving ART and healthy controls
Vujkovic-Cvijin et., al (2013)	22 HIV positive individuals (6 viremic untreated, 16 on HAART) and 9 seronegative	Phylochip 16S rDNA profiling	Rectosigmoid biopsies	• Dysbiotic mucosal adherent microbiota. • Enrichment of *Proteobacteria* and depletion of *Bacteroidia* in HIV infected individuals. • Dysbiosis in individuals on HAART was correlated with the kynurenine pathway of tryptophan catabolism and plasma concentrations of IL-6
Dillon et., al (2014)	18 HIV+ subjects and 14 seronegative	Illumina sequencing of the V4 region 16S rRNA gene	Colon biopsies, mucosa samples, fecal aspirates or rectal swabs	• Increased abundance of Proteobacteria and decreased abundance of Firmicutes in colon biopsies of HIV infected individuals compared with seronegative individuals. • Within the Bacteroidetes phylum, an increase in Prevotellaceae and a decrease in Bacteroidaceae, Prophyromonadaceae and Rikenellaceae was observed HIV positive individuals. • Within the Firmicutes phylum, Lachnospiraceae, Christensenellaceae and Ruminococcaceae were decreased in HIV-infected patients. • The increased abundance in Proteobacteria seen in mucosal samples of HIV positive individuals, was not observed in fecal aspirates or rectal swabs. • The decrease in mucosal Firmicutes in HIV infected individuals was observed in fecal aspirates, but not in stool samples. • Trends seen in mucosal abundances of Proteobacteria and Firmicutes families and genera were not consistent in stool samples and fecal aspirates.
Nowak et., al (2014)	31 HIV positive individuals (28 viremic, 3 elite controllers), 9 HIV seronegative individuals	Illumina sequencing of the V3-V4 region of 16S rRNA gene	Stool samples	• Decreased alpha diversity in untreated HIV infected before and after the ART. • *Prevotella* genus significantly reduced during ART in HIV positive individuals. • Higher relative abundance of Bacteroidetes in elite controllers compared to viremic patients. • Increased abundance of Actinobacteria and Proteobacteria in viremic patients compared to elite controllers. • Elite controllers did not differ significantly from seronegative controls at the phylum level. • Increased relative abundance of Lactobacillus in viremic patients compared to seronegative individuals. • *Lachnobacterium, Faecalibacterium*, and *Haemophilus* were significantly reduced in viremic patients compared to seronegative individuals.
Lozupone et., al (2014)	40 HIV positive individuals (of them 28 on ART), 15 HIV seronegative individuals.	Illumina sequencing of the V4 region of 16S rRNA gene	Stool samples	• *Prevotella* genus, the Paraprevotellaceae family, and *Eubacterium biforme* increase with HIV infection, abundance varies in individuals undergoing ART (do not reach typical low levels of HIV-negative individuals). • Peptococcus genus increased in untreated HIV infected individuals and decreases with ART. • *Desulfovibrio* and *Catenibacterium* genera increase in untreated HIV infected individuals, while in individuals on ART they trended back to levels seen in HIV seronegative individuals. • The microbiota composition of individuals on ART was more similar to that of individuals with untreated HIV infection than seronegative Individuals.
Dinh et al., (2015)	21 HIV positive individuals on ART, 16 seronegative individuals	Pyrosequencing of the V3-V5 regions of the 16S rRNA gene	Stool samples	• Greater abundance of Proteobacteria in HIV positive individuals compared to controls. • Enrichment in Gammaproteobacteria, Enterobacteriales and Enterobacteriaceae in the Proteobacteria phylum in HIV positive individuals compared to controls. • Enrichement in Erysipelotrichi, Erysipelotrichales, and Erysipelotrichaceae in the Firmicutes phylum in HIV positive individuals compared to controls. • Enrichement in *Barnesiella* and reduction in Rikenellaceae and *Alistipes* in the phylum Bacteroidetes in HIV positive individuals compared to controls.
Noguera-Julian et al., (2016)	HIV-1-infected subjects and HIV-1- negative controls in Barcelona (n=156) and Stockholm(n=84)	Illumina sequencing of the V3-V4 regions of 16S rRNA gene	Stool samples	• Sexual preferences could be related to the HIV-gut asosciated dysbiosis, increasing the abundance species from the *Prevotella* genus.
Ling et al., (2016)	67 HIV-1-infected patients (32 receiving HAART and 35 HAART naïve) and 16 healthy controls	Pyrosequencing of the V1-V3 regions of the 16S rRNA gene	Stool samples	• Prevotella, were prevalent in HIV infected patients *Phascolarctobacterium, Clostridium XIVb, Dialister* and *Megamonas* were significantly correlated with systemic inflammatory cytokines
Monaco et al. (2016)	40 HIV positive individuals on ART, 42 HIV positive individuals not on ART, 40 HIV seronegative individuals	Illumina sequencing of the V4 region of 16S rRNA gene and Viral shotgun metagenomics	Stool samples	• Low peripheral CD4 T-cell concentration associated with an increase in a group of bacteria such as *Enterbacteriaceae, Enterococcaceae* and *Lactobacillaceae*.

1.4. Clinical intervention to the microbiota

The prebiotics and probiotics are modulators of the microbiota composition that contribute to the well-being of their host. Currently, multiple studies have investigated their use to modify the gut microbiota to replace or augment conventional therapies. The microbiota-based therapy has grown markedly, due in large part to an improved safety profile with fewer side effects when compared to traditional therapy.

1.4.1. Prebiotics

Prebiotics are nondigestible food ingredient that promotes or induce the growth of beneficial microorganisms in the intestine. Prebiotics are normally non-digestible fibers that pass through the GIT until the large intestine, where commensal bacteria use it as a substrate. The overgrowth of these bacterial groups induces changes in SCFA production in healthy individuals [292], in patients with IBS [293,294] and in obese individuals [295]. The use of prebiotics has also been tested in the HIV infection. For instance, HIV-infected individuals not receiving ART have maintained the immune functions using a combination of probiotics and/or prebiotics or a mixture (Symbiotics) [296–298]. The trials have been focused on the use of the fructo-oligosaccharides (FOS) [297], mixtures of galacto-oligosaccharides (GOS) and poly-unsaturated acids [296].One of the pioneer studies in the prebiotic use was the one performed by Gori et al., (2011) [299]. In a pilot double-blind, randomized, placebo-controlled study they measured the effectiveness of a prebiotic oligosaccharide mixture (15 or 30 g short chain galactooligosaccharides/long chain fructooligosaccharides/pectin hydrolysate-derived acidic oligosaccharides (scGOS/lcFOS/pAOS) daily), or a placebo for 12 weeks in ART-naïve HIV-1-infected adults. Their results showed that the prebiotic administration increased the abundance of *Bifidobacteria* species and reduced the abundance of the pathogenic species *Clostridium lituseburense* and *Clostridium histolyticum* group. Additionally, the levels of the T-cell activation (%CD4+ CD25+ T-cells) and sCD14 levels decreased after the prebiotic intake [299]. Nevertheless, this work has been performed using qPCR so that only a fraction of the microbial species could be characterized leaving aside the effect of the prebiotic over the whole bacterial community.

1.4.2. Probiotics

The probiotics are living microorganisms that are known to provide benefits to the host [300]. They are used to restore microbial dysbiosis such as the one produced by the *C. difficile* infection [301],

acute infectious diarrhea, Crohn's disease and IBS. Additionally, it has been observed that lactic acid bacteria have the potential for reducing serum cholesterol levels [302,303] and may help lactose-intolerant individuals [302]. Normally, species from the *Lactobacillus* and *Bifidobacterium* are the most widely used as probiotics. For the HIV gut-associated dysbiosis, *Lactobacillus rhamnosus, Lactobacillus reuteri* and *Bifidobacterium lactis Bi-07* [297,298] have been used as probiotics, given their anti-inflammatory properties. Additionally, *Bifidobacteria* species has shown to normalize circulating LPS levels in a mouse model of endotoxemia (the presence of endotoxins in the blood stream) [304]. Moreover, Stiksrud et al., (2015) [305] and Villar-García et al. (2015) [306] showed that the use of probiotics in a cohort of HIV+ individuals reduced significantly the inflammation [305,306].

The use of a mixture of prebiotics/probiotics in SIV-infected macaques beginning ART has resulted in an increased of the frequency of gastrointestinal antigen-presenting cells, enhanced reconstitution and functionality of Th17 cells, and reduced fibrosis of lymphoid follicles in the colon [307].

1.4.3. Fecal transplantation

A new way of restoring the whole microbiota is the fecal microbial transplantation (FMT) that consists in the transfer of the fecal microbiota from a healthy donor in the gastrointestinal tract of a subject with a dysbiosis. Although it was first described in the scientific literature in 1958 [308], there are records that fecal transplantations have been used as a remedy for recurrent diarrhea for at least 1700 years [309].

The FMT has shown excellent results for the control of the antibiotic-resistance *C. difficile* infection. The idea behind the microbial transplantation is that the healthy fecal microbes suppress *C. difficile* blooms through niche competition, however, the exact mechanism is still unknown (reviewed in [85]). Moreover, it is arguably the most complete and ideal probiotic, one can see its potential value in other bacteria-mediated diseases such as UC, IBD, celiac disease, allergies, neurodevelopmental disorders, endocrinopathies, and other autoimmune disorders [310]. Indeed, more than a half a dozen clinical trials have been registered to study FMT in IBD (reviewed in [311]). In the same way, in a study of 10 children with UC treated with FMT showed 78% clinical response within a week [312]. Similarly, a study performed in a small cohort of six subjects reported the success of FMT documenting the complete clinical and histological reversal of UC [313].

However, in some cases, it has been observed adverse events such as transient fever, abdominal pains, bloating, and even no clinical improvement with only transient effects on the host's fecal microbial composition in patients with Crohn's disease [314]. Furthermore, there are risks that the FMT can spread infectious diseases such as HIV or hepatitis [311] or there are also theoretical hazards that FMT could change the microbiome to make people more susceptible to chronic conditions such as obesity or autoimmune disorders [311]. Then, more long-term studies based on larger cohorts are needed to determine the effects of the FMT in different microbiome related diseases and human health.

2. OBJECTIVES

Nowadays, the use of ART in HIV+ subjects has increased considerably their life expectancy restoring the CD4+ T-cell counts and maintaining at low levels the viral load. However, their life expectancy is 10 years lower than the average population. This reduction is due to unrelated AIDS illness, such as cardiovascular diseases and atherosclerosis that are caused by a persistent immune activation and chronic inflammation. This phenomenon could be due to a constant bacterial translocation from the intestinal lumen to the systemic circulation because of the prior disruption of the GALT by HIV. Moreover, the loss of the lymphoid tissue would lead to a microbial imbalance that could be related to the systemic immune activation.

In the last decade, the high-throughput molecular techniques as metagenomics, metatranscriptomics, meta-metabolomics, and metaproteomics allow the whole characterization of the structure, functions, and interactions of the human gut microbiome. However, several pathologies related to the disturbance of the gut homeostasis remains to be investigated, such as the HIV infection. For this reason, the general aim of this thesis was to gain insight into the effect of the HIV infection on the gut microbiome and its implications in human health applying these approaches and different multivariable analysis. This main objective was divided into three specific objectives, which were focused on different aspects of the HIV gut-associated dysbiosis.

Objective 1: Study the contribution of the altered metabolism of the gut microbiota in the chronic immune activation in the HIV-infected ART-responder individuals.

Since long-term ART incompletely reverses the profound CD4+ T cell depletion in the GALT, the immune activation and inflammation continue long after peripheral CD4+ T-cell restoration. Although it has been characterized the dysbiotic microbiota in ART HIV+ subjects, the extent of the community perturbation has not been fully understood. Current studies lack from a functional characterization of the gut-microbiome as well as the clinical impact of the host health.

The first objective was to characterize the structure (diversity and microbial composition) and functions (metagenome) of the microbial ecosystem in an HIV and ART environment in order to identify the interactions between functional and structural dysbiosis and host health.

Objective 2: Study the effects of prebiotics on microbial dysbiosis, butyrate production and immunity in HIV-infected subjects.

Although different studies in HIV patients have described changes in the immunity and in the intestinal bacteria after prebiotic, probiotic, or symbiotic administration, a deeper understanding of the ecological effects of such interventions on the structure of gut bacterial communities is needed, especially in the context of HIV-infected individuals who are undergoing ART.

The second objective was to study the effect of a dietary supplementation with prebiotics and glutamine on the HIV-associated dysbiosis and the innate and adaptive immunity.

Objective 3: Implement a holistic characterization of the HIV gut-associated dysbiosis using a "multiomic" approach and the determination of the ecological and functional network structure.

Despite a large number of studies focused on HIV-associated dysbiosis, the studies of the gene content and gene expression in the gut-associated dysbiosis have not been well characterized. Additionally, the gut-microbiome is a complex community in which exists several relationships between the microbial species and genes. In that sense, the network analysis is a powerful tool to study the combination of a tightly interlinked complex system that happens in the biological systems. Finally, although nowadays exist several studies based on different "omic" technologies, there is no an integrative effort to study the HIV gut-associated dysbiosis in a holistic perspective.

The objectives of chapter three are the characterization of the gut-microbiota using different "omic" technologies, the study of the ecological and functional networks in the HIV-dysbiotic community and the implementation of a "multiomic" approaches to predict the effect of the metagenomic, metatranscriptomic and metabolomic data in the immune response.

3. MATERIAL AND METHODS

3.1. Cohort

A total of 95 subjects were recruited from HIV clinics of two University hospitals in Madrid, Spain (University Hospital Clínico San Carlos and University Hospital Ramón y Cajal) and randomly assigned to the active or placebo group, using centralized allocation via a telephone randomization system. Patients and researchers were blinded to study allocation until statistical analysis. Inclusion criteria were serologically documented HIV infection, age 18 years or older. Controls were healthy HIV-uninfected volunteers (most of them staff working in either institution) and were recruited aiming to reach a group of similar ages.

Exclusion criteria were: use of concomitant medications; use of systemic antibiotics during the previous three months; and any acute or chronic condition other than chronic HIV infection, including gastrointestinal symptoms (constipation, bloating or diarrhea) or co-infections by hepatitis B or C viruses. Study subjects completed a dietary survey detailing the number of servings of various diet-related food groups per day during a five-day framework, including the weekend.

To capture a wide spectrum of HIV immunopathogenesis, we recruited viremic untreated (VU) HIV-infected subjects, immunological ART responders and non-responders (IR and INR, ≥350 and <350 CD4+ T-cell counts after >2 years of viral suppression, respectively) [1], and unmatched HIV-uninfected (HIV-) individuals. From the 95 individuals screened in the study, 35 were not eligible and 60 underwent randomization to the prebiotic intervention or placebo (2:1). A total of 44 individuals completed the 6-week-treatment. 34 subjects were in the active arm (20g mixture of prebiotics, including 5g of short chain galacto-oligosaccharides (Purimune®), 10g of long chain fructo-oligosaccharides (Orafti-HP® and Actilight®), and 5g of glutamine (Nutrición Médica®) and 10 individuals in the placebo arm (20g of maltodextrin): 12 VU (9/3), 8 INR (8/0), 15 IR (10/5), and 9 (7/2) controls (HIV-). Nine patients dropped out of the study (3 in the active arm and 6

in the placebo arm). See **Figure M.1** for a detailed schema of all the process. Fecal and plasma samples were obtained for each subject in the cohort.

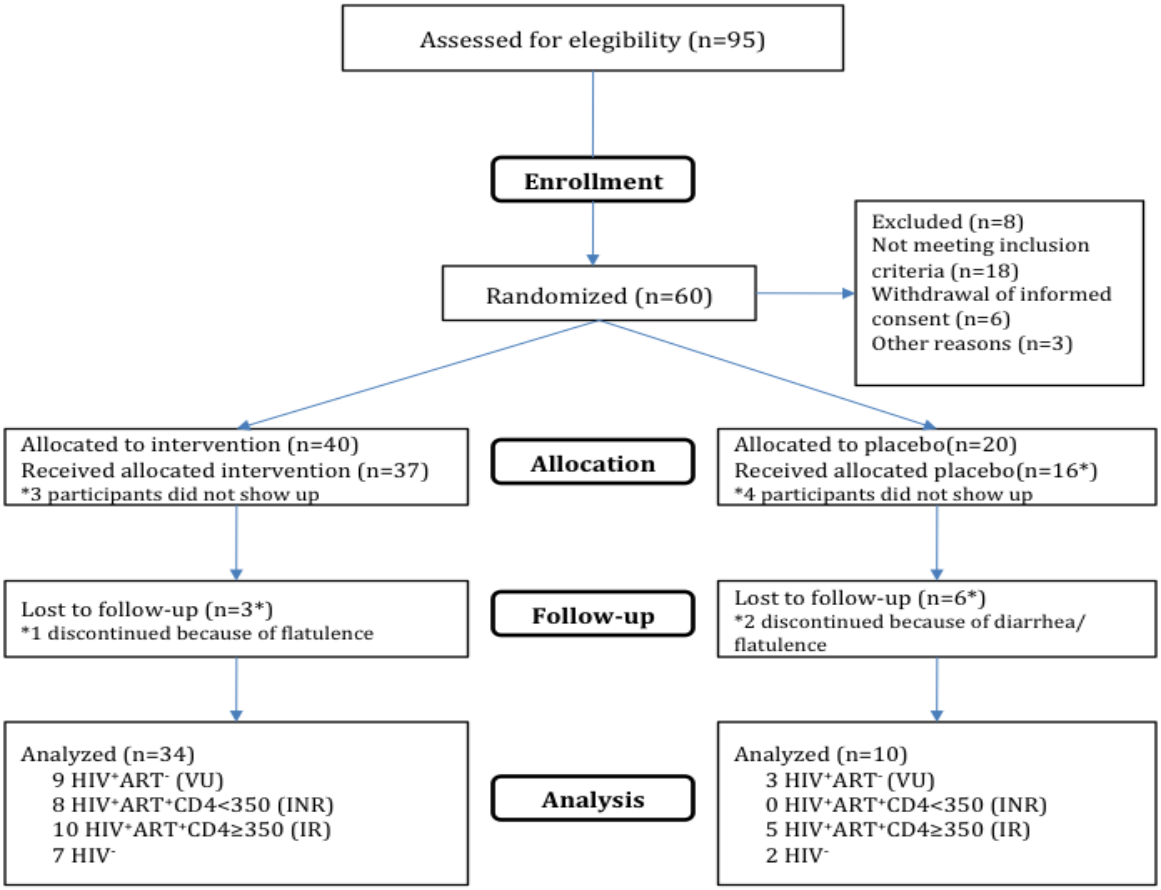


Figure M.1 Study profile. INR, immunological non-responder; IR, immunological responder; VU, viremic untreated. Reproduced from Serrano-Villar et al. (2012) [315] with permission from © 2017 Society for Mucosal Immunology.

This study was approved by the Ethics Committee at both recruiting institutions (University Hospital Clínico San Carlos [approval number 11/284], ceic.hcsc@salud.madrid.org and University Hospital Ramón y Cajal, ceic.hrc@salud.madrid.org). The fecal and blood samples were obtained for each patient before and after the prebiotic administration. Briefly, 20 ml of blood was drawn into EDTA tubes by fasting sterile venipuncture for at least 8 hours and two stool samples were collected in Falcon tubes with 10 ml of later RNA, which were immediately frozen at -80°C.

All participants signed an informed consent prior to initiation of study procedures and all blood and fecal samples analyzed in the study was obtained from participants who signed the consent from Clinical Trials Registry Identification Number Identifier (clinicaltrials.gov): NCT01838915.

3.2. Clinical measurements of the systemic biomarkers of disease progression

As mention before, the HIV gut-associated dysbiosis is related to the bacteria translocation and a chronic immune activation that is associated with cardiovascular diseases and atherosclerosis. In collaboration with Hospital and Clinical centers from Madrid and Sevilla, we measured different markers from the thymic function, T-cell markers, bacterial translocation, endothelial functions and several plasma biomarkers related to systemic inflammation from the blood samples.

3.2.1. Flow mediated dilation

The flow mediated dilation (FMD) measures the dilation of the brachial artery during reactive hyperemia after brief arterial occlusion, involves endothelium-dependent mechanisms and predicts coronary events [316], therefore the FMD marker is a cardiovascular disease marker. Since the gut microbiota has demonstrated to promote atherosclerosis through the metabolism of L-carnitine [317], in the present thesis we measure the effect of the prebiotic treatment to the improvement of a surrogate marker of cardiovascular disease.

3.2.2. Markers of innate immune activation and bacterial translocation

All the plasma biomarkers related to the systemic inflammation, thrombosis, bacterial translocation and the disease progression markers were measured in collaboration with the department of infectious diseases from the Hospital Universitario Ramón y Cajal (Universidad de Alcalá, Madrid Spain).

The clinical markers related to the systemic inflammation: the interleukin-6 (IL6), the high-sensitivity C-reactive protein (hs-CRP) were measured using classic enzymatic methods and nephelometry (VISTA System, Siemens Healthcare Diagnostics Inc, Deerfield, IL) respective. Similarly, the plasma biomarkers related to cardiovascular diseases: the concentrations of glucose, the total cholesterol, high-density lipoprotein and cholesterol and triglycerides levels were measured in the blood samples using standard enzymatic methods. The marker of thrombosis: the fibrin degradation product and D-dimers were measured using turbidimetry method (Beckman-Coulter, Inc, Munster, Germany). The bacterial translocation markers, bactericidal permeability increasing protein (BPI, Hycult Biotech) and soluble CD14 (sCD14, Biovendor), were determined by

immunoassay. The BPI is a 50kDa protein [318] with a potent killing activity against Gram-negative bacteria which bind to the lipopolysaccharides of the bacteria membrane; likewise the sCD14 acts as a co-receptor for the detection of bacterial lipopolysaccharide.

Finally, the disease progression marker, the plasma viral load (VL) was measured using the Cobas Taq-Man HIV-1 assay (Roche Diagnostics Systems, Inc, Branchburg, NJ).

3.2.3. Markers of adaptive immune activation

The systemic chronic immune activation is considered today as the driving force of CD4+ T-cell depletion and acquired AIDS [319]. The adaptive immune activation markers measure the infection progression [320,321], immunosenescence [322] and the response of the innate immune system to the microbial translocation (reviewed in [323]). For these reasons in collaboration with the Hospital, Universitario Virgen del Rocío (Sevilla, Spain) and the Department of Infectious Diseases from the University Hospital Ramón y Cajal (Madrid, Spain) different markers of the adaptive immune activation were measured from the blood samples.

3.2.3.1. T-cell immunophenotyping

In collaboration with the Department of Infectious Diseases from the University Hospital Ramón y Cajal (Madrid, Spain), we obtain the clinical variables related to the HIV immune cell harassment, the CD4+, and CD8+ T-cells count. The percentage of cells expressing markers of activation (CD25+, CD38+, HLADR+ or CD38+/HLA-DR+) and senescence (CD57+). These values were measured from the fresh EDTA anticoagulated whole blood using the following antibody combination: CD3-eFluor450, CD4-peridinin chlorophyll protein complex-Cy5.5 (PerCP-Cy5.5), CD8- Horizon V500, CD25-Allophycocyanin (APC), CD8-phycoerythrin-Cy7 (PE-Cy7), CD38-APC700 and HLA-DR-APC-Cy7. Antibodies were from Becton Dickinson (Becton Dickinson, New Jersey, USA), and unstained controls were performed for all samples. Cells were collected using a Gallios flow cytometer (Beckman Coulter, California, USA) and analyzed with Kaluza software (Beckman Coulter) to quantify the percentage of CD4+ and CD8+ T-cells expressing markers of activation (CD25+, CD38+, HLADR+ or CD38+/HLA-DR+) and senescence (CD57+).

3.2.3.2. sj/β-TREC ratio quantification

In collaboration with the Hospital Universitario Virgen del Rocío (Sevilla, Spain) the thymic function was calculated in peripheral blood mononuclear cell DNA using the ratio of two different

T-cell receptor excision circles (TRECs): The signal-joint T-cell receptor excision circles (sj-TREC) and the DβJβ-TREC. The TRECs are small circles of DNA created in T-cells during their passage through the thymus as they rearrange their T-cell receptor (TCR) genes. Therefore, the ratio of such TRECs, often expressed as the sj/β-TREC ratio, is a direct estimator of the thymic function which has the advantage to take into account the DβJβ-TREC, which is the product of the β chain TCR rearrangement at the most immature thymocyte subset and the sj-TREC, product of the α chain TCR rearrangement. The sj/β-TREC ratio was estimated by the subsequent PCR reactions as previously described [324] with minor modifications as shown in [315].

3.2.4. Plasma concentrations of trimethylamine N-oxide

High levels of trimethylamine N-oxide (TMAO) are associated with an increased risk of major adverse cardiovascular events and atherosclerosis [325,326]. Interestingly, the gut microbiota has the potential to metabolized choline or carnitine to TMAO increasing its levels in the blood stream. Therefore, in collaboration with the Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad CEU San Pablo, Campus Montepríncipe, (Madrid, Spain) using liquid chromatography techniques, as described in [315], the blood levels of TMAO were measured in the cohort participants.

3.3. Metagenomic and metatranscriptomic sequence analysis

In order to characterize the colonic microbiome fecal samples were taken from each of the participants of the cohort, before and after the prebiotic administration. The fecal sampling has the advantage of being an easy-to-sample non-invasive method which has been widely used for the gut-microbiome study. Additionally, it has been reported that luminal microbial contents of the colon, in which reduced transit time and high nutrient availability are observed, correlate with feces in terms of species diversity and bacterial abundance [10–12]. For these reasons, nucleic acids from the colonic bacteria populations, retrieved from the fecal samples, were used to characterize the taxonomic and functional profile of the microbiome.

3.3.1. Nucleic acid purification

Fecal samples were stored in RNAlater (Life Technologies, Carlsbad, CA) at -80ºC until use. The fecal samples were defrosted and homogenized and 5 ml of each were diluted with 5 ml of phosphate buffered saline (PBS) (containing, per liter, 8 g of NaCl, 0.2 g of KCl, 1.44 g of Na2HPO4, and 0.24 g of KH2PO4 [pH 7.2]). Then, they were centrifuged at 2000 rpm at 4ºC for 2 min to remove fecal debris. The supernatant was centrifuged at 13000 rpm for 5 min to pellet the cells Total DNA was extracted from pelleted cells using QIAamp DNA Stool Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Total RNA was extracted using the RiboPureTM Bacteria kit (Ambion, Austin, TX, USA) and then treated with Baseline-ZEROTM DNAse (Illumina, Eindhoven, Netherlands) as described in Pérez-Cobas et al., (2013) [327]. The efficiency of the DNase treatment was checked by amplifying each RNA sample by PCR. Then, the rRNA removal was performed using Ribo-Zero Magnetic Kit (Bacteria) (Illumina, Eindhoven, Netherlands).

3.3.2. Analysis of the 16S rRNA gene

3.3.2.1. 16S rRNA gene amplification and sequencing

For each sample, the hypervariable V1-V3 region of the 16S rRNA gene was amplified by polymerase chain reaction (PCR) with the universal primers E8F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 530R (5′-CCGCGGCKGCTGGCAC-3′). The E8F

primer included a sample-specific Multiplex Identifier (MID) to be multiplexed and sequenced simultaneously. To amplify the 16S rRNA gene we used the following protocol.

For each sample, a 50µl PCR mix was prepared to contain: 5µl of Buffer Taq (10X) with 20mM MgCl2, 2µl of dNTPs (10mM), 1µl of each primer (10mM), 0.4µl of Taq FastStart polymerase (5u/µl), 39.6µl of nuclease-free water and 1µl of DNA template. PCR was run under the following conditions: 95º for 2 min followed by 25 cycles of 95º for 30 s, 52º for 1 min and 72º for 1 min and a final extension step at 72º for 10 min. The amplification process was checked by electrophoresis in agarose gel (1.4%). PCR products were purified using NucleoFast® 96 PCR Clean-Up Kit (Macherey-Nagel) and quantified with Nanodrop-1000 Spectrophotometer (Thermo Scientific) and with the QuantiT PicoGreen dsDNA Assay Kit (Invitrogen). The PCR products were purified using the NucleoFast 96 PCR Clean-Up Kit (Macherey-Nagel, Duren, Germany). The pooled PCR products were directly pyrosequenced using a Roche GS FLX sequencer and Titanium chemistry in the Centre for Public Health Research (FISABIO-Salud Pública, Valencia Spain. The raw sequences retrieved from the Roche GS FLX sequencer were subject to quality trimming using the prinseq-lite.pl script [328] (parameters: -min_len 170 -min_qual_mean 20 -ns_max_n 1 -trim_qual_left 20 -trim_qual_right 25 -trim_qual_type min -trim_qual_rule lt -trim_qual_window 5 -trim_qual_step 1). The reads that did not accomplish one of the following criteria: read length above the 170 nucleotides (nts), mean quality score above 20 or contain less than the 1% of the ambiguous base were removed. Additionally, the prinseq-little script trims the low quality 3´ end of the sequences by the sliding window method, which takes five bases at the end of the sequences and removes them if at least one of the bases has a quality score below 25. If the 5-base window is removed then the window moves one base to the left and redoes the process.

3.3.2.2. 16S OTU characterization

The Amplicon data were analyzed following the recommendations of the metagenomic state-of-the-art pipeline QIIME (v 1.8.0.) [329]. The master-script pick_otus_through_otu_table.py (master-script) from the QIIME pipeline [329] was used to picks Operational Taxonomic Units (OTUs) from the 16S rDNA amplicons. Briefly, an OTU is an operational definition used to classify groups of closely related individuals [330]. In the metagenomic analysis, the OTUs are defined as clusters of 16S small subunit (SSU) rRNA gene which normally shares at least the 97% of sequence identity. The OTUs are often considered a synonym of bacteria species; however, the identity threshold to separate the sequences in different species OTUs could vary depending on the species.

In the present thesis, it is referred as an OTU a set of sequences which share at least a 97% of sequence identity. The classification of species or genus was carried by adding all the OTUs who share the same taxonomic label at the determined taxonomic level.

OTUs are the backbone of established workflows such as the QIIME pipeline which implement the pick_otus_through_otu_table.py master script to quantify the OTUs from the 16S rDNA amplicons following the next steps:

1) OTU creation: The script pick_otus.py performs sequence denoising, chimera detection, and OTU-clustering. The USEARCH [331] pipeline was first used for filtering chimeric sequences given PCR amplification. Two strategies of chimera detection were used: The reference based approximation (parameters: usearch --usersort –chimeras –uchime) using as template the curated 16S rDNA database rRNA16S.gold.fasta (from 2010-04-29) and the *de novo* chimera detection (usearch --usersort –chimeras --abskew 2 –uchime). Next the USEARCH pipe-line cluster all sequences who share at least 97% of similarity (parameters: usearch –cluster --id 0.97 --w 64 --maxrejects 32). Finally, those sequences which were not allocated in a cluster were label as putative sequence artifacts and were removed from the dataset.

2) Obtain the reference sequences for each OTU: the script pick_rep_set.py took the reference sequences for each of the clusters in the dataset.

3) Assign OTU taxonomy: the script assign_taxonomy.py assign the taxonomy from each of the OTUs reference sequences. The script calls the Ribosomal Database Project-II (RDP) script [332], which infers the taxonomy using as reference the Greengenes database [333] (gg_13_8_otus cluster at 97% of identity available in the Qiime package v1.8). The annotation was accepted when the bootstrap confidence estimation value was over 0.8, stopping the assignation at the last well-identified phylogenetic level.

4) Alignment of the reference sequences: the reference sequences were carried out by the align_seqs.py script. The script calls the PyNastAligner alignment software [334] (parameters: algorithm: NAST, min_len: 246, min_pct: 75, pairwise_alignment_method: UCLUST) using as reference the greengenes core_set_aligned.fasta (available in the Qiime package v1.8). Then, the filter_alignment.py removes positions, which are gaps in every sequence.

5) Make a phylogeny: The script make_otu_table.py used the FastTree [335] (parameters:-root_method midpoint) to generate a phylogenetic tree using as template the alignment generated by PyNastAligner.

6) Finally, the abundance OTU table was created using the script make_otu_table.py, the script merged the reference OTU taxonomic assignation with its relative abundance in the samples. The OTU table was also summarized in the genus taxonomic level using the script summarize_taxa_through_plots.py.

Rarefaction curves of fecal microbiota at genus, species and OTU taxonomic levels were constructed (library "vegan" function "rarecurve") to assess the saturation of the 16S rRNA gene sequencing. As shown in **Figure M.2**, the curves reached a plateau for all three taxonomic levels, indicating that the species level was well represented in the study. Additionally, 5 rarefactions of 2185 sequences per sample were performed to validate the 16S-amplicon sequencing analysis. The results were compared with the analysis using the entire dataset using a Procrustes test (library "vegan" function "protest") and the Mantel test (library "vegan" function "mantel.rtest"). We did not find any significant difference (mantel.rtest = 0.001 and protest p-value < 0.001) between the OTU table obtained with the whole dataset and that obtained for each of the rarefied samples.

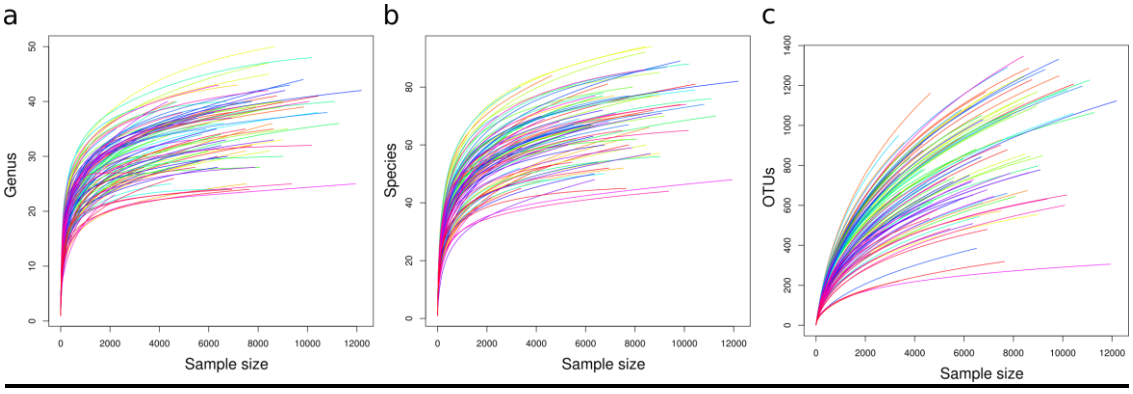


Figure M.2 Rarefactions curves. Rarefactions curves from the total samples used in the cohort at genus (**a**), species (**b**) and OTUs (**c**) levels. Modified from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

3.3.2.3. Microbial quantification by quantitative PCR

The qPCR reactions were performed using LightCycler 480 instrument (Roche) and KAPA SYBR® FAST qPCR Kit (Kapa Biosystems). Amplification reactions were run on total DNA purified from the fecal sample as stated before using 0.2 µM of the universal bacterial primer 8F (5′-AGAGTTTGATCCTGGCTCAG) and the broad-range bacterial primer 338R (5′-

TGCTGCCTCCCGTAGGAGT-3′). Final assay volumes of 20μL were dispensed in duplicate in 96-well plates. We used 25 ng average of genomic DNA per 20 μL reaction as template. Standard curves were prepared by serial dilution of the PCR product of *Enterococcus faecalis* 16S rRNA gene obtained using the primers described above. The reaction conditions were 95°C for 10 min followed by 40 cycles of 95°C for 30 s, 52°C for 30 s, and 72°C for 1 min. The results were expressed as the number of 16S rDNA copies per ng of total DNA.

3.3.3. Metagenomics and metatranscriptomics sequencing

In the present thesis, we used two different sequence technologies to obtain the taxonomic and functional assignation within the gut-microbiome. The 454 sequencer Genome Sequencer FLX with GS FLX Titanium technology and the Illumina MiSeq Reagent Kit v3 (600 cycles) sequencing. The two sequence technologies were used given that at the time that we started the project the Illumina sequencing was not affordable and the read length was no longer than 100 bp. However, with the acquisition of a MiSeq sequencer in the research center in which I realized the thesis, the "Fundación para el Fomento de la Investigación Sanitaria y Biomédica", the Illumina sequencing was affordable for my thesis project. With these sequencing technologies available, I took advantages of the read-length of the 454 sequences and the coverage of the Illumina MiSeq reads to characterize the gene content and the taxonomic profile of the microbiome. Specifically, the metagenomic data (DNAseq) included both sequence technologies while the metatranscriptomic data (RNAseq) only includes the MiSeq sequencing. The sequencing analysis and treatment are specified in the "Material and Methods" section of Chapter 1, Chapter 2 and Chapter 3 according to the sequencing technology used.

3.3.3.1. 454 pyrosequencing

The 454 libraries for shot-gun pyrosequencing were performed following the protocol of "Rapid Library preparation for FLX Titanium, Roche" using the Roche GS FLX Titanium Sequencing Kit XL. Briefly, the total DNA from fecal samples was fragmented by nebulization and purified by column centrifugation (Qiagen MinElute PCR Purification Kit). The DNA-end fragments were repaired and ligated to the sequencing adaptor and the Multiplex Identifier (MID, Roche) by incubating for 10 min at 25ºC with the "RL MID Adaptor mix". Then, the small DNA fragments were removed by AMPure magnetic beads. Finally, the library was quantified with the QuantiT PicoGreen dsDNA Assay Kit (Invitrogen). The DNA libraries were then pyrosequenced using the

Roche GS FLX sequencer and Titanium chemistry in the Centre for Public Health Research (CSISP-FISABIO, Valencia Spain)

3.3.3.2. Illumina sequencing

The Illumina libraries for shot-gun V3 sequencing were performed following the protocol of "Nextera® XT DNA Sample Preparation Guide". Briefly, the total DNA was tagged and fragmented by the Nextera XT transposome. The Nextera XT transposome simultaneously fragments the input DNA and adds adapter sequences to the ends, allowing amplification by PCR in which the index primers are added to each sequence. Similarly, to the 454 libraries construction, the Illumina libraries were purified by the magnetic Ampure XP beads. Finally, the library was quantified with the QuantiT PicoGreen dsDNA Assay Kit (Invitrogen).

The RNA was processed with ScriptSeq V2RNA-Seq Library Preparation kit (Illumina, Eindhoven, Netherlands) to generate the double strand cDNA and the sequencing libraries according to the manufacturer's instructions. Briefly, the fragmented RNA was reverse transcribed into cDNA using the anneal cDNA synthesis primers and the StarScript AMV Reverse Transcriptase. Then, the 3´end of each cDNA fragment is tagged with a terminal tagging oligo which later was used to perform a PCR which added the sequencing adapters and the index primers to each sequence. The Illumina libraries were purified by the magnetic Ampure XP beads. Finally, the library was quantified with the QuantiT PicoGreen dsDNA Assay Kit (Invitrogen). The metatranscriptomes and metagenomes were sequenced using the Kit V3 (2X300 cycles) in MiSeq platform (Illumina, Eindhoven, Netherlands). The sequencing was performed at the Centre for Public Health Research (FISABIO-Salud Pública, Valencia, Spain).

3.4. Phylogenetic analysis, alpha diversity, beta diversity, and clustering

3.4.1. Alpha diversity

The alpha diversity estimators for the 16S rDNA OTUs were calculated doing a rarefaction of 2185 sequences. The Shannon index, the richness estimators Chao1 and ACE and the total number of taxa were calculated to assess the OTUs and genus diversity within the community using the alpha_diversity.py script from the Qiime v1.8.0 pipeline.

The diversity estimators for the RNAseq and DNAseq were estimated from their taxonomic abundance tables (See Chapter 3 Materials and Methods section) by performing a rarefaction of 16476 and 91719 sequences respectively. The diversity Shannon index was calculated using the "diversity" function (Library "vegan"), similarly, the Chao1 estimator was calculated using the "chao1" function (Library "fossil") for all the data matrix.

3.4.2. Beta diversity and clustering

The beta diversity measures the different species composition among habitats by means of a dissimilarity/similarity index or the use of specific ecological distances. Within the gut microbiome, the ecological analyses normally take advantage of the use of clustering and ordination techniques, which define the structure and the cluster configuration of different human gut associated habitats. In the present thesis, different beta diversity distances/index, cluster algorithms, and ordination analyses were implemented to better characterize the different "omic-data" retrieved from the cohort subjects.

The beta diversity analysis from the 16S rDNA OTUs data was estimated using a combination of the Qiime v1.8.0 pipeline and the free statistical package R (version 3.0.1). The clustering analysis of the samples was performed with the total OTU table and the table summarized at the genus level using the free statistical package R (version 3.0.1) as described by Arumugam et., al (2011) [14] and Koren et., al (2013) [336]. Prior to performing the clustering analysis, the samples were standardized by the total number of sequences. The Partitioning Around Medoids (PAM) algorithm (library "cluster", function "pam") was used to identify the potential cluster in our data set testing four different distances: Bray-Curtis (library "Vegan", function "vegdist"), Jensen-Shannon

divergence (library “phyloseq”, function “distance”), Jensen-Shannon distance, calculated as Arumugam et., al (2011) and weighted UniFrac distance [337] (implemented by the beta_diversity.py script in the Qiime1.8 pipeline, using the previous calculated FastTree reference OTU phylogenetic tree). The last one was used only for the OTUs cluster analysis. The optimal cluster configuration was defined as the distance that maximized the silhouette index (library “cluster”, function “silhouette”); enhance the variance explained by the first component of the Principal Coordinates Analysis (PCoA) (function “dudi.pco” function “ad4”) and the distances that were based on ecological or phylogenetic principles. Considering those conditions, the weighted UniFrac (wUniFrac) distance and the Bray-Curtis index were used to estimate the dissimilarities between samples in the OTU table and the genus table, respectively. Clusters were validated applying the Permutational Multivariate Analysis of Variance Using Distance Matrices (ADONIS test) (library “Vegan” function “adonis”) using the weighted UniFrac distance and default 999 permutations.

The annotations from the KEGG orthologous (KO) database, the Comprehensive Antibiotic Resistance Database (CARD) and the taxonomic annotations were standardized by the total number of sequences (see Chapter 3 “Materials and Methods” section), this performed for the RNAseq and the DNAseq shot-gun data. Then, to avoid biases given the data composition we normalize data using the Hellinger transformation (Library “vegan” function “decostand”). The Hellinger distance (Library “vegan” function “vegdist”) and the Bray-Curtis index (Library “vegan” function “vegdist”) were calculated from the abundance matrix from the functional and taxonomic composition, respectively, to quantify the compositional dissimilarity between two different communities. Similar to the 16S rDNA analyses, the PAM algorithm was used to determine the cluster configuration. The optimal cluster configuration for the RNAseq/DNAseq KO functional and taxonomic annotations were established as the one that maximized the silhouette index. Those cluster configurations were validated using the ADONIS test.

For all the different “omic” approaches, the compositional differences between the groups of the cohort as well as the one given by the prebiotic administration were tested using the ADONIS test (library “Vegan” function “adonis”).

The non-metric multidimensional analysis (Library “vegan” function “metaMDS”), as well as the PCoA (function “dudi.pco” function “ad4”) were performed to reduce dimensionality in the taxonomic and functional distance matrix (Library “vegan” function “metaMDS”). The comparison

between the ordination analysis was performed using the Procrustes test (Library “vegan” function “protest”), similarly, the Mantel test was performed to measure the correlation between the distance matrix (Library “vegan” function “mantel.rtest”). The representation of each of the ordination analysis and the clustering analysis was given by the use of the s.class (Library “ade4” function “s.class”) and the vegan functions Ordihull and Ordispider (Library “vegan” functions “Ordihull” and “Ordispider”); those which represented as ellipses which represent the percentage of the samples that belongs to a cluster.

3.5. Biomarker discovery

The linear discriminative analysis (LDA) effect size (LEfSe) [338] is an algorithm for high-dimensional biomarker discovery of genomic features (species, pathways, and genes) between two or more groups. LEfSe first identifies the significant differences in taxa/function composition between groups by applying the Kruskal-Wallis test. Then, the Wilcoxon test was used to check all pairwise comparisons within the groups. Finally, the LDA is estimated for those taxa/function whose Kruskal-Wallis test and Wilcoxon test p-value were below 0.05, the threshold used to consider a discriminative feature for the logarithmic LDA score was set to >2. The LDA is supported by bootstrapping over 30 cycles, each sampling two-thirds of the data with replacement and subsequent averaging. In the present thesis, LEfSe was used to identify specific taxa, genes, and metabolic pathways as biomarkers for the HIV+ subjects and the HIV- controls and within the four groups.

3.6. Generalized linear models

The Generalized Linear Models (GLM) are a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution. The GLMs were estimated using the glment R package [339], which fits a generalized linear model via penalized maximum likelihood. The regularization path is computed for the least absolute shrinkage and selection operator (lasso) or elasticnet penalty at a grid of values for the regularization parameter lambda. Glmnet fits models by solving the following problem:

$$\min_{\beta_0,\beta} \sum_{i=1}^{N} w_i l\left(y_i, \beta_0 + \beta^T x_i\right) + \lambda \left[(1-\alpha)\beta \frac{2}{2} + \alpha \beta \right]_1,$$

over a grid of values of λ covering the entire range. Here $l(y,\eta)$ is the negative log-likelihood contribution for observation l; e.g. for the *Gaussian* case, it is $1/2(y-\eta)^2$. The elastic-net penalty is controlled by α and bridges the gap between lasso ($\alpha = 1$, the default) and ridge ($\alpha = 0$). The tuning parameter λ controls the overall strength of the penalty [340].

The ridge penalty shrinks the coefficients of correlated predictors towards each other while the lasso tends to pick one of them and discard the others. The elastic-net penalty mixes these two; if predictors are correlated in groups, an $\alpha = 0.5$ tends to select the groups in or out together. This is a higher-level parameter, and users might pick a value up front, else experiment with a few different values. One use of α is for numerical stability; for example, the elastic net with $\alpha = 1 - \varepsilon$ for some small $\varepsilon > 0$ performs much like the lasso but removes any degeneracies and wild behavior caused by extreme correlations [340].

In the present thesis, several GLM were implemented to associate the effect of the dysbiosis over the immune system and the association between metabolites and pathways with specific taxa. Those GLM implementations are described in the method sections of its corresponding chapter.

3.7. Bayesian networks

The Bayesian networks (BN) are probabilistic directed acyclic graphical (DGA) models in which the nodes represent random variables and the edges conditional dependencies. In a DGA the connecting edges are represented by arrows ($Xi \rightarrow Xj$) and represent a statistical dependence between the corresponding variables as can be seen in the example represented in **Figure M.3a** [341]. This means that a value taken by variable X_j depends on the value taken by variable X_i. In this case, the X_i is then referred to as a *parent* of X_j and, similarly, X_j is referred to as the *child* of X_i. The structure of the acyclic graph guarantees that there is no node that can be its own *parent* or its own *child* [342].

The BN enable an effective representation and computation of the joint probability distribution (JPD) over a set of random variables [341] and its corresponding factorization. This is achieved by the fact that in a BN each variable node is independent of its non-descendants in the graph given the state of its *parents'* nodes. This is consistent with the Markovian property [343], where the conditional probability distribution (CPD) at each node depends only on its parents. Thus, the joint distribution of a collection of variables can be determined uniquely by these local conditional probabilities. This property is used to reduce the number of parameters that are required to characterize the JPD and allow the estimation of BN when the datasets are highly multivariable, such as the ones observed in the metagenomic studies. Bayesian Networks are formally defined as:

A Bayesian network *BN* is an annotated acyclic graph that represents a JPD over a set of random variables *V*. The network is defined by a pair $BN = \{G, \Theta\}$, where *G* is the DAG whose nodes X_1, X_2, $X_3 \ldots, X_n$ represents random variables, and the edges represent the direct dependencies between these variables. The graph *G* encodes independence assumptions, by which each variable X_i is independent of its nondescendants given its parents in *G*. The second component Θ denotes the set of parameters of the network. This set contains the parameter $\theta_{xi|\pi i} = P_B(x_i|\pi_i)$ for each realization x_i of X_i conditioned on π_i, the set of parents of X_i in *G*. Accordingly, *BN* defines a unique *JPD* over *V*, namely*:* (Definition taken from [342]).

$$PB(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P_B(X_i \mid \pi_i) = \prod_{i=1}^{n} \theta_{X_i|\pi_i}$$

With cross-sectional data, the connecting arrows represent mutual associations rather than causality [59] those networks are also called Markov networks [342] and provide a simple definition of

independence between any two distinct nodes based on the concept of a "Markov blanket" (MB) (**Figure M.3b**). The MB of a node *A*, noted *Mb(A),* is the set of nodes that shield the node *A* from the rest of the nodes of the BN. This implies that every set of nodes in the BN is conditionally independent of node *A* when conditioned on the set of nodes from the *Mb(A)*. The MB of the node *A* has the Markov property; formally, for distinct nodes *A* and *B.*

$$P(A \mid Mb(A), B) = P(A \mid Mb(A))$$

Indicating that only the nodes from the *Mb(A)* are the only ones' conditional dependents from the node *A.* The MB of a node *A* is conformed by its parents, children, and children's other parents (**Figure M.3b**).

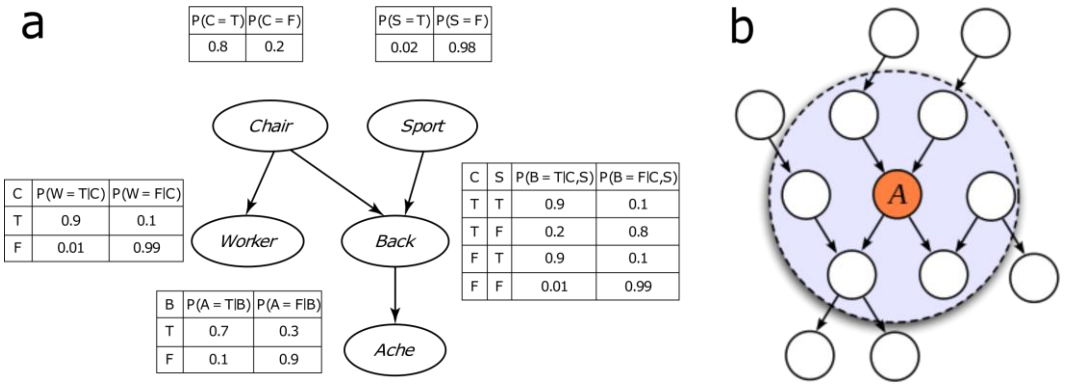


Figure M.3 Bayesian network. Panel **a**: Example of the backache BN. It considers a person who might suffer from a back injury, an event represented by the variable Back (*B*). Such an injury can cause a backache, an event represented by the variable Ache (*A*). The back injury might result from a wrong sport activity, represented by the variable Sport (*S*) or from new uncomfortable chairs installed at the person's office, represented by the variable Chair (*C*). In the latter case, it is reasonable to assume that a coworker will suffer and report a similar backache syndrome, an event represented by the variable Worker (*W*). All variables are binary; thus, they are either true ("T") or false ("F"). The conditional probability tables of each node are listed beside the node. In this example, the parents of the variable Back are the nodes Chair and Sport. The child of Back is Ache, and the parent of Worker is Chair. Following the BN independence assumption, several independence statements can be observed in this case. For example, the variables Chair and Sport are marginally independent, but when Back is given they are conditionally dependent. This relation is often called explaining away. When Chair is given, Worker and Back are conditionally independent. When Back is given, Ache is conditionally independent of its ancestor's Chair and Sport. The conditional independence statement of the BN provides a compact factorization of the JPD. Instead of factorizing the joint distribution of all the variables by the chain rule, i.e., P(C,S,W,B,A) = P(C)P(S|C)P(W|S,C)P(B|W,S,C)P(A|B,W,S,C),the BN defines a unique JPD in a factored form, i.e. P(C,S,W,B,A) = P(C)P(S)P(W|C)P(B|S,C)P(A|B). Panel **b**: The MB of a node into a BN is the set of nodes composed of its parents, its children's and the children's other parents. The MB of a node contains all the variables that shield the node from the rest of the network. This means that the MB of a node is the only knowledge needed to predict the behavior of that node [Judea Pearl 1998]. In the current example the blue circle encompasses all the nodes that belong to the MB of the orange node *A.* Text and image reproduced from Faltin and Kenett (2007) [342] with permission of John Wiley & Sons, Inc., and from https://en.wikipedia.org/wiki/Markov_blanket with permission of Wikipedia® Wikimedia Foundation, Inc.

BN has been used in metagenomic studies [315,344–347] given they can deal with noisy data [344] in multivariate complex modeling. This approach offers two attractive properties. First, it allows discovering hidden relationships among multiple variables, and second, the network is structured in

modules represented by statistically related subgraphs, namely, the Markov Blankets that predict the behavior of a node given its neighborhood.

In the current manuscript, we used three BN in order to find a putative association between the immune system and different conditions and elements of the microbiome.

3.8. General statistical analysis

All the correlation analyses were performed using the Spearman's rank correlation coefficient (library ‘‘stats’’, function ‘‘cor.test’’). To evaluate differences between groups in continuous variables, the Kruskal–Wallis test was used. Between-group comparisons of continuous variables were analyzed using Wilcoxon rank-sum test. All p-values were adjusted using the Benjamini-Hochberg's correction (library “stats”, function “p.adjust” from R package).

3.8.1. Statistical robustness

Given the high-dimensional data and to control the false discovery rate, we validated the statistical tests using the following strategies. First, we estimated the power (R library “pwr” functions: “power.anova.test” and “power.t.test”) of each of the statistical tests to verify the robustness of the results (**Table S12.2.8,** Appendix section). Second, a permutation test based on 9999 Monte-Carlo steps over each of the statistical test used in the current work was performed (**Table S12.2.8,** Appendix section**).** These kinds of tests are used when the sample size is low or when the distribution of data is unknown. Permutation tests are able to construct sampling distributions by resampling the observed data. These analyses were performed using the R package, coin library, (library “coin”, function “wilcoxsign_test”) and (library “coin”, function “oneway_test”).

4. CHAPTER 1

ALTERED METABOLISM OF GUT MICROBIOTA CONTRIBUTES TO CHRONIC IMMUNE ACTIVATION IN HIV INFECTED INDIVIDUALS

The results of the present chapter have been published in:

Vazquez-Castellanos JF*, Serrano-Villar S*, Latorre A, Artacho A, Madrid N, Vera M, *et al*. **Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals**. Mucosal Immunol. 2014;8:760–72. [348].

4.1. Introduction

Although HIV-infected adults (HIV+) with access to modern ART regimens will presumably be able to suppress HIV replication indefinitely, the profound CD4+ T-cell depletion in gut-associated lymphoid tissue is incompletely reversed by long-term ART, and microbial translocation continues long after peripheral CD4+ T-cell restoration [267,349,350]. This residual activation of both the innate and the adaptive immune systems during treated HIV infection is also associated with markers of inflammation and coagulation and decreased thymic output. Furthermore, it is an independent predictor of morbidity and mortality [265,266,268]. In fact, in most of HIV individuals on long-term ART, the risk of non-AIDS disorders [266] such as cardiovascular diseases and atherosclerosis is higher than expected. However, the way in which the immune system shapes the microbiome and contributes to disease is poorly understood. As reviewed in the general introduction, mounting evidence suggested that disruption of gut immunity in HIV infection favored the dysbiosis of the gut microbial community, which negatively affected critical pathways for healthy immune homeostasis [16,105,351–353].

In other chronic inflammatory diseases, such as IBD, metagenomic [105] and metaproteomic surveys [352] have shown that the extent of dysbiosis is not limited to a shift in commensal organisms, but that it is also associated with up- or downregulation of pathways related to oxidative stress, virulence, and secretion [352]. In this context, the functional metagenomic analysis is useful to understand the metabolic influence of the dysbiotic bacterial community in the maintenance of persistent immune dysfunction during ART. For this reason, the present chapter determines from fecal samples the functional capacity profile of the intestinal microbiota of HIV+ patients without comorbidities during effective ART using shot-gun metagenomic sequences and its relationship with the immune dysfunction and the disease progression. Additionally, the 16S rRNA gene analysis was used to obtain a deep taxonomic characterization of the ART-gut associated microbiome and to examine its associations with bacterial translocation and immune activation. Finally, in the current chapter, a Bayesian network [341] was estimated to model the interactions of the main factors in HIV infection allowing the identification of potential targets for intervention.

4.2. Materials and Methods

4.2.1. Study design, participants, setting, and eligibility

In the present chapter, a subset of 30 participants was selected to perform a case-control study (15 cases and 15 controls) on the HIV gut-associated dysbiosis. Cases were HIV-infected subjects with an effective response to the ART attending the HIV clinics of two University hospitals in Madrid, Spain (University Hospital Clínico San Carlos and University Hospital Ramón y Cajal). The inclusion criteria were serologically documented HIV infection, age 18 years or older, at least 2 years under ART-mediated HIV RNA suppression with a regimen containing at least three antiretroviral drugs, and a CD4+ T-cell count ≥ 350 cells per ml. The controls were healthy non-HIV-infected volunteers, who were recruited to form an age-matched control group. The exclusion criteria were the use of concomitant medications, use of systemic antibiotics during the previous 3 months, and any acute or chronic condition other than chronic HIV infection, including gastrointestinal symptoms (constipation, bloating, or diarrhea) or co-infections by hepatitis B or C viruses. The fecal samplings were only collected for 12 controls and 9 HIV+ subjects, this given to the patients' indisposition. The nucleoid acid purification from the fecal samples was performed as described in Material and Methods section 3.3.1 and sequencing according to Material and Methods section 3.3.2.1. All sequences were deposited in the public European Nucleotide Archive server under accession number PRJEB5185.

4.2.2. Metagenome analysis

Sequence trimming, dereplication, and removal of host sequences were performed using the MG-RAST pipeline (Release version 3.2) (default parameters) (http://metagenomics.anl.gov/) [354]. Functional assignments were obtained from the MG-RAST pipeline using BLAT software (https://genome.ucsc.edu/FAQ/FAQblat.html) (e-value e^{-5}, minimum identity 60%, and minimum alignment length 15 amino acids) against the KO database for each hierarchical level (Level 1, Level 2, Pathway level, and KO group).

4.2.3. Clustering and ordination analysis

The ordination analysis for the 16S rDNA amplicons was performed as described in the general method section “3.4.2. Beta diversity and clustering”.

A hierarchical clustering analysis (Library “stats” function “hclust” method = ward) was performed for those LEfSe biomarkers from the genus and pathway annotations and was represented using in a heatmap using the functions heatmap (library "stats" function heatmap) and heatmap.2 (library "gplots" function “heatmap.2”) respectively.

4.2.4. Correlation analyses

The correlations between markers of innate and adaptive immune activation, markers of innate immunity, and the first component of the principal coordinates analysis (PCoA) of microbiota composition, were assessed as described elsewhere [285]. Linear regression coefficients (library ‘‘stats’’, function ‘‘lm’’) were calculated for all correlations with a significant p-value ($\alpha < 0.05$) in the Spearman correlation index (library ‘‘stats’’, function ‘‘cor.test’’). Correlation analyses were also performed between clinical variables and metabolic pathways. All p-values were adjusted using the Benjamini–Hochberg correction (library ‘‘stats’’, function ‘‘p.adjust’’). Functional pathway abundances were correlated using the Spearman correlation index (library ‘‘stats’’, function ‘‘cor.test’’) with a p-value cutoff of $\alpha < 0.1$, with the markers of innate and adaptive immune activation and innate immunity.

4.2.5. Bayesian network

The statistical R package “Bayesian network learning and inference” (bnlearn) [355] was used to estimate a probabilistic graph model among bacterial genera, functional composition, and clinical parameters. The network topology was created using a hill-climbing (HC) score-based learning algorithm. The algorithm states the optimal network and in consequence the “father” to “child” node relationships as the one that maximizes the Bayesian Information Criterion (BIC).

For the analyses in HIV-infected individuals, the input variables were as clinical variables: markers of adaptive immunity, thymic function and bacterial translocation; as bacterial genera: biomarker genera and taxa with relative abundance above 0.5%; as metabolic functions: pathways biomarkers, glutathione metabolism pathway (ko00480) and the D-glutamine and D-glutamate metabolism (ko00471). We did not include the HIV+ sample H02 in network estimation given that it was the most dissimilar sample in terms of genus composition.

The underlying graphical structure of the network and the conditional probability, given the model parameters, were estimated using the HC algorithm, utilizing the BIC as the criteria for model selection (function “hc”, package “bnlearn”). The option blacklist (R Package “bnlearn” function

"hc") was used to define the set of arcs that was not included in the model, excluding those variables with a correlation Benjamini-Hochberg (BH) adjusted p-value above 0.1 (R Package "stats", function "cor", method "Spearman"). Similarly, those arcs with significant correlations that were not included in the final graph were incorporated by the "withelist" option (function "hc", package "bnlearn"). Function mb (R Package "bnlearn" function "mb") was employed to estimate the Markov Blanket from the lipopolysaccharide biosynthesis, zeatin biosynthesis, *Coprococcus*, and the markers of immune activation/senescence %CD4+CD38+ T-cells and %CD8+CD57+ T-cells.

4.3. Results

4.3.1. Differences in the clinical variables between ART-treated HIV+ individuals and healthy subjects

From the cohort of 15 chronically HIV-infected individuals on suppressive ART and 15 healthy controls, clinical measure comparisons were performed. Patients had a median CD4+ T-cell nadir of 203 cells/ml, median cumulative ART exposure of 6 years, good CD4+ T-cell reconstitution (584 cells/ml (466–794)), and had almost reached normal CD4/ CD8 ratios (1,2 (0.9–1.3)). No meaningful differences were detected in age, body mass index, or glycemic and lipid plasma profiles, and there were non-statistically significant higher proportion of women in the control group. As for plasma biomarkers, only sCD14 levels were higher in patients than in controls (p-value=0.05). As shown in **Table S12.1.1 (**Appendix section**)**, ART-treated HIV-infected individuals displayed lower CD4+ T-cell counts and lower CD4/CD8 ratios than controls and significantly increased frequencies of CD4+ and CD8+ T-cells expressing markers of T-cell activation/senescence (HLA-DR/+CD38+, CD38+, CD25+) and senescence (CD57+).

4.3.2. Differences in gut microbiota composition between HIV-infected individuals under ART and healthy subjects

An average of 5,392 16S rRNA gene sequences per sample was obtained from the multiplex pyrosequencing from DNA extracted from fecal samples of a subset of our cohort (12 controls and 9 ART-treated HIV+ patients). The remaining nine participants initially consented to participate in the study but then refused to provide stool samples. Taxonomical assignation was performed at operational taxonomic units (OTUs) (97%) and at the genus level, as they allowed a higher discriminatory power between samples. We used the weighted UniFrac distance and Bray–Curtis dissimilarity index for the cluster analysis, because these metrics gave the optimal cluster configuration at OTU (97%) and genus level, respectively (**Figure S12.1.1,** Appendix section). The clustering showed that the samples formed two clear groups for both taxonomic levels (**Figure 4.1**). These clusters and the group category (HIV- vs. HIV+) were validated by an ADONIS test at OTU (97%) and genus level p-value=0.001 and p-value=0.04 for the cluster analysis and p-value=0.001

and p-value=0.01 for the group category, respectively). The PCoA retrieved from the OTUs (**Figure 4.1a**) showed three patients (H14, H18, and H24) located in the control cluster. The heat map showed in **Figure 4.2** also revealed two main clusters. One cluster (in red) was composed mainly of HIV+ individuals, with a high abundance of *Prevotella* (44.1%) and *Succinivibrio* (14.6%). Subject C49, who harbored a high proportion of *Prevotella* genus, was also included in this cluster. The other cluster (in blue) was composed mainly of HIV- individuals in whom *Bacteroides* (27.5%) and *Faecalibacterium* (16.7%) were the most abundant genera (**Figure 4.2**). Analysis of the clustering at the genus level showed that the subject H02 clustered with the control group (**Figure 4.1b**). This patient presented an unusual microbiota composition with a high abundance of *Bacteroides* and low level of *Faecalibacterium*. Intriguingly, the bacterial community of HIV+ individuals had a much higher proportion of Gram-negative bacteria than HIV- individuals (ratio %Gram negative/%Gram positive: 71/18 vs. 45/44, respectively).

Comparison of richness estimators, ACE and Chao1, and the Shannon index revealed statistically significant differences only at OTU (97%) level (**Table 4.1**), being the bacterial community of HIV- individuals more diverse. However, the overall bacterial load (expressed as a number of 16S rRNA gene copies) was significantly higher in HIV+ subjects than in HIV- subjects (**Table 4.1**).

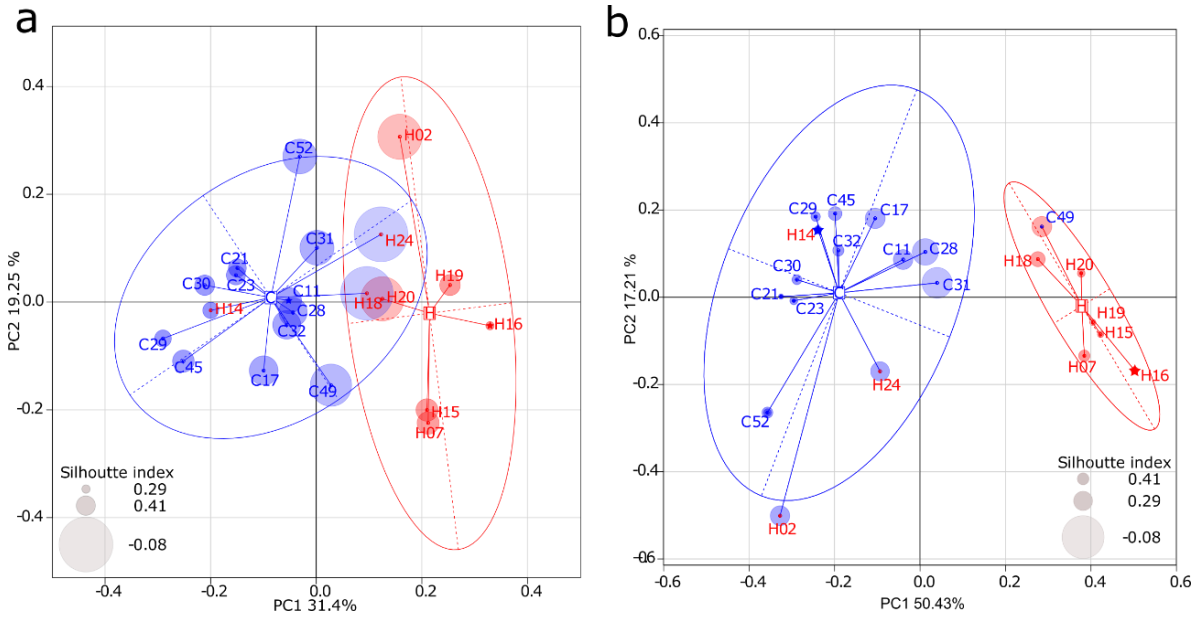


Figure 4.1 Comparison of microbiota between HIV+ART and uninfected subjects. Principal coordinates analysis of the bacterial composition in controls (blue dots) and cases (red dots) at (**a**) operational taxonomic unit (OTU) (97%) level and (**b**) the genus level. The stars in blue and red correspond to the medoid retrieved from the PAM algorithm. The centroid is represented by a capital letter denoting the condition of the cluster (C for controls and H for HIV+ subjects), while the blue and red ellipses represent the 95% of the samples belonging to each condition. Each point contains a halo proportional to its silhouette index value: as larger is the halo, more dissimilar is the element to its corresponding object. Modified from Vazquez-Castellanos et al. (2014) [348] with permission from © 2017 Society for Mucosal Immunology.

Table 4.1 Diversity parameters of microbiota.

	Patients on HAART [a]	Controls [a]	p-value [b]	q-value [c]
OTU level				
Shannon index	5.96 ± 1.03	7.00 ± 0.51	0.01	0.04
Chao1 estimator	567.69 ± 175.21	776.27 ± 166.63	0.02	0.05
Ace estimator	563.69 ± 176.44	794.90 ± 172.46	0.01	0.04
Genus level				
Shannon index	1.82 ± 0.32	2.03 ± 0.23	0.43	0.60
Chao1 estimator	29.28 ± 7.22	27.39 ± 6.03	0.62	0.72
ACE estimator	30.57 ± 7.56	29 ± 5.68	0.86	0.86
Bacterial density				
Number of 16S RNA gene copies/ngDNA	1451434.41 ± 899075.31	762212.50 ± 317670.42	0.03	0.05

[a]Values are expressed as the mean ± standard deviation (SD).

[b]Analysis was performed using a Wilcoxon rank-sum test. P is probability at α=0.05.

[c]p-value adjusted according to the Benjamini-Hochberg method.

The LEfSe biomarker discovery tool was used to elucidate which genera were driving divergences between the groups. It was found 11 biomarkers for the HIV+ cluster: 7 belonged to the Firmicutes phylum: *Acidaminococcus* (p-value=0.01), *Butyrivibrio* (p-value=0.00), *Eubacterium* (p-value=0.02), *Mitsuokella* (p-value=0.00), *Bulleidia* (p-value=0.00), *Megasphaera* (p-value=0.01), and *Catenibacterium* (p-value=0.00); 3 belonged to the Proteobacteria phylum: *Succinivibrio* (p-value=0.00), *Trabulsiella* (p-value=0.03), and *Desulfovibrio* (p-value=0.02); and finally, a single genus from the Bacteroidetes phylum: *Prevotella* (p-value=0.00). In the control group, it was observed nine biomarkers: six belonged to the Firmicutes phylum: *Faecalibacterium* (p-value=0.00), *Roseburia* (p-value=0.00), *Ruminococcus* (p-value=0.04), *Blautia* (p-value=0.00), *Coprococcus* (p-value=0.03), and Anaerostipes (p-value=0.00); two belonged to Bacteroidetes phylum: *Bacteroides* (p-value=0.00) and *Parabacteroides* (p-value=0.00); and a single genus representing Proteobacteria phylum: *Escherichia* (p-value=0.00) (**Figure 4.3a**). These biomarkers presented high LDA scores (LDA = 4.35) and generated sample clustering similar to those obtained using all taxa (**Figure 4.2 and Figure 4.3b**). The *Prevotella* biomarker had the highest LDA score

(5.2) and was 7.8 times more abundant in HIV+ subjects than in HIV- subjects (1542 ± 641.9 vs. 196.9 ± 288.2), whereas *Succinivibrio* (LDA score = 4.8) was not present in HIV- subjects. In control group, *Bacteroides* and *Faecalibacterium* presented the highest discriminative power (LDA score = 5.1 and 4.8, respectively).

To elucidate whether specific clinical parameters might correlate with the extent of dysbiosis, the correlations (Spearman rank correlation coefficient (ρ)) between the first component of the principal coordinates analysis (PC1) and the CD4+ T-cell counts, CD8+ T-cell counts, CD4/CD8 ratio, CD4 nadir, time from HIV diagnosis to ART initiation, and duration of ART were calculated. Only the CD4/CD8 ratio correlated with the PC1 (ρ = -0.4, p-value=0.045), although statistical significance was lost after adjustment for multiple comparisons (adjusted p-value=0.117).

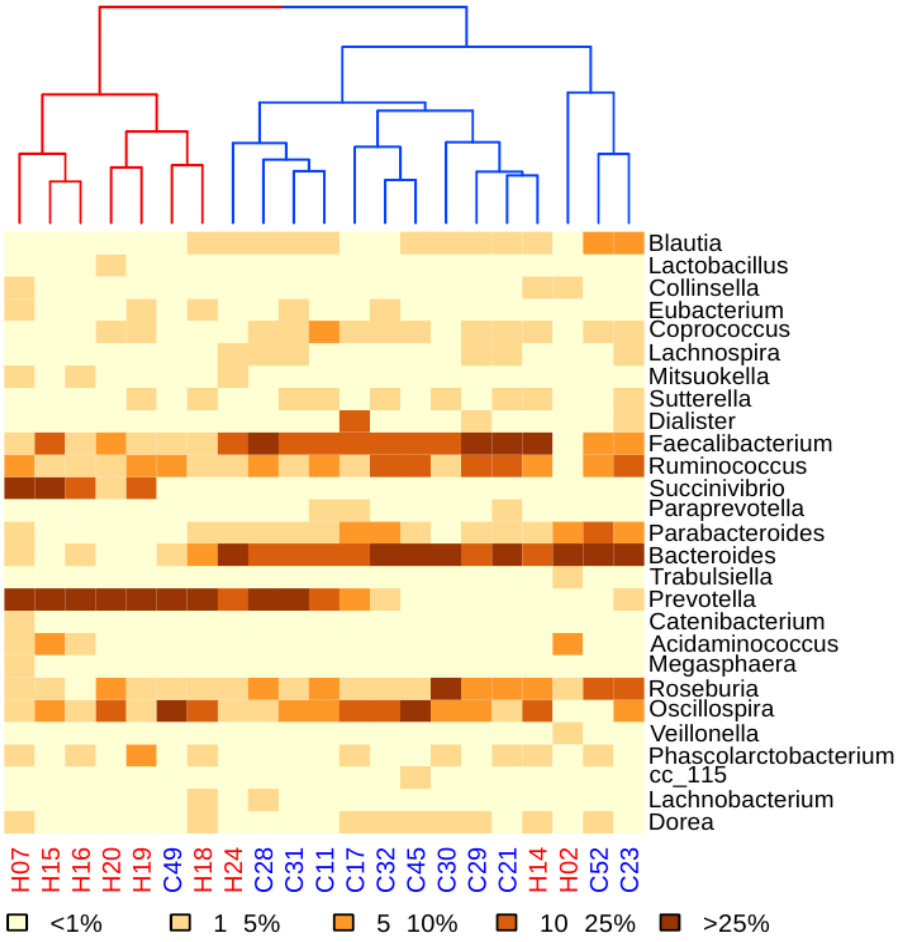


Figure 4.2 Heat map of the samples at the genus level. HIV+ subjects are marked in red and controls in blue. The top dendrogram is divided in two main sub trees highlighted in red or blue, according to the predominance of samples from HIV+ individuals or controls, respectively. In the heat map the percentage range of sequences assigned to main taxa (abundance >1% in at least one sample) is represented by a color gradient. Reproduced from Vazquez-Castellanos et al. (2014) [348] with permission from © 2017 Society for Mucosal Immunology.

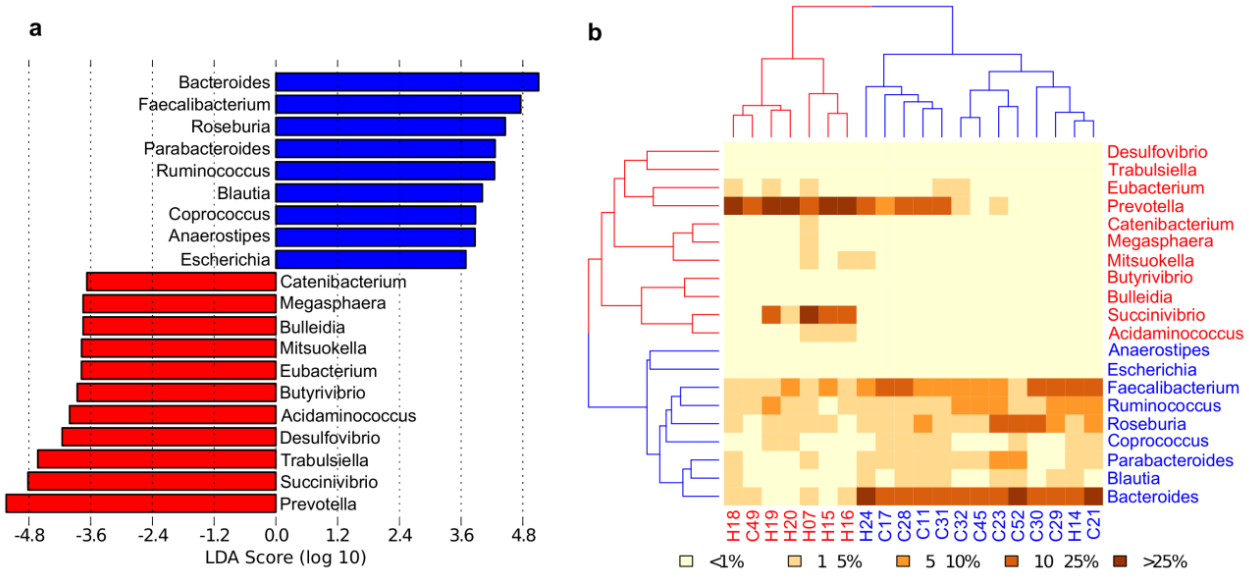


Figure 4.3 Taxonomic biomarkers. (**a**) linear discriminative analysis (LDA) effect size LEfSe) analysis between the case cluster (in red) and control cluster (in blue). LDA scores (log 10) for the most prevalent taxa in controls are represented on the positive scale, whereas LDA-negative scores indicate enriched taxa in cases. (**b**) Heat map of genus biomarkers. Biomarkers are represented in red and blue for cases and controls, respectively. In the heat map, the percentage range of sequences assigned to main taxa (abundance 41% in at least one sample) is represented by a color gradient. Reproduced from Vazquez-Castellanos et al. (2014) [348] with permission from © 2017 Society for Mucosal Immunology.

4.3.3. The impact of total microbiota on immunological predictors of disease progression

Correlation analysis was performed between the first component of the PC1 at OTU level (97%) and markers of bacterial translocation (BPI, sCD14), monocyte activation (T-cell activation %HLA DR+/CD38+/CD25+) and senescence (%CD57+), thymic function (sj/b-TREC ratio), inflammation (hs-CRP and IL6), and thrombosis (D dimers). PC1 correlated positively with the inflammation marker hs-CRP and with markers of T-cell activation, including %CD4+HLA-DR+CD38+ T-cells, %CD4+CD25+ T-cells, %CD8+HLA-DR+CD38+ T-cells, and %CD8+CD38+ T-cells (**Figure 4.4a–e**). No significant correlation was observed between PC1 and the markers of bacterial translocation (sCD14, $\rho = 0.2$, adjusted p-value=0.56 and BPI, $\rho = 0.3$, adjusted p-value=0.45), thrombosis (D-dimers, $\rho = 0.3$, p-value=0.41), inflammation (IL6, $\rho = -0.2$, adjusted p-value=0.45), and thymic function (sj/b-TREC ratio, adjusted $\rho = -0.25$, adjusted p-value=0.43).

As lipopolysaccharide is one of the principal antigens translocated from the gut to the bloodstream driving chronic immune activation, the correlation between PC1 and the number of sequences

assigned to LPS biosynthesis pathway (see metagenome analyses below) were assessed, showing a significant positive correlation (**Figure 4.3f**).

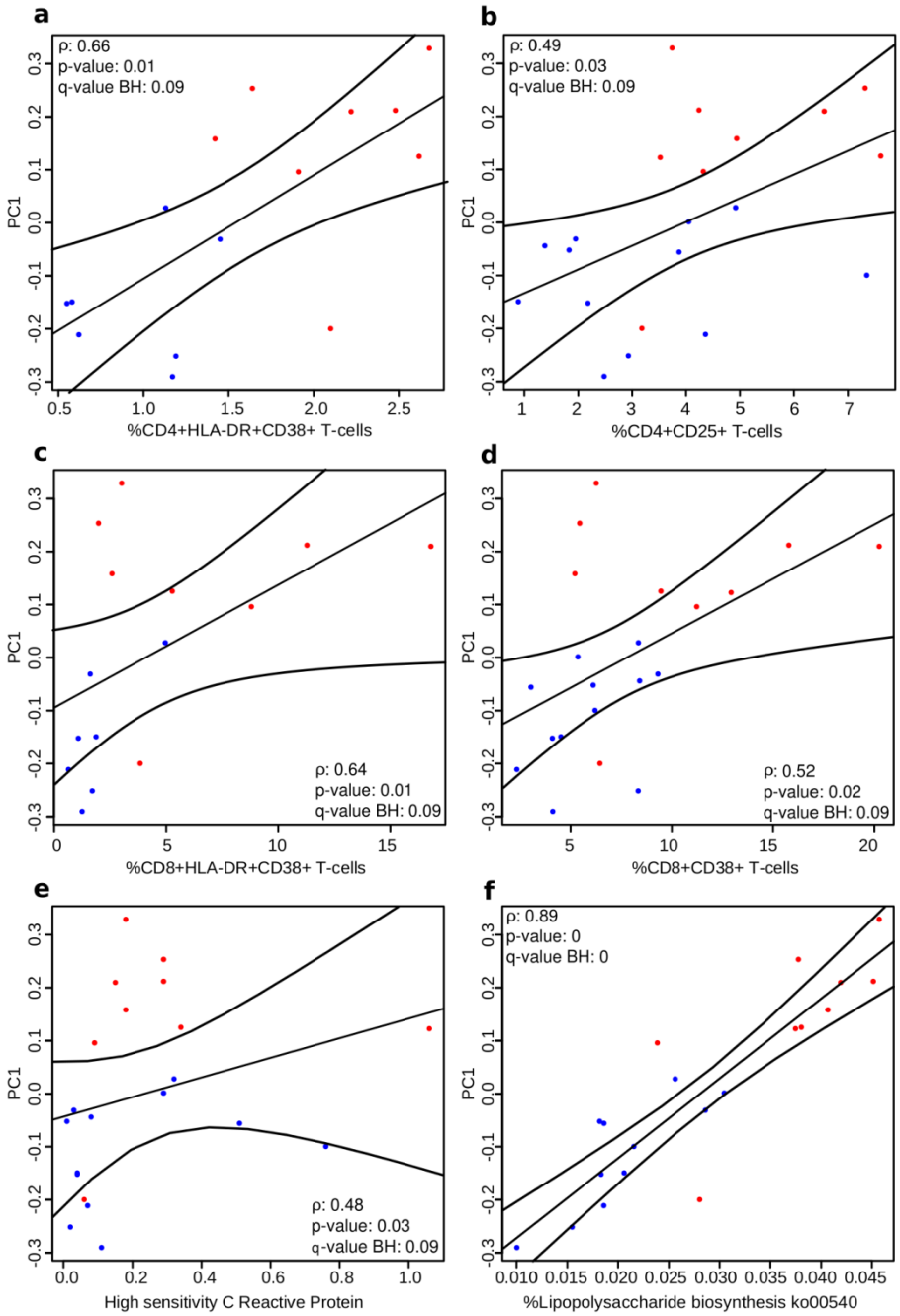


Figure 4.4 Associations between PC1 and markers of immune activation, inflammation, and bacterial translocation. (**a–d**) Correlation between PC1 and %CD4+HLA-DR+CD38+, %CD4+CD25+, %CD8+HLA-DR+CD38+, and %CD8+CD38+ T-cells. (e) Correlation between PC1 and hs-CRP. (**f**) Correlation between PC1 and the lipopolysaccharide (LPS) biosynthesis pathway, as a marker of bacterial translocation. The best-fitted linear model that predicted T-cell markers and the LPS pathway is represented as a solid line. Dashed lines represent the 95% confidence interval for the linear regression coefficients. In blue are represent the controls while in red the HIV+ individuals. ρ represents the Spearman correlation coefficient and its corresponding *p-value*. *q-value* BH is the P-value adjusted using the Benjamini–Hochberg correction. Reproduced from Vazquez-Castellanos et al. (2014) [348] with permission from © 2017 Society for Mucosal Immunology.

4.3.4. Differences in microbiota metabolic functions between HIV- and ART-treated HIV+ subjects

To explore functional hallmarks of the HIV-associated dysbiotic bacterial community, the shot-gun metagenome data was analyzed in both groups. Pyrosequencing of the samples yielded a total of 659Mb with an average length of 655±32 bp. All the high-quality reads (average 18.5Mb per sample) were compared with the KO database at different hierarchical levels (level 1, top level; level 2, subcategories of the top level; pathway level, and KO, gene level), 16 giving a high functional assignment to 34% of the reads (49,517 open reading frames per sample).

At level 2, the functional profiles were fairly homogeneous for all individuals. The most abundant categories were ‘‘carbohydrate metabolism’’ (19.2%), ‘‘amino acid metabolism’’ (13.8%), ‘‘energy metabolism’’ (9.2%), and ‘‘nucleotide metabolism’’ (8.8%), thus highlighting the importance of the gut microbiota in these metabolic pathways. Similarly, major pathways in both groups were related to purine and pyrimidine metabolism, amino and nucleotide sugar metabolism, alanine, aspartate and glutamate metabolism, and transport systems.

The LEfSe analysis identified significant variations in the functional profile of both groups at different hierarchical levels. At level 2, HIV+ individuals showed a unique differential category, i.e., ‘‘infectious diseases’’ (Wilcoxon test adjusted p-value=0.001), while ‘‘carbohydrate metabolism’’ (Wilcoxon test adjusted p-value=0.02) and ‘‘endocrine system’’ (Wilcoxon test adjusted p-value=0.04) were significantly more abundant in the HIV- group. However, 173 KOs were significantly different between both groups. In order to gain insights in the metabolic routes that had high discriminative power, the KEGG pathways were also subject to the LEfSe analysis. **Figure 4.5** shows the biomarkers found at pathway level: 12 in the HIV+ group and 23 in the HIV- group. All biomarkers presented similar percentages of sequencing coverage in both groups (**Table S12.1.2,** Appendix section). For HIV+ subjects, the pathways with the highest discriminative power were the ‘‘ribosome’’ and ‘‘LPS biosynthesis’’ pathways (LDA score = 3.2 and p-value=0.01 and LDA score = 3.2 and p-value=0.00, respectively), followed by the ‘‘phenylalanine tyrosine and tryptophan biosynthesis’’ pathway (ko00400) (LDA score = 2.8, p-value=0.00). In this group, functional biomarkers were mainly involved in biosynthetic pathways as ‘‘terpenoid backbone biosynthesis’’, ‘‘fatty acid biosynthesis’’, ‘‘ubiquinone and other terpenoid-quinone biosynthesis,’’ and ‘‘zeatin biosynthesis’’. The KO-annotated genes belonging to pathogenesis processes as ‘‘Legionellosis’’ and ‘‘Vibrio cholerae pathogenic cycle’’ pathways were overrepresented in the

HIV+ group. In the HIV- subjects, nine biomarkers were related to metabolite degradation and eight to metabolism, with the ‘‘starch and sucrose metabolism’’ pathway presenting the highest LDA score (LDA score = 3.1, p-value=0.03).

One of the biomarkers in this group, the peroxisome proliferator-activated receptor signaling pathway (LDA score = 2.6, p-value=0.01), contained phosphotransferases that have been related to anti-inflammatory responses. The heat map obtained using the pathway biomarkers found for HIV+ patients and controls is shown in **Figure 4.6**. In contrast to that observed in the compositional clustering analysis, subject C49 (HIV-) showed a healthy functional profile, whereas subject H24 (HIV+) clustered with the HIV+ group and had a dysfunctional microbiota. Only subject H14 remained clustered with the HIV- group (sample H02 was not included in this analysis). In the HIV+ group, the microbiota was depleted in genes belonging to main energetic processes as pyruvate metabolism, glycolysis, and gluconeogenesis. In addition, a decrease in amino acid metabolism (glycine, serine, threonine, tryptophan, and histidine) was identified in this group (ko00260, ko00260, ko00260, ko00380, and ko00346, respectively). However, it was detected the enrichment for pathways involved in the metabolism of cofactors and vitamins (ubiquinone and other terpenoid–quinone biosynthesis, thiamine metabolism and nicotinate and nicotinamide metabolism) in the dysbiotic bacterial community.

Finally, as inflammation has been associated with oxidative stress, we investigated pathways that could be specifically related to this process. Even though the differences did not reach statistical significance, the relative abundance of genes belonging to glutathione metabolism and D-glutamine and D-glutamate metabolism was higher in the HIV+ group compared with the HIV- group (0.66 vs. 0.28%; 0.26 vs. 0.21%, respectively).

4.3.5. Correlations between markers of innate and adaptive immunity, and gut microbiota metabolic pathways in HIV+ subjects on effective ART

We explored the correlations between metabolic functions of the intestinal microbiota and markers of T-cell activation/senescence (% of CD4+ and CD8+ T-cells expressing HLA-DR/+CD38+, CD38+, CD25+, or CD57+), bacterial translocation (BPI and sCD14), inflammation (hs-CRP and IL6), thrombosis (D-dimers), and thymic function. Different associations were found (**Table 4.2**), although no relation remained statistically significant after adjusting for multiple comparisons. Positive correlations were detected between the percentage of CD4+HLA-DR+CD38+, and the ‘‘LPS biosynthesis’’ and ‘‘glutathione metabolism’’ pathways (**Table 4.2**). Interestingly, the

‘‘zeatin biosynthesis’’ pathway correlated negatively with bacterial translocation markers, sCD14 and BPI, as well as with the sj/b-TREC ratio. hs-CRP and the thrombosis D dimers correlated negatively with the ‘‘thiamine metabolism pathway’’ and only the ‘‘glutathione metabolism’’ pathway correlated positively with hs-CRP (**Table 4.2**).

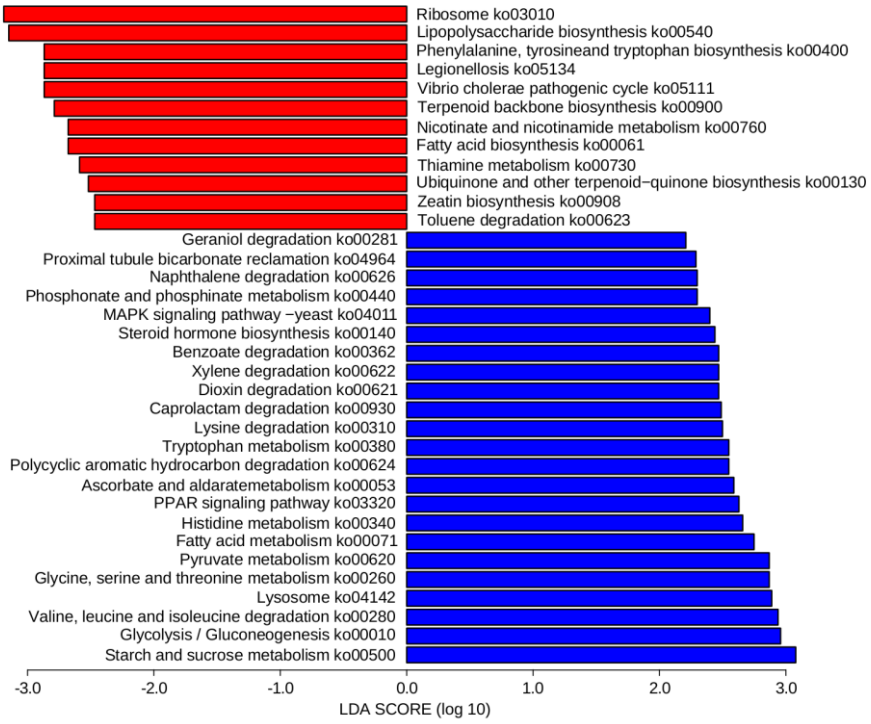


Figure 4.5 Linear discriminative analysis (LDA) effect size (LEfSe) analyses of statistically significant KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. Negative LDA scores (red) are enriched in patients while positive LDA scores (blue) are enriched in controls. Reproduced from Vazquez-Castellanos et al. (2014) [348] with permission from © 2017 Society for Mucosal Immunology.

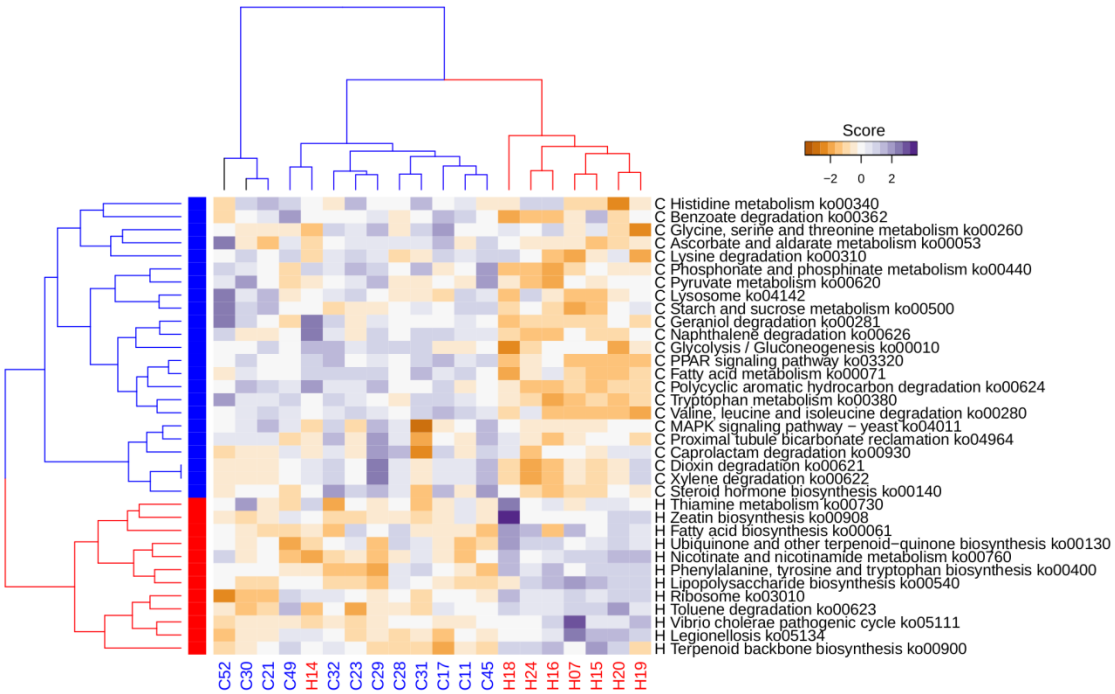


Figure 4.6 Heat map of the functional biomarkers for patients (in red) and controls (in blue). Over/underrepresentation is depicted by a color gradient. On the right, ''H'' identifies those pathways predicted as biomarkers for cases and ''C'' those pathways predicted as biomarkers for controls. The dendrogram of the pathway abundances is divided into two main groups, a blue cluster (controls) and a red cluster (patients). Reproduced from Vazquez-Castellanos et al. (2014) [348] with permission from © 2017 Society for Mucosal Immunology.

Table 4.2 Significant correlation between pathways and clinical variables.

	Spearman correlation index	q-value[a]
T-cell markers		
Ribosome ko03010|CD4 T-cells	-0.71	0.06
Terpenoid backbone biosynthesis ko00900 | CD4 T-cells	0.67	0.08
Glutathione metabolism ko00480 | %CD4+HLA-DR+CD38+ T-cells	0.71	0.05
Lipopolysaccharide biosynthesis ko00540 | %CD4+HLA-DR+CD38+ T-cells	0.71	0.05
D-Glutamine and D-glutamate metabolism ko00471 | %CD4+CD38+ T-cells	-0.71	0.06
Nicotinate and nicotinamide metabolism ko00760 | %CD4+CD25+ T-cells	0.95	0.00
Ribosome ko03010 | %CD4+CD25+ T-cells	0.83	0.02
Glutathione metabolism ko00480 | %CD8+CD25+ T-cells	0.79	0.03
Toluene degradation ko00623 | %CD4+CD25+ T-cells	0.67	0.08
D-Glutamine and D-glutamate metabolism ko00471|%CD8+CD57+ T-cells	0.64	0.10
Bacterial translocation		
Legionellosis ko05134|BPI[b]	0.86	0.01
Glutathione metabolism ko00480|BPI[b]	0.79	0.03
Lipopolysaccharide biosynthesis ko00540|BPI[b]	0.71	0.06
Zeatin biosynthesis ko00908|BPI[b]	-0.71	0.06
Zeatin biosynthesis ko00908|sCD14[c]	-0.86	0.01
Ubiquinone and other terpenoid quinone biosynthesis ko00130|sCD14[c]	-0.64	0.10
Thymic function		
Zeatin biosynthesis ko00908 |sj/β-TREC ratio	-0.75	0.03
Thiamine metabolism ko00730|sj/β-TREC ratio	-0.66	0.08
Inflammation		
Thiamine metabolism ko00730 |hs-CRP[d]	-0.65	0.08
Glutathione metabolism ko00480|hs-CRP[d]	0.62	0.10
Thrombosis		
Thiamine metabolism ko00730 | D-dimers	-0.67	0.08

[a] p-value adjusted according to the Benjamini-Hochberg method. α=0.1.

[b] Bactericidal-permeability increasing protein.

[c] Soluble CD14.

[d] High sensitivity C Reactive Protein.

4.3.6. Bayesian networks and Markov blankets estimation

Bayesian networks are probabilistic models in which the nodes correspond to random variables and the arcs represent causal relationships [341]. With cross-sectional data, the connecting arrows

represent mutual associations rather than causality [356]. To unravel the complex interactions between microbiota and metabolic pathways contributing to T-cell activation, thymic function, and bacterial translocation, we fitted a Bayesian model. This analysis demonstrated a complex network in which most pathways and genera are interconnected. Most of these variables were associated with at least one genus or pathway, with the exception of BPI and the percentage of CD8+CD57+ T-cells that exhibited a greater number of interactions with other network components (**Figure S12.1.2,** Appendix section).

The set of nodes that predicts the behavior of another node in a Bayesian network is named the ‘‘Markov Blanket’’ [341]. We estimated the Markov Blanket for the subset of nodes that showed a significant correlation with the immunological predictors. Three Markov Blankets were selected as follows: (i) one that included lipopolysaccharide biosynthesis and zeatin biosynthesis pathways, hereafter referred to as lipopolysaccharide–zeatin Markov Blanket (**Figure 4.7a**); (ii) that included the %CD8+CD57+ and %CD4+CD38+ clinical variables, hereafter referred to as %CD8+CD57+ and %CD4+38+ Markov Blanket (**Figure 7b**); and (iii) the Markov Blanket of the *Coprococcus* genus (**Figure 4.7c**). The lipopolysaccharide–zeatin Markov Blanket contained positive correlations between genera of Gram-negative bacteria and pathways related to inflammation as Legionellosis, *V. cholera* pathogenic cycle or lipopolysaccharide biosynthesis. Together with glutathione metabolism, the lipopolysaccharide biosynthesis pathway correlated positively with BPI and the percentage of the CD4+HLADR+CD38+ T-cells. The sCD14 marker correlated negatively with the zeatin biosynthesis pathway, and unexpectedly with *Prevotella* abundance. Finally, we observed no positive correlations with the sj/b-TREC ratio (**Figure 4.7a**). Then, we explored the Markov Blanket for activated (%CD4+CD38+) and senescent (%CD8+CD57+) T-cells (**Figure 4.7b**). The depletion of *Faecalibacterium* was associated with an overgrowth of *Eubacterium* and a higher percentage of CD8+CD57+ T-cells, which in turn correlated positively with the genus *Dorea* genus, and D-Glutamine and D-glutamate metabolism pathway. The percentage of CD4+CD38+ T-cells showed a single negative correlation with the latter metabolic pathway. For *Coprococcus*, the Markov Blanket illustrated how a low abundance of this genus had an impact on different phenotypes of activated T-cells (%CD8+CD38+, %CD8+HLA-DR+ CD38+, and %CD8+CD25+). The glutathione metabolism pathway again correlated positively with markers of T-cell activation (%CD8+CD25+) (**Figure 4.7c**).

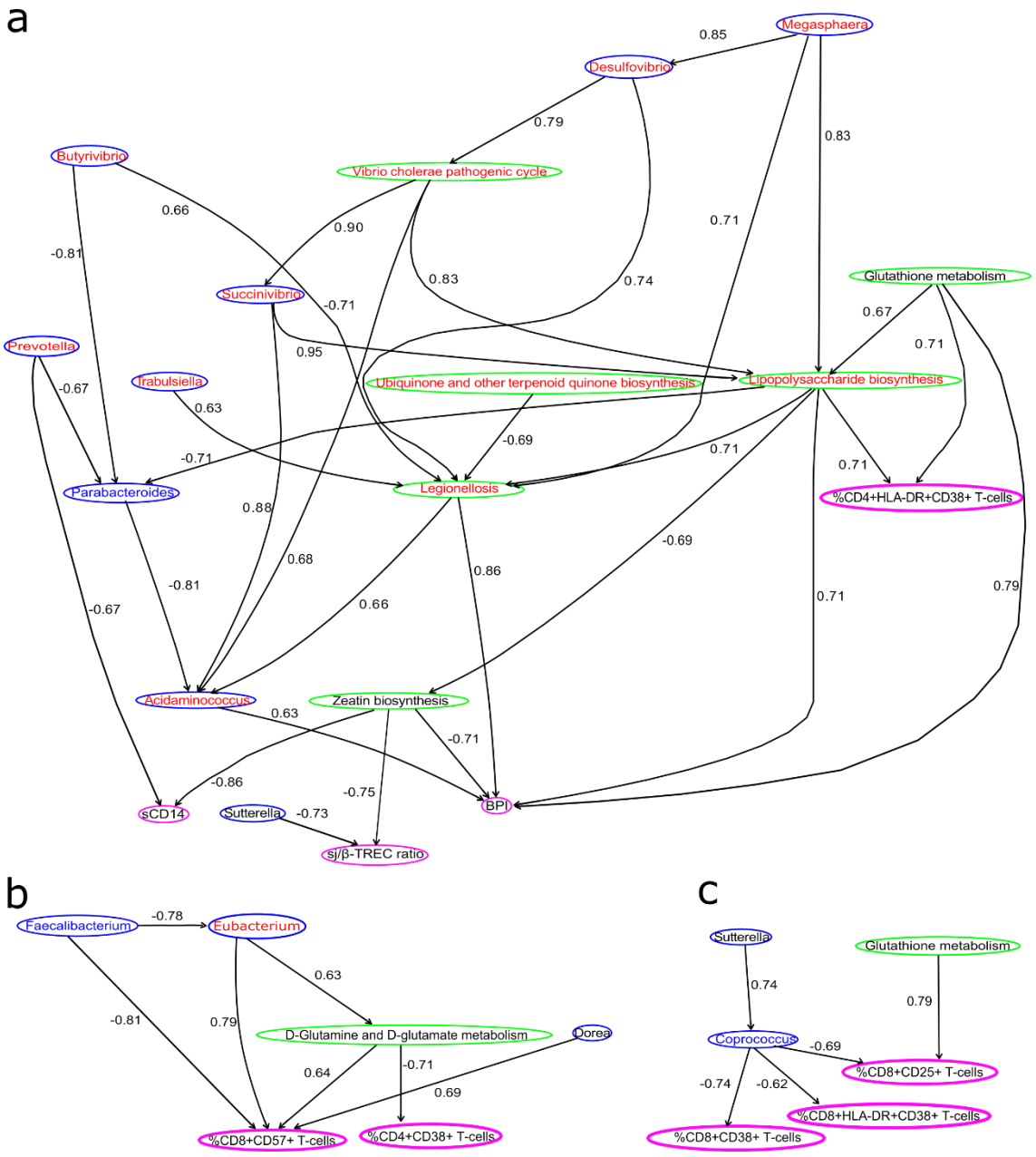


Figure 4.7 Markov blankets integrating genera, markers of adaptive immunity, and metabolic pathways. The Markov blanket subgraph of the Bayesian networks represent the relationships between genera abundance (blue ellipses), pathway abundances (green ellipses) and markers of adaptive immunity, thymic function, and bacterial translocation (pink ellipses) related to (**a**) the lipopolysaccharide and zeatin biosynthesis pathways (**b**) the %CD8+CD57+ T-cells and %CD4+CD38+ T-cells (**b**) and the *Coprococcus* genus. Taxa and pathway biomarkers of HIV+ group are represented in the red while the control biomarkers are represented in blue. Arrows indicate conditional dependencies between variables. The Spearman correlation coefficient is indicated next to the lines. Modified from Vazquez-Castellanos et al. (2014) [348] with permission from © 2017 Society for Mucosal Immunology.

4.4. Discussion

Most studies show that despite modern ART, HIV+ individuals have reduced life expectancy, mainly owing to the increase in morbidity and mortality associated with non-AIDS-related diseases, which are driven in part by persistent inflammation and immune activation [357]. Several recent studies [280,281,284,285,287,288,353,358] have explored the relationships between gut microbiota composition and HIV infection using different sampling techniques (stool, anal swab, biopsy, and sponge collection) and bacterial classification methods (16Sr RNA gene PCR, quantitative PCR, and microarray). Even though stool microbiota might not be an exact reflection of the gut resident bacterial community [359] amplification of the16S rRNA gene in stools will probably be sufficiently informative, as it recovers bacteria from mucosal desquamation and is the easiest sampling technique to be considered standard for these investigations. Indeed, the results are congruent with those of previous studies based on colon biopsies that report similar dysbiotic microbiota depleted in *Bacteroides* and enriched in *Prevotella* [284,288].

Given the broad range of clinical consequences of heightened immune activation in HIV+ individuals, the study was focused on HIV+ subjects without comorbidities receiving long-term effective ART. These patients are probably good candidates for strategies aimed at shaping the gut microbiome. In this study, HIV+ individuals presented a distinctive microbiota composition characterized by high abundance of *Prevotella* and *Succinivibrio,* and depletion of *Bacteroides*, *Faecalibacterium,* and *Roseburia*. Some of these taxa were recently reported to be members of the gut-resident microbiota in HIV+ individuals [281,284,285,353]. The Enterobacteriaceae family has been associated with bacterial translocation and immune activation, [360,361] although, in contrast with Vujkovic-Cvijn et al., [285] and consistent with Lozupone et al., [281] we did not detect a high prevalence of this bacterial taxon in our HIV-infected cohort. This discrepancy is most likely due to the use of mucosal samples rather than luminal samples for these studies. Vujkovic-Cvijn and Dunham [285] examined colonic mucosal samples, which are widely colonized by Enterobacteriaceae, whereas Lozupone et al., [281] examined stool, as did in this study. However, our HIV-infected cohort showed a gut microbiota composition dominated by Gram-negative bacteria—representing 71% of the bacterial community—increased bacterial load and decreased diversity at the OTU (97%) level. Intriguingly, the two individuals (H14 and H24) with a ‘‘normal’’

microbiota composition showed the highest CD4/CD8 ratio in the cohort, and this ratio was the only variable available in clinical practice that correlated with the extent of dysbiosis, suggesting that normalization of this biomarker of immunological dysfunction during ART may also predict normalization of gut microbiota, at least in terms of composition. In ART-treated individuals, the fact that the CD4/CD8 ratio correlates with activity of the kynurenine pathway of tryptophan catabolism, an established marker of disease progression [362], provides indirect support for the recent observation by Vujkovic-Cvijin and Dunham [285], who suggested that dysbiosis of the gut microbial community affects negatively this critical pathway in healthy immune system [285].

In other studies, reduced diversity has been observed in treated HIV+ individuals [281,288] indicating that the combination of HIV infection and its treatment leads to decreased microbial diversity. This particular microbiota composition, characterized by *Prevotella* enrichment and *Bacteroides* depletion, has been described mainly in human populations with a carbohydrate-rich diet [15,56,138,363] and recently in HIV-infected subjects [281]. In addition, Lozupone et al., [281] showed that the microbiota of HIV+ subjects is closer to those of people from agrarian cultures than to that of healthy people from the United States. All subjects in our cohort followed a similar western diet, indicating that HIV infection is the likely factor driving changes in the gut microbiota. The observed depletion of *Faecalibacterium* and *Roseburia*, the major butyrate producers, might result in the lower local production of short-chain fatty acids. In the healthy human gut, short-chain fatty acids such as butyrate, propionate, and acetate are an important energy source for the maintenance of homeostasis in the colonic mucosa and display anti-inflammatory properties [139,364,365]. Alterations in the short-chain fatty acid ratio have been related to an amplification of inflammatory responses in diseases that are typically associated with bacterial dysbiosis, such as ulcerative colitis and bacterial vaginosis [105,366]. In fact, a drastic decrease in the number of genes involved in glycolysis and pyruvate metabolism was observed in the metagenomes from HIV-infected individuals. Further short-chain fatty acid quantification should be performed to address its involvement in the inflammation process. On the other hand, depletion of the anti-inflammatory commensal genus *Faecalibacterium* has been reported in the anal microbiota of treated HIV+ patients [288] and in patients with Crohn's disease [103]. Interestingly, the Markov Blankets of the %CD8+ CD57+ T-cells (**Figure 4.7b**) parallels previous observations in centenarians, in whom the depletion of *Faecalibacterium* genus correlates with an increase in *Eubacterium* species [367] and with an increase in %CD8+CD57+ T-cell, a hallmark of immunosenescence in HIV infection [265,368]. Furthermore, given that certain *Bacteroides* species

are required for differentiation of the Th17 T-cells, the significant loss of this bacterial genus in treated-HIV individuals might either aggravate the Th17 loss secondary to HIV infection or impair their reconstitution under ART [369].

The high number of significantly abundant pathways and KOs in the bacterial community of HIV+ individuals in the study revealed a functional dysbiosis that might explain, at least in part, the situation of chronic inflammation observed during treated HIV infection. Subject H24 (HIV+) had an altered metabolic profile despite clustering with individuals with a healthy bacterial composition. The fact that this subject was overweight (body mass index=29 kg/m^{-2}) suggests that other proinflammatory factors, such as obesity-related factors, might have a role. In addition, control C49 had an altered gut microbiota composition dominated by *Prevotella* genus "*Prevotella* enterotype" [14] but showed healthy functional capacity.

The significant correlation found between PC1 and the LPS biosynthesis pathway (p-value=0, adjusted p-value=0, $\rho = 0.89$) would indicate that Gram-negative-enriched microbiota was responsible for the high abundance of KOs belonging to the LPS biosynthesis pathway in HIV+ individuals. These lipopolysaccharides are microbe-associated molecular patterns, which are potent immune activators that act via Toll like receptor (TLR4) by promoting the inflammatory response [370]. In other gut microbiota-associated diseases and *in silico* predictions of the HIV microbiota, an increase in antioxidant pathways such as riboflavin, glutathione, and glutamine metabolism has been described and interpreted as the bacterial compensatory mechanism that attenuates the oxidative stress caused by epithelial damage [105,353,365,371]. Likewise, the microbiota of treated HIV+ individuals was functionally enriched for ubiquinone and other terpenoid–quinone biosynthesis, nicotinate and nicotinamide metabolism, glutathione metabolism, and thiamine metabolic pathways. This latter pathway correlated negatively with hs-CRP and D-dimers.

These results are congruent with the anti-inflammatory effect of thiamine that has been described in mammals [372,373] and with the fact that a depletion of this vitamin exists in other gastrointestinal illnesses, such as Crohn's disease [374]. Unlike the observations in a previous study [285], we did not detect differences in genes involved in the kynurenine pathway of tryptophan catabolism. This inconsistency could be given by the different nature of the sample (colorectal biopsies vs. stool samples) or by the fact that the species capable to perform tryptophan catabolism are less frequent in the stool microbiota. In our view, the depletion of genes involved in amino acid metabolism and energy processes might be upregulating inflammatory pathways in HIV-infected individuals.

Altogether, in the present chapter, it is proposed a complex network that integrates the different interactions between gut microbiota, metabolic functions, and host immunity. Although no relationship was found between microbial composition and sCD14, we found a correlation between bacterial translocation markers (sCD14 and BPI) and the zeatin biosynthesis pathway. As zeatins belong to a class of phytocytokine involved in the complex cell signaling pathway, this observation suggests that the interaction between the microbiota and bacterial translocation could occur indirectly by cytokine signaling. Importantly, we detected a strong correlation between bacterial genera composition and the LPS pathway, suggesting that the abundance of Gram-negative bacteria in the dysbiotic microbiota of HIV infected subjects might contribute to the burden of translocated bacterial antigens and, consequently, to chronic immune activation. This hypothesis is also supported by the significant association between the LPS biosynthesis pathway and both BPI and T-cell activation (%CD4+HLA-DR+CD38+ T-cells). Similarly, pathways related to bacterial antioxidant response, glutathione metabolism, and D-glutamine and D-glutamate metabolism correlated with different immune activation markers.

Considering our findings and current knowledge in the field, we believe that the profound disruption of GALT by the HIV infection would generate a dysbiotic microbiota, both in terms of its composition Gram-negative bacteria-enrichment and in terms of its altered metabolic profile with many genes involved in the LPS biosynthesis pathway, pathogenic pathways, and processes related to oxidative stress. This compositional and functional dysbiosis seems to fuel chronic innate and adaptive immune dysfunction and may be a viable target for interventions. Further longitudinal studies should be performed to provide evidence of causality in the correlations. A larger sample size and the inclusion of HIV+ patients with a different immunological response will be necessary to obtain a comprehensive understanding of the role of dysbiotic microbiota in HIV infection. However, this last point will be addressed in the following chapters of the thesis.

5. CHAPTER 2

THE EFFECTS OF PREBIOTICS ON MICROBIAL DYSBIOSIS, BUTYRATE PRODUCTION AND IMMUNITY IN HIV-INFECTED SUBJECTS

The results of the present chapter have been published in:

Serrano-Villar S*, Vázquez-Castellanos JF*, Vallejo A, Latorre A, Sainz T, Ferrando-Martínez S, et al. **The Effects of Prebiotics on Microbial Dysbiosis, Butyrate Production and Immunity in HIV-Infected Subjects**. Mucosal Immunol. 2016; (Epub ahead). [315]

5.1. Introduction

HIV infection induces a compositional shift of the gut microbiota, with enrichment of bacterial populations that are either pro-inflammatory or potentially pathogenic [281,283–285,287,353] and changes in microbial diversity that are correlated with immune status [286]. This perturbation of the gut ecosystem might be reflected in the bacterial functional capabilities, with enrichment of genes involved in various pathogenic processes and inflammatory pathways (as shown in Chapter 1) and the production of metabolites that directly influence host health, such as the immunomodulatory derivatives kynurenine [279] and SCFAs [175,375,376]. SCFAs are the products of bacterial anaerobic fermentation of dietary fiber and might influence host health by different mechanisms. In particular, butyrate is considered the preferred energy source of colon epithelial cells because 70% of their energy is derived from the oxidation of this acid [377]. This metabolite activates genes involved in different pathways, such as colonic gluconeogenesis, or genes responsible for the proliferation and differentiation of epithelial cells [168]. Moreover, butyrate has been shown to contribute to maintaining homeostasis in the gut by promoting immunotolerance to commensal bacteria via the downregulation of lipopolysaccharide-induced proinflammatory mediators [175].

Although different studies in HIV patients [296,299,305,306,378,379] have described changes in immunity and in intestinal bacteria after prebiotic, probiotic, or symbiotic administration, a deeper understanding of the ecological effects of such interventions on the structure and function of gut bacterial communities is needed, especially in the context of HIV-infected individuals who are undergoing treatment. These therapies included fructo-oligosaccharides (FOS) [297], mixtures of galacto-oligosaccharides (GOS) and polyunsaturated acids [296], and the bacterial strains *Lactobacillus rhamnosus*, *Lactobacillus reuteri* and *Bifidobacterium lactis Bi-07* [297,298]. In HIV-infected individuals not receiving ART, such interventions have been shown to contribute to the maintenance of immune functions over a longer period [296–298]. Gori et al. demonstrated, in a controlled study of ART-naive HIV-infected subjects, significant declines of markers of inflammation sCD14 and T-cell activation (%CD4+CD25+ T-cells) in association with increased bifidobacteria and decreased pathogenic *Clostridium lituseburense* and *Clostridium histolyticum* levels [299]. Subsequently, in SIV-infected macaques beginning ART, a symbiotic treatment (a mixture of prebiotics and probiotics) resulted in increased frequency and functionality of gastrointestinal antigen-presenting cells, enhanced reconstitution and functionality of CD4+ T-cells, and reduced fibrosis of lymphoid follicles in the colon [307]. Recently, in two placebo-controlled

trials with treated HIV-infected individuals, a short probiotic intervention elicited a significant reduction in inflammatory predictors of mortality, i.e., interleukin-6 and D-dimers [305,306]. Outside the field of HIV infection, glutamine has been shown to be a crucial metabolic fuel source for enterocytes, enhancing cell turnover, attenuating bacterial translocation and, conceivably, altering gut microbiota structure [380–382]. In the present chapter, we study the effect of a dietary supplementation with prebiotics and glutamine on the HIV-associated dysbiosis, the innate and adaptive immunity and the production of SCFAs.

5.2. Materials and Methods

5.2.1. Study design, participants, setting, and eligibility

To capture a wide spectrum of HIV immunopathogenesis, we recruited viremic untreated (VU) HIV-infected subjects, immunological ART responders and non-responders (IR and INR, ≥350 and <350 CD4+ T-cell counts after >2 years of viral suppression, respectively). A total of 44 individuals completed the 6-week study on-treatment, 34 in the active arm (20g mixture of prebiotics, including 5g of short chain galacto-oligosaccharides (Purimune®), 10g of long chain fructo-oligosaccharides (Orafti-HP® and Actilight®), and 5g of glutamine (Nutrición Médica®) and 10 in the placebo arm (20g of maltodextrin). The distribution by groups was: 12 VU (9/3), 8 INR (8/0), 15 IR (10/5), and 9 (7/2) controls (HIV-). Nine patients dropped out of the study (3 in the active arm and 6 in the placebo arm). The full description of the cohort is in the General Material and Method section "3.1. Cohort".

5.2.2. Short chain fatty acids measurements

This work has been carried out in collaboration with the Food Technology Department, Agrotecnio Center, University of Lleida, Spain. They used the bacterial suspension obtained as previously described in the General Material and Method section "3.3.1". The bacterial fraction was acidified with 45 μL of orthophosphoric acid in order to facilitate the liberation of SCFA from the fecal matrix. Then, the samples were centrifuged (9000 rpm, 10 min at room temperature) and the supernatants were directly injected into the gas chromatograph (GC). The analysis of acetic, propionic, butyric, isobutyric, isovaleric and valeric acids was performed by GC (Agilent 7890A Series) using a capillary BP-21 column (30 m, 0.25 mm, 0.25 μm) (SGE, Cromlab SL, Barcelona, Spain), coupled to a flame ionization detector (FID). Identification of the SCFA was made according to the retention time of standard compounds (Sigma-Aldrich, ST. Louis, MO, USA) and their quantification was determined with reference to the peak side of internal standard. The concentration was expressed as mM of SCFA.

5.2.3. Short chain fatty acids clustering and GLM analysis

An NMDS analysis (library "Vegan" function "metaMDS") was used to determine the clustering of the samples based on the relative abundances of SCFAs. The input distance matrix was calculated using the Bray-Curtis dissimilarity index (library "Vegan" function "vegdist") and the statistical difference between the HIV+ group and the HIV- controls were compared with the ADONIS test

(library “Vegan” function “adonis”). The envfit (library “Vegan” function “envfit”) function was used to determine the correlations between the NMDS analysis of each SCFA in the distance matrix with the assessed variables. The core of species determining the abundance of each of the six SCFAs was analyzed. First, we calculated the log10 ratio for the SCFA abundance and the bacterial taxa after/before the intervention. The resulting matrix was used to estimate the correlation between the bacterial species and SCFA using the Spearman's rank correlation coefficient (library “stats” function “cor.test”). Finally, a GLM was estimated using the log10 ratio values for each of the six SCFA and the relative abundance of the species using the ridge regression and 40 steps of cross-validation. The model was calculated setting the SCFA as the response variable and the species matrix as the predictors. Only those regressors that showed a significantly Spearman correlation index (p-value < 0.01) were considered to determine SCFAs abundance.

5.2.4. Generalized linear model of the prebiotic effect

To assess the interactions between changes in genera contributing to HIV-associated dysbiosis changes in peripheral markers of disease progression and changes in SCFA abundance, we used GLM (penalized linear “elasticnet” models). As response variables, we selected those SCFAs and peripheral markers with statistically significant variations (p-value < 0.05) after the intervention (**Table S12.2.6,** Appendix section) and the inflammatory predictors of disease progression CRP and sCD14. As predictors, we considered all SCFA whose abundance was found significantly modified after the prebiotic treatment, the species-core-responders to the SCFA, all the LEfSe biomarkers related to the HIV+ condition and all the LEfSe biomarkers related to the prebiotic intake. The magnitude of the effect was measured using the difference between the values before/after of each of these variables measured in the study. To avoid over-fitting problems, all the models were validated by means of n-fold cross validation (library “glment” function “cv.glmnet”). All the models were represented as a network (library “Igraph” function “graph.data.frame” and “plot”). Only those variables included in the GLM and showed a significant Spearman correlation coefficient (p-value < 0.1) were included in the network.

5.2.5. Prebiotic effect Bayesian network

The BN was estimated using the R statistical software (package “bnlearn”, function “hc”) setting as input the difference (before/after prebiotic intake) of the microbial composition, the relative abundance of the SCFAs and the systemic biomarkers of disease progression. The white list (i.e., all those arcs that must be included in the network and possess a Spearman p-value < 0.05) the

interactions across the SCFA, the species-core-responders related to the butyric acid, the species that were significantly modified with the prebiotic administration and the systemic biomarkers were used to define the setting. The option blacklist (‘‘bnlearn’’ function ‘‘hc’) was used to define the set of arcs not included in the model, excluding those variables with a correlation p-value above 0.1. Function mb (R Package ‘‘bnlearn’’ function ‘‘mb’’) was used to estimate the Markov Blanket from butyric acid in the VU group in order to find all the variables significantly associated with the butyrate after the prebiotic intervention. We used the Bayes Information criteria and the Spearman-correlation index to control for spurious associations. The network was plotted using the R package Igraph.

5.2.6. Effect of the sexual orientation

As reviewed in the introduction, the sexual preferences could lead the shift in the microbiota species composition [289]. To ponder the effects of HIV infection and sexual orientation, we used the ADONIS test to prove which variable can explain the variance observed in the weighted UniFrac distance matrix. The model formula was specified as:

$$wUnifrac \sim HIV_{status} + Group * Transmission$$

Where the "*" operator represents the interaction between the variables “study group” and “sexual orientation” and the "+" separates the terms “HIV status” from the study group and sexual orientation variables.

5.3. Results

5.3.1. General characteristics of the study population and safety data

A total of 44 individuals completed the 6-week course of treatment, with 34 receiving prebiotics and 10 receiving a placebo. These included 12 viremic, ART-untreated (VU) HIV-infected subjects, 15 immunological ART responders (IR), 8 non-responders (INR), and 9 HIV-uninfected (HIV) controls. A characteristic of the 44 individuals and main parameters studied in this work are seen in **Table 5.1.** HIV-infected individuals on ART were representative of a medium-aged population undergoing long-term treatment without comorbidities (**Table 5.1**). No statistically significant differences were observed among the groups in dietary habits (**Figure 5.1**). The average compliance to the intervention in both arms was superior to 85%, and overall, the intervention was well tolerated.

Table 5.1 Characteristics of the study population.

	HIV+ART- (VU) (N=12)	**HIV+ART+ CD4<350 (INR) (N=8)**	**HIV+ART+ CD4>350 (IR) (N=15)**	**HIV- (N=9)**	**P-value**
Age (years, IQR)	34 (33-35)	48 (41-53)	40 (33-48)	47 (31-60)	0.096
Male gender (No., %)[a]	11 (91.7)	8 (100)	13 (81.2)	6 (67)	0.001
CD4+ T-cell count (cells/mm^3, IQR)	558 (432-646)	291 (230-324)	561 (426-794)	762 (653-878)	<0.001
CD4/CD8 ratio (IQR)	0.53 (0.44-0.65)	0.71 (0.35-0.89)	1.13 (0.96-1.26)	1.51 (1.20-1.78)	<0.001
HIV RNA level (copies/mL, IQR)	22198 (9955-40621)	<20	<20	-	0.0016
Nadir CD4+ T-cell count (cells/mm^3, IQR)	540 (540-540)	131 (30-171)	247 (144-284)	-	0.098
Time since HIV diagnosis (years, IQR)	5.4 (2.9-9.9)	11.8 (5.6-20.8)	8.0 (4.5-13.9)	-	0.124
Duration of ART (years, IQR)	-	8.9 (4.6-12.7)	6.1 (4.3-11.7)	-	0.56
Framingham Risk Score (%, IQR)	2 (1-2.5)	4 (4-7)	4 (1-7)	2 (1-4)	0.134
Body mass index (kg/m^2)	24.3 (22.6-27.0)	25.6 (24.0-26.6)	23.7 (22.4-24.5)	23.0 (20.8-24.3)	0.926
Glucose (mg/dL)	86 (84-92)	92 (88-101)	93 (81-97)	95 (87-99)	0.067
Total cholesterol (mg/dL)	164 (152-187)	152 (142-201)	198 (153-209)	184 (147-213)	0.008
HDL cholesterol (mg/dL)	47 (42-54)	50 (43-60)	54 (50-61)	62 (48-75)	0.02
LDL cholesterol (mg/dL)	89 (87-102)	78 (67-126)	112 (92-124)	97 (75-133)	0.046

Active arm (No., %)	9 (75)	8 (100)	10 (62.5)	7 (77.8)	0.682

Abbreviations: ART, antiretroviral therapy; HIV-, HIV-uninfected healthy controls; INR, HIV infected patients' immunological non-responders to antiretroviral therapy; IR, HIV- infected patients' immunological responders to antiretroviral therapy; VU, HIV-infected subjects viremic untreated. All values are expressed as P50 (P25-P75) unless otherwise expressed. P-values were calculated using the Kruskal–Wallis to compare median values across all groups and the Mann–Whitney test was used to compare the values between two groups.
[a]Percentages are calculated over the total number of subjects in each group.

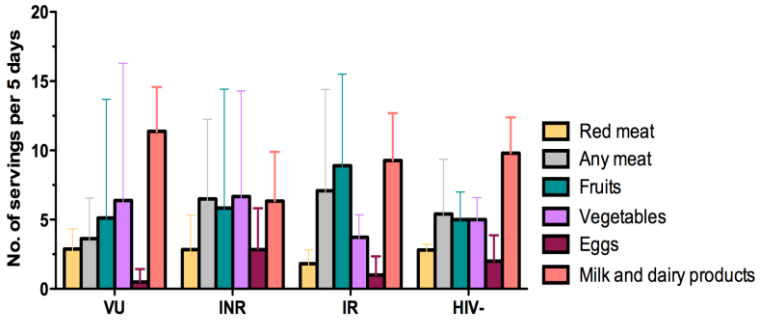


Figure 5.1 Dietary habits of study participants. Standardized values are shown as "number of servings per 5 days" and mean ± SD. No comparison reached the threshold of statistical significance (p-value<0.05). Abbreviations: VU, viremic untreated; INR, immunological non-responder, IR, immunological responder. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

The frequency of adverse effects is summarized in **Table 5.2**. During the intervention, a non-statistically significant higher proportion of patients were lost to in the placebo arm for reasons other than adverse effects. One patient in the active arm and two in the placebo arm dropped out of the study because of gastrointestinal-related effects. A higher proportion of patients reported mild gastrointestinal effects in the active arm, but the difference did not reach the threshold of statistical significance. No other adverse effects were registered during the study.

Table 5.2 Description of adverse events.

	Active arm (N=37)	**Placebo arm (N=16)**	**P value (Fisher's exact test)**
Gastrointestinal symptoms leading to discontinuation	1 (2.7%)	2 (12.5%)	0.213
Diarrhea/Loose stools	8 (21.6%)	1 (6.2%)	0.248
Bloating/Flatulence	11 (29.7%)	1 (6.2%)	0.080
Others	-	-	-

5.3.2. Effects of the nutritional prebiotic intervention on gut microbiota structure

Alpha diversity is used to measure the richness and evenness of bacterial taxa within a community. We found that, before intervention, VU subjects had the highest microbial richness of the cohort for each of the four-different metrics (ACE, Chao1, Shannon, and the total of observed species), followed by HIV- individuals, while INR subjects showed the lowest diversity (**Figure 5.2a and Table S12.2.1**, Appendix section). The nutritional intervention did not result in a significant variation of the alpha diversity parameters, as the microbiota of VU individuals was still the most

diverse at the end of the study, whereas that of INR individuals remained the least diverse (**Figure 5.2a and Table S12.2.2,** Appendix section). Analysis of changes in beta-diversity showed that the prebiotic intervention significantly modified the microbiota structure in all the groups (**Figure 5.2b and Tables S12.2.3 and S12.2.4**, Appendix section).

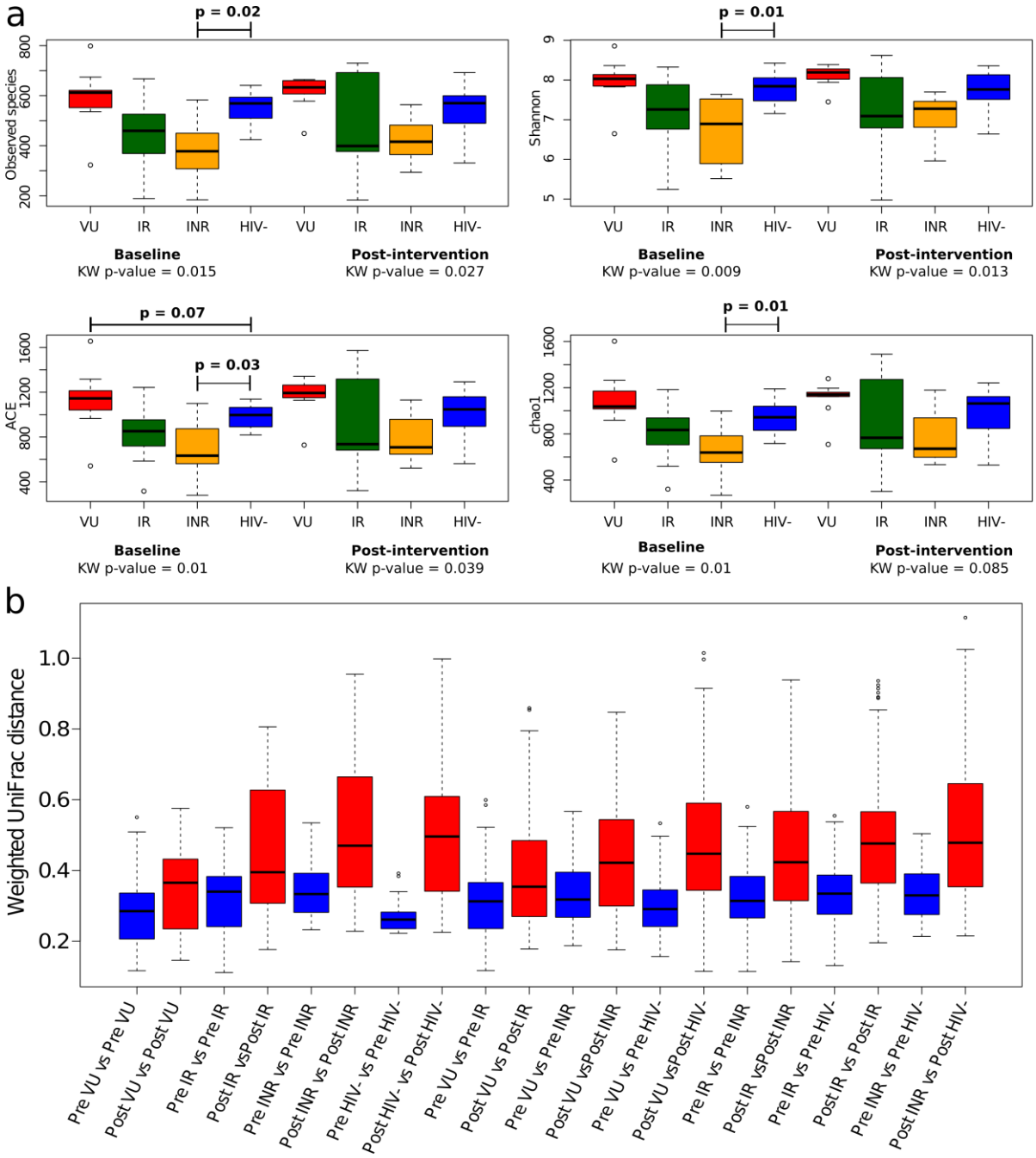
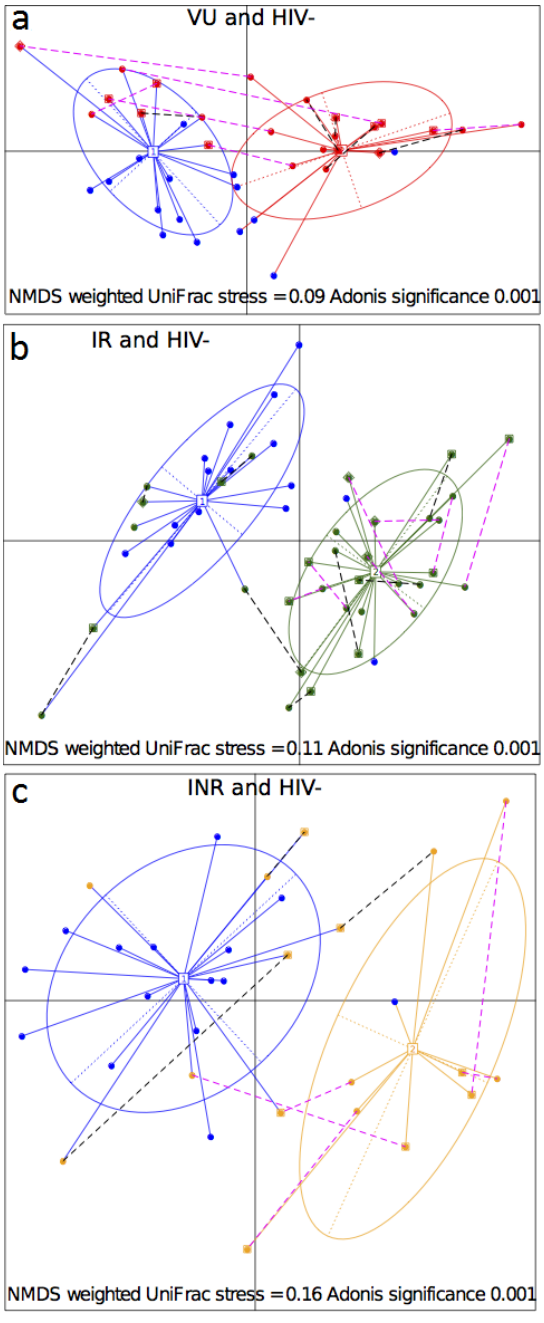


Figure 5.2 Diversity parameters of microbiota. (**a**) Alpha diversity at baseline and after prebiotics. (**b**) Beta diversity parameters of microbiota at baseline and after the prebiotic intervention. HIV+, HIV-uninfected control; INR, immunological non-responder, IR, immunological responder; VU, viremic untreated. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

Then, we asked whether the differences in the gut microbiota before and after the treatment might be representative of the disease status. Before treatment, partitioning around medoids analysis revealed distinct clustering in the subject groups, representing the extremes of health and disease, the HIV- and HIV+ individuals (**Figure S12.2.1a;** ADONIS; p-value=0.001, Appendix section). In addition, all HIV+ individuals harbored a dysbiotic microbiota compared with HIV- individuals (**Figure S12.2.1d,f,g;** ADONIS, p-value<0.01, Appendix section). Given the evidence that HIV-associated dysbiosis might be explained by sexual practices [289], we performed a permutational multivariate analysis to simultaneously adjust for this covariate. In keeping with recent observations, sexual orientation appeared as an environmental influence on microbiota composition, explaining 9% of the variance in the microbiota composition across groups (ADONIS p-value<0.001). However, this effect was weaker than the impact elicited by HIV itself, which explained 12% of the microbiota variability (ADONIS p-value<0.001) after adjusting for sexual orientation (**Table S12.2.5,** Appendix section).



Unsupervised clustering analyses revealed that although the intervention did not cause the bacterial communities in IR individuals to approximate those of the HIV- group, VU individuals and, to a lesser extent, INR individuals experienced a compositional shift towards the control group (**Figure 5.3**). Among the patients receiving the placebo, only one migrated to the HIV- cluster, indicating that the observed alterations were attributable to the prebiotic intervention.

Figure 5.3 Changes in gut bacterial communities of each group in response to the intervention (prebiotic and placebo arms). NMDS analysis of the composition distribution at OTU level based on the weighted UniFrac distance matrix. (**a**) VU (red dots) vs. HIV- at baseline (blue dots). (**b**) IR (green dots) vs. HIV- at baseline (blue dots). (**c**) INR (orange dots) vs. HIV- at baseline (blue dots). Each point is connected with a dashed line to another dot corresponding to the sample at the end of follow-up. Purple dash lines indicated those samples that significantly migrated from their baseline species profile, black dash lines indicated those samples that did not significantly change their baseline species profile based on the weighted UniFrac significance test. Dots framed by a rectangle represent individuals after treatment in the active arm; dots

framed by a diamond represent individuals after treatment in the placebo arm. Centroids for HIV- individuals are represented as ‘‘1’’ and those for HIV+ patients are represented as ‘‘2’’. Ellipses represent 70% of the samples belonging to each condition. Community differences were calculated using ADONIS test. INR, immunological non-responder; IR, immunological responder; VU, viremic untreated. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

We further investigated which genera determined baseline divergences of microbial communities across groups using the LEfSe biomarker discovery tool (**Figure 5.4 and Figure S12.2.2a,** Appendix section). As observed in previous studies [281,283–285,288], before the intervention, *Prevotella copri* was the most enriched species in all three groups of HIV+ subjects (**Figure S12.2.2a**). Other genera in the Firmicutes phylum, including *Eubacterium*, *Acidaminococcus*, and *Mitsuokella*, were consistently enriched. In contrast, the most depleted genus was *Bacteroides*. *Faecalibacterium* was also consistently depleted among HIV+ individuals (**Figure 5.4 and Figure S12.2.2a–c**, Appendix section). This genus includes *Faecalibacterium prausnitzii*, an important butyrate producer in the human gut that has anti-inflammatory effects and induces the novel immunoregulatory T-cell subset CD4CD8αα [383]. LEfSe analysis confirmed the depletion of *F. prausnitzii* in the fecal microbiota of HIV+ individuals and the depletion of other butyrate-producing genera, such as *Lachnospira*, *Anaerostipes*, *Butyricimonas*, *Coprococcus* (**Figure S12.2.2a**, Appendix section) and *Roseburia* (**Figure 5.4**) [375]. When we evaluated which specific species' abundance differed after the prebiotic treatment, microbiota variation was apparent only in the VU individuals (**Figure 5.5**), with an increase in Firmicutes (*Faecalibacterium*, *Catenibacterium*, *Blautia*, *Eubacterium*) and Actinobacteria (*Collinsella* and *Corynebacterium*).

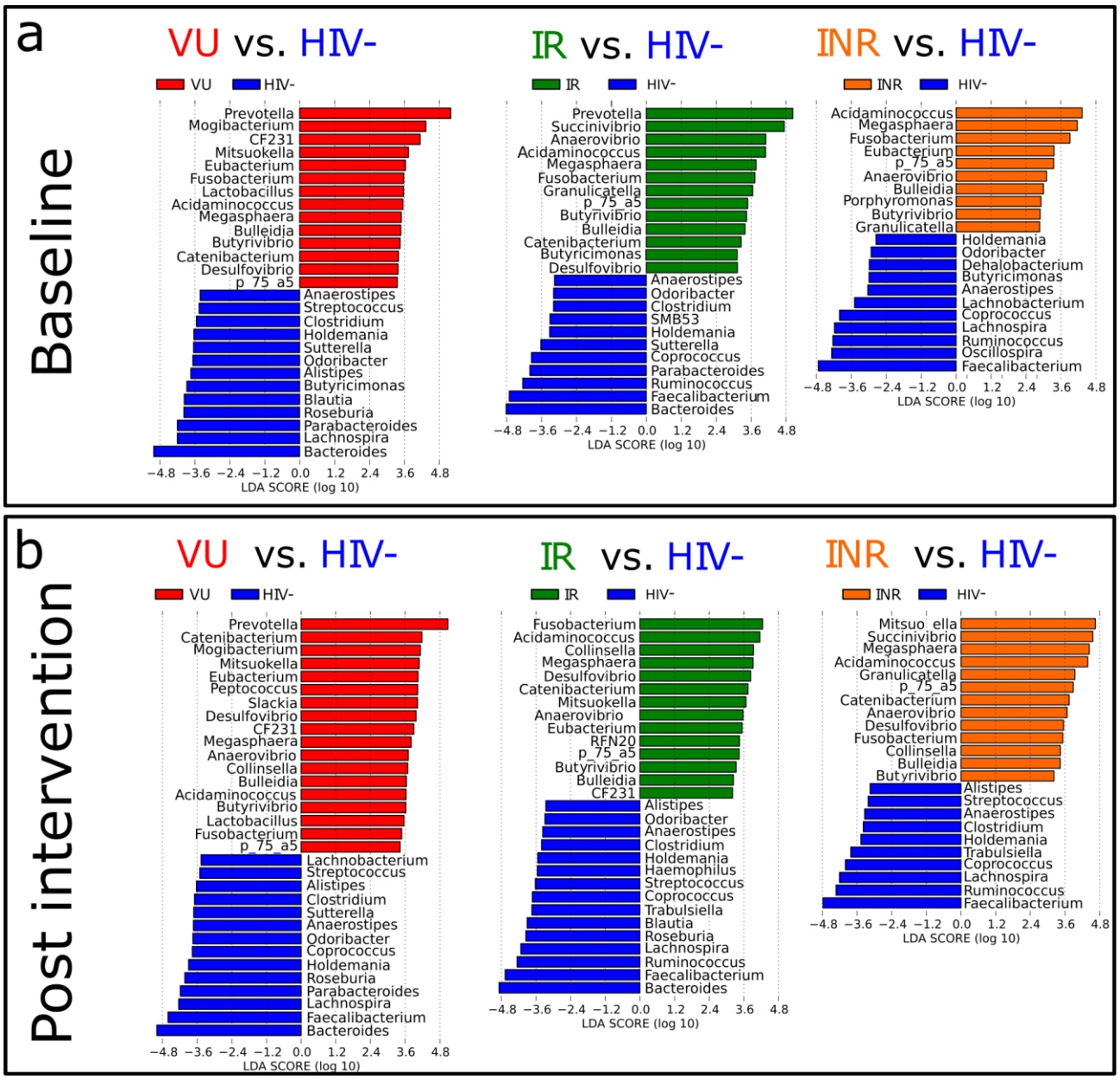


Figure 5.4 Bacterial taxa driving HIV-associated dysbiosis using LEfSe analysis in each group. (**a**) At baseline. (**b**) After the prebiotic intervention. The linear discriminative analysis (LDA) scores (log 10) for the most prevalent taxa among HIV+ individuals are represented on the positive scale, whereas LDA-negative scores indicate those taxa enriched in HIV- individuals at baseline. HIV- HIV-uninfected controls (blue); INR, immunological non-responders (orange); IR, immunological non-responders (green); VU, viremic untreated (red). Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

Taken together, alpha diversity analyses showed that the number of taxa is not influenced by the dietary intervention. However, the intervention significantly modified microbiota structure, as shown by changes in beta diversity (**Figures 5.2b and 5.3**). Interestingly, compared with healthy controls, the VU group treated with the prebiotic showed attenuated HIV-associated dysbiosis and increased growth of beneficial genera, which was not observed among ART-treated patients (**Figure 5.5**).

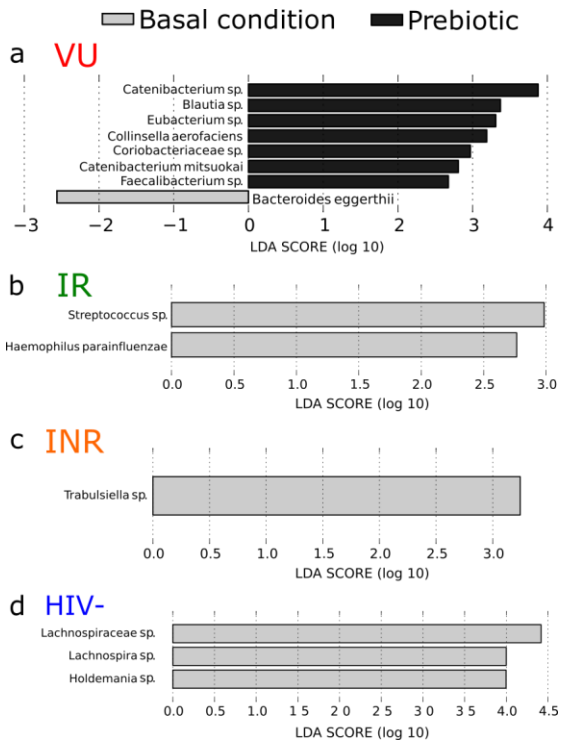


Figure 5.5 Changes in bacterial communities after the prebiotic intervention using LEfSe biomarker discovery tool at baseline and after the prebiotic intervention. We used the linear discriminative analysis (LDA) effect size (LEfSE) algorithm between the basal condition and the prebiotic administration in the (**a**) viremic untreated group (VU); (**b**) immunological responder group (IR); (**c**) immunological non-responder group (INR); (**d**) uninfected group (HIV-) LDA scores (log 10) (log 10 LDA > 2) for the most prevalent taxa in the basal samples are represented on the positive scale, whereas the prebiotic related species are represented on the negative scale. This analysis identified bacterial biomarkers after the prebiotic intervention only in the VU group, indicating that the intervention promoted the growth of determined bacteria in this group. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

5.3.3. Short chain fatty acid profile of HIV-dysbiotic bacteria and effects of the intervention

The above analysis implicated butyrate bacterial producers as potential key bacteria in HIV immunopathogenesis. To provide mechanistic insights into the influence of the altered bacterial communities on the host's health, we measured the abundance of SCFAs, including acetate, propionate, butyrate, valerate, isobutyrate, and isovalerate, in fecal bacteria (**Figure 5.6**). At baseline, a different SCFA profile characterized the HIV-infected group (**Figure S12.2.3**; ADONIS p-value=0.019, Appendix section) with more abundant propionate and lower levels of acetate (p-value=0.05 and 0.036, respectively) and no significant variation in the other four acids, including butyrate. After the prebiotic intervention, the propionate abundance increased in HIV-uninfected individuals, and only butyrate significantly increased in VU individuals (**Figure 5.7**, p-value=0.05). It should be noted that among all of the SCFAs, butyrate seems to exert the most profound effects on gut health [64].

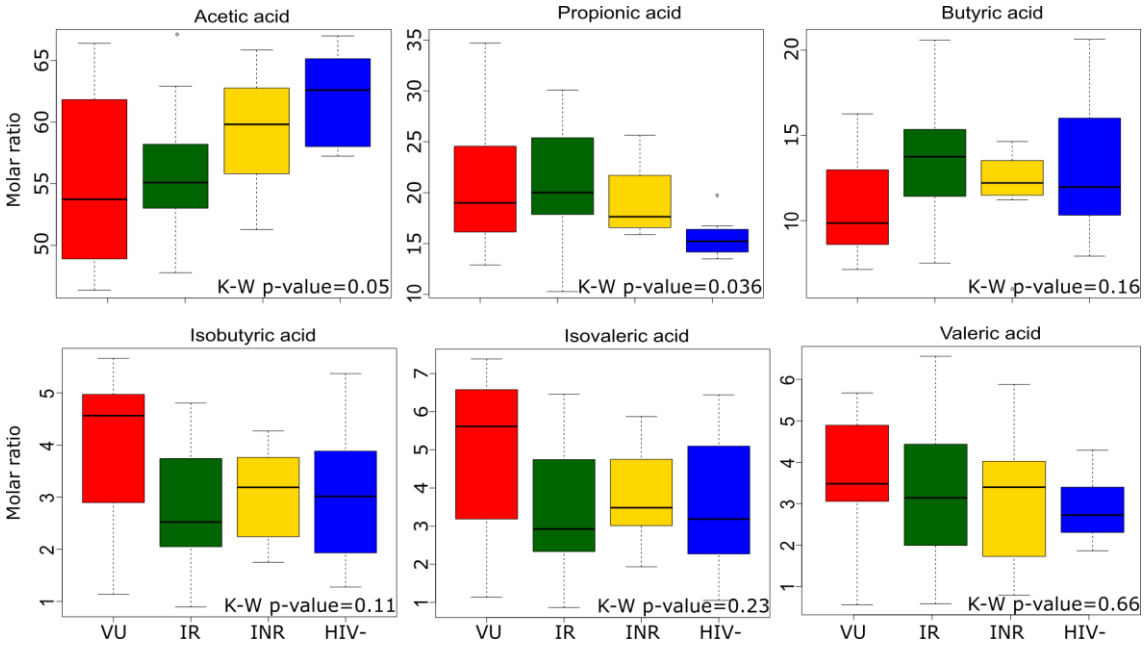


Figure 5.6 Relative abundance of SCFA in fecal bacteria per groups. The p-value reflects the differences in the relative abundance for the grouped comparison using a Kruskal–Wallis test. The concentration is expressed as millimolar of SCFA. INR, immunological non-responder, IR, immunological responder; SCFA, short-chain fatty acid; VU, viremic untreated. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

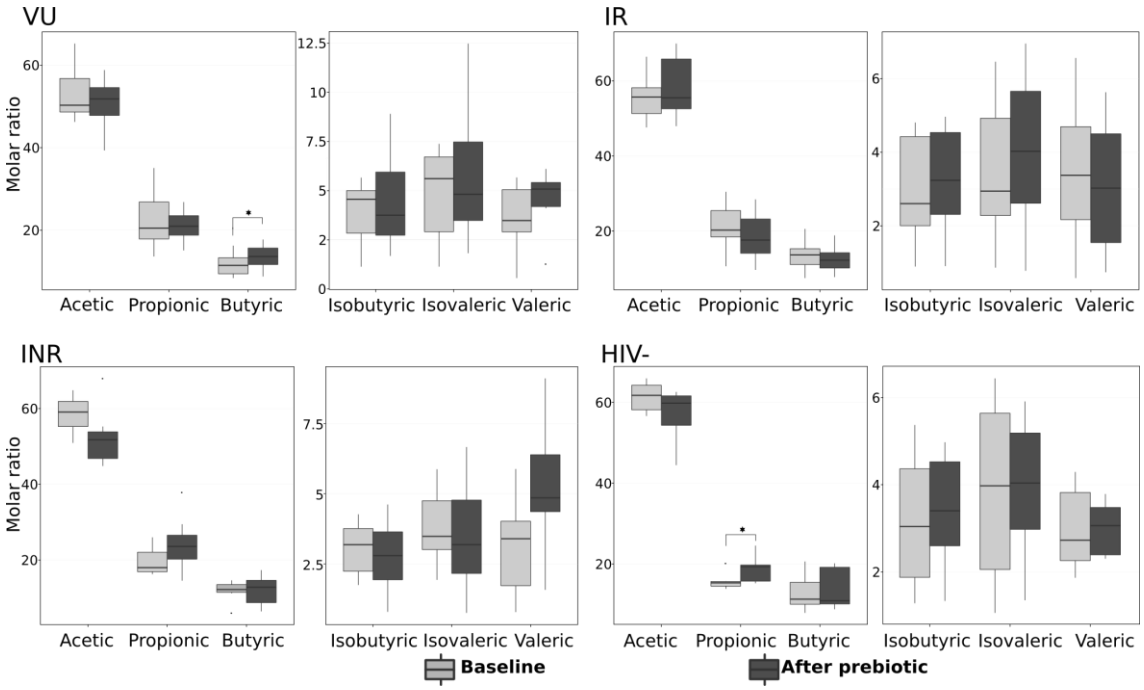


Figure 5.7 Relative abundance of SCFA in fecal bacteria. The abundance of each SCFA before (light gray) and after (dark gray) the prebiotic intervention. To calculate differences across groups in continuous variables, we used the Kruskal–Wallis test; and for differences between groups, we used Wilcoxon rank-sum tests. The concentration is expressed as millimolar of SCFA. INR, immunological non-responder, IR, immunological responder; SCFA, short-chain fatty acid; VU, viremic untreated. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

The bacterial species whose abundance changed in response to the prebiotic intervention (**Figure 5.5**) are salient candidates to the metabolism of the SCFAs, directly, or in association with a third species. Hence, we examined whether shifts in species abundance determined changes in SCFA

abundance. We found that butyrate abundance correlated positively with butyrate-producer species, including *Roseburia faecis* (p-value=0.0003), *Lachnospira* (p-value=0.0016), *Ruminococcus torques* (p-value=0.0093) and *F. prausnitzii* (p-value=0.0121; **Table S12.2.6**, Appendix section).

5.3.4. Changes in plasma biomarkers of activation of innate immunity

A number of bacterial species have been suggested to enhance pro-inflammatory pathways during chronic HIV infection [284,285] and that structural changes of the epithelial barrier persist even during successful ART [384], we hypothesized that an intervention aimed at restoring the gut microbiota and enhancing the gut barrier would affect markers of innate immunity activation, i.e., markers involved in inflammatory signaling (interleukin-6 and hs-CRP), D-dimers indicating thrombosis, asymmetric dymethilargininine indicating vascular dysfunction and sCD14 and BPI indicating bacterial translocation; (**Figure 5.8a**). No significant decreases were observed after the intervention, except for the BPI levels. This protein is produced by neutrophils and epithelial cells in response to proinflammatory stimuli and neutralizes the lipopolysaccharide. Hence, it is considered an indirect marker of bacterial translocation and a predictor of severe atherosclerosis [37]. The BPI levels decreased in all the groups to the levels observed at baseline in the HIV- group (Delta change: VU, -24±11, p-value=0.154; INR, -21±14, p-value=0.074; IR, -56±24, p-value=0.009).

5.3.5. Changes in plasma trimethylamine n-oxide concentrations

Metabolism of dietary choline by gut bacterial communities is a pathway with the potential to promote atherosclerosis in non-HIV-infected patients by producing proatherogenic TMAO [317]. After the prebiotic treatment, we found no significant differences in the median TMAO concentrations at baseline among groups, and a significant increase was noted in the IR group (Delta change: VU, 1.2±0.6, p-value=0.298; INR, -0.2±0.9, p-value=0.803; INR, 7.3±0.5, p-value <0.001; HIV-, 18.5 = 16.6, p-value=0.134; **Figure 5.8a**).

5.3.6. Changes in markers of activation of adaptive immunity

HIV-associated dysbiosis has also been linked to activation of adaptive immunity, which can predict disease progression during ART [39]. Hence, we evaluated changes in T-cell activation Markers (HLA-DR+ CD38+ and CD25+) and senescence (CD57+). At baseline, T-cell activation markers were higher in HIV-infected patients than in healthy individuals. After the intervention a decrease was observed especially for the percentage of the HLA-DR+CD38+ CD4+ T-cells (Delta change: VU, -1.6±0.7, p-value=0.006; INR, -0.2±0.5, p-value=0.552; IR, -0.4±0.2, p-value=0.040;

HIV-, 0.2±0.2, p-value=0.402) and the percentage of HLA-DR+ CD38+ CD8+ (Delta change: VU, -3.2±1.5, p-value=0.015; INR, 0.1±0.7, p-value=0.788; IR, 0.5±0.5, p-value=0.533; HIV-, -0.5±0.3, p-value=0.104) in the VU and IR groups. The percentage of CD25+CD8+ T-cells (Delta change: VU, -0.4±0.7, p-value=0.015; INR, -0.4±0.3, p-value=0.646; IR, -0.05±0.8, p-value=0.854; HIV-, 0.1±0.2, p-value=0.821) in VU showed as well a decrease. In contrast, we did not observe changes in the percentage of CD57+ T-cells (**Figure 5.8b**). Impaired thymic function is a hallmark of immunosenescence and is associated with a proinflammatory state and poor immunological recovery during treated HIV infection [40]. We studied thymic output by measuring levels of T-cell receptor rearrangement excision circles (the TRECratio). Overall, after treatment, thymic output improved in all groups of HIVs+ individuals, although it remained impaired compared with that in the HIV- controls, and the increase only reached statistical significance in VU individuals (Delta change: VU, 61±51, p-value=0.002; INR, 47±38, p-value=0.179; IR, 51±47, p-value=0.827; HIV-, 60±78, p-value=0.748; **Figure 5.8b**).

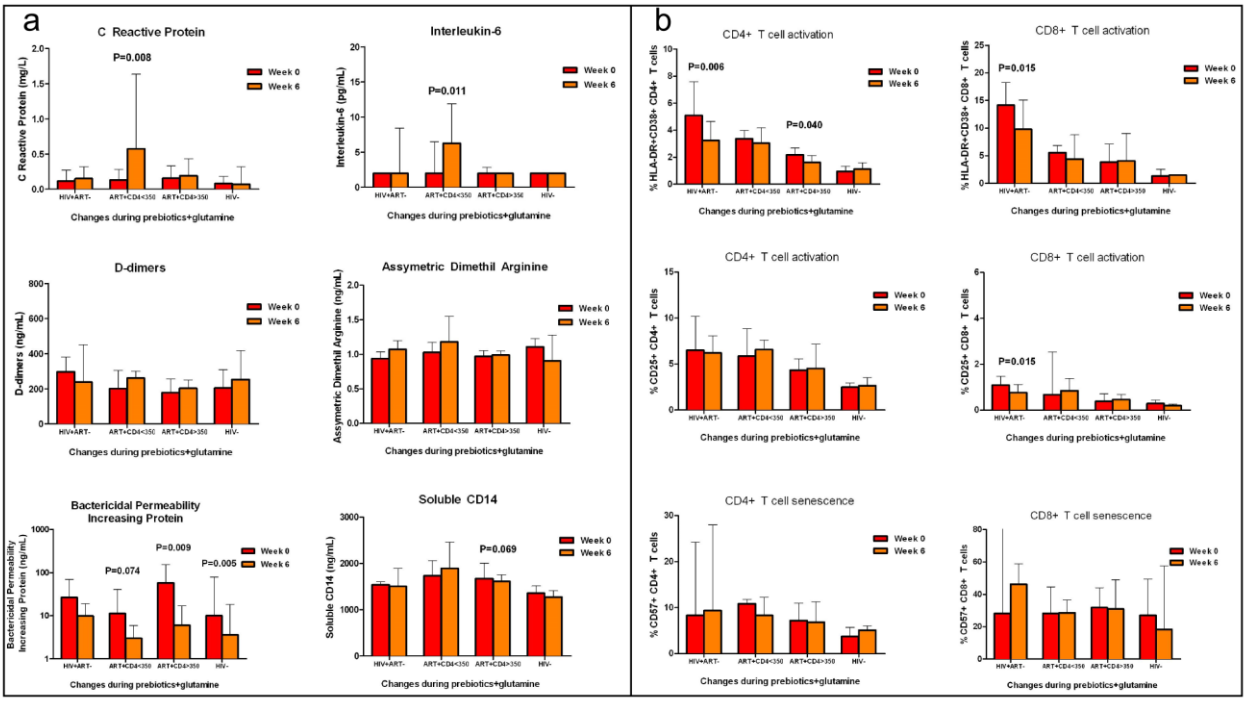


Figure 5.8 (**a**) Changes in systemic inflammatory markers in the prebiotic arm. (**b**) Changes in systemic markers of immune activation and thymic output in the prebiotic arm. INR, immunological non-responder, IR, immunological responder; VU, viremic untreated. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

5.3.7. Potential microbial targets for treatment of immune dysfunction in patients

The above analyses helped us identify taxa that might ameliorate the sustained proinflammatory state characteristic of HIV infection through the butyrate synthesis pathway and are thus of therapeutic interest. To untangle the impact of the prebiotic intervention on the complex interactions among key microbiota components, butyrate production, and circulating immunological predictors of disease progression, we speculated that the observed effects on systemic markers might be a consequence of changes in a subset of key bacteria, which could serve as potential microbial targets for interventions. Hence, we investigated whether changes in genera contributing to HIV-associated dysbiosis were associated with significant differences in SCFA abundance and peripheral markers of disease progression using generalized linear models and Bayesian networks (**Figure 5.10 and Table S12.2.7**, Appendix section). First, the analysis suggested that the disturbances in some of the taxa driving HIV-associated dysbiosis might contribute to the maintenance of persistent immune dysfunction, while changes in the abundance of other genera seemed a mechanism to alleviate it (**Figure 5.10a**). For example, *Bacteroides* was the most depleted genus in VU and IR (**Figure 5.4**), and their decrease strongly correlated with an increase in T-cell activation (**Figure 5.10a**). Similarly, *Prevotella* and *Acidaminococcus* were among the most enriched genera in HIV-infected individuals, and their increase further led to increased T-cell activation (**Figures 5.4 and 5.10a**). Hence, *Bacteroides*, *Prevotella*, and *Acidaminococcus* represent central components of the HIV-associated dysbiosis that seems to contribute to chronic immune dysfunction. In contrast, other taxa followed the reverse pattern of associations between their abundance at baseline and the effects of their changes on peripheral markers, likely representing a beneficial adaptation of the gut ecosystem. For example, all the groups of HIV-infected individuals were consistently enriched for *Butyrivibrio* species, and an increase in this species further correlated with a decrease in T-cell activation. These findings argue that some adaptive changes of the gut microbial ecosystem might help alleviate the consequences of the original perturbation (i.e., acute HIV infection) in the intestinal habitat, but others might perpetuate a vicious circle of dysbiosis and immune dysfunction.

As described above, the analysis of UniFrac distances suggested that the microbiota configuration was easier to modulate in VU individuals, who experienced a significant increase in butyrate production. Thus, in this group, we modeled the interactions among the bacterial biomarkers, SCFA production, and the peripheral biomarkers in a Bayesian network, and we estimated the Markov Blanket for the butyrate (**Figure 5.10b**). This method is able to infer multiple causal relationships

for a given effect. We found that increases in the abundance of *F. prausnitzii* determined an increase in butyrate, which in turn correlated with a decrease in the levels of two inflammatory predictors of mortality, high-sensitivity CRP, and soluble CD14[2]. Finally, because the analyses repeatedly implicated butyrate as a key SCFA, we aimed to identify the most likely bacterial determinants of butyrate abundance in VU individuals, and we modeled the interaction using generalized linear models. As shown in **Figure 5.10c**, the generalized linear model analysis identified *F. prausnitzii* and *Lachnospira* as major drivers of butyrate abundance.

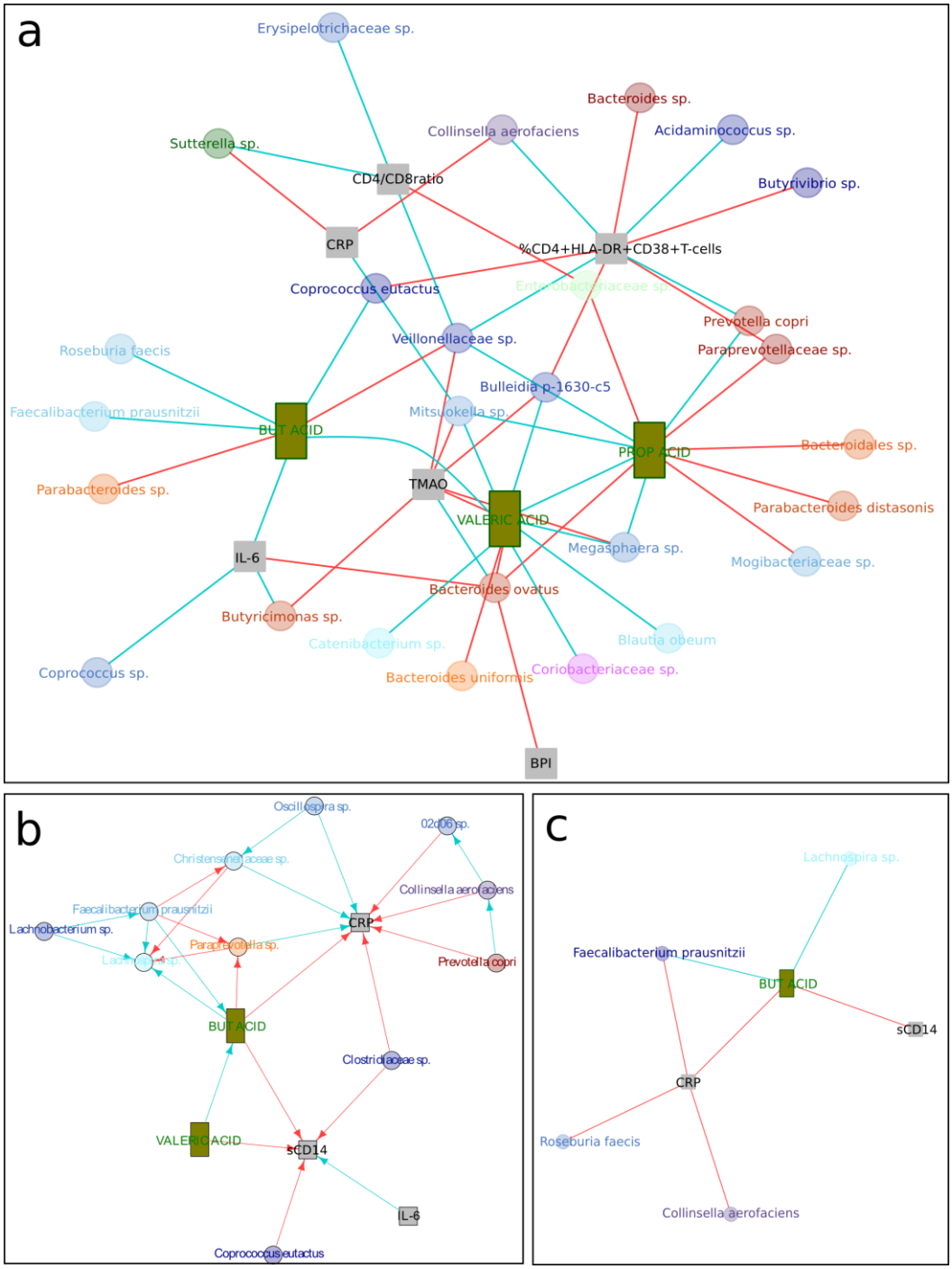


Figure 5.10 Network integrating genera abundance, SCFA abundance, and markers of innate and adaptive immunity. (**a**) Interactions between changes in genera contributing to HIV-associated dysbiosis, SCFAs, and peripheral markers of disease progression using generalized linear models are represented in (**a**). Red edges represent negative correlations and blue edges positive correlations. Gray squares represent the peripheral markers of disease progression and the green rectangles represent the SCFAs. Blue gradient circles correspond to species to the Firmicutes phylum, red gradient circles correspond to species to the Bacteroidetes, purple gradient corresponds to the Actinobacteria phylum and gold gradient corresponds to the Fusobacteria phylum. In (**a**,) the generalized linear model (GLM)-based network of all the HIV+ subjects of the cohort and illustrate the interactions between the microbiota, immunological markers, and SCFA production are shown. (**b**, **c**) Highlight the positive findings in the VU group with the prebiotic intervention by means of a Bayesian Network and a GLM, respectively. Arrows indicate conditional dependencies between variables in the Bayesian Network plotted in (**b**). The networks (**b** and **c**) highlight the interactions of *F. prausnitzii*, as well as other known butyrate producers, with butyrate production and the inflammatory markers high-sensitivity C-reactive protein (CRP) and soluble CD14 (sCD14). The increase of the butyrate predicts an increase of species related to SCFA production and a decrease of the inflammatory markers CRP and sCD14. SCFA, short-chain fatty acid. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

5.4. Discussion

In the present chapter 5, it is reported the results from a pilot study on HIV-infected individuals analyzing the effects of a nutritional intervention with prebiotics on both gut bacterial composition using next generation sequencing and the impact on the intestinal SCFA abundance in feces. The results show evidence of interplay between intestinal microbial ecology and the levels of inflammatory predictors of disease progression and characterized key components of the microbiota. This interaction occurred at least partially via the production of butyrate by fecal bacteria. It is also showed that the response to prebiotics depends on the stability and resilience of the whole bacterial community, which appears to depend on the control of HIV replication and the extent of ART-mediated immunological restoration. Furthermore, the findings argue that some adaptive changes in the gut microbial ecosystem might serve to alleviate the consequences of the perturbation in the intestinal habitat secondary to acute HIV infection, whereas others might perpetuate a vicious circle of dysbiosis and immune dysfunction.

The human gut microbiota comprises an extremely complex bacterial ecosystem that has co-evolved with its host, establishing entangled symbiotic relationships. Carbohydrates, which are a major component of the human diet, include plant derived polysaccharides (cellulose, hemicellulose, and pectin), starches, and sugars [385]. Human cells are able to hydrolyze some disaccharides and absorb monosaccharides as glucose or galactose, but they cannot break down most of the polysaccharides. Thus, a considerable number of dietary carbohydrates and proteins pass to the colon after having escaped digestion in the small bowel, and they are then metabolized by the gut microbiota, which contains a large repertoire of hydrolytic enzymes, producing mainly SCFAs [16,385–387]. It is now widely accepted that SCFAs have an important role in maintaining health and that they are associated with a wide range of diseases, including IBD, type 2 diabetes, obesity, celiac disease, allergies, colon cancer, and autoimmune disease [103,179]. As observed under other proinflammatory conditions [388], dietary modulation of gut microbiota with a diet rich in non-digestible but fermentable carbohydrates significantly promoted beneficial groups of bacteria and increased butyrate-producers, specifically the *F. prausnitzii* and *Lachnospira* species, which contributed to the alleviation of inflammation in HIV-infected patients. Hence, the dietary modulation of gut microbiota holds promise as a viable intervention strategy for HIV infection to attenuate persistent immune defects. The results also suggest, however, that the modulatory capacity of these interventions might vary depending on the fermentative capability of the baseline

altered bacterial community and on the type and chemical structure of prebiotics [388]. For example, whereas GOS are mainly degraded by Actinobacteria (*Bifidobacterium*, *Collinsella*), FOS can be metabolized by many microorganisms of different phyla (Firmicutes, Bacteroidetes, and Actinobacteria), and the observed effects would depend on the microbial digestion of each oligosaccharide [385].

We observed a spectrum of HIV-associated dysbiosis, reflecting the disease status of the study participants and determined by the extent of immunovirological control. Consistent with previous studies in HIV-infected subjects [283,284,288], we observed enrichment for *Prevotella*, a member of the *Bacteroidetes* phylum that has been linked with heightened T-cell activation in HIV+ individuals [284], and depletion of *Bacteroides* [281,285,288]. In addition, we found decreased levels of butyrate-producers, such as *Faecalibacterium*, *Lachospira* and *Roseburia*, and enrichment of pro-inflammatory genera, such as *Succinivibrio*, *Desulfovibrio*, and *Fusobacterium*. In chapter 4 was shown that in treated HIV-infected individuals that this pattern of dysbiosis is characterized by an enrichment of genes involved in lipopolysaccharide biosynthesis and a depletion of genes involved in amino acid metabolism and energy processes, suggesting that an altered functional profile could contribute to chronic bacterial translocation and immune dysfunction (mention in Chapter 1). In contrast to previous studies in HIV-infected subjects [281,285,389], we found decreased levels of Erysipelotrichaceae. Given the strong impact of diet on this family [390] and its correlation with obesity, this could have been affected by the Mediterranean diet and normal body mass index that characterized our study population. Still, whether HIV-associated dysbiosis is a result of HIV infection itself or an effect of a combination of several putative drivers, including ART and lifestyle factors, remains debated [289,391,392]. Indeed, at least two different studies in SIV infected macaques have failed to detect differences in the intestinal microbiota composition [393–395], in striking contrast with human studies, although a recent study of untreated acute SIV infection of non-human primates demonstrated depletion of Lactobacilli, supporting previous findings in HIV-infected individuals [396].

Because FOS stimulate the growth of bacteria such as the *F. prausnitzii*, *E. rectale*, *R. inulinivorans*, and *Lachnospira* species, all of them butyrate-producer species [139,385], we expected an increase in these taxa after administration of prebiotics. In VU individuals, the nutritional supplement increased the abundance of the depleted *Faecalibacterium*, a bacterium that is also depleted in Crohn's disease and ulcerative colitis, another inflammatory condition, [179,397] as well as other genera belonging to the Firmicutes (*Catenibacterium*, *Blautia* and *Eubacterium*)

and Actinobacteria (*Collinsella* and *Corynebacterium*) phyla. These changes indicate a beneficial impact of the intervention on HIV-associated dysbiosis. In addition to the important role of *Faecalibacterium* in inducing regulatory T-cells [383] and decreasing intestinal permeability [398], it has been shown that *Blautia* increases following fecal microbiota transplantation from healthy donors to individuals with recurrent *C. difficile* infections, who represent a subgroup of individuals with extremely impaired gut bacterial composition [397]. Moreover, and *Catenibacterium*, *Eubacterium*, and *Blautia* have been associated with SCFA production [399–403]. We also modeled the relationships among the bacteria that most likely influence the maintenance of chronic immunological abnormalities and found evidence that butyrate-producer species, particularly the fiber-fermenter and SCFA producers *F. prausnitzii* and *Lachnospira* species, might be viable targets for interventions. Although there is a significant increase in the butyrate levels after the prebiotic intervention in the VU group, it should be noted that no significant differences were noted between groups at baseline. The fact that the butyrate concentrations were one-third lower in VU than in the other groups suggests that this could be explained by the insufficient statistical power to detect changes at this level.

The effects on the microbiota composition were less apparent among treated HIV+ individuals, yet we detected some significant effects on markers of innate and adaptive immunity in peripheral blood (i.e., a decrease in BPI, decreased T-cell activation, improved thymic output). The younger median age of the VU group might have influenced in the higher impact of the nutritional intervention in the bacterial community in this group. We think, however, that a more compelling explanation is related to the longer median exposure to HIV disease of ART-treated participants (IR, 8.9 years; INR, 6.1 years) compared with that of VU individuals (5.4 years). Given results in the present chapter a model suggests, based on ecological principles, in which shorter exposure to the HIV disease might have elicited dysbiosis with a degraded transient state of the bacterial community. The combined and perhaps synergistic effects of a prolonged situation of chronic immunologic dysfunction, and ART itself could result in a second transition of microbiota structure to a degraded stable state capable of resisting colonization by the presumably beneficial species selected by the prebiotics [22,404]. Based on this model, only the VU-associated microbiota, in a transient state and with decreased resistance to an external nutritional intervention, was able to respond to the prebiotics. Although the landscape of stable states for the human gut is far from being determined, these findings suggest that HIV-associated dysbiosis might achieve a stable

configuration over time and that the dysbiosis of patients who have lived with HIV for many years might prove more difficult to reverse than that of their counterparts.

Some limitations must be considered when interpreting the results. First, this is a pilot study, and the limited sample size warrants caution regarding the conclusions. We measured many parameters in small groups, so we cannot rule out that some of the statistically significant changes are due to randomness rather than the intervention. In addition, the small number of patients receiving the placebo was useful in assessing the tolerability of the intervention but insufficient for comparisons of the assessed markers among arms. Because this study concluded before the effect of sexual orientation on the microbiota was appreciated [289], we had to conduct post hoc analyses to control for this potential confounder. Second, a limitation of most studies trying to understand the effects of shaping the microbiota, including ours, is that nutritional interventions could alter bacterial metabolism, which was not measured in our study, without altering the bacterial composition. Third, SCFA abundance was measured in feces, so the measurements represent the difference between production of SCFAs by the bacterial community and their uptake by the gut mucosa. Fourth, we measured bacterial translocation using only indirect markers (i.e., BPI and soluble CD14), and we did not sample the intestinal mucosa, which would have provided further insight into the bacterial communities driving mucosal disruption, immune defects, and increased gut permeability, so we cannot provide experimental evidence that the intervention affected gut immunity. Finally, the study was not designed to assess the effect of dietary changes on gut microbiota beyond the prebiotic intervention, and limited information on diet was collected. Therefore, given the strong dependence of TMAO on diet, it is difficult to interpret the significant increase of this metabolite that was observed in the IR group [317].

The field has only begun to scratch the surface of the potential implications of HIV-associated dysbiosis in the maintenance of immune dysfunction, and the study of gut microbiota modulation in this setting is in its early stages. So far, it is unclear how to best shape the gut-microbiome to modulate immunity. Although influencing the microbiota with a probiotic might prove harder than doing so with a prebiotic, studies comparing the effects of different probiotic strains and dietary interventions are lacking. There is, however, emerging consensus that a number of bacterial-derived metabolites, particularly SCFA, can impact systemic metabolism and immunity. Our study expands previous knowledge on the effects of microbiome-targeted interventions in HIV+ individuals by a comprehensive analysis of the influence of changes in gut bacterial composition and in SCFA production on innate and adaptive immunity. It is shown that a short and well-tolerated nutritional

intervention influenced gut microbiota structure and elicited downstream effects on a number of markers of clinical progression, although its impact would depend on the stability of the gut microbial community. The data show a distinct intestinal microbial makeup related to the extent of immunovirological control and suggest that the gut-microbiota is a viable target for interventions. Along with recent studies using other nutritional interventions aiming to affect gut microbiota [305,306], our data encourage testing new strategies aiming to affect HIV-associated dysbiosis. Future trials should focus on patients undergoing stable ART and analyze the long-term effects of therapies aiming to ameliorate HIV-associated dysbiosis. Personalized therapies based on microbiota structure would help to understand how we might exploit microbial–host interactions to restore normal health to individuals living with HIV.

6. CHAPTER 3

INTERPLAY BETWEEN GUT MICROBIOTA METABOLISM AND INFLAMMATION IN THE HIV INFECTION

6.1. Introduction

Most of the studies performed in the HIV gut-associated dysbiosis have been mainly focused on the characterization of the taxonomic profile, leaving out the potential role of functional shifts over the host health. Although in Chapter 1 we show that some of the metabolic pathways increased in the HIV+ subjects were related to the immune activation and bacterial translocation, little is known about the full potential of the virus infection over the metabolic capacities of the microbial communities and its effect on the host health. In 2016, Serrano-Villar et al., [405] showed applying meta-proteomic and metabolomic approaches that the HIV infection drastically modified the metabolism of three amino acids (proline, phenylalanine, and lysine). In congruence with the work of Vukovij [285], they also found that the 3-hydroxyanthranilate, a product of the kynurenine pathway, was accumulated in the microbiota of the HIV+ infected subjects [405]. Moreover, it was shown an association between an optimal response to ART and the active fraction of the gut microbiome, suggesting a role of the microbiota in the immune recovery [405,406]. However, nowadays gene expression studies in the gut-associated dysbiosis have not been addressed. Metatranscriptomics analysis provides key information about the microbiota metabolism, assessing what predicted genes are expressed in the community and in what conditions and to what extent [6,407–413].

Additionally, we should not ignore that the gut-microbiome is a complex community with highly interacted microbial species [414] and that the imbalance of these interactions could be an important factor for the dysbiosis establishment and maintenance. Such community structure has not been described in pathologies in which the immune system is compromised. In this context, the network analysis is a powerful tool to study the tightly interlinked biological systems [415–417]. For instance, Greenblum et al, [418] based on a community-level metabolic network of the gut-microbiome deduced that there are topological differences associated with obesity and IBD where the genes enriched in those conditions were located in the periphery of the metabolic network. Moreover, the use of ecological co-occurrence networks has been helpful to elucidate coexistence patterns spanning from pairs of microbial taxa in several ecosystems [419–421]. The co-occurrence networks allow us to measure the importance of the species within a community (e.g., degree, betweenness, measures of centrality), and possibly identifying keystone species within an ecosystem [419,422–424]. More importantly, recent studies suggested that the microbial

relationships shown in correlation-networks can be used to determine their contribution to specific conditions, as health status or disease. [14,24,75,414,418,424–426].

In the present chapter, we used a combination of 454 pyrosequencing and Miseq Illumina V3 sequencing to characterize the gene content by metagenomics and its expression level by metatranscriptomics of the microbiome in HIV patients. We found that in the HIV+ microbiota genes related to anti-inflammatory metabolic processes such as propanoate (ko00640) and butanoate (ko00650) pathways were under expressed while the genes related to stress resistance mechanisms (ko00730, ko00521 and ko4141) were overexpressed. This gene expression profile allows the adaptation to the inflammatory environment of the *Prevotella*, *Acidaminococcus*, and *Streptococcus* species, which were enriched in the active and total microbiota fraction. Indeed, the co-occurrence network retrieved from the HIV microbial community state *Prevotella copri* as one of the most important hub-species. In fact, the different discriminative biomarker species appeared as essential in the ecological network structure. Likewise, the KO biomarkers associated with HIV infection showed higher eigenvector centrality than the rest of the network enzymes indicating that the HIV infection causes dramatic changes in the metabolic structure of the gut microbiota.

Finally, we implement a “multiomic” network approach which accounts metagenomic, metatranscriptomic, metabolomic and clinical data retrieved from HIV+ subjects to model a Bayesian network to understand the immune response to dysbiosis associated with HIV infection.

6.2. Materials and Methods

6.2.1. Study cohort and clinical features

The subjects involved in the present chapter were the ones that participate in the "Chapter 2", before the prebiotic intake. The metagenomic samples and clinical measurements were for all complete set of subjects. The metatranscriptomic sequencing was performed in a subset of 33 samples. VU (6), IR (10), INR (4) and HIV- controls (13).

6.2.2. Sequence quality filtering and trimming

Illumina paired-end sequences retrieved from metagenomic (DNAseq) and metatranscriptomic (RNAseq) sequences were trimmed for adapters and the transposase sequence using the cutadapt software (v1.6) [427]. Sequence artifacts and low quality reads from the raw sequences were removed using the fastx_artifacts_filter software and the fastq_quality_trimmer (parameters: -t 20 -l 50) available into the fastxtoolkit [428]. Then, we used the prinseq-lite.pl script (parameters: -min_len 50 -min_qual_mean 20 -ns_max_n 1 -trim_qual_left 20 -trim_qual_right 25 -trim_qual_type min -trim_qual_rule lt -trim_qual_window 5 -trim_qual_step 1) to trim the 5' end and the 3' ends of each sequence, removing the trimmed reads that did not accomplish one of the following criteria: read length above the 170 nucleotides, mean quality score above 20 or contain less than the 1% of ambiguous base.

The possible reads derived from the host genome was removed using the deconseq.pl script (v0.4.3) [429] using as reference the human genome (Human Reference GRCh38). Finally, the paired-end reads were assembled using the pandaseq tool (v4.0.3) [430] (parameters: -N -l 50 -o 10) to obtain paired-end sequences.

6.2.3. Metagenomic functional annotation strategies

Different variation in the KEGG functional annotation was performed to avoid bioinformatics biases when analyzing the whole metagenomic dataset. This because the great variety of different assembly software and the whether or not previous use of an open reading frame (ORF) prediction to assign the functional annotations. In that sense, the metagenomic pipeline benchmarking was focused mainly on the use of the assembler (Ray-Meta vs IDBA-Meta) and the previous use of an ORF prediction tool versus the direct mapping sequences (coverage) approximation. Additionally, we compare the functional annotations retrieved by the PICRUST [431] pipeline, using the 454 16S

rRNA gene data set from the Chapter 2, and the pipeline used in the work of Nielsen et al., (2014) [432] modifying the assembler to the Ray-Meta assembler [433]. This due to the Ray-Meta assembler was the one that obtains the greater contig length (based on the N50 statistic) based on simulated datasets [434] and observed in the current study. Additionally, the 454 reads were included to improve the ORF prediction. In summary, we obtained eight different KEGG functional annotations that can be seen in **Table 6.1.** The results for this section are shown in the supplementary results "S12.3.1. Functional metagenomic pipeline comparison".

6.2.3.1. Read assembly

To obtain the contigs, the metagenomic Illumina paired-end sequences were assembled using two different assemblers: Ray-Meta [433] (V. 2.3.1 parameters: -k 31 -minimum-contig-length 300) and Meta-IDBA [435] (parameters: --pre_correction --mink 10 --step 10 -o assemblie --min_contig 300). The K-mer size was selected as the one that maximizes the N50 value for the Ray-Meta assembler.

6.2.3.2. ORF prediction

The ORF prediction from the metagenomic contigs and the unassembled Illumina and the 454 GS FLX Titanium sequences were obtained through the MetaGeneMark software [436] (prokaryotic version 3.25).

All the ORF were translated into amino acids (aa) sequences, and the ORF predicted from the metatranscriptomics data (see section 6.2.7.1. Metatranscriptomic functional annotation) were clustered in a non-redundant database (ORFaanr) using the USEARCH software (v8.1.1831 parameters: -id 0.95 -threads 2 -strand both --query_cov 0.9).

6.2.3.3. Sequence functional annotation

The ORFaanr database was compared against the KEGG orthologous group (KO) database [437] and the comprehensive antibiotic resistance database (CARD) [438] using rapsearch2 software [439] (parameters: min align length > 60, log e-value < -5 and identity >60). The assignation criteria were based by the ORF best-hit, prioritizing first the e-value, then the bit-score, the percentage of identity and finally the alignment length. In the case of two or more KO database entries contain the same best-hit values, then the ORF kept the annotation of all the KO entries. However, most of the ORFs that contained more than one gene annotation usually belonged to the

same KO group. Finally, the HUMAnN pipeline [440] was used to determine the presence/absence and abundance of microbial pathways retrieved from the KEGG database.

The relative abundance of each ORF was calculated mapping all the reads against the nucleotide sequences of the ORFaanr database using the soap aligner (version 2.21) [441] and the soap.coverage script (2.7.7) [441] according to the pipeline of Nielsen et al., (2014) [432].

By contrast, the coverage approximation was performed with the aim to observe if the ORF prediction introduces biases in the functional characterization. This is based on the direct comparison of the sequences against the database without a previous ORF prediction. The annotations for the coverage approximation was done by first translating in amino acid sequences the metagenomic unassembled reads and the contigs predicted by the Meta-IDBA and Ray-Meta assemblers and then mapping against the KEGG database using the rapsearch2 software (parameters: -e -5 -b 0 -t n). The assignation criteria cut-off was the same used for the ORF approximation.

The gene quantification for the coverage approximation was performed based on the work of Arumugam et al., (2011) [14] for the metagenomic unassembled reads and the contigs predicted by the Meta-IDBA and Ray-Meta, following the current equations:

Equation 1

$$abundance(g) = \frac{\sum_{r \in R} baseoverlap(g,r)}{baselength(g)}$$

The abundance of each gene *g* is calculated based on the sum of reads that overlap over *g* divided by the gene length.

Equation 2

$$abundance(KO) = \sum_{g \in KO} abundance(g)$$

Table 6.1 Annotation strategies.

Annotation strategy	Assembler	Sequence technology	ORF prediction
Ray_Illumina_454_ORF	Ray-Meta	Illumina nextera V3 and 454 Genome FLX Sequencer	MetaGene Mark
ray_ORF	Ray-Meta	Illumina miseq nextera V3	MetaGene Mark

ray_coverage	Ray-Meta	Illumina miseq nextera V3	No ORF prediction
idba_ORF	Meta-IDBA	Illumina miseq nextera V3	MetaGene Mark
idba_coverage	Meta-IDBA	Illumina miseq nextera V3	No ORF prediction
Unassembled_ORF	No assembly	Illumina miseq nextera V3	MetaGene Mark
Unassembled_coverage	No assembly	Illumina miseq nextera V3	No ORF prediction
Picrust	No assembly	No sequencing	No ORF prediction

6.2.3.4. Statistical analyses

In order to determine the consistency of the KEGG functional annotations a non-metric multidimensional analysis (Library "vegan" function "metaMDS") based on the Hellinger distance (Library "vegan" function "vegdist") was performed for each of the functional taxonomic assignation. The comparison between the ordination analyses was performed using the Procrustes test (Library "vegan" function "protest"), similarly, the Mantel test was performed to measure the correlation between the distance matrix (Library "vegan" function "mantel.rtest").

6.2.4. Sensitive metabolic pathway detection

The remote homolog detection was carried out by Hidden Markov Models (HMM) for each of the KEGG maps of the tryptophan metabolism (M00038) and the enzymes: Tryptophanase (K01667), Catalase-peroxidase (K03781) and the o-aminophenol oxidase (K20219). Briefly, we search a Pfam family (a collection of related protein regions) for each of the enzymes mentioned above in the Pfam database (http://pfam.xfam.org/). The Pfam database is a large collection of protein families. Each family consists of a curated "seed" alignment, an alignment small set of representative sequences of the family, and an HMM profile based on the seed alignment. We download the corresponding seed alignment for each of the enzymes, mention before, to construct its respective HMM. Those seed alignments are: PF06052 (KO: K00452), PF03301 (KO: K00453), PF01231 (KO: K00463), PF01494 (KO: K00486), K01432 (KO: PF07859), PF00266 (KO: K01556), PF04909 (KO: K03392), PF04199 (KO: K07130), PF00171 (KO: K10217) and PF07859 (KO: K14263).

Each of the seed alignments was used to create an HMM using the hmmbuild (v3.1b2) software [442]. Finally, all the HMM from the tryptophan metabolism were used to detect remote homologs in the ORFaanr database using the hmmsearch tool (v3.1b2) [442]. The homology detection was given if the full sequence e-value were below 0.001.

In a similar way, the enzymes related to the formation of trimethylamine (TMA) from choline and L-carnitine; the oxygenase component *YeaW/cntA* (ACCESSION APQ93983 and APP31394 respectively) and the reductase component *YeaX/cntB* (ACCESSION WP_001287026 and APQ93983 respectively) were searched into the ORFaanr database using Blastp (v2.2.31+) (alignment coverage >= 0.7 and e-value < 0.00001).

6.2.5. Taxonomic assignments

First, we constructed our own non-redundant genome database. The sequences of the reference genomes from the species found in the gastrointestinal tract were downloaded from the HMP web site (http://downloads.hmpdacc.org/data/reference_genomes/body_sites/Gastrointestinal_tract.nuc.fsa). Similarly, all the reference genomes and the latest assemblies of the genomes of the species related to the HMP genome database were downloaded from NCBI RefSeq database (ftp://ftp.ncbi.nlm.nih.gov/genomes/all) using the assembly_summary.txt file as a reference (https://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/assembly_summary.txt). The resulting database contains 1,113,533 sequence entries in 39, 537, 222, 120 bp. The sequences that share the same taxa-id were grouped in a fasta file. Each file was then compared against itself using megablast (parameters: blastn -task megablast -perc_identity 95). Those sequences that share a 95% of sequence identity and an alignment that covers at least the 90% of the smallest of the sequences were clustered, taking as the reference the longest one. For this task, we excluded all the plasmid sequences from the database. The taxonomy of the clustered database was obtained using the acc_taxid.zip, taxdump.zip files available on the NCBI website. The final genome clustered (Genomeclust) database contains 2463 species, 76,478 sequence entries in 10,545,961,079 bp.

The taxonomic assignation of the metagenomic and the metatranscriptomic sequences was carried out by mapping the unassembled reads in the Genomeclust database using the blastn tool (v2.2.31 parameters: -task megablast -perc_identity 75). Then, all reads whose alignment length was below 100 nts, or the percentage of identity was below 75%, or e-value was above 0.00001 were discharged. The assignation criterion was based by the read's best-hit, prioritizing first the e-value, then the bit-score, the percentage of identity and finally the alignment's length. In the case of two or more database genomes contain the same best-hit values; the read coverage was divided by the number of hits with the same score.

6.2.6. Generalized linear model for the taxonomic and functional biomarker associations

The putative associations between the pathways and species biomarkers were estimated by means of Generalized Linear Models (GLM). The GLM was estimated setting as the response variable the pathway abundance and as the predictors the species matrix. In order to avoid over-fitting, each model was cross-validated 54 times and the least absolute shrinkage and selection operator (lasso) was used to perform variable selection and regularization (library "glment" function "cv.glmnet" alpha = 1). Additionally, the Spearman's rank correlation coefficient (library ''stats'', function ''cor.test'') was calculated for each of pairwise taxonomic-pathway variables selecting only those associations whose adjusted p-value were below 0.05.

Once the statistical associations were determined, we look whether the genomes of those biomarker species contained the genes of the biomarker metabolic pathways from which they had been associated. In order to achieve this, we downloaded the ORFs fasta file from the NCBI RefSeq genomes database (ftp://ftp.ncbi.nlm.nih.gov/genomes/all) from all the selected species. Then, the program makeblastdb (v2.2.31) produces BLAST databases for each of the species ORF translated to amino acid sequences fasta files. Finally, the blastp tool (v2.2.31) was used to search all the genes related to those metabolic pathways in each genome ORF translated database. We took a good hit if the alinement's length were above 60 aa, the e-value were below to 0.000001 and if the percentage of identity were above the 60%.

6.2.7. Metatranscriptomic analyses

Metatranscriptomics is a relatively new field, with the first mention around 2008 [443], which it is advancing rapidly [7,51,409,444,445]. Due to the novelty of this technique, there is no consensus on how to analyze this "omic" approach. Although nowadays there are new approaches for the analysis [446,447] at the time that I performed the analyses, very few bioinformatic pipelines were suited for the assembly and ORF annotation [448,449]. For these reason in the present thesis based on the recommendation of [450–452], we develop an in-house pipeline for the analysis of the metatranscriptomic sequences.

6.2.7.1. Metatranscriptomic functional annotation

Following the literature recommendation [450–452], we assemble the metatranscriptomic Illumina sequences using the Trinity assembler [448] (parameters: Trinity --seqType fa --no_cleanup –normalize_reads). The ORF prediction from the metatranscriptomic reads was done using the TransDecoder.LongOrfs [453] and its expression was measured by mapping the metatranscriptomic sequences against the ORFaanr using the RSEM tool implemented into the rsem-synthesis-reference-transcripts [448,454]. The results were expressed in Fragments Per Kilobase of target transcript length per Million reads mapped (FPKM). The idea of the FPKM is to normalize for sequencing depth and gene length. Briefly, the total number of reads is divided by 1,000,000, the ratio result is known as the "per million scaling factor". Then, the abundance of each gene, in their respective sample, is divided by its respective "per million scaling factor" to normalized for the sequencing depth. Finally, the resulting value is divided by the gene length.

6.2.7.2. Diversity and homogeneity

As described in Franzosa et al., (2014) [455] the homogeneity of the genus and the functional KO composition within the metatranscriptomic and the metagenomic matrix was calculated using the Bray-Curtis index (Library "vegan" function "vegdist"). Similarly, the evenness within groups was measured using the Pielou's index [456] estimated as:

$$Pielou = \frac{H}{\log(S)}$$

Where H is the Shannon index (library "vegan" function "diversity") and S is the expected number of species (library "vegan" function "specnumber").

6.2.7.3. Biomarkers pathway expression

The LEfSe biomarkers (See in General Methods section 3.5) from the metatranscriptomic and metagenomic pathways were selected to observe differential expression. A pathway was over or infra expressed if the logarithm of the RNAseq / DNAseq ratio was statistically different from zero (T-student adjusted p-value < 0.01). Then, we compared the relative expression of those pathways between the HIV+ and the HIV- condition using the Wilcoxon test (library "stats" function "wilcox.test"). The hierarchical clustering was done (Library "stats" function "hclust" method = ward) using the log (RNAseq/DNAseq) from those pathways that were differentially expressed.

6.2.8. Gut-bacterial-metabolome

The metabolic data and the method of extraction were described in Serrano-Villar et., al (2016) [376] and Serrano-Villar et., al (2016) [405].

6.2.9. Metabolite data treatment, statistical analysis, and identification

The metabolomic data treatments were described in Serrano-Villar et., al (2016) [376] and Serrano-Villar et., al (2016) [405].

The identity of the compounds selected according to their significance in the class separation was further confirmed using LC-MS/MS in the same LC-ESI-QTOF-MS. Ions were targeted for collision-induced dissociation (CID) fragmentation based on the previously determined accurate masses and retention times. Their identity was confirmed by comparing the fragments that were obtained with the structure of the proposed compound in the MS/MS spectra in a public database (METLIN: https://metlin.scripps.edu/metabolites_list.php) or against commercially available standards. In the case in which the metabolite contains more than one annotation, we prioritized the KEGG metabolite annotation, if exist, otherwise, we select one from the most abundant annotation of the METLIN database or we took one from the most abundant Sub Class annotation of the Lipidomics Gateway database (http://www.lipidmaps.org/data/structure/).

The abundance matrix retrieved from the metabolomic data was standardized by the total number of metabolites. Then, to avoid biases given species composition we normalized data using the Hellinger transformation (Library “vegan” function “decostand”). The NMDS (Library “vegan” function “metaMDS” was performed to reduce dimensionality in the metabolomic data) and the ADONIS (Library “vegan” function “adonis”) test was employed to verify if exists a significant difference in the metabolic composition between the HIV+ groups. Finally, the Wilcoxon test (library “stats” function “wilcox.test”) was employed to detect the over and infra represented metabolites in the HIV+ groups (adjusted p-value < 0.01).

6.2.10. Ecological and metabolic networks

The ecological network was calculated for all the HIV+ subjects using all the species that were present in at least 70% of the samples and whose average relative abundance was above the 0.01%. The correlation matrix was estimated using the SparCC.py script [457] (parameters: -i 10). The statistical support was carried out by performing 1000 bootstrap resampling using the script

MakeBootstraps.py [457]. We only took as significant correlation if the adjusted p-value were below 0.01 and that the absolute value of the correlation coefficient was above 0.1. The co-occurrence network was estimated using the R package Igraph (function “graph.adjacency” mode "undirected") removing all loops and unconnected nodes (Igraph functions “simplify” and “delete_vertices”).

The walktrap community algorithm (Igraph function “walktrap.community”) was used to detect all the densely connected subgraphs, also called communities, within the ecological network. Those subgraphs that possess at least three nodes where later plotted as polygons in the ecological network. The taxonomic composition of such community’s sub-graphs was plotted as bar plots (R function “barplot”). The enzymatic reactions metabolic network was created using the R library “KEGGgraph”. First, all the KGML files from the pathways that were involved in the KEGG orthologous groups found in the HIV condition were downloaded from the KEGG: Kyoto Encyclopedia of Genes and Genomes website. The list of selected metabolic pathways was then parsed using the R function “parseKGML2Graph” (library “KEGGgraph”) and the direct network was assembly using the “mergeKEGGgraphs” function (Library “KEGGgraph”, edgemode = "directed"). The total network was then plotted using the Igrpah function plot.igraph.

The centrality network statistics, betweenness (Library “Igraph” function "betweenness"), degree (Library “Igraph” function “degree”) and eigenvector centralities (Library “Igraph” function “evcent”), were calculated for both the metabolic and ecological networks. Similarly, the distribution comparison test Kolmogorov-Smirnov (function “power.law.fit” and “degree.distribution”); the calculation of the shortest path (function “shortest.paths”) were calculated using the Igraph library.

The modularity was calculated by first defining the modules within the networks using the walktrap community algorithm (Igraph function “walktrap.community”). Then the fraction of the edges that fall within the given modules minus the expected fraction was calculated using the Igraph function “modularity”.

The network fragmentation is defined as the relative fraction of disconnected compartments within a co-occurrence network [417]. Based on the work of Widder et al., (2014) [417] we compute the fragmentation modifying the number of unconnected subgraph from the network to the number of clusters within the network. The fragmentation was calculated using the number of clusters derived for the function “clusters” and the walktrap.community”.

$$Fragmentation = \frac{\log(Num\_clusters)}{Num\_vertix}$$

In order to test the consistency of the results of the ecological and the functional networks, we simulate 10000 random networks using the Erdős–Rényi model (library Igraph function "erdos.renyi.game"), the probability of drawing an edge between two arbitrary nodes was sampled from a random distribution (mean = 0.5, sd = 0.3) for each of 10000 networks. The results are expressed in the **Table 10.3.2** (Appendix section).

6.2.11. Multiomic Bayesian network

A multiomic BN was estimated using the log normalized relative abundance of the metagenome, metatranscriptome and metabolome data from the HIV+ infected subjects in order to create a model which could predict the effect of the microbiota in the markers of innate and T-cell activation variables, thymic function, and bacterial translocation.

The network was estimated using the hill-climbing score-based learning algorithm (R library "bnlearn" function "hc"). The algorithm states the optimal network and in consequence the ''father'' to ''child'' node relationships which maximizes the Bayes information criterion (BIC). The whitelist option was used to define the set of arcs that will be included in the model. For this task we performed a GLM using the lasso variable selection and regularization (library "glment" function "cv.glmnet" alpha = 1) As response variables, we selected those RNAseq and DNAseq pathways, species and metabolites that were overrepresented in the HIV condition and as predictor the clinical variables related to the markers of innate and T-cell activation variables, thymic function, and bacterial translocation as done in the multivariate statistical framework MaAsLin pipeline [458]. To avoid over-fitting problems, all the models were validated by means of n-fold cross validation (library "glment" function "cv.glmnet"). Additionally, we included all the Spearman correlation adjusted p-value was above 0.01. We used the blacklist option (R library ''bnlearn'' function ''hc'') to exclude all the edges whose Spearman correlation adjusted p-value were above 0.1.

The function mb (R Package ''bnlearn'' function ''mb'') was used to dissect all the Markov Blankets in the HIV+ network to perform the statistical analysis and to plot the butanoate metabolism ko00650 Markov blanket.

All the network related statistics and plots were performed using the R package Igraph.

6.3. Results

6.3.1. Functional characterization of the HIV-associated metagenome

In order to characterize the HIV+ functional dysbiosis, we sequenced an average 98 Mb per sample (total 9.5 Gb) using a combination of Illumina and 454 sequencing technologies.

The unsupervised PAM clustering analysis represented in **Figure 6.1a** showed a cluster configuration in which most of the HIV+ patients are separated by important changes in their gene composition from the HIV- subjects (ADONIS test, p-value=0.001). This differential composition was also found when the HIV- group was compared against each of the three different HIV+ groups (ADONIS test, p-value=0.001) (**Figure 6.1b**).

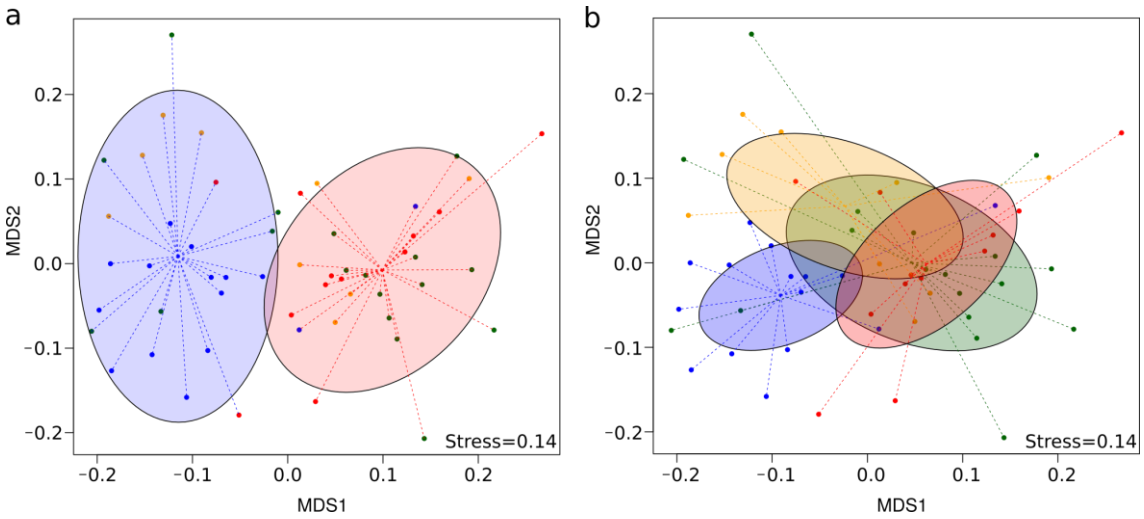


Figure 6.1 Comparison of the microbiota gene composition between HIV infected and uninfected subjects. NMDS analysis of the KEGG Orthology (KO) gene composition retrieved from the metagenomic sequences. The VU (red), IR (green), INR (orange) and HIV- (blue) subjects are represented by the ellipses which represent 70% of the samples. (**a**) Optimal clustering configuration retrieved from the PAM algorithm, red ellipses represent a cluster mainly composed of HIV infected subjects while the blue ellipse represents the cluster mainly conformed by HIV- subjects, the cluster configuration was validated using the ADONIS test (p-value=0.001). (**b**) Group cluster configuration from the four groups of the cohort, the VU (red), IR (green), INR (orange) and HIV- (blue) subjects. The cluster configuration was validated using the ADONIS test by the four groups of the cohort (p-value=0.001). INR, immunological non-responder; IR, immunological responder; VU, viremic untreated.

The diversity of the gene composition, measured by the Shannon index, showed that the VU group presented the lowest gene diversity significantly different to those of healthy group. No significant differences were found between HIV+ on ART and HIV- subjects (**Figure S12.3.2a**, Appendix section). However, the functional homogeneity analysis within groups showed VU and HIV- groups presented higher homogeneity than HIV+ on ART individuals (**Figure S12.3.2b**, Appendix section).

To address the gene and pathways that consistently explain the differences between the groups, we applied the linear discriminant analysis effect size (LEfSe) method. We found 34 KEGG pathways (ko) and 186 KEGG orthology groups (KO) were statistically different between the HIV-infected patients and HIV- subjects (**Figure 6.2 and S12.3.3**, Appendix section). HIV+ bacterial communities presented an increase of pathways related to lipopolysaccharide (LPS) biosynthesis and peptidoglycan biosynthesis, in keeping with our previous findings (Chapter 1). Two pathways, alanine, aspartate and glutamate metabolism and zeatin biosynthesis, showed a significantly higher abundance in the HIV+ condition, suggesting that the HIV-associated microbiota might be adapted to mitigate a pro-oxidative environment. Also, the HIV+ bacterial communities presented an increase of pathways related to infectious diseases (ko05111, ko05120) and the metabolism of amino acids (ko00473, ko00400). On the other hand, this altered microbiota is depleted of pathways related to the signal transduction and membrane transport (**Figure 6.2**). Surprisingly, the HIV- subjects possessed a microbiota enriched in genes related to the antibiotic beta lactam resistance pathway (**Figure 6.2**). However, when we compared the total number of antibiotic resistance annotations retrieved from the Comprehensive Antibiotic Resistance Database (CARDB), we did not find statistical differences between groups (Kruskal-Wallis test, p-value > 0.05).

The differences were also stated for the controls versus each HIV+ groups. Thus, the VU group showed the greatest number of different enriched KOs (62 VU, 28 IR, 27 INR) but we found a core of metabolic pathways associated with HIV infection with high abundance of KOs related to the resistance to the oxidative stress (**Figure S12.3.3b, c, d**, Appendix section).

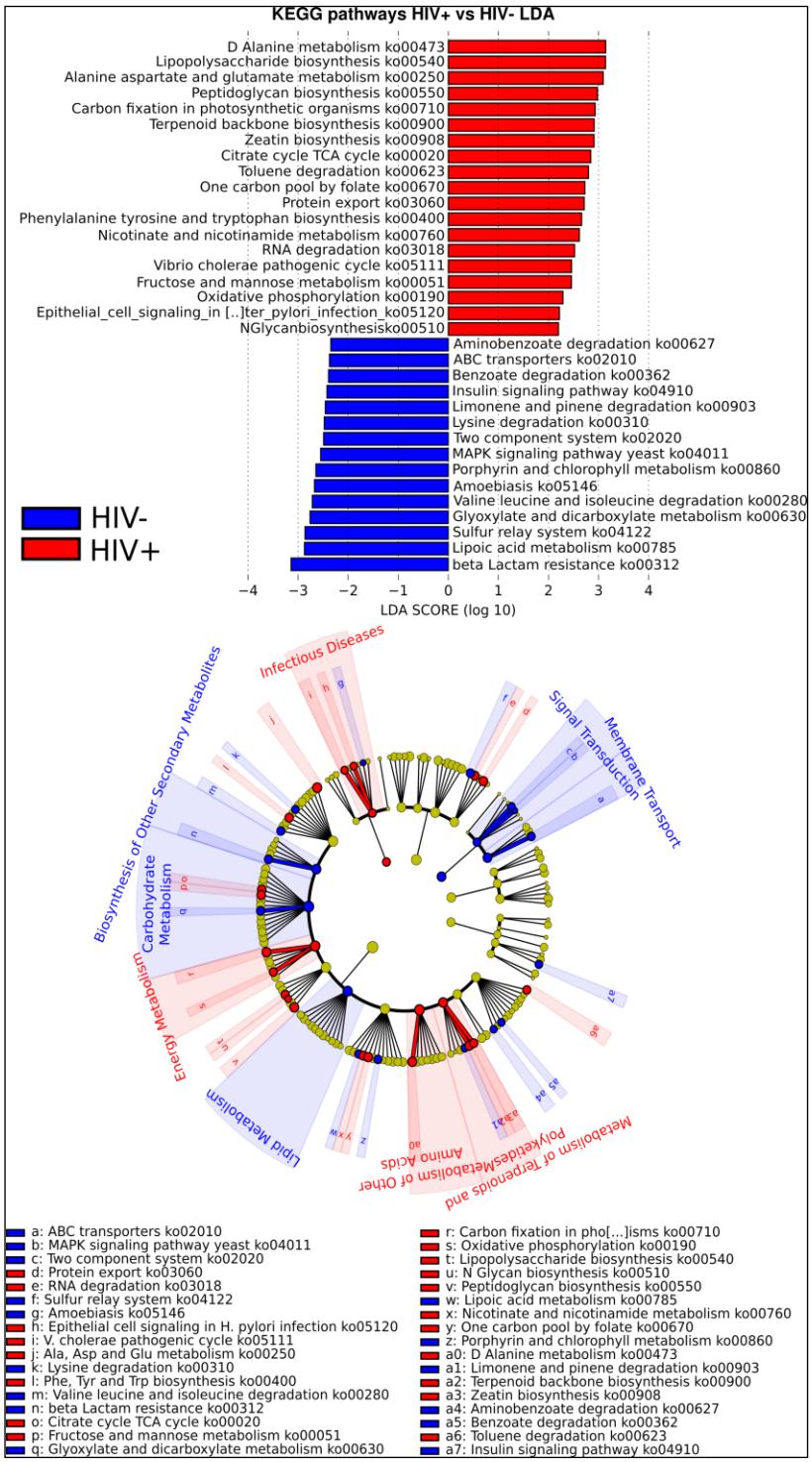


Figure 6.2 Metagenomic pathways biomarkers. Linear discriminative analysis (LDA) effect size LEfSe analysis between the HIV+ (in red) and HIV- (in blue) subjects. LDA scores (log 10) for the most discriminative pathways in HIV- are represented on the negative scale, whereas LDA-positive scores indicate enriched pathways in HIV +subjects. The cladogram represents the biomarkers of the upper hierarchical classes within the KEGG database.

6.3.2. Bacteria involved in the dysbiotic metabolism

The taxonomic annotation from the metagenomes was performed by mapping all the reads against Genomeclust database. The HIV- group segregated independently from the HIV+ patients, as could be seen in the cluster analysis represented in **Figure S12.3.4a** (Appendix section). The differences of the HIV-infected patients were given by an increase of bacterial species, mainly, from the genera *Prevotella Acidaminococcus* and *Streptococcus* and a decrease of commensal species such as *Bacteroides, Bifidobacterium, Akkermansia, Odoribacter* and *Alistipes* (**Figure S12.3.4b,** Appendix section). Strikingly, *Faecalibacterium prausnitzii* was not detected as a biomarker of the HIV- subjects as was described previously (Chapter 2). This result could be because of the reference sequences for this species were just a draft genome leading to a miss-annotation. We also found in the controls different members of Enterobacteriaceae family due to these species are overrepresented in the database used.

The alpha diversity, measured by the Shannon index and Chao1 richness estimators and the expected number of species showed no statistical differences between the four groups of the cohort (**Figure S12.3.4c**, Appendix section). The analysis of the bacterial composition homogeneity within each group showed that the HIV- group possessed the most homogeneous composition in the cohort and that IR group was the one with the greatest taxonomic homogeneity in HIV+ groups (**Figure S12.3.4d**, Appendix section).

In order to relate the bacterial taxa to the metabolic pathways that lead the functional dysbiosis, we performed a Generalized Linear Model (GLM) by setting the pathway biomarkers as the response variable and the species matrix as the predictors. We found a significant positive correlation between the respective bacterial taxa and pathway biomarkers for both groups, HIV+ and HIV- (**Table S12.3.3**, Appendix section). To determine whether the species biomarkers carried the genes involved in the pathways or if they promoted the growth of other members of the microbiota that was responsible for the metabolic functions, we mapped all the reads of each pathway biomarkers against the reference genomes. Then, we found that both HIV and healthy microbiota presented the genes involved in their respective metabolic pathways (**Table S12.3.4**, Appendix section). Thus, we detected that the dysbiotic metabolism could be mostly performed by species of *Prevotella* genus. Other Gram-negative bacteria belonging to *Acidaminococcus*, *Dialister*, *Fermentomonas* and *Bibersteinia* genera also presented genes for the different metabolic pathways including LPS biosynthesis pathway (ko00540).

6.3.3. Metatranscriptome of the HIV-associated microbiota

We explored the gene expression of the gut-microbiota in a subset of samples (8 VU, 13 IR, 6 INR and 14 HIV-) that did not differ in their general characteristics from the whole sample. The total mRNA was purified and the cDNA sequenced, obtaining an average 16 Mb per sample (a total of 715 Mb).

We observed that 964 KOs from the metagenome (DNA-KO) were not present in the metatranscriptomic data while 157 KOs from the metatranscriptome (RNA-KO) were missing in the metagenome. This difference could be due to a transcriptional regulation of the gene, but it could not be discarded the low sequencing coverage. To address these differences, we calculated the Bray-Curtis dissimilarity and the evenness indexes for the HIV+ and HIV- groups as described in [455]. We found, in both groups, that RNA-KO composition between individuals was significantly more variable than DNA-KO composition (**Figure S12.3.5a**, Appendix section) but the evenness index was similar for both functional profiles (**Figure S12.3.5b**, Appendix section). Thus, the differences observed could be explained by a transcriptional regulation of the metagenome based on the environmental requirements. Additionally, at the transcriptional level (RNA-KO), the dissimilarity between subjects and the evenness for HIV- group was significantly higher than those for HIV+ group. These results suggested that the similar gene expression profile of the microbiota found in HIV+ subjects would be a reflection of the functional adaptation to the HIV environment. Moreover, unique metabolic functions would be preferentially expressed reducing the evenness of this group.

The clustering analysis showed the metatranscriptomes, RNA-KO, were statistically different between the HIV+ individuals and the controls (ADONIS test, p-value=0.006) and across all the groups (ADONIS test, p-value=0.001) (**Figure 6.3a**). In the LEfSe analysis (**Figure 6.3b**), we observed that HIV-associated microbiota showed a significantly higher abundance of transcripts related to the infection diseases (ko05111) and to stress response (ko04141). Also, we found enrichment in pathways related to the metabolism of cofactors and vitamins (ko00730, ko00053), biosynthesis of other secondary metabolites (ko00521), folding sorting and degradation of proteins (ko03018) and those involved in glycolysis (ko00010). On the other hand, the dysbiotic microbiota showed a depletion of pathways related to the signaling adaptation (ko04910, ko04626), protein transporters (ko02010, ko02020) as well as pathways related to the metabolism of amino acids (ko00280, ko00300, and ko00260). Furthermore, HIV-associated microbiota showed a significantly

lower abundance of transcripts related to anti-inflammatory metabolic processes such as butanoate metabolism (ko00650), propanoate metabolism (ko00640) or fatty acid metabolism (ko00071). These results showed a good congruence with the low abundance of Short Chain Fatty Acid (SCFA) in HIV-associated microbiota described in Chapter 2.

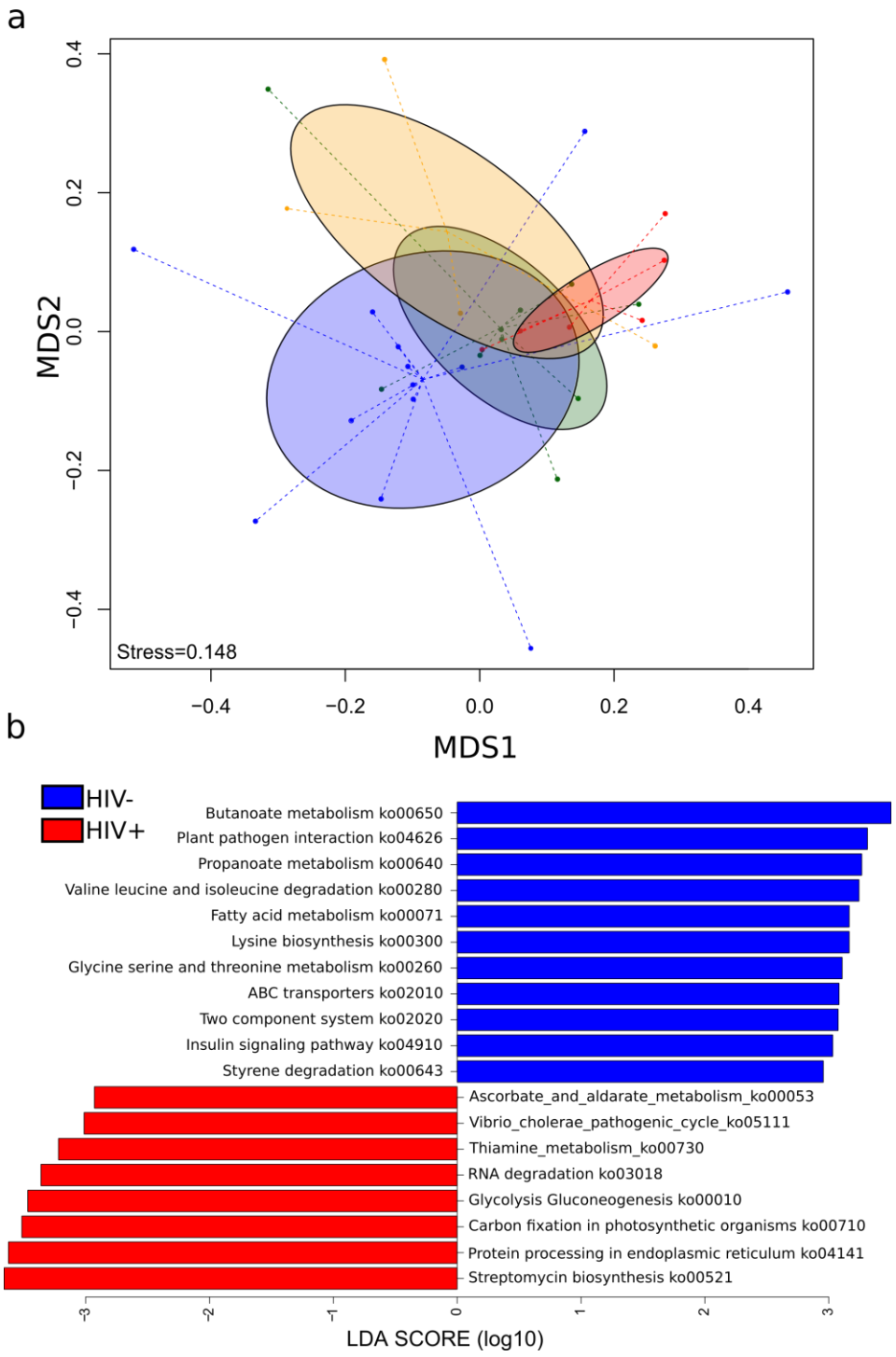


Figure 6.3 Comparison of the microbiota gene expression between HIV infected and uninfected subjects. (**a**) NMDS analysis of the KEGG Orthology (KO) gene expression retrieved from the metatranscriptomic sequences. The VU (red), IR (green), INR

(orange) and HIV- (blue) subjects are represented within the ellipses which represent 70% of the samples; the group's cluster configuration was validated using the ADONIS test (p-value=0.001). (**b**) The linear discriminative analysis (LDA) effect size LEfSe) analysis between the HIV+ (in red) and HIV- (in blue) subjects. LDA scores (log 10) for the most discriminative pathways in controls are represented on the positive scale, whereas LDA-negative scores indicate enriched pathways in HIV+ individuals. INR, immunological non-responder; IR, immunological responder; VU, viremic untreated.

We determined that the differentially expressed genes were 49.08 % and 40.61% in HIV+ and HIV- groups, being overexpressed 13.89% and 11.90 %, respectively. **Figure 6.4** showed the heatmap of those metagenomic and metatranscriptomic biomarkers that were significantly over- or under-expressed (T-student test for log RNA/DNA abundance ratio different from zero, adjusted p-values < 0.05) and also presented a significant difference of the expression level between HIV+ and HIV- groups (Wilcoxon test, adjusted p-value ≤ 0.05). We found that in HIV+ individuals while the anti-inflammatory metabolic processes such as propanoate (ko00640) and butanoate (ko00650) pathways were under expressed, the genes related to stress resistance mechanisms (ko00730, ko00521 and ko4141) were overexpressed. Thus, the transcriptional profile indicated a pro-inflammatory microbiota with a preferential expression of those metabolic pathways that attenuated the oxidative stress caused by the inflammation.

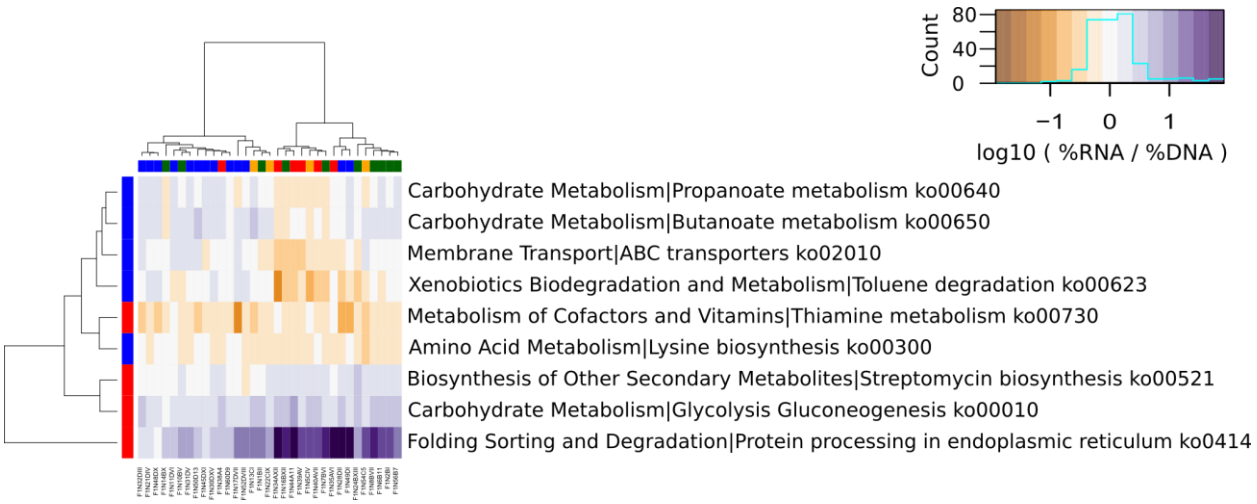


Figure 6.4 Heatmap of the relative gene expression of the metatranscriptomic biomarkers. Hierarchical clustering representation of the biomarker gene expression ratio (|log10 RNA/DNA| >0). Only those significantly over/infra expressed pathways (BH corrected p-value <= 0.05) and that were statistically differentially expressed between HIV+ (red) and controls (blue) subjects (Wilcoxon test BH corrected p-value <= 0.05) were included in the analysis. In red is represented those biomarkers that were over represented in the HIV+ subjects, meanwhile in blue the ones that were represented in the controls. The VU (red), IR (green), INR (orange) and HIV- (blue) subjects were represented as tips of the column's cladogram. The brown to purple gradient represent the relative gene expression level.

6.3.4. The active microbiota is distinct in HIV-infected individuals

The metatranscriptomes were mapped against the Genomeclust database in order to obtain a taxonomic assignation for the active microbiota. We observed that the dominant phyla were Firmicutes and Bacteroidetes for both the RNA and the DNA datasets, but the relative abundance of Firmicutes phylum was higher in the active microbiota (mean 77.68 ± 14.27% Firmicutes and 17 ± 12.17% Bacteroidetes) (**Figure 6.5**). The members of the Firmicutes phylum were commensal genera of the gut microbiota as *Ruminococcus*, *Eubacterium*, *Streptococcus*, *Faecalibacterium*, *Blautia*, *Leuconostoc*, *Anaerostipes*, *Blautia* or *Roseburia* (**Figure S12.3.6,** Appendix section). In the active microbiota, we also found an archaeon species belonging to the genus *Methanobrevibacter* (**Figure 6.5**). However, the Enterobacteriaceae family, found with the metagenomic assignment, was not detected as an active member.

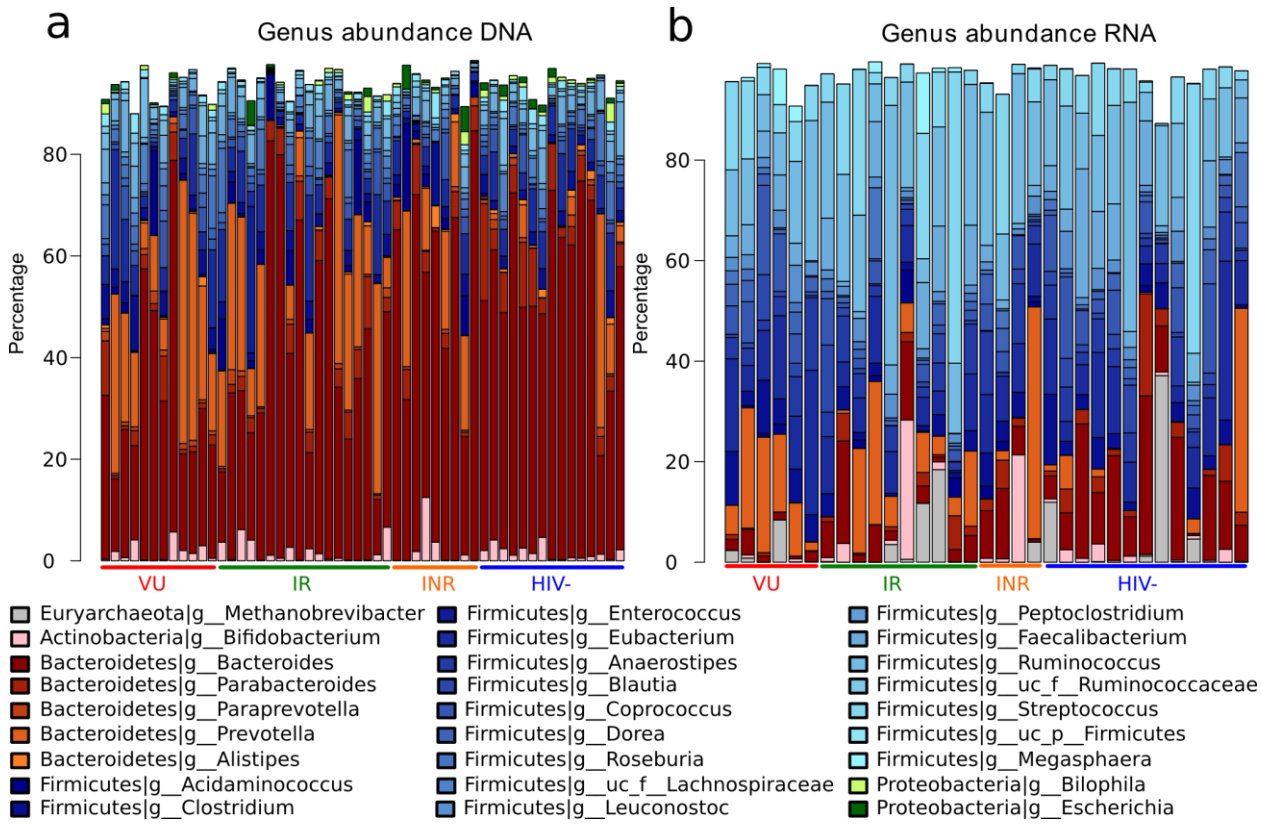


Figure 6.5 Taxonomic abundance. Relative abundance of the most abundant genera from the metagenomic data (mean relative abundance > 0.05 and be present in at least 70% of the samples) (**a**) and the metatranscriptomic data (**b**) for the VU (red), IR (green), INR (orange) and HIV- (blue) subjects. In gray are represented the Euryarchaeota, in pink the Actinobacteria, in brown gradient the Bacteroidetes, in blue gradient the Firmicutes and in green gradient the Proteobacteria.

Moreover, the species profile of the HIV+ subjects was statistically different from the HIV- individuals (ADONIS test, p-value=0.005), being the genera *Prevotella* (LEfSe p-value=0.007),

Acidaminococcus (p-value=0.017), *Coprobacillus* (p-value=0.01) and *Streptococcus* (p-value=0.037) differentially increased in the active HIV-associated microbiota. On the other hand, *Bacteroides, Parabacteroides, Odoribacter,* and *Alistipes* appeared as depleted genera in the active dysbiotic bacterial community.

6.3.5. Microbial metabolism of dietary tryptophan and choline

The degradation of the essential amino acid, tryptophan, in mammals is mainly carried out through the aerobic kynurenine pathway but in prokaryotes, it has only been described in few aerobic bacteria. In order to find a group of bacteria capable of metabolizing tryptophan via the kynurenine pathway, we conducted a sensitive search of remote gene homologs and species that could contain this pathway in the metagenome and metatranscriptome data. We found, in very low abundance, the gene and transcripts of three of the five steps of the metabolic route (**Figure 6.6a**). However, neither the gene content nor the transcript abundance was statistically different among the groups (Kruskal-Wallis test, p-value > 0.1). The missing enzymes were those that catabolize oxygen-dependent reactions such as the enzymes of the first step, indoleamine 2,3-dioxygenase (IDO1) or the tryptophan 2,3-dioxygenase and the enzyme, the 3-hydroxyanthranilate 3,4-dioxygenase, that catabolize the 3-hydroxyanthranilate in amino-3-carboxymuconate semi aldehyde. Neither genes nor expression was found for the degradation of the 3-hydroxyanthranilate. Thus, this metabolite could only be transformed in 3-methoxyanthranilate by a methylation reaction catalyzed by a wide-range of bacterial or human methyltransferases (**Figure 6a**). As well, the species that were able to perform the degradation of the tryptophan via the kynurenine pathway were a minor bacterial group belonging mainly to Proteobacteria phylum (**Table S12.3.5,** Appendix section). More importantly, none of these species correlates significantly with the Kynurenine/Tryptophan ratio (**Table S12.3.5,** Appendix section).

The anaerobic gut environment would favor the non-oxidative degradation of tryptophan in indole and derivatives via bacterial tryptophanase. These catabolites play a crucial role in gut immunity and in mucosa homeostasis via aryl hydrocarbon receptor (AhR) promoting anti-inflammatory processes [149]. We have investigated the gene and transcripts content involved in the tryptophan fermentation in HIV+ and HIV- microbiota (**Figure 6.6b**). We found the abundance of transcription products was 100 times higher than those of the genes in both groups, indicating an overexpression of this pathway. Intriguing, the expression level of the tryptophanase gene was higher, although not significant, in HIV- subjects than in HIV+ (HIV- 2.56/0.78 vs HIV+ 2.212/0.741, p-value=0.2).

These results suggested an impaired expression of the tryptophanase gene would have repercussions in the gut inflammation.

The trimethylamine (TMA) is a bacterial metabolite of the choline fermentation via choline TMA-lyase, that is converted in the liver to trimethylamine-N-oxide (TMAO), being both metabolites related to cardiovascular disease and atherosclerosis in both the general population [325] and HIV-infected subjects [459]. We looked for the genes and transcripts of the choline catabolism in the metagenomes and metatranscriptomes of HIV- and HIV + individuals (**Figure 6.6c**). We did not find significant differences among the groups for *cutC* gene (Choline TMA Lyase) (p-value=0.65) and *cutD* gene (Choline TMA−Lyase activating protein) (p-value=0.55) from the metagenome and metatranscriptome data set, although both genes were overexpressed in all the individuals. We found few gene homologous that catabolize the transformation of L-carnitine into TMA.

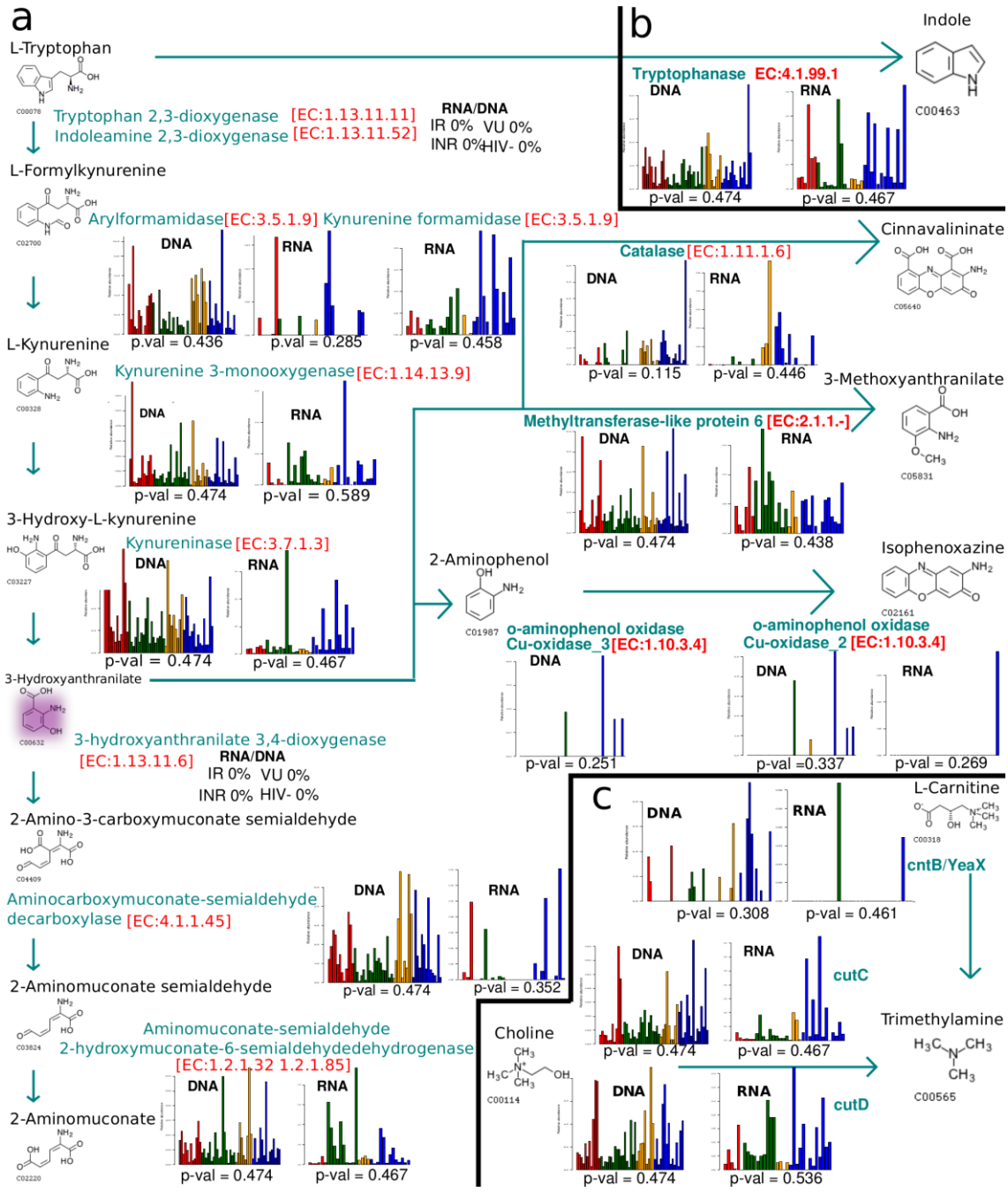


Figure 6.6 Metabolism of dietary tryptophan and choline. Remote homology detection (hmmsearch e-value < 0.001 and Blastp alignment coverage >= 0.7 and e-value < 0.00001) from the (**a**) metabolic routes of the IDO1 tryptophan catabolism, (**b**) the tryptophan fermentation into indole and (**c**) the fermentation of L-carnitine and choline to TMA. In cyan are represented the enzymes and the reaction flow arrows from one metabolite to the next one in the metabolic pathway. The bar plots represented the relative abundance of the gene and transcripts for each of the enzymes within each step from the VU (red), IR (green), INR (orange) and HIV- (blue) subjects. The statistical test between the relative abundance of the four groups of the cohort was carried by the Kruskal-Wallis test. In pink is represented the metabolite 3-Hydroxyanthranilate which has been previously reported in the work of Vujkovic [285] and Serrano-Villar [405].

6.3.6. Ecological and functional networks of the dysbiotic bacterial community

In order to study the ecological interactions in dysbiotic bacterial community, we obtained a co-occurrence network, calculating a pairwise correlation matrix, using the Sparcc tool (p-value < 0.001 and |rho| > 0.1), for those species whose presence in the cohort were at least of the 70% to avoid miss-correlations given a high proportion of zero values [460].The resulting adjacency matrix was transformed into a network (see Materials and Methods) in which the nodes correspond to species and the edges to positive or negative correlations between the species (**Figure 6.7a**). We first confirmed that the HIV+ network presented the properties of the biological network for the connectivity distribution and the small-world effect. The connectivity distribution represents the distribution of edge per node and in a biological network fits a power law distribution, meaning that few nodes presented a high number of interactions while most nodes possessed sparse connections (Kolmogorov Smirnov p-value > 0.1, **Table S12.3.6,** Appendix section). The small world effect was reflected in the low shortest path mean (4.07) (**Table S12.3.6,** Appendix section) indicating that the network presented a high number of clusters with shortcuts between them. We also found a high connectivity degree (average of the numbers of edges per species) but the negative interactions were more abundant than the co-occurrence relationships (43% positive vs 57% negative connections).

Modularity is another important characteristic of the biological networks. The modules or clusters represent link-dense areas separated by regions of low connectivity. Species within a module are more tightly linked than they are to species in other modules. The modularity coefficient of HIV+ network was 0.428 which implies a high degree of modularity. We used the "walktrap community" algorithm from the Igraph R package to define modules into the total network. We determined 20 modules that contained at least three species each but the module size arrived at 34 (**Figure 6.7b**). The modules were dominated by Firmicutes and Bacteroidetes phyla and to a lesser extent by Actinobacteria and Proteobacteria. Interestingly in the modules in which there are members of the Actinobacteria phylum, there are not often species of the Proteobacteria, indicating competence between these two phyla.

The betweenness centrality of a node or species is the measure of its relevance in the community structure allowing identifying the highest connected ones that were referred to as hub-species. Besides different commensal bacteria as *Bacteroides* or *Eubacterium*, *Prevotella copri,* HIV+

biomarker, was stated as important hubs of the HIV+ community (**Figure S12.3.7a,** Appendix section). We also found *Myroides odoratimimus*, a bacterium that causes infection in immunocompromised patients [461–463], as a hub-species. Complementary to the betweenness centrality, the eigenvector centrality measures the importance of a node based on the closeness to highly connected ones. We found the eigenvector centrality coefficient revealed more hub-species (**Figure S12.3.7b,** Appendix section) than the ones observed with the betweenness centrality because there are bacteria that possessed direct edges with the nodes highly connected. Thus, the eigenvector centrality hub-species encompassed all the hub-species retrieved from the betweenness centrality and included several species related to the SCFA producers and others commensal species of the human microbiome (**Figure S12.3.7b,** Appendix section). Moreover, we found a significant positive correlation between the hub-species and the LDA score of the biomarkers (**Figure S12.3.8,** Appendix section).

We also applied a network approach to the bacterial metabolism of the HIV dysbiotic microbiota. Thus, we took all the metabolic pathways for the HIV+ groups predicted by the HUMAnN pipeline [440] and with the enzymatic reactions we estimated a directed enzymatic metabolic network in which nodes (3700 nodes) represented enzymes and those enzymes that catalyze successive reactions (product/substrate) were connected by directed edges (31725 edges) (**Figure S12.3.9,** Appendix section). We found the enzymatic network possessed the biological network properties such as the small world effect and the node power law distribution (**Table S12.3.6,** Appendix section). Moreover, it showed a high degree of compartmentalization reflected into the high modularity value and into the fragmentation index as it has been found in others metabolic network studies [4, 5] (**Table S12.3.6,** Appendix section). To assess the relevance of the enriched and depleted KO biomarkers (**Figure S12.3.3,** Appendix section) found for the HIV+ microbiota, we located them in the enzymatic metabolic network. We found their degree and eigenvector centrality indexes were significantly higher than those of the rest of enzymes within the network (**Figure S12.3.10,** Appendix section).

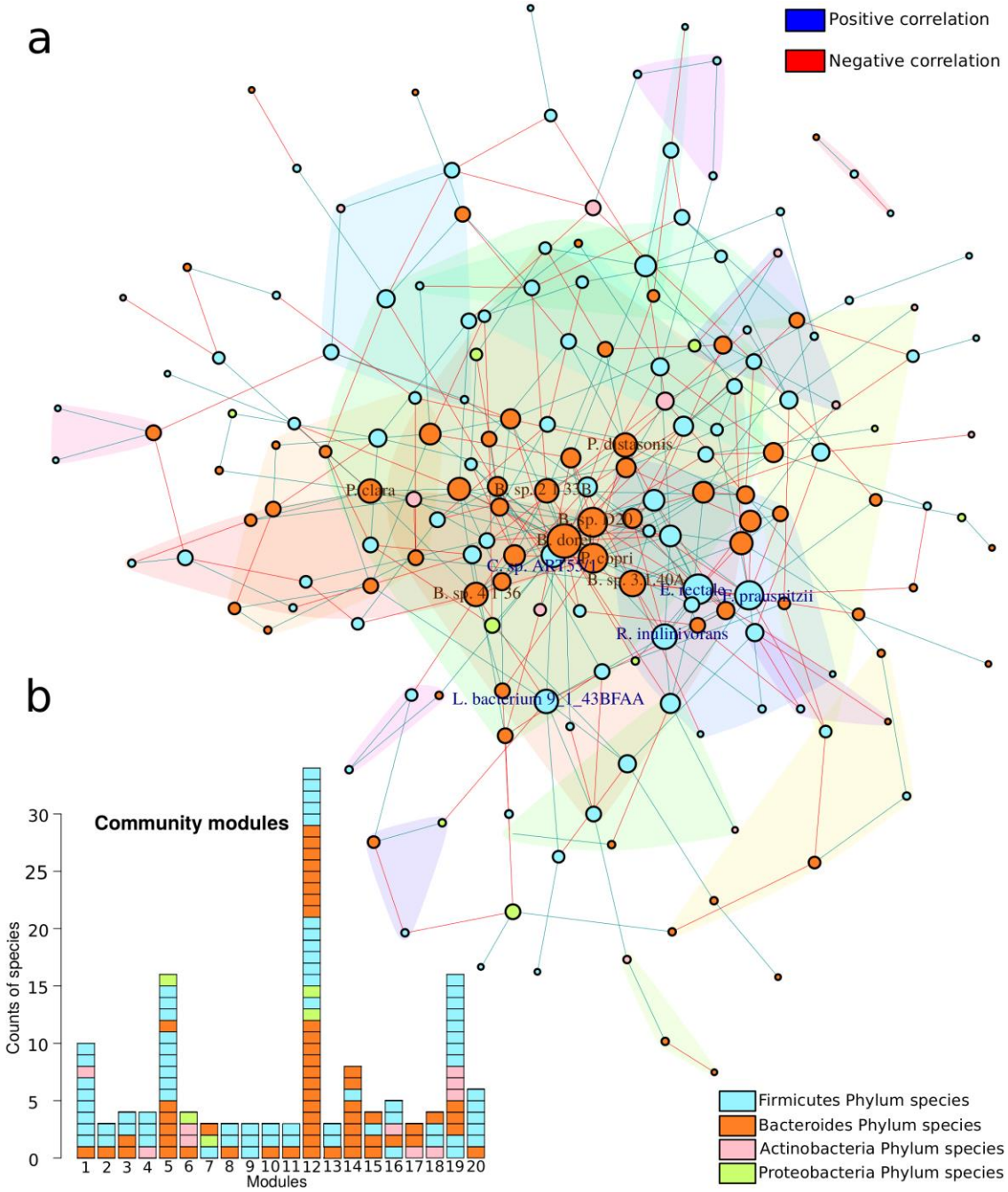


Figure 6.7 Co-occurrence ecologic network. (**a**) Co-occurrence network inferred from the correlation matrix obtained using the SparCC algorithm (two-sided pseudo p-value ≤ 0.001 based on bootstrapping of 1000 repetitions < 0.01 and a coef > 0.1) with those species that were present in at least 70% of the samples. The nodes represent species belonging to the Firmicutes (cyan), Bacteroides (orange), Actinobacteria (pink) and Proteobacteria (green) phylum. The size of the node was scaled to the logarithm of its degree centrality and only the nodes belonging to the quantile 95 were labeled. The edges represent the positive correlations (blue) and the negative correlations (red) for each pair of species. The color polygons represent one of the cluster communities defined by the walktrap community algorithm. (**b**) Barplot representation of the phylum of each species for the communities defined by the walktrap community algorithm.

6.3.7. HIV-associated dysbiosis and host health: Bayesian network

In order to explain the overall effect of the HIV-associated dysbiosis into the deterioration of the host health, we integrate the data from metagenomics (DNAseq), metatranscriptomics (RNAseq), metabolomics [376] and clinical variables of the HIV+ subjects into a Bayesian Network (BN).

Briefly, BN is a graphic probabilistic model in which the nodes correspond to random variables and the arcs represent conditional dependencies [341]. The network contained 190 nodes and 548 edges (**Figure S12.3.11**, Appendix section) being the clinical variables and to a lesser extent the pathways retrieved from the metagenomic data those that showed a higher degree connectivity with the rest of the nodes (**Figure S12.3.12a**, Appendix section). The adaptive immunity markers, CD4+ T-cells and the %CD4+HLADR+CD38+ T-cells, the bacterial translocation marker, BPI, and the thymic function showed a high degree of connectivity (**Table S12.3.7**, Appendix section). The metabolic pathways that presented a high degree of direct links (**Table S12.3.7**, Appendix section) were pathways related to xenobiotics biodegradation and metabolism category as terpenoid backbone biosynthesis (ko00900) and aminobenzoate degradation (ko00627). Other pathways related to the resistance to the oxidative stress as nicotinate and nicotinamide metabolism (ko00760) and alanine aspartate and glutamate metabolism (ko00250) showed a high connectivity degree. Moreover, the expression of the metabolic pathway related to the butanoate metabolism (ko00650) (hereafter butanoate metabolism-RNAseq) also shows a high number of direct links. Complementary to the direct degree of connectivity of a node, the betweenness centrality was used to estimate the relevance of a node within the network. The betweenness centrality analysis showed that most of the central nodes (quantile 95) of the BN are related to *Prevotella* species such as *P. sp. Oral taxon 299, P. melaninogenica* and *P. salivae,* to metabolic pathways related to the amino acid metabolism category (ko00400, ko00280) and to the resistance to the oxidative stress (ko00250).

An interesting propriety of BN is that it could be dissected into MB. MB of a node contains all the variables needed to predict the behavior of that node. In order to identify the nodes that are related to a large number of variables, we estimate the MB size of all the variables within the network. We found that the MB of the nodes corresponding to metabolic pathways obtained from the DNAseq and RNAseq datasets have the higher size (**Figure S12.3.12b**, Appendix section) with a considerable number of clinical variables (**Figure S12.3.12c**, Appendix section). The MB of the butanoate metabolism-RNAseq (hereafter butanoate MB) appeared as the one which included a greater number of nodes, bearing 89 nodes of which 16 were clinical variables. Thus, to assess the

immune response due to the HIV-associated dysbiosis, we deeply analyzed the butanoate MB (**Figure 6.8**). We observed the butanoate metabolism-RNAseq possessed direct links with systemic markers related to inflammation (hs-CRP), bacterial translocation (BPI), endothelial dysfunction (ADMA) and coagulation (D-dimers). We also detected significant negative correlations with Nadir CD4+ T-cells counts but significant positive correlations with other metabolic pathways, retrieved from the RNAseq, related to the propanoate (ko00640) and fatty acids metabolism (ko00071). We found that *P. copri*, a central node, correlated positively in the HIV-infected subjects with important HIV+ associated pathways related to resistance to oxidative stress (ko00250, ko00900) and the metabolism of amino acids (ko00400, ko00473) and correlated negatively with the lipoic acid metabolism (ko00785) which is an antioxidant pathway. The CD4+ T-cell counts showed high connectivity but most of the links were with different types of unknown metabolites. The bile salt, muricholic acid, that was overrepresented in the HIV subjects showed positive correlations with the CD4+ T-cell counts. We also found that the oleanane triterpene-related metabolite and different membrane-structural lipids showed significant negative correlations with the immune activation markers, CD8+HLA−DR+CD38+ T−cell counts, and %CD4+HLA−DR+CD38+ T−cell counts. Interestingly the IR group was the one that possesses the highest abundance of such metabolites (Kruskal-Wallis p-value=0).

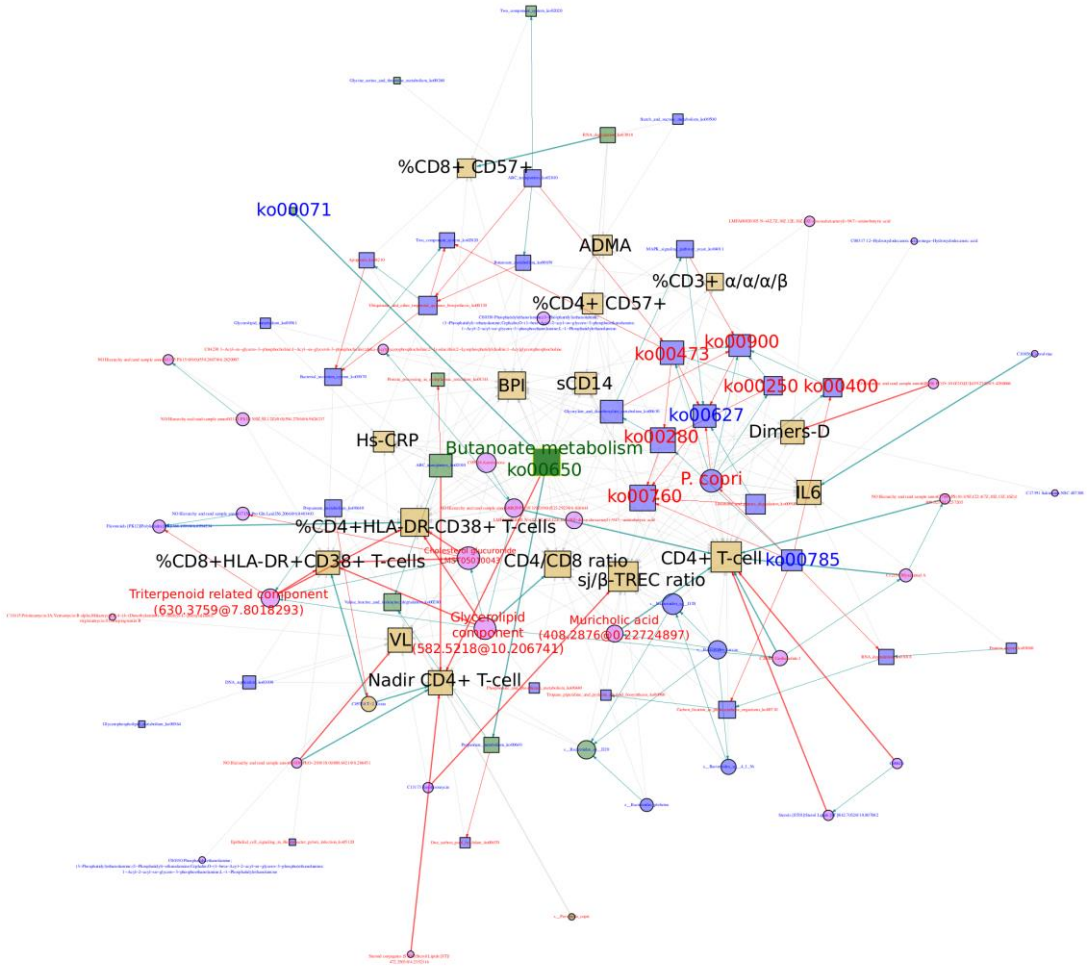


Figure 6.8 Butanoate Markov Blanket. A subgraph of the "multiomic" BN composed by the metagenomic (blue nodes), metatranscriptomic (green nodes), metabolomic (pink circle nodes) and the clinical variables (golden square nodes) from the HIV+ subjects. The data from the metagenomic and metatranscriptomic include the information of the relative abundance of the species (circles) and the pathway (squares). The labels of the nodes represent if the variable is overrepresented in the HIV- subjects (blue labels) or the HIV+ subjects (red label). Arrows indicate conditional dependencies between variables. The Spearman correlation coefficient is represented by the arrow's color, blue for a significant positive (PH adjusted p-value < 0.1), red for a significantly negative correlation (PH adjusted p-value < 0.1) and gray for a non-significant correlation.

6.4. Discussion

The HIV-associated dysbiotic microbiota has been widely studied at the structural level and related to the human health by identifying associations between bacterial species and systemic markers of innate and adaptive immunity [280,281,284,285,287,288,290,291]**.** However, the knowledge on the metabolic shifts associated with the altered microbiota and its effects on the human health is scarce. To gain insight into the functional consequences of HIV-associated dysbiosis, we determined the gene content by metagenomics and its expression level by metatranscriptomics of the gut microbiome in HIV patients.

The HIV+ microbial community presented differential gene content with unique metabolic functions in congruence with our previous metagenomic study based on 454 shot-gun sequencing (Chapter 1) and with the recently published results derived from the metabolome and proteome data [376,405]. The deleterious effects of the HIV infection on the immune system and gut mucosa result in an inflammatory environment with the production of highly reactive oxygen metabolites. Thus, the dysbiotic microbiota presented a significant high abundance of the pathways and metabolites related to the resistance to the oxidative stress. On the other hand, as the HIV-associated microbiota is characterized by a large proportion of Gram-negative bacteria (as shown in Chapters 1,2 and [280,281,283,391,464]), an enrichment in metabolic functions as LPS biosynthesis was found in the metagenomic analysis. Also, the genes involved in pathogenesis processes were more abundant in the altered microbiota as occurred in another inflammatory disease [105]. However, the HIV infection produced a decrease in the functional diversity respect to the health status but no change in the diversity parameters has been detected for the microbiota composition. Besides the characteristic imbalance between *Prevotella* and *Bacteroides* genera of the HIV-associated microbiota, the species biomarkers for the dysbiosis, mainly *Prevotella* and *Acidominococcus* species, presented the genetic potential for the dysbiotic metabolism.

A deeper understanding of the functional changes in the microbiome in response to HIV infection requires a metatranscriptomic analysis to gain information on the gene expression of the metabolic pathways. Thus, the homogenous gene content observed in the metagenomic dataset was subjected to a transcriptional regulation giving a more variable functional profile. Curiously, in both group, most of the bacteria expressing genes belong to Firmicutes phylum, indicating that the active genera (*Streptococcus*, *Leuconostoc*, *Anaerostipes*, and *Blautia*), although no abundant, would play an

important role in the metabolism. Moreover, *Methanobrevibacter smithii*, a low abundant archaeon, emerged as an active member of the microbiota in agreement with a previous study [407]. *M. smithii* is the predominant archaeon in the human gut and plays an important role in the efficient digestion of polysaccharides by consuming the residual hydrogen derived from bacterial fermentation. The HIV infection was also reflected in the gene expression profile which was shaped forward the metabolic functions allowing the adaptation to the inflammatory environment, being *Prevotella*, *Acidaminococcus*, *Coprobacillus* and *Streptococcus* the transcriptional active genera of the HIV-associated microbiota. Although the dysbiotic microbiota presented a high abundance of genes related to the lipopolysaccharide biosynthesis, no differential expression was found. LPS constitutes the external part of the outer membrane of the Gram-negative bacteria and the gene expression of the LPS biosynthesis pathway would respond to bacterial growth rather than to the inflammatory environment. But, LPS as a microbe-associated molecular pattern (MAMP) can be recognized by Toll-like receptor 4 (TLR4) of different immune cells that promote the inflammatory response. Then, the inflammation induced by HIV infection could be enhanced by the high number of pro-inflammatory bacteria. A similar scenario has been described in rheumatoid arthritis by Scher et al., (2013) [106]. Although no clear pro-inflammatory responses of *P. copri* have been found [282], this bacteria has been associated with inflammation in different studies [280,281,283,288,391,464,465]. In turn, this dysbiotic microbiota overexpressed genes related to the resistance to oxidative stress gaining a selective advantage. Another important factor in the maintenance of the inflammation would be the depletion of anti-inflammatory bacteria as *Bacteroides* [281] and the decrease of those species related to SCFA biosynthesis (Chapter 2). Striking, the dysbiotic metagenome presented the genes involved in butanoate and propanoate metabolism but they were under-expressed in HIV-associated microbiota.

The IDO1 involved in the tryptophan catabolism through kynurenine pathway has been correlated with epithelial barrier disruption and bacterial translocation in HIV infection [209,279]. Vujjkovick-Cvijin et al., (2013) [285] suggested, based on an indirect approach, an important role of *Pseudomonas* genus promoting the immunosuppressive response by metabolizing the 3-hydroxyanthranilic acid through the kynurenine pathway. Nevertheless, we found neither IDO1 gene nor its expression in the bacterial metagenome or the metatranscriptome from HIV-infected individuals. However, Serrano-Villar et al., (2016) [405] recently described a statistical increase in the abundance of the 3-hydroxyanthranilic acid in the gut metabolome of HIV patients. Taking into account that the human IDO1 is up regulated in HIV infection [279] a metabolic complementation

of the bacterial pathway could be occurred giving the accumulation of 3-hydroxyanthranilic acid. Also, a microbial-mammalian co-metabolism has been described for the choline fermentation [466]. A recent study on macaques infected by SIV showed an association between the depletion of *Lactobacillus* genus and the increase of IDO1 activity [467]. In congruence with others works [280,281,284,285,291,353,405,406], we did not find a shift of this genus in HIV-associated microbiota but the observed enrichment of the active *Streptococcus* genus in the dysbiotic microbiota could be the responsible for IDO1 induction as occurred in rats [468]. Further studies will be needed to address specific host-microbe interactions and which mechanisms are involved.

In anaerobiosis, the gut microbiota metabolized the tryptophan to indole and derivates as indole-3-acetic acid (IAA) or indole-3-aldehyde (I3A). These metabolites play a beneficial role for the human health activating the aryl hydrocarbon receptor (AhR) and hence increasing interleukin 22 (IL-22) secretion involved in the maintenance of the barrier integrity. However, unlike Lamas et al., [469] results in IBD, the HIV-associated microbiota no showed a differential gene and transcript abundance for the tryptophan metabolism.

The gut microbiota metabolized the choline to trimethylamine (TMA) that is converted in the liver to the pro-atherogenic metabolite, trimethylamine-N-N-oxide (TMAO). The microbial enzymes involved in the choline catabolism were early described in *Desulfovibrio* genus [470] but a recent study revealed that is widely but unevenly distributed among Proteobacteria, Firmicutes (Clostridia and Bacilli), Actinobacteria and Fusobacteria but no homologous gene clusters were found in Bacteroidetes phylum [471]. No differences were found in the metagenomes and metatranscriptomes for this catabolic pathway between HIV-infected and healthy subjects in congruence with the no differential plasma TMAO concentration described previously (Chapter 2).

The shift of the bacterial composition into the HIV-associated community has been widely studied and related to the host health [280,281,284–289,291,353,396,472]. However, the gut microbiome is a complex community in which exists several relationships between the microbial species [1,414,418,424,473] and that the imbalance of these interactions could be an important factor for the dysbiosis establishment and maintenance. The HIV-associated ecological network presented a high degree of connectivity with *Prevotella copri* as one of the most important hub-species. In fact, the different discriminative biomarker species appeared as essential in the ecological network structure. Moreover, the depleted Bifidobacterium genus was externally located showing a co-exclusion with Proteobacteria species. Likewise, the KO biomarkers associated with HIV infection,

with high centrality degree, were the core of the metabolic network of the dysbiotic community. Thus, the HIV infection causes dramatic changes in the metabolic structure of the gut microbiota losing and gaining important core-metabolic enzymes. However, in another gut dysbiosis the highly abundant enzymes tend to be located at the periphery of the metabolic network [418].

In the present work, we implemented a multiomic bayesian network to reveal the importance of each component within the HIV-associated microbiome. The dissection of HIV-associated BN into MB revealed that the metabolic pathways from the microbiome, the total and the expressing pathways, were the ones that included more variables into their MB. Thus, the functional potential of the microbiome was the most determinant factor to maintain the microbiome structure being congruent with the notion of the functional-core [16,21,474] that establishes that any species could be lost as long as its functional role can be replaced by a metabolic equivalent species. In this context, we observed that the expression of the butanoate metabolism-related genes was an important factor in the interplay between HIV-associated microbiome and the host health due to its known beneficial effect on the colonocytes [146,475], the mucin production [169], the tight-junction integrity [167], the differentiation of T-regulatory cells [176,177] and the regeneration of the epithelial barrier [476]. In the butanoate MB, we also detected associations between the important taxonomic biomarker, *P. copri*, and several pathways increased into the HIV+ condition such as the alanine aspartate and glutamate metabolism which would confer resistance to the oxidative stress.

We observed that the microbial metabolome also interacted with different inflammation markers in the butanoate MB. Particularly, the immune activation markers, %CD8+HLA−DR+CD38+ T−cells and %CD4+HLA−DR+CD38+ T−cells correlated negatively with several membrane-structural lipids as triterpenoid related component, the glycerolipid related component, and the cholesterol glucuronide related metabolite**.** All these components could be involved in the structure of the cellular membrane. Considering the fact that all those metabolites were preferentially enriched in the IR group, the ART-responder group, it could be possible that the bacterial members of gut-microbiota acted as a transporter of those membrane precursors to the enterocytes facilitating the recovery of the gut environment.

Finally, we observed a metabolite similar to the muricholic acid correlated positively with CD4+ T-cell counts. Moreover, it has been shown that the muricholic, induced the systemic CD4+ T-cells [477–481]. This metabolite is a potent farnesoid X receptor (FXR) antagonist [482]**.** The FXR

mediates anti-inflammatory and immunomodulatory responses as well as bacterial overgrowth control and mucosal integrity and permeability maintenance in the small intestine [483,484]. On the other hand, the accumulation of bile acids stimulates the secretion of pro inflammatory molecules [477,483,484] and generated tissue damage giving an oxidative stress environment [483–487]. The increase of such metabolite in the HIV subjects is congruent with the proinflammatory state promoted by the dysbiotic microbiota; however, its positive correlation with the increase of the CD4+ T-cells is incongruent with the potential healthier state given the immune recovery. The muricholic acid metabolite is statistically increased in the IR group, followed by the VU and the INR group, the group with the lower CD4+ T-cells counts and probably the group in which the virus has caused more damage into the GALT. The Muricholic acids are a group of bile acids, this makes sense with the fact that in enteropathy that afflicts HIV-infected individuals it has been observed the malabsorption of bile acid [256,257,488]. The severe damage caused by the HIV infection in the INR group could be the reason why primary bile acids, the putative precursors of the related muricholic acid metabolite, are not been absolved by the mucosa and consequently its low abundance. However, deeper metabolome analysis is needed to identify the metabolites that interacted with the adaptive immune system.

7. GENERAL DISCUSSION

The HIV infection causes a disruption of the GALT leading the dysbiosis of the microbial community that cannot be restored by the ART. Moreover, the infection time would affect the diversity of the microbiota and the ecosystem stability. This dysbiotic community is enriched in Gram-negative species which are adapted to the inflammatory environment of the gut produced by HIV infection. Then, the gene expression and metabolic profiles from the HIV-dysbiotic microbiome were characterized by an increase of genes related to infectious diseases and to the oxidative-stress and by a depletion in the expression of genes related to anti-inflammatory metabolic pathways such as butanoate metabolism, propanoate metabolism or fatty acid metabolism. These results showed congruence with the low abundance of SCFAs and the decrease of butyrate-producer species.

Prebiotic administration produces in the VU group an increase of the *F. prausnitzii* related to the production of the butyric acid, which in turn possesses negative correlations with the bacterial translocation and the systemic inflammation. This indicates that the prebiotic administration, probably by bacteria-cross-feeding, stimulates the growth of butyrate-producer species such as *F. prausnitzii* and *Lachnospira sp.* Butyrate has been shown to contribute to maintaining homeostasis in the gut by promoting immunotolerance to commensal bacteria via the downregulation of lipopolysaccharide-induced proinflammatory mediators [175], in that sense the increase of the butyrate production may improve the GALT integrity and in consequence reducing the bacterial translocation and the systemic inflammation. It is important to highlight that such species might be viable targets for interventions.

The Bayesian network has shown the butyric metabolism is the factor with the major influence on the growth and metabolism of the bacterial population and on the interactions with the host. It is consistent with previous studies [149–153] which establish the important role of this SCFA in the intestinal homeostasis, immune regulation and in the control of the commensal microbiota.

Although, further longitudinal studies should be performed increasing the number of participants; in the present thesis, we describe in a holistic way the fundamental role of the microbiome in the pathogenesis of HIV infection. More importantly, we suggest that the microbiota may be a new target for clinical interventions in patients infected with HIV and proposed putative candidates for been viable targets for such interventions.

8. GENERAL CONCLUSIONS

1. The intestinal microbiota of HIV+ individuals presents a clear dysbiosis at taxonomic level. This dysbiosis is characterized by an increase of Gram-negative species mainly belonging to the *Prevotella* genus and pro-inflammatory species related to the genera *Succinivibrio*, *Desulfovibrio* and *Fusobacterium* and by a depletion of *Bacteroides* genus and the *Faecalibacterium, Lachospira* and *Roseburia* butyrate-producer genera.

2. The HIV+ gut-associated microbiota is enriched in genes related to the Gram-negative bacteria, such as the presence of genes from the LPS biosynthesis pathway as well as genes related to pathogenic pathways and the oxidative stress.

3. The long-term effective ART response does not recover a healthy microbiome composition. HIV+ ART positive responder shows a microbial composition similar to the other HIV+ groups.

4. In the ART responder group, the systemic immune activation, inflammation, and bacterial translocation are related to the Gram-negative HIV-dysbiosis.

5. The VU group present bacterial diversity similar to the one observed in the HIV- controls.

6. The prebiotic has an effect on a community whose original configuration is receptive to the nutritional intervention, this was seen in the VU group which the shortest median exposure to HIV.

7. The prebiotic intervention increases the butyrate levels by means of the increase of SCFA-producer species such as *Faecalibacterium sp.* The increment of the levels of the butyrate is related to the decrease of the bacterial translocation and systemic inflammation.

8. HIV-infection leads to a functional dysbiosis detected by metagenomic, metatranscriptomic and metabolomic approaches. The functional dysbiosis is congruent with the taxonomic alteration in a sense that the bacteria species that are overgrowth in the HIV+ subjects

possess the metabolic pathways enriched in the HIV-associated microbiome. This indicates that the dysbiosis is led by a complete set of bacteria rather than by a metabolic shift into their expression.

9. The Gram-negative-enriched HIV-microbiome promotes a proinflammatory environment, which disrupts the GALT integrity allowing the bacterial translocation and the consequent systemic immune activation and inflammation. This was observed in the metagenomic, metatranscriptomic, metabolomic and the 16S rDNA amplicons.

10. Although the dysbiotic microbiota presented a high abundance of genes related to the LPS biosynthesis, no differential expression was found. LPS constitutes the external part of the outer membrane of the Gram-negative bacteria and the gene expression of the LPS biosynthesis pathway would respond to bacterial growth rather than to the inflammatory environment.

11. The HIV-dysbiotic community is depleted in the expression of genes related to anti-inflammatory metabolic pathways such as butanoate metabolism, propanoate metabolism or fatty acid metabolism.

12. The ecological network of the HIV gut associated-community shows “biological network” properties. This implies that this dysbiotic-community is able to conform a stable-community which is associated with the deterioration of the patient's health.

13. The species and genes, enriched or depleted in the HIV, are important components of the ecological and metabolic networks, respectively. This first indicates that the species that are overrepresented into the HIV-condition heavily influence the rest of the bacteria community. Second, the HIV infection causes a dramatic change in the metabolic structure of the gut microbiota losing and gaining important core-metabolic enzymes.

14. The functional potential of the microbiome was the most determinant factor to maintain the microbiome network-structure being congruent with the notion of the functional-core that establishes that any species could be lost as long as its functional role can be replaced by a metabolic equivalent species.

15. The BN analysis shows that the expression of the butanoate metabolism was the factor that most influence the immune system. This is congruent with the current literature that shows

the beneficial effects of the butyrate over the host health and with the positive effect of butyrate in the reduction of the systemic immune activation, inflammation and bacterial translocation shown in Chapter 2.

9. RESUMEN EN ESPAÑOL

Introducción

Desde la aplicación de la Terapia Antirretroviral de Gran Actividad (TARGA) la supervivencia de los sujetos infectados con el virus de la inmunodeficiencia humana (VIH) se ha incrementado considerablemente; sin embargo, su expectativa de vida es todavía 10 años menor a la de la media de la población. Esta reducción es debida a enfermedades no relacionadas con el síndrome de inmunodeficiencia adquirida (SIDA). Entre las principales causantes de este descenso se encuentra el aumento de la incidencia de problemas cardiovasculares. Recientemente, se ha descrito que la activación inmune persistente y la inflamación crónica son dos factores implicados en la morbimortalidad de estos pacientes.

Un acontecimiento central en la fisiopatología del VIH es la destrucción, en fases muy tempranas de la infección, de los linfocitos Th17 del tejido linfoide asociado a mucosa (TLAM), órgano en el que residen alrededor del 90% del acervo de linfocitos totales. Junto al daño del TLAM tiene lugar apoptosis de las células epiteliales y la pérdida de la integridad de la mucosa que conduce a una translocación bacteriana anormal que se cree responsable de la inmunoactivación sistémica observada en los sujetos con VIH. Fisiológicamente las células Th17 juegan un papel crucial en la defensa frente a la translocación bacteriana, ya que estimulan la proliferación de células epiteliales y la expresión de defensinas antibacterianas, al mismo tiempo que promueven la quimiotaxis de neutrófilos hacia el TLAM para eliminar los productos bacterianos. Por tanto, la pérdida masiva de linfocitos Th17 determina probablemente el aumento de translocación bacteriana observado en la infección por VIH. Recientemente, por otro lado, se ha observado que el cambio en la composición de la microbiota intestinal asociada a la infección por el VIH está relacionada con la pérdida del TLAM, lo que podría, a su vez, ser una de las principales causas de la inflamación sistémica.

Objetivos

El objetivo principal de la presente tesis es caracterizar la comunidad microbiana intestinal disbiótica tanto en su composición como en su función, así como estudiar su efecto en la

immunoactivación sistémica. Como objetivo secundario, abordaremos el estudio del efecto de una intervención nutricional dirigida a modificar la composición bacteriana intestinal hacia una comunidad menos inflamatoria.

Materiales y métodos

Para estudiar el efecto de la infección del VIH sobre la microbiota intestinal humana, se reclutó una corte de sujetos HIV+ sin TARGA, sujetos HIV+ con respuesta positiva al TARGA, sujetos HIV+ con respuesta negativa al TARGA y sujetos sanos como controles. Adicionalmente, se realizó una intervención nutricional de seis semanas de 5 g de galacto-oligosacáridos de cadena corta (Purimune®), 10 g de fructooligosacáridos de cadena larga (Orafti-HP® y Actilight®) y 5 g de glutamina (Nutrición Médica®) y placebo (20 g de maltodextrina). De cada integrante de la cohorte se recogieron muestras fecales y sanguíneas, antes y después de la intervención nutricional, con el fin de poder caracterizar la microbiota intestinal, así como medir marcadores de activación inmune, inflamación sistémica y translocación bacteriana. La microbiota intestinal se caracterizó recurriendo a técnicas de metagenómica, metatranscriptómica y meta-metabolómica.

Para cada una de las muestras sanguíneas se obtuvieron distintos marcadores de activación inmune sistémica innata, adaptativa, translocación bacteriana, así como un análisis completo de la química sanguínea de cada uno de los participantes de la cohorte. La medición de dichos marcadores se llevó a cabo en colaboración con el Hospital Universitario Virgen del Rocío (Sevilla, España) y el Departamento de Enfermedades Infecciosas del Hospital Universitario Ramón y Cajal (Madrid, España)

De las muestras fecales se extrajo el ADN y el ARN bacterianos para posteriormente poder secuenciarlo por medio de la combinación de las tecnologías de pirosecuenciación (Roche GS FLX y química de Titanium) y secuenciación de cadenas pareadas de Illumnina (Miseq química V3). Del ADN bacteriano se amplificó el gen ribosómico 16S con el fin de poder llevar a cabo luego la caracterización taxonómica de la comunidad bacteriana correspondiente. La secuenciación masiva fue realizada en el área de Genómica y Salud de la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO)

En colaboración con el Centro de Metabolómica y Bioanálisis (CEMBIO de la, Universidad CEU San Pablo (Madrid, España), obtuvimos el perfil metabólico de las comunidades bacterianas residentes del intestino humano. Los meta-metabolomas se determinaron por medio de la extracción

de los metabolitos totales para, posteriormente, separarlos mediante la aplicación de una cromatografía líquida de alta eficacia acoplada a un espectrómetro de masas ESI-QTOF. Los metabolitos fueron identificados por medio del servidor web METLIN, las bases de datos Lipidomics Gateway database y KEGG.

Del mismo modo, en colaboración con el Departamento de Tecnología Alimentaria del Centro Agrotécnico, Universidad de Lleida (Lleida, España), obtuvimos las mediciones de los ácidos grasos de cadena corta (AGCC).

La anotación funcional de las secuencias de metagenomas se realizó mediante el ensamblaje de los metagenomas, recurriendo al programa Ray-Meta. La predicción de los marcos de lectura abiertos (MLA) la determinamos recurriendo al programa MetaGeneMark.

De forma similar, los metatranscriptomas se ensamblaron recurriendo al programa Trinity y sus MLA se predijeron por el programa TransDecoder.LongOrfs. A partir de los MLA de los metagenomas y de los metatranscriptomas creamos una base de datos no redundante de MLA aplicando el programa USEARCH. Esta base de datos se comparó, recurriendo al programa rapsearch2 con la base de datos funcionales KEGG y la base de datos de resistencias a antibióticos CARD. La cuantificación de cada MLA se llevó a cabo de forma separada para metagenomas y metatranscriptomas, recurriendo en el primer caso al programa soap.coverage, y el RSEM para el segundo.

La anotación taxonómica de las secuencias se realizó utilizando el mapeo de las secuencias de los metagenomas y metatranscriptomas contra una base de datos no redundante de genomas de referencia de especies residentes del microbioma humano. Para el caso de las secuencias de los amplicones del gen ribosómico 16S recurrimos al paquete de programas de análisis ecológico Qiime.

La identificación de los biomarcadores relativos a las condiciones HIV+ o HIV-, así como los subgrupos correspondientes de los HIV+, se identificaron por medio del programa LEfSe, de forma indistinta para las anotaciones taxonómicas y funcionales.

Recurrimos a Qiime y el programa de análisis estadístico R (v.3.3) para llevar a cabo análisis de diversidad alfa y beta. Así mismo, el programa de análisis estadístico R se utilizó para realizar los análisis de correlación, regresión lineal, pruebas de rango, los modelos generalizados lineales, las

redes bayesianas, las redes ecológicas y funcionales, así como el resto de los análisis estadísticos de la presente tesis.

Resultados y discusión

La alteración de la mucosa intestinal provocada por el VIH genera una disbiosis que se detecta a partir de la comparación de los datos provenientes de la metagenómica, la metatranscriptómica y la meta-metabolómica en las diferentes cohortes del estudio. Del mismo modo se observa una correlación con las variables clínicas de activación inmune e inflamación sistémica. La disbiosis se caracteriza por el incremento de bacterias Gram-negativas capaces de resistir el estrés oxidativo generado por la inflamación intestinal. Varios compuestos metabólicos de esta comunidad bacteriana, tales como los lipopolisacáridos de membrana, son potentes activadores de la respuesta inmune e inflamatoria. Por otro lado, varias especies bacterianas conocidas por tener un papel antinflamatorio o ser importantes productoras de AGCC ven disminuida su abundancia en el microbioma asociado a los pacientes infectados por el VIH. La pérdida de dichas especies se ve reflejada en una reducción de la concentración de los AGCC en el intestino. Los AGCC, en especial el ácido butírico, poseen efectos beneficiosos sobre la salud del hospedador. Estos ayudan a la producción de mucina, la integridad de las uniones ocluyentes, la diferenciación de las células T reguladoras y la regeneración de la barrera epitelial.

La disbiosis asociada al VIH no se ve aminorada por el uso de TARGA y la incorporación del prebiótico mostró un efecto moderado en la mayoría de los participantes. Sin embargo, dicho efecto fue más notorio en los individuos del grupo HIV+ sin TARGA. La administración del prebiótico mostró un incremento en la abundancia de especies productoras de ácido butírico junto con una mayor producción de este ácido graso. Este incremento correlacionaba con una reducción en la translocación bacteriana y la reducción de los marcadores de inflamación sistémica.

El análisis de redes reveló que la comunidad bacteriana asociada al VIH muestra propiedades de "red biológica". Esto implica que esta comunidad disbiótica es capaz de conformar una comunidad estable que está asociada al deterioro de la salud del paciente. Adicionalmente, las especies y genes que se han enriquecido o perdido en la comunidad bacteriana intestinal asociada al VIH son componentes fundamentales que mantienen la estructura de las redes ecológicas y metabólicas correspondiente. Así, las especies que están sobrerrepresentadas en la condición del VIH influyen fuertemente en el resto de la comunidad de bacterias. Por otro lado, la infección del VIH causa

cambios dramáticos en la estructura metabólica del microbioma intestinal perdiendo y ganando importantes enzimas metabólicas.

Aunque se deben realizar estudios longitudinales adicionales, así como incrementar el tamaño muestral de participantes, en la presente tesis mostramos en forma holística el papel fundamental que el microbioma intestinal tiene en la patogénesis de la infección por el VIH. Más importante aún, proponemos, por un lado, que la microbiota puede ser objeto de intervención clínica en pacientes infectados con el VIH y, por otro, sugerimos posibles candidatos bacterianos probióticos capaces de dar respuestas viables en las correspondientes intervenciones.

10. RESUM EN VALENCIÀ

Introducció

Des de l'aplicació de la Teràpia Antiretroviral de Gran Activitat (TARGA) la supervivència dels individus infectats amb el virus de la immunodeficiència humana (VIH) s'ha incrementat considerablement; no obstant això, la seva expectativa de vida és encara 10 anys menor a la de la mitjana de la població. Aquesta reducció és deguda a malalties no relacionades amb la síndrome d'immunodeficiència adquirida (SIDA). Entre les principals causes d'aquest descens es troba l'augment de la incidència de problemes cardiovasculars. Recentment, s'ha descrit que l'activació immune persistent i la inflamació crònica són dos factors implicats en la morbimortalitat d'aquests pacients.

Un esdeveniment central en la fisiopatologia del VIH és la destrucció, en fases molt primerenques de la infecció, dels limfòcits Th17 del teixit limfoide associat a mucoses (TLAM), òrgan on resideixen al voltant del 90% del total de limfòcits. A més del dany del TLAM, té lloc l'apoptosi de les cèl·lules epitelials i la pèrdua de la integritat de la mucosa, fets que condueixen a una translocació bacteriana anormal, que es creu responsable de la immunoactivació sistèmica observada en els individus amb VIH. Fisiològicament les cèl·lules Th17 juguen un paper crucial en la defensa davant la translocació bacteriana, ja que estimulen la proliferació de cèl·lules epitelials i l'expressió de defensines antibacterianes, al mateix temps que promouen la quimiotaxi de neutròfils cap al TLAM, per tal d'eliminar els productes bacterians. Per tant, la pèrdua massiva de limfòcits Th17 determina, probablement, l'augment de translocació bacteriana observat en la infecció per VIH. Recentment, per una altra part, s'ha observat que el canvi en la composició de la microbiota intestinal, associada a la infecció pel VIH, està relacionat amb la pèrdua del TLAM, cosa que podria ser una de les principals causes de la inflamació sistèmica.

Objectius

L'objectiu principal de la present tesi és caracteritzar la comunitat microbiana intestinal disbiótica, tant en la seva composició com en la seva funció, així com estudiar el seu efecte en la immunoactivació sistèmica. Com a objectiu secundari, abordarem l'estudi de l'efecte d'una intervenció nutricional dirigida a modificar la composició bacteriana intestinal cap a una comunitat menys inflamatòria.

Materials i mètodes

Per tal d'estudiar l'efecte de la infecció del VIH sobre la microbiota intestinal humana, es va reclutar una cohort d'individus VIH+ sense TARGA, d'individus VIH+ amb resposta positiva al TARGA, d'individus VIH+ amb resposta negativa al TARGA i d'individus sans com a controls. Així mateix, es va realitzar una intervenció nutricional de sis setmanes de 5 g de galactooligosacàrids de cadena curta (Purimune®), 10 g de fructooligosacàrids de cadena llarga (Orafti-HP® i Actilight®) i 5 g de glutamina (Nutrición Médica®) i placebo (20 g de maltodextrina). De cadascun dels integrants de la cohort es van prendre mostres fecals i sanguínies, abans i després de la intervenció nutricional, amb la finalitat de poder caracteritzar la microbiota intestinal, així com també es van mesurar marcadors d'activació immune, inflamació sistèmica i translocació bacteriana. La microbiota intestinal es va caracteritzar mitjançant tècniques de metagenòmica, metatranscriptòmica i meta-metabolòmica.

Per a cadascuna de les mostres sanguínies es van obtenir diferents marcadors d'activació immune sistèmica innata, adaptativa i translocació bacteriana, i a més a més, es va realitzar una anàlisi completa de la química sanguínia de cadascun dels participants de la cohort. La presa de mesures d'aquests marcadors es va dur a terme amb la col·laboració de l'Hospital Universitari Verge del Rocío (Sevilla, Espanya) i el Departament de Malalties Infeccioses de l'Hospital Universitari Ramón y Cajal (Madrid, Espanya).

De les mostres fecals es van extraure l'ADN i l'ARN bacterians per, posteriorment, poder seqüenciar-los mitjançant la combinació de les tecnologies de piroseqüenciació (Roche GS FLX i química de Titanium) i de seqüenciació de cadenes aparellades de Illumnina (Miseq química V3). De l'ADN bacterià es va amplificar el gen ribosòmic 16S amb la finalitat de dur a terme la caracterització taxonòmica de la comunitat bacteriana corresponent. La seqüenciació massiva va ser realitzada a l'àrea de Genòmica i Salut de la Fundació per al Foment de la Recerca Sanitària i Biomèdica de la Comunitat Valenciana (FISABIO).

En col·laboració amb el Centre de Metabolòmica i Bioanàlisi (CEMBIO, Universitat San Pablo CEU (Madrid, Espanya), es va obtenir el perfil metabòlic de les comunitats bacterianes residents a l'intestí humà. Els meta-metabolomes es van determinar per mitjà de l'extracció dels metabòlits totals per a, posteriorment, separar-los mitjançant l'aplicació d'una cromatografia líquida d'alta resolució acoblada a un espectròmetre de masses ESI-QTOF. Els metabòlits van ser identificats per mitjà del servidor web METLIN, i les bases de dades Lipidomics Gateway database i KEGG.

De la mateixa manera, en col·laboració amb el Departament de Tecnologia Alimentària del Centre Agrotècnic, Universitat de Lleida (Lleida, Espanya), es van obtenir les mesures dels àcids grassos de cadena curta (AGCC).

L'anotació funcional de les seqüències dels metagenomes es va realitzar mitjançant l'assemblatge d'aquests, utilitzant el programa Ray-Meta. La predicció dels marcs oberts de lectura (MOL) es va determinar amb el programa MetaGeneMark.

De manera semblant, es va procedir a l'assemblatge dels metatranscriptomes mitjançant el programa Trinity i els seus MOLes van predir amb el programa TransDecoder.LongOrfs. A partir dels MOL dels metagenomes i dels metatranscriptomes es va crear una base de dades no redundant de MOL aplicant el programa USEARCH. Aquesta base de dades es va comparar, per mitjà del programa rapsearch2 amb la base de dades funcionals KEGG i la base de dades de resistències a antibiòtics CARD. La quantificació de cada MOL es va dur a terme de forma separada per metagenomes i per metatranscriptomes, utilitzant en el primer cas el programa soap.coverage, i en el segon cas el programa RSEM.

L'anotació taxonòmica de les seqüències es va realitzar utilitzant el mapatge de les seqüències dels metagenomes i metatranscriptomes contra una base de dades no redundant, de genomes de referència d'espècies residents al microbioma humà. Per al cas de les seqüències dels amplicons del gen ribosòmic 16S es va fer servir el paquet de programes d'anàlisi ecològica Qiime.

La identificació dels biomarcadors relatius a les condicions VIH+ o VIH-, així com els subgrups corresponents dels VIH+, es va realitzar per mitjà del programa LEfSe, de manera indistinta per a les anotacions taxonòmiques i funcionals.

Es va emprar Qiime i el programa d'anàlisi estadística R (v.3.3) per dur a terme les anàlisis de diversitat alfa i beta. Així mateix, el programa d'anàlisi estadística R es va utilitzar per realitzar les anàlisis de correlació, regressió lineal, proves de rang, els models generalitzats lineals, les xarxes

bayesianes, les xarxes ecològiques i funcionals, així com la resta de les anàlisis estadístiques de la present tesi.

Resultats i discussió

L'alteració de la mucosa intestinal provocada pel VIH genera una disbiosi, que es detecta a partir de la comparació de les dades provinents de la metagenòmica, la metatranscriptòmica i la meta-metabolòmica, en les diferents cohorts de l'estudi. De la mateixa manera, s'observa una correlació amb les variables clíniques d'activació immune i d'inflamació sistèmica. La disbiosi es caracteritza per l'increment de bacteris gramnegatius capaços de resistir l'estrès oxidatiu generat per la inflamació intestinal. Diversos compostos metabòlics d'aquesta comunitat bacteriana, tals com els lipopolisacàrids de membrana, són potents activadors de la resposta immune i inflamatòria. En canvi, diverses espècies bacterianes, conegudes per tenir un paper antiinflamatori o ser importants productors de AGCC, veuen disminuïda la seva abundància en el microbioma associat als pacients infectats pel VIH. La pèrdua d'aquestes espècies es veu reflectida en una reducció de la concentració dels AGCC a l'intestí. Els AGCC, especialment l'àcid butíric, posseeixen efectes beneficiosos sobre la salut de l'hoste. Aquests ajuden a la producció de mucina, la integritat de les unions oclusives, la diferenciació de les cèl·lules T reguladores i la regeneració de la barrera epitelial.

La disbiosi associada al VIH no es veu minorada per l'ús de TARGA, i la incorporació del prebiòtic va mostrar un efecte moderat en la majoria dels participants. No obstant això, aquest efecte va ser més notori en els individus del grup VIH+ sense TARGA. L'administració del prebiòtic va mostrar un increment en l'abundància d'espècies productores d'àcid butíric, juntament amb una major producció d'aquest àcid gras. Aquest increment correlacionava amb una reducció en la translocació bacteriana i la reducció dels marcadors d'inflamació sistèmica.

L'anàlisi de xarxes va revelar que la comunitat bacteriana associada al VIH mostra propietats de "xarxa biològica". Això implica que aquesta comunitat disbiótica és capaç de conformar una comunitat estable, que està associada a la deterioració de la salut del pacient. A més a més, les espècies i els gens que s'han enriquit o perdut en la comunitat bacteriana intestinal associada al VIH són components fonamentals, que mantenen l'estructura de les xarxes ecològiques i metabòliques corresponents. Així, les espècies que estan sobrerepresentades en la condició del VIH influeixen

fortament en la resta de la comunitat de bacteris. Per una altra part, la infecció del VIH causa canvis dramàtics en l'estructura metabòlica del microbioma intestinal, perdent i guanyant importants enzims metabòlics.

Encara que s'han de realitzar estudis longitudinals addicionals, així com incrementar la grandària mostral de participants, en la present tesi vam mostrar de manera holística el paper fonamental que el microbioma intestinal té en la patogènesi de la infecció pel VIH. Més important encara, proposem, d'una banda, que la microbiota pot ser objecte d'intervenció clínica en pacients infectats amb el VIH i, de l'altra, suggerim possibles candidats bacterians probiòtics capaços de donar respostes viables en les corresponents intervencions.

11. REFERENCES

1. The NIH HMP Working Group. The NIH Human Microbiome Project. Genome Res. 2009;19:2317–23.

2. McCray LJ. 'Ome Sweet 'Omics—a genealogical treasury of words. Scientist. 2001;15.

3. Jones B V. The human gut mobile metagenome: a metazoan perspective. Gut Microbes. 2010;1:415–31.

4. OpenStax College. Overview of the Digestive System. 25 April 2. Anat. Physiol. OpenStax College; 2013.

5. Contributors W. Gastrointestinal tract. 13 Novembe. Wikipedia. Wikipedia, The Free Encyclopedia.; 2016.

6. Mondot S, Lepage P. The human gut microbiome and its dysfunctions through the meta-omics prism. Ann. N. Y. Acad. Sci. 2016;1372:9–19.

7. Leimena MM, Ramiro-Garcia J, Davids M, van den Bogert B, Smidt H, Smid EJ, et al. A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. BMC Genomics. 2013;14:530.

8. Claesson MJ, O'Sullivan O, Wang Q, Nikkilä J, Marchesi JR, Smidt H, et al. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. PLoS One. 2009;4.

9. Sherwood L, Willey J, Woolverton C. Prescott's Microbiology. 9th ed. New York: McGraw Hill; 2013.

10. van den Bogert B, de Vos WM, Zoetendal EG, Kleerebezem M. Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples. Appl. Environ. Microbiol. United States; 2011;77:2071–80.

11. Couch RD, Navarro K, Sikaroodi M, Gillevet P, Forsyth CB, Mutlu E, et al. The approach to

sample acquisition and its impact on the derived human fecal microbiome and VOC metabolome. PLoS One. United States; 2013;8:e81163.

12. Lepage P, Seksik P, Sutren M, de la Cochetière M-F, Jian R, Marteau P, et al. Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. Inflamm. Bowel Dis. 2005;11:473–80.

13. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the Human Intestinal Microbial Flora. Science (80-.). 2006;308:1635–8.

14. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473:174–80.

15. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486:222–227.

16. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464:59–65.

17. Rajilić-Stojanović M, Heilig HGHJ, Molenaar D, Kajander K, Surakka A, Smidt H, et al. Development and application of the human intestinal tract chip, a phylogenetic microarray: Analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. Environ. Microbiol. 2009;11:1736–51.

18. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. Nature Publishing Group; 2009;457:480–4.

19. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature. Nature Publishing Group; 2010;466:334–8.

20. Eloe-Fadrosh EA, Rasko DA. The human microbiome: from symbiosis to pathogenesis. Annu. Rev. Med. 2013;64:145–63.

21. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. Genome Medicine; 2016;8:51.

22. Lozupone C, Stombaugh J, Gordon J, Jansson… J. Diversity, stability and resilience of the human gut microbiota. Nature. 2012;489:220–30.

23. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science. United States; 2016;352:560–4.

24. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. Elsevier Inc.; 2014;159:789–99.

25. Benson AK, Kelly S a, Legge R, Ma F, Low SJ, Kim J, et al. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. U. S. A. 2010;107:18933–8.

26. O'Connor A, Quizon PM, Albright JE, Lin FT, Bennett BJ. Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics. Mamm. Genome. 2014;25:583–99.

27. McKnite AM, Perez-Munoz ME, Lu L, Williams EG, Brewer S, Andreux PA, et al. Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS One. 2012;7.

28. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y. Genome-wide association studies of the human gut microbiota. PLoS One. 2015;10:1–22.

29. Erwin G. Zoetendal Wilma M. Akkermans-van Vliet, J. Arjan G. M. de Visser, Willem M. de Vos ADLA. The Host Genotype Affects the Bacterial Community in the Human Gastronintestinal Tract. Microb. Ecol. Health Dis. Taylor & Francis; 2001;13:129–34.

30. Org E, Parks BW, Joo JWJ, Emert B, Schwartzman W, Kang EY, et al. Genetic and environmental control of host-gut microbiota interactions. Genome Res. 2015;25:1558–69.

31. Sommer F, Nookaew I, Sommer N, Fogelstrand P, Bäckhed F. Site-specific programming of the host epithelial transcriptome by the gut microbiota. Genome Biol. 2015;16:62.

32. Zackular JP, Baxter NT, Iverson KD. The Gut Microbiome Modulates Colon Tumorigenesis. MBio. 2013;4:1–9.

33. Gariboldi M, Manenti G, Canzian F, Falvella FS, Pierotti MA, Porta G Della, et al. Chromosome Mapping of Murine Susceptibility Loci to Liver Carcinogenesis 1. Cancer. 1993;209–11.

34. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. Sci. Transl. Med. 2014;6:237ra65.

35. Zheng J, Xiao X, Zhang Q, Mao L, Yu M, Xu J. The placental microbiome varies in association with low birth weight in full-term neonates. Nutrients. 2015;7:6924–37.

36. Gosalbes MJ, Llop S, Vall??s Y, Moya A, Ballester F, Francino MP. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. Clin. Exp. Allergy. 2013;43:198–211.

37. Ardissone AN, De La Cruz DM, Davis-Richardson AG, Rechcigl KT, Li N, Drew JC, et al. Meconium microbiome analysis identifies bacteria correlated with premature birth. PLoS One. 2014;9:1–8.

38. Del Chierico F, Vernocchi P, Petrucca A, Paci P, Fuentes S, Pratico G, et al. Phylogenetic and metabolic tracking of gut microbiota during perinatal development. PLoS One. 2015;10:1–26.

39. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. U. S. A. 2010;107:11971–5.

40. Grölund M-M, Lehtonen O-P, Eerola E, Kero P. Fecal Microflora in Healthy Infants Born by Different Methods of Delivery: Permanent Changes in Intestinal Flora After Cesarean Delivery. J. Pediatr. Gastroenterol. Nutr. 1999;28:19–25.

41. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy. Pediatrics. 2006;118:511–21.

42. Biasucci G, Benenati B, Morelli L, Bessi E, Boehm G. Cesarean delivery may affect the early biodiversity of intestinal bacteria. J. Nutr. 2008;138:1796S–1800S.

43. Biasucci G, Rubini M, Riboni S, Morelli L, Bessi E, Retetangos C. Mode of delivery affects the bacterial community in the newborn gut. Early Hum. Dev. Elsevier Ltd; 2010;86:13–5.

44. Jakobsson HE, Abrahamsson TR, Jenmalm MC, Harris K, Quince C, Jernberg C, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. Gut. 2014;63:559–66.

45. Fallani M, Young D, Scott J, Norin E, Amarri S, Adam R, et al. Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. J. Pediatr. Gastroenterol. Nutr. 2010;51:77–84.

46. Song SJ, Dominguez-Bello MG, Knight R. How delivery mode and feeding can shape the bacterial community in the infant gut. Cmaj. 2013;185:373–4.

47. Lee SA, Lim JY, Kim BS, Cho SJ, Kim NY, Kim O Bin, et al. Comparison of the gut microbiota profile in breast-fed and formula-fed Korean infants using pyrosequencing. Nutr. Res. Pract. 2015;9:242–8.

48. Hollister EB, Riehle K, Luna RA, Weidler EM, Rubio-Gonzales M, Mistretta T-A, et al. Structure and function of the healthy pre-adolescent pediatric gut microbiome. Microbiome. Microbiome; 2015;3:36.

49. Claesson MJ, Cusack S, O'Sullivan O, Greene-Diniz R, de Weerd H, Flannery E, et al. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. Proc. Natl. Acad. Sci. U. S. A. 2011;108:4586–91.

50. Mariat D, Firmesse O, Levenez F, Guimarães V, Sokol H, Doré J, et al. The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. BMC Microbiol. 2009;9:123.

51. Cho I, Blaser MJ. The Human Microbiome: at the interface of health and disease. Nat Rev Genet. 2008;12:260–270.

52. Voreades N, Kozil A, Weir TL. Diet and the development of the human intestinal microbiome. Front. Microbiol. Switzerland; 2014;5:494.

53. Tongeren S van. Fecal microbiota composition and frailty. Appl. Envirionmental Microbiol. 2005;71:6438–42.

54. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, et al. Gut microbiota composition correlates with diet and health in the elderly. Nature. England; 2012;488:178–84.

55. Greenhalgh K, Meyer KM, Aagaard KM, Wilmes P. The human gut microbiome in health: establishment and resilience of microbiota over a lifetime. Environ. Microbiol. 2016;18:2103–16.

56. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. U. S. A. 2010;107:14691–6.

57. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, et al. Diet drives

convergence in gut microbiome functions across mammalian phylogeny and within humans. Science. 2011;332:970–4.

58. Haydon DT, Cuddington K. Alternative Stable States in Ecology. Front. Ecol. Environ. 2003;1:376–82.

59. PJ T, Backhed F, Fulton L, JI G. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe. 2008;3:213–23.

60. Haukioja E, D'Antonio CM. Elton revisited: a review of evidence linking diversity and invasibility. Oikos. 2009;87:441–50.

61. Belizario JE, Napolitano M. Human microbiomes and their roles in dysbiosis, common diseases, and novel therapeutic approaches. Front. Microbiol. Switzerland; 2015;6:1050.

62. Nathan C. Antibiotics at the crossroads. Nature. England; 2004;431:899–902.

63. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc. Natl. Acad. Sci. U. S. A. 2011;108:4554–61.

64. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol. United States; 2008;6:e280.

65. Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, et al. The microbiome of uncontacted Amerindians. Sci. Adv. 2015;1:e1500183–e1500183.

66. Raymond F, Ouameur A a, Déraspe M, Iqbal N, Gingras H, Dridi B, et al. The initial state of the human gut microbiome determines its reshaping by antibiotics. ISME J. 2015;1–14.

67. Wakamoto Y, Dhar N, Chait R, Schneider K, Signorino-Gelo F, Leibler S, et al. Dynamic Persistence of Antibiotic-Stressed Mycobacteria. Science (80-.). 2013;339:91–6.

68. Cochetière MFD La, Durand T, Lepage P, Bourreille A, Galmiche JP, Doré J, et al. Resilience of the Dominant Human Fecal Microbiota upon Short-Course Antibiotic Challenge. J. Clin. Microbiol. 2005;43:5588.

69. Barc MC, Bourlioux F, Rigottier-Gois L, Charrin-Sarnel C, Janoir C, Boureau H, et al. Effect of amoxicillin-clavulanic acid on human fecal flora in a gnotobiotic mouse model assessed with

fluorescence hybridization using group-specific 16S rRNA probes in combination with flow cytometry. Antimicrob. Agents Chemother. United States; 2004;48:1365–8.

70. Sullivan A, Edlund C, Nord CE. Effect of antimicrobial agents on the ecological balance of human microflora. Lancet. Infect. Dis. United States; 2001;1:101–14.

71. Lofmark S, Jernberg C, Jansson JK, Edlund C. Clindamycin-induced enrichment and long-term persistence of resistant Bacteroides spp. and resistance genes. J. Antimicrob. Chemother. England; 2006;58:1160–7.

72. Nyberg SD, Osterblad M, Hakanen AJ, Lofmark S, Edlund C, Huovinen P, et al. Long-term antimicrobial resistance in Escherichia coli from human intestinal microbiota after administration of clindamycin. Scand. J. Infect. Dis. England; 2007;39:514–20.

73. Buffie CG, Jarchum I, Equinda M, Lipuma L, Gobourne A, Viale A, et al. Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis. Infect. Immun. 2012;80:62–73.

74. Bassis CM, Theriot CM, Young VB. Alteration of the murine gastrointestinal microbiota by tigecycline leads to increased susceptibility to clostridium difficile infection. Antimicrob. Agents Chemother. 2014;58:2767–74.

75. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, et al. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature. Nature Publishing Group; 2015;517:205–8.

76. Sekirov I, Tam NM, Jogova M, Robertson ML, Li Y, Lupp C, et al. Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. Infect. Immun. 2008;76:4726–36.

77. Jakobsson HE, Jernberg C, Andersson AF, Sjölund-Karlsson M, Jansson JK, Engstrand L. Short-term antibiotic treatment has differing long- term impacts on the human throat and gut microbiome. PLoS One. 2010;5:e9836.

78. Jakobsson H, Wreiber K, Fall K, Fjelstad B, Nyren O, Engstrand L. Macrolide resistance in the normal microbiota after Helicobacter pylori treatment. Scand. J. Infect. Dis. England; 2007;39:757–63.

79. Roduit C, Scholtens S, de Jongste JC, Wijga AH, Gerritsen J, Postma DS, et al. Asthma at 8 years of age in children born by caesarean section. Thorax. 2009;64:107–13.

80. Mårild K, Ye W, Lebwohl B, Green PHR, Blaser MJ, Card T, et al. Antibiotic exposure and the development of coeliac disease: a nationwide case-control study. BMC Gastroenterol. 2013;13:109.

81. Morgun A, Dzutsev A, Dong X, Greer RL, Sexton DJ, Ravel J, et al. Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks. Gut. 2015;64:1732–43.

82. Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, et al. Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea. J Infect Dis. 2008;197:435–8.

83. Willing BP, Russell SL, Finlay BB. Shifting the balance: antibiotic effects on host-microbiota mutualism. Nat. Rev. Microbiol. England; 2011;9:233–43.

84. Woodmansey EJ, McMurdo MET, Macfarlane GT, Macfarlane S. Comparison of compositions and metabolic activities of fecal microbiotas in young adults and in antibiotic-treated and non-antibiotic-treated elderly subjects. Appl. Environ. Microbiol. United States; 2004;70:6113–22.

85. Langdon A, Crook N, Dantas G. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. Genome Med. Genome Medicine; 2016;8:39.

86. Cho I, Yamanishi S, Cox L, Methe BA, Zavadil J, Li K, et al. Antibiotics in early life alter the murine colonic microbiome and adiposity. Nature. 2012;488:621–6.

87. Young VB, Schmidt TM. Antibiotic-associated diarrhea accompanied by large-scale alterations in the composition of the fecal microbiota. J. Clin. Microbiol. United States; 2004;42:1203–6.

88. Walk ST, Young VB. Emerging Insights into Antibiotic-Associated Diarrhea and Clostridium difficile Infection through the Lens of Microbial Ecology. Interdiscip. Perspect. Infect. Dis. Egypt; 2008;2008:125081.

89. Kamada N, Chen GY, Inohara N, Nunez G. Control of pathogens and pathobionts by the gut microbiota. Nat. Immunol. United States; 2013;14:685–90.

90. Hammami R, Fernandez B, Lacroix C, Fliss I. Anti-infective properties of bacteriocins: an update. Cell. Mol. Life Sci. Switzerland; 2013;70:2947–67.

91. Marteyn B, West NP, Browning DF, Cole JA, Shaw JG, Palm F, et al. Modulation of Shigella virulence in response to available oxygen in vivo. Nature. England; 2010;465:355–8.

92. Gantois I, Ducatelle R, Pasmans F, Haesebrouck F, Hautefort I, Thompson A, et al. Butyrate specifically down-regulates salmonella pathogenicity island 1 gene expression. Appl. Environ. Microbiol. United States; 2006;72:946–9.

93. Hsiao A, Ahmed AMS, Subramanian S, Griffin NW, Drewry LL, Petri WA, et al. Members of the human gut microbiota involved in recovery from Vibrio cholerae infection. Nature. Nature Publishing Group; 2014;515:423–6.

94. Khan CMA. The Dynamic Interactions between *Salmonella* and the Microbiota, within the Challenging Niche of the Gastrointestinal Tract. Int. Sch. Res. Not. 2014;2014:846049.

95. Matsuoka K, Kanai T. The gut microbiota and inflammatory bowel disease. Semin. Immunopathol. 2015;37:47–55.

96. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. Nature Publishing Group; 2012;491:119–24.

97. Balzola F, Bernstein C, Ho GT. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes: Commentary. Inflamm. Bowel Dis. Monit. Elsevier Inc.; 2011;11:166.

98. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc. Natl. Acad. Sci. U. S. A. 2007;104:13780–5.

99. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut. 2006;55:205–11.

100. Tong M, Li X, Parfrey LW, Roth B, Ippoliti A, Wei B, et al. A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. PLoS One. 2013;8:1–14.

101. Gophna U, Sommerfeld K, Gophna S, Doolittle WF, Veldhuyzen Van Zanten SJO.

Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. J. Clin. Microbiol. 2006;44:4136–41.

102. Peterson DA, Frank DN, Pace NR, Gordon JI. Metagenomic Approaches for Defining the Pathogenesis of Inflammatory Bowel Diseases. Cell Host Microbe. 2008;3:417–27.

103. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux J-J, et al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc. Natl. Acad. Sci. U. S. A. 2008;105:16731–6.

104. Wang W, Chen L, Zhou R, Wang X, Song L, Huang S, et al. Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. J. Clin. Microbiol. 2014;52:398–406.

105. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward D V, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13:R79.

106. Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, et al. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife. 2013;1–20.

107. Turnbaugh PJ, Ley RE, Mahowald M a, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444:1027–31.

108. Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan T-J, et al. Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota. Science (80-.). 2012;

109. Meakins TS, Jackson AA. Salvage of exogenous urea nitrogen enhances nitrogen balance in normal men consuming marginally inadequate protein diets. Clin. Sci. (Lond). England; 1996;90:215–25.

110. Neis EPJG, Dejong CHC, Rensen SS. The role of microbial amino acid metabolism in host metabolism. Nutrients. Switzerland; 2015;7:2930–46.

111. Smith EA, Macfarlane GT. Dissimilatory amino Acid metabolism in human colonic bacteria. Anaerobe. England; 1997;3:327–37.

112. Gibson GR, Cummings JH, Macfarlane GT. Growth and activities of sulphate-reducing bacteria in gut contents of healthy subjects and patients with ulcerative colitis. FEMS Microbiol. Ecol. Blackwell Publishing Ltd; 1991;9:103–11.

113. Attene-Ramos MS, Wagner ED, Plewa MJ, Gaskins HR. Evidence that hydrogen sulfide is a genotoxic agent. Mol. Cancer Res. United States; 2006;4:9–14.

114. Ridlon JM, Kang D-J, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. J. Lipid Res. United States; 2006;47:241–59.

115. Jones B V, Begley M, Hill C, Gahan CGM, Marchesi JR. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. Proc. Natl. Acad. Sci. U. S. A. United States; 2008;105:13580–5.

116. Gilliland SE, Speck ML. Deconjugation of bile acids by intestinal lactobacilli. Appl. Environ. Microbiol. United States; 1977;33:15–8.

117. Stamp DH. Three hypotheses linking bile to carcinogenesis in the gastrointestinal tract: certain bile salts have properties that may be used to complement chemotherapy. Med. Hypotheses. United States; 2002;59:398–405.

118. Hill MJ. Intestinal flora and endogenous vitamin synthesis. Eur. J. Cancer Prev. England; 1997;6 Suppl 1:S43-5.

119. Noda H, Akasaka N, Ohsugi M. Biotin production by bifidobacteria. J. Nutr. Sci. Vitaminol. (Tokyo). Japan; 1994;40:181–8.

120. Pompei A, Cordisco L, Amaretti A, Zanoni S, Matteuzzi D, Rossi M. Folate production by bifidobacteria as a potential probiotic property. Appl. Environ. Microbiol. United States; 2007;73:179–85.

121. LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M. Bacteria as vitamin suppliers to their host: A gut microbiota perspective. Curr. Opin. Biotechnol. Elsevier Ltd; 2013;24:160–8.

122. Perkins J, Pero J. Vitamin Biosynthesis. Bacillus subtilis its Closest Relat. from Genes to Cells. Washington, DC: ASM Press; 2002. p. 271–286.

123. Bacher A, Eberhardt S, Richter G. Biosynthesis of riboflavin. Escherichia coli Salmonella Cell. Mol. Biol. (edn 2). ASM Press; 1996. p. 657–664.

124. Martens JH, Barg H, Warren MJ, Jahn D. Microbial production of vitamin B12. Appl. Microbiol. Biotechnol. Germany; 2002;58:275–85.

125. Roth JR, Lawrence JG, Bobik TA. Cobalamin (coenzyme B12): synthesis and biological significance. Annu. Rev. Microbiol. United States; 1996;50:137–81.

126. Smith AG, Croft MT, Moulin M, Webb ME. Plants need their vitamins too. Curr. Opin. Plant Biol. England; 2007;10:266–75.

127. Taranto MP, Vera JL, Hugenholtz J, De Valdez GF, Sesma F. Lactobacillus reuteri CRL1098 produces cobalamin. J. Bacteriol. United States; 2003;185:5643–7.

128. Saulnier DM, Santos F, Roos S, Mistretta T-A, Spinler JK, Molenaar D, et al. Exploring metabolic pathway reconstruction and genome-wide expression profiling in Lactobacillus reuteri to define functional probiotic features. PLoS One. United States; 2011;6:e18783.

129. Champagne CP, Tompkins TA, Buckley ND, Green-Johnson JM. Effect of fermentation by pure and mixed cultures of Streptococcus thermophilus and Lactobacillus helveticus on isoflavone and B-vitamin content of a fermented soy beverage. Food Microbiol. England; 2010;27:968–72.

130. Davidson RT, Foley AL, Engelke JA, Suttie JW. Conversion of dietary phylloquinone to tissue menaquinone-4 in rats is not dependent on gut bacteria. J. Nutr. United States; 1998;128:220–3.

131. Suttie JW. The importance of menaquinones in human nutrition. Annu. Rev. Nutr. United States; 1995;15:399–417.

132. Cook SI, Sellin JH. Review article: short chain fatty acids in health and disease. Aliment. Pharmacol. Ther. England; 1998;12:499–507.

133. Hijova E, Chmelarova A. Short chain fatty acids and colonic health. Bratisl. Lek. Listy. Slovakia; 2007;108:354–8.

134. Binder HJ. Role of colonic short-chain fatty acid transport in diarrhea. Annu. Rev. Physiol. United States; 2010;72:297–313.

135. Cummings JH, Pomare EW, Branch WJ, Naylor CP, Macfarlane GT. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. Gut. England; 1987;28:1221–7.

136. Topping DL, Clifton PM. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. Physiol. Rev. United States; 2001;81:1031–64.

137. Cuervo A, Salazar N, Ruas-Madiedo P, Gueimonde M, Gonzalez S. Fiber from a regular diet is directly associated with fecal short-chain fatty acid concentrations in the elderly. Nutr. Res.

United States; 2013;33:811–6.

138. Ou J, Carbonero F. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. Am J Clin Nutr. 2013;98:111–20.

139. Louis P, Scott KP, Duncan SH, Flint HJ. Understanding the effects of diet on bacterial metabolism in the large intestine. J. Appl. Microbiol. 2007;102:1197–208.

140. Macfarlane S, Macfarlane GT. Regulation of short-chain fatty acid production. Proc. Nutr. Soc. 2003;62:67–72.

141. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, et al. Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature. England; 2011;469:543–7.

142. Frost G, Sleeth ML, Sahuri-Arisoylu M, Lizarbe B, Cerdan S, Brody L, et al. The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. Nat. Commun. England; 2014;5:3611.

143. Salonen A, Lahti L, Salojarvi J, Holtrop G, Korpela K, Duncan SH, et al. Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. ISME J. England; 2014;8:2218–30.

144. Flint HJ, Duncan SH, Scott KP, Louis P. Links between diet, gut microbiota composition and gut metabolism. Proc. Nutr. Soc. England; 2015. p. 13–22.

145. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nat. Rev. Microbiol. England; 2014;12:661–72.

146. Reichardt N, Duncan SH, Young P, Belenguer A, McWilliam Leitch C, Scott KP, et al. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. ISME J. England; 2014;8:1323–35.

147. Louis P, Duncan SH, Mccrae SI, Jackson MS, Flint HJ, Millar J. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. J. Bacteriol. 2004;186:2099–106.

148. den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud DJ, Bakker BM. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. J

Lipid Res. 2013;54:2325–40.

149. Zhang LS, Davies SS. Microbial metabolism of dietary components to bioactive metabolites: opportunities for new therapeutic interventions. Genome Med. Genome Medicine; 2016;8:46.

150. Macfarlane GT, Macfarlane S. Bacteria, colonic fermentation, and gastrointestinal health. J. AOAC Int. United States; 2012;95:50–60.

151. Macia L, Tan J, Vieira AT, Leach K, Stanley D, Luong S, et al. Metabolite-sensing receptors GPR43 and GPR109A facilitate dietary fibre-induced gut homeostasis through regulation of the inflammasome. Nat. Commun. England; 2015;6:6734.

152. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. Nat. Rev. Immunol. Nature Publishing Group; 2016;16:341–52.

153. Breuer RI, Soergel KH, Lashner BA, Christ ML, Hanauer SB, Vanagunas A, et al. Short chain fatty acid rectal irrigation for left-sided ulcerative colitis: a randomised, placebo controlled trial. Gut. England; 1997;40:485–91.

154. Vernia P, Marcheggiano A, Caprilli R, Frieri G, Corrao G, Valpiani D, et al. Short-chain fatty acid topical treatment in distal ulcerative colitis. Aliment. Pharmacol. Ther. England; 1995;9:309–13.

155. Vernia P. Butyrate in the treatment of ulcerative colitis. Dig. Liver Dis. Editrice Gastroenterologica Italiana S.r.l.; 2007;1:27–30.

156. Trompette A, Gollwitzer ES, Yadava K, Sichelstiel AK, Sprenger N, Ngom-Bru C, et al. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. Nat. Med. 2014;20:159–66.

157. AN T, CI M, Shen S, Stanley D, Macia L, LJ M. Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. Nat Commun. 2015;6:7320.

158. Vieira AT, Macia L, Galvao I, Martins FS, Canesso MCC, Amaral FA, et al. A Role for Gut Microbiota and the Metabolite-Sensing Receptor GPR43 in a Murine Model of Gout. Arthritis Rheumatol. (Hoboken, N.J.). United States; 2015;67:1646–56.

159. Andrade-Oliveira V, Amano MT, Correa-Costa M, Castoldi A, Felizardo RJF, de Almeida DC, et al. Gut Bacteria Products Prevent AKI Induced by Ischemia-Reperfusion. J. Am. Soc.

Nephrol. United States; 2015;26:1877–88.

160. Ho KJ, Xiong L, Hubert NJ, Nadimpalli A, Wun K, Chang EB, et al. Vancomycin treatment and butyrate supplementation modulate gut microbe composition and severity of neointimal hyperplasia after arterial injury. Physiol. Rep. United States; 2015;3.

161. Brown AJ, Goldsworthy SM, Barnes AA, Eilert MM, Tcheang L, Daniels D, et al. The Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. J. Biol. Chem. United States; 2003;278:11312–9.

162. Le Poul E, Loison C, Struyf S, Springael J-Y, Lannoy V, Decobecq M-E, et al. Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. J. Biol. Chem. United States; 2003;278:25481–9.

163. Thangaraju M, Cresci GA, Liu K, Ananth S, Gnanaprakasam JP, Browning DD, et al. GPR109A is a G-protein-coupled receptor for the bacterial fermentation product butyrate and functions as a tumor suppressor in colon. Cancer Res. United States; 2009;69:2826–32.

164. Clausen MR, Mortensen PB. Kinetic studies on colonocyte metabolism of short chain fatty acids and glucose in ulcerative colitis. Gut. England; 1995;37:684–9.

165. Davie JR. Inhibition of histone deacetylase activity by butyrate. J. Nutr. United States; 2003;133:2485S–2493S.

166. Siavoshian S, Segain JP, Kornprobst M, Bonnet C, Cherbut C, Galmiche JP, et al. Butyrate and trichostatin A effects on the proliferation/differentiation of human intestinal epithelial cells: induction of cyclin D3 and p21 expression. Gut. England; 2000;46:507–14.

167. Peng L, Li Z-R, Green RS, Holzman IR, Lin J. Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers. J. Nutr. United States; 2009;139:1619–25.

168. Burger-van Paassen N, Vincent A, Puiman PJ, van der Sluis M, Bouma J, Boehm G, et al. The regulation of intestinal mucin MUC2 expression by short-chain fatty acids: implications for epithelial protection. Biochem. J. 2009;420:211–9.

169. Jung T-H, Park JH, Jeon W-M, Han K-S. Butyrate modulates bacterial adherence on LS174T human colorectal cells by stimulating mucin secretion and MAPK signaling pathway. Nutr. Res.

Pract. Korea (South); 2015;9:343–9.

170. Inan MS, Rasoulpour RJ, Yin L, Hubbard AK, Rosenberg DW, Giardina C. The luminal short-chain fatty acid butyrate modulates NF-kappaB activity in a human colonic epithelial cell line. Gastroenterology. United States; 2000;118:724–34.

171. Segain JP, Raingeard de la Bletiere D, Bourreille A, Leray V, Gervois N, Rosales C, et al. Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease. Gut. England; 2000;47:397–403.

172. Menzel T, Luhrs H, Zirlik S, Schauber J, Kudlich T, Gerke T, et al. Butyrate inhibits leukocyte adhesion to endothelial cells via modulation of VCAM-1. Inflamm. Bowel Dis. United States; 2004;10:122–8.

173. Luhrs H, Gerke T, Muller JG, Melcher R, Schauber J, Boxberge F, et al. Butyrate inhibits NF-kappaB activation in lamina propria macrophages of patients with ulcerative colitis. Scand. J. Gastroenterol. England; 2002;37:458–66.

174. Luhrs H, Gerke T, Boxberger F, Backhaus K, Melcher R, Scheppach W, et al. Butyrate inhibits interleukin-1-mediated nuclear factor-kappa B activation in human epithelial cells. Dig. Dis. Sci. United States; 2001;46:1968–73.

175. Chang P V, Hao L, Offermanns S, Medzhitov R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. Proc. Natl. Acad. Sci. U. S. A. United States; 2014;111:2247–52.

176. Arpaia N, Campbell C, Fan X, Dikiy S, van der Veeken J, deRoos P, et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. Nature. England; 2013;504:451–5.

177. Furusawa Y, Obata Y, Fukuda S, TA E, Nakato G, Takahashi D. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. Nature. 2013;504:446–50.

178. Bird JJ, Brown DR, Mullen AC, Moskowitz NH, Mahowald MA, Sider JR, et al. Helper T cell differentiation is controlled by the cell cycle. Immunity. United States; 1998;9:229–37.

179. Machiels K, Joossens M, Sabino J, De Preter V, Arijs I, Eeckhaut V, et al. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in

patients with ulcerative colitis. Gut. England; 2014;63:1275–83.

180. Gao Z, Yin J, Zhang J, Ward RE, Martin RJ, Lefevre M, et al. Butyrate improves insulin sensitivity and increases energy expenditure in mice. Diabetes. United States; 2009;58:1509–17.

181. Thibault R, Blachier F, Darcy-Vrillon B, De Coppet P, Bourreille A, Segain JP. Butyrate utilization by the colonic mucosa in inflammatory bowel diseases: A transport deficiency. Inflamm. Bowel Dis. 2010;16:684–95.

182. Holm E, Hagmuller E, Staedt U, Schlickeiser G, Gunther HJ, Leweling H, et al. Substrate balances across colonic carcinomas in humans. Cancer Res. United States; 1995;55:1373–8.

183. Zhang Y, Zhou L, Bao YL, Wu Y, Yu CL, Huang YX, et al. Butyrate induces cell apoptosis through activation of JNK MAP kinase pathway in human colon cancer RKO cells. Chem. Biol. Interact. Ireland; 2010;185:174–81.

184. Canani RB, Costanzo M Di, Leone L, Pedata M, Meli R, Calignano A. Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. World J. Gastroenterol. United States; 2011. p. 1519–28.

185. Lin H V, Frassetto A, Kowalik EJJ, Nawrocki AR, Lu MM, Kosinski JR, et al. Butyrate and propionate protect against diet-induced obesity and regulate gut hormones via free fatty acid receptor 3-independent mechanisms. PLoS One. United States; 2012;7:e35240.

186. Al-Lahham SH, Peppelenbosch MP, Roelofsen H, Vonk RJ, Venema K. Biological effects of propionic acid in humans; metabolism, potential applications and underlying mechanisms. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids. Elsevier B.V.; 2010;1801:1175–83.

187. Sun Y, Oliver JD. Antimicrobial action of some GRAS compounds against Vibrio vulnificus. Food Addit. Contam. England; 1994;11:549–58.

188. Levison ME. Effect of colon flora and short-chain fatty acids on growth in vitro of Pseudomonas aeruginsoa and Enterobacteriaceae. Infect. Immun. United States; 1973;8:30–5.

189. Lawhon SD, Maurer R, Suyemoto M, Altier C. Intestinal short-chain fatty acids alter Salmonella typhimurium invasion gene expression and virulence through BarA/SirA. Mol. Microbiol. England; 2002;46:1451–64.

190. Dannhardt G, Lehr M. Nonsteroidal antiinflammatory agents, XVII: Inhibition of bovine

cyclooxygenase and 5-lipoxygenase by N-alkyldiphenyl-pyrrolyl acetic and propionic acid derivatives. Arch. Pharm. (Weinheim). Germany; 1993;326:157–62.

191. Ritsema T, Smeekens S. Fructans: beneficial for plants and humans. Curr. Opin. Plant Biol. England; 2003;6:223–30.

192. Versteeg HH, van Bergen en Henegouwen PM, van Deventer SJ, Peppelenbosch MP. Cyclooxygenase-dependent signalling: molecular events and consequences. FEBS Lett. England; 1999;445:1–5.

193. Perry RJ, Peng L, Barry NA, Cline GW, Zhang D, Cardone RL, et al. Acetate mediates a microbiome-brain-beta-cell axis to promote metabolic syndrome. Nature. England; 2016;534:213–7.

194. Everard A, Lazarevic V, Gaia N, Johansson M, Stahlman M, Backhed F, et al. Microbiome of prebiotic-treated mice reveals novel targets involved in host response during obesity. ISME J. England; 2014;8:2116–30.

195. Subbarao P, Anand SS, Becker AB, Befus AD, Brauer M, Brook JR, et al. The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma. Thorax. England; 2015;70:998–1000.

196. MC A, LT S, PA D, Thorson L, Russell S, Yurist-Doutsch S. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci Transl Med. 2015;7:307ra152.

197. Faderl M, Noti M, Corazza N, Mueller C. Keeping bugs in check: The mucus layer as a critical component in maintaining intestinal homeostasis. IUBMB Life. 2015;67:275–85.

198. Turner JR. Intestinal mucosal barrier function in health and disease. Nat. Rev. Immunol. England; 2009;9:799–809.

199. Helander HF, Fandriks L. Surface area of the digestive tract - revisited. Scand. J. Gastroenterol. England; 2014;49:681–9.

200. Brown EM, Sadarangani M, Finlay BB. The role of the immune system in governing host-microbe interactions in the intestine. Nat. Immunol. United States; 2013;14:660–7.

201. Ganz T. Paneth cells--guardians of the gut cell hatchery. Nat. Immunol. United States; 2000;1:99–100.

202. Mabbott NA, Donaldson DS, Ohno H, Williams IR, Mahajan A. Microfold (M) cells: important immunosurveillance posts in the intestinal epithelium. Mucosal Immunol. Society for Mucosal Immunology; 2013;6:666–77.

203. Johansson ME. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proc. Natl Acad. Sci. USA. Nature Publishing Group; 2008;105:15064–9.

204. Pilewski JM, Frizzell RA. Role of CFTR in airway disease. Physiol. Rev. Nature Publishing Group; 1999;79:S215–55.

205. Gunawardene AR, Corfe BM, Staton CA. Classification and functions of enteroendocrine cells of the lower gastrointestinal tract. Int. J. Exp. Pathol. England; 2011;92:219–31.

206. Cornes JS. Number, size, and distribution of Peyer's patches in the human small intestine. Gut. 1965;6:225–9.

207. Kiyono H, McGhee JR, Wannemuehler MJ, Frangakis M V, Spalding DM, Michalek SM, et al. In vivo immune response to a T-cell-dependent antigen by cultures of disassociated murine Peyer's patch. Proc. Natl. Acad. Sci. U. S. A. United States; 1982;79:596–600.

208. Barré-Sinoussi F, Ross AL, Delfraissy J-F. Past, present and future: 30 years of HIV research. Nat. Rev. Microbiol. Nature Publishing Group; 2013;11:877–83.

209. Sandler NG, Douek DC. Microbial translocation in HIV infection: causes, consequences and treatment opportunities. Nat. Rev. Microbiol. Nature Publishing Group; 2012;10:655–66.

210. Liu JZ, Pezeshki M, Raffatellu M. Th17 cytokines and host-pathogen interactions at the mucosa: dichotomies of help and harm. Cytokine. ; 2009;48:156–60.

211. Sakaguchi S, Sakaguchi N, Asano M, Itoh M, Toda M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. J. Immunol. United States; 1995;155:1151–64.

212. Sakaguchi S. Naturally arising Foxp3-expressing CD25+CD4+ regulatory T cells in immunological tolerance to self and non-self. Nat. Immunol. United States; 2005;6:345–52.

213. Sakaguchi S. Naturally arising CD4+ regulatory t cells for immunologic self-tolerance and negative control of immune responses. Annu. Rev. Immunol. United States; 2004;22:531–62.

214. Shevach EM. Regulatory T cells in autoimmmunity*. Annu. Rev. Immunol. United States; 2000;18:423–49.

215. Shevach EM. CD4+ CD25+ suppressor T cells: more questions than answers. Nat. Rev. Immunol. England; 2002;2:389–400.

216. Goto Y, Lamichhane A, Kamioka M, Sato S, Honda K, Kunisawa J, et al. IL-10-producing CD4(+) T cells negatively regulate fucosylation of epithelial cells in the gut. Sci. Rep. England; 2015;5:15918.

217. Kurashima Y, Kiyono H. Mucosal Ecological Network of Epithelium and Immune Cells for Gut Homeostasis and Tissue Healing. 2016;119–47.

218. Joller N, Lozano E, Burkett PR, Patel B, Xiao S, Zhu C, et al. Treg cells expressing the coinhibitory molecule TIGIT selectively inhibit proinflammatory Th1 and Th17 cell responses. Immunity. United States; 2014;40:569–81.

219. Venuprasad K, Kong Y-CM, Farrar MA. Control of Th2-mediated inflammation by regulatory T cells. Am. J. Pathol. United States; 2010;177:525–31.

220. Elphick DA, Mahida YR. Paneth cells: their role in innate immunity and inflammatory disease. Gut. England; 2005;54:1802–9.

221. Scheibe K, Backert I, Wirtz S, Hueber A, Schett G, Vieth M, et al. IL-36R signalling activates intestinal epithelial cells and fibroblasts and promotes mucosal healing in vivo. Gut. England; 2016;

222. Lelouard H, Fallet M, de Bovis B, Meresse S, Gorvel J-P. Peyer's patch dendritic cells sample antigens by extending dendrites through M cell-specific transcellular pores. Gastroenterology. United States; 2012;142:592–601.e3.

223. Bonnardel J, Da Silva C, Henri S, Tamoutounour S, Chasson L, Montanana-Sanchis F, et al. Innate and adaptive immune functions of peyer's patch monocyte-derived cells. Cell Rep. United States; 2015;11:770–84.

224. Hsieh C, Macatonia S, Tripp C, Wolf S, O'Garra A, Murphy K. Development of TH1 CD4+ T cells through IL-12 produced by Listeria-induced macrophages. Science (80-.). 1993;260:547–9.

225. Siegal FP, Kadowaki N, Shodell M, Fitzgerald-Bocarsly PA, Shah K, Ho S, et al. The nature of the principal type 1 interferon-producing cells in human blood. Science. United States;

1999;284:1835–7.

226. Macpherson AJ, Uhr T. Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. Science. United States; 2004;303:1662–5.

227. Sommer F, Bäckhed F. The gut microbiota--masters of host development and physiology. Nat. Rev. Microbiol. Nature Publishing Group; 2013;11:227–38.

228. Reinoso Webb C, Koboziev I, Furr KL, Grisham MB. Protective and pro-inflammatory roles of intestinal bacteria. Pathophysiol. Off. J. Int. Soc. Pathophysiol. Netherlands; 2016;23:67–80.

229. van den Elsen LW, Poyntz HC, Weyrich LS, Young W, Forbes-Blom EE. Embracing the gut microbiota: the new frontier for inflammatory and infectious diseases. Clin. Transl. Immunol. Nature Publishing Group; 2017;6:e125.

230. Ichinohe T, IK P, Kumamoto Y, DR P, JH H, TS M. Microbiota regulates immune defense against respiratory tract influenza A virus infection. Proc Natl Acad Sci USA. 2011;108:5354–9.

231. Forchielli ML, Walker WA. The role of gut-associated lymphoid tissues and mucosal defence. Br. J. Nutr. Hosp Universitario LA FE; 2005;93:S41.

232. JL R, SM L, Li J, Tran G, Jabri B, TA C. The toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. Science (80-.). 2011;332:974–7.

233. Yin Y, Wang Y, Zhu L, Liu W, Liao N, Jiang M, et al. Comparative analysis of the distribution of segmented filamentous bacteria in humans, mice and chickens. ISME J. England; 2013;7:615–21.

234. Davis CP, Savage DC. Habitat, succession, attachment, and morphology of segmented, filamentous microbes indigenous to the murine gastrointestinal tract. Infect. Immun. United States; 1974;10:948–56.

235. Ivanov I, Atarashi K, Manel N, EL B, Shima T, Karaoz U. Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell. 2009;139:485–98.

236. Bisgaard H, Li N, Bonnelykke K, BL C, Skov T, Paludan-Muller G. Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. J Allergy Clin Immunol. 2011;128:645–6.

237. Ling Z, Li Z, Liu X, Cheng Y, Luo Y, Tong X. Altered fecal microbiota composition

associated with food allergy in infants. Appl Env. Microbiol. 2014;80:2546–54.

238. KE F, AR S, Havstad S, DL L, Levan S, Fadrosh D. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. Nat Med. 2016;22:1187–91.

239. Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, et al. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science. United States; 1998;282:2085–8.

240. Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature. Macmillan Publishers Limited. All rights reserved; 2009;458:1191–5.

241. Beutler B, Poltorak A. The sole gateway to endotoxin response: how LPS was identified as Tlr4, and its role in innate immunity. Drug Metab. Dispos. United States; 2001. p. 474–8.

242. Leitner T, Hahn B, Mullins J, Wolinsky S, Foley B, Apetrei C, et al. HIV Sequence Compendium 2016 Editors. Los Alamos, New Mexico 87545 U.S.A.; 2016.

243. Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M. Nucleotide sequence of the AIDS virus, LAV. Cell. 1985;40:9–17.

244. Sanchez-Pescador R, Michael D. P, Philip J. B, Kathelyn S. S, Michelle M. S, Sheryl L. B-S, et al. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science (80-.). 1985;227:484–92.

245. Deeks SG, Overbaugh J, Phillips A, Buchbinder S. HIV infection. Nat. Rev. 2015;1:1–22.

246. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. The early spread and epidemic ignition of HIV-1 in human populations. Science (80-.). 2014;346:56–61.

247. Keele BF, Heuverswyn F Van, Li Y, Bailes E, Santiago ML, Bibollet-ruche F, et al. Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1. Science (80-.). 2006;313:523–6.

248. Barre-Sinoussi F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (80-.). 1983;220:868–71.

249. ONUSIDA. ONUSIDA [Internet]. 2017. Available from: www.unaids.org/es

250. AIDS.gov. What Is HIV/AIDS? [Internet]. 2016. Available from: https://www.aids.gov/hiv-aids-basics/hiv-aids-101/what-is-hiv-aids/

251. Vogel M, Schwarze-Zander C, Wasmuth J-C, Spengler U, Sauerbruch T, Rockstroh JK. The treatment of patients with HIV. Dtsch. Arztebl. Int. Germany; 2010;107:507–15; quiz 516.

252. Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature. 1984;312:763–7.

253. Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, Hercend T, et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature. 1984;312:767–8.

254. Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell. 1986;47:333–48.

255. El Hed A. Susceptibility of human Th17 cells to human immunodeficiency virus and their perturbation during infection. J. Infect. Dis. ; 2010;201:843–54.

256. Kapembwa MS. Altered small-intestinal permeability associated with diarrhoea in human-immunodeficiency-virus-infected Caucasian and African subjects. Clin. Sci. (Lond.). ; 1991;81:327–34.

257. Bjarnason I, Sharpstone DR, Francis N, Marker A, Taylor C, Barrett M, et al. Intestinal inflammation, ileal structure and function in HIV. AIDS. England; 1996;10:1385–91.

258. Epple HJ. Acute HIV infection induces mucosal infiltration with CD4+ and CD8+ T cells, epithelial apoptosis, and a mucosal barrier defect. Gastroenterology. ; 2010;139:1289–300.

259. Mehandru S. Lack of mucosal immune reconstitution during prolonged treatment of acute and early HIV-1 infection. PLoS Med. ; 2006;3:e484.

260. World Health Organization. HIV/AIDS [Internet]. 2017. Available from: http://www.who.int/hiv/topics/treatment/en/

261. World Health Organization. Consolidated ARV guidelines. 2013.

262. Hammer S, Squires K, Hughes M, Grimes J, Demeter L, Currier J, et al. A Controlled Trial of

Two Nucleoside Analogues plus Indinavir in Persons with Human Immunodeficiency Virus Infection and CD4 Cell Counts of 200 per Cubic Millimeter or Less. N. Engl. J. Med. 2003;337:1–7.

263. Palella FJ, KM D, AC M, MO L, J F, GA S, et al. Declining Morbidity and Mortality among Patients with Advanced Human Immunodeficiency Virus Infection. N. Engl. J. Med. 1998;338:853–60.

264. Ho DD, Neumann a U, Perelson a S, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature. 1995;373:123–6.

265. Appay V, Sauce D. Immune activation and inflammation in HIV-1 infection: causes and consequences. J. Pathol. 2008;214:231–41.

266. Deeks SG, Lewin SR, Havlir D V. The end of AIDS: HIV infection as a chronic disease. Lancet. Elsevier Ltd; 2013;382:1525–33.

267. Neuhaus J, Jacobs DR, Baker J V, Calmy A, Duprez D, La Rosa A, et al. Markers of inflammation, coagulation, and renal function are elevated in adults with HIV infection. J. Infect. Dis. 2010;201:1788–95.

268. Sandler NG, Wand H, Roque A, Law M, Nason MC, Nixon DE, et al. Plasma levels of soluble CD14 independently predict mortality in HIV infection. J. Infect. Dis. 2011;203:780–90.

269. Hansson GK, Hermansson A. The immune system in atherosclerosis. Nat. Immunol. ; 2011;12:204–12.

270. Mantovani A, Cassatella MA, Costantini C, Jaillon S. Neutrophils in the activation and regulation of innate and adaptive immunity. Nat. Rev. Immunol. ; 2011;11:519–31.

271. Pacifici R. The immune system and bone. Arch. Biochem. Biophys. ; 2010;503:41–53.

272. Estes J. Collagen deposition limits immune reconstitution in the gut. J. Infect. Dis. ; 2008;198:456–64.

273. Funderburg NT. Increased tissue factor expression on circulating monocytes in chronic HIV infection: relationship to in vivo coagulation and immune activation. Blood. ; 2010;115:161–7.

274. Marmur JD. Identification of active tissue factor in human coronary atheroma. Circulation. ; 1996;94:1226–32.

275. Davi G, Patrono C. Platelet activation and atherothrombosis. N. Engl. J. Med. ; 2007;357:2482–94.

276. Lyons JL. Plasma sCD14 is a biomarker associated with impaired neurocognitive test performance in attention and learning domains in HIV infection. J. Acquir. Immune Defic. Syndr. ; 2011;57:371–9.

277. Ancuta P. Microbial translocation is associated with increased monocyte activation and dementia in AIDS patients. PLoS One. ; 2008;3:e2516.

278. Reeves RK. Gut inflammation and indoleamine deoxygenase inhibit IL-17 production and promote cytotoxic potential in NKp44+ mucosal NK cells during SIV infection. Blood. ; 2011;118:3321–30.

279. Favre D. Tryptophan catabolism by indoleamine 2,3-dioxygenase 1 alters the balance of TH17 to regulatory T cells in HIV disease. Sci. Transl. Med. ; 2010;2:32ra36.

280. Ellis C, Ma Z-M, Mann S, Al. E. Molecular Characterization of Stool Microbiota in HIV-Infected Subjects by Panbacterial and Order-Level 16S Ribosomal DNA (rDNA) Quantification and Correlations with Immune Activation. J. Acquir. Immune Defic. Syndr. 2011;57:363–370.

281. Lozupone C a, Li M, Campbell TB, Flores SC, Linderman D, Gebert MJ, et al. Alterations in the Gut Microbiota Associated with HIV-1 Infection. Cell Host Microbe. Elsevier Inc.; 2013;14:329–39.

282. Lozupone CA, Rhodes ME, Neff CP, Fontenot AP, Campbell TB, Palmer BE. HIV-induced alteration in gut microbiota. Gut Microbes. 2014;5:562–70.

283. Ling Z, Jin C, Xie T, Cheng Y, Li L, Wu N. Alterations in the Fecal Microbiota of Patients with HIV-1 Infection: An Observational Study in A Chinese Population. Nat. Publ. Gr. Nature Publishing Group; 2016;1–12.

284. Dillon S, Lee E, Kotter C, Austin G. An altered intestinal mucosal microbiome in HIV-1 infection is associated with mucosal and systemic immune activation and endotoxemia. Mucosal Immunol. Nature Publishing Group; 2014;0:mi.2013.116.

285. Vujkovic-Cvijin I, Dunham RM, Iwai, M. C. Maher SS, Albright RG, Broadhurst MJ, Hernandez RD, et al. Dysbiosis of the gut microbiota is associated with HIV disease progression

and tryptophan catabolism. Sci Transl Med. 2013;5:193ra91.

286. Nowak P, Troseid M, Avershina E, Barqasho B, Neogi U, Holm K, et al. Gut microbiota diversity predicts immune status in HIV-1 infection. Aids. 2015;29:2409–18.

287. Pérez-Santiago J, Gianella S, Massanella M. Gut lactobacillales are associated with higher CD4 and less microbial translocation during HIV infection. AIDS. 2013;27:1921–31.

288. Mutlu E a., Keshavarzian A, Losurdo J, Swanson G, Siewe B, Forsyth C, et al. A Compositional Look at the Human Gastrointestinal Microbiome and Immune Activation Parameters in HIV Infected Subjects. Relman DA, editor. PLoS Pathog. 2014;10:e1003829.

289. Noguera-Julian M, Rocafort M, Guillén Y, Rivera J, Casadellà M, Nowak P, et al. Gut Microbiota Linked to Sexual Preference and HIV Infection. EBioMedicine. The Authors; 2016;5:135–46.

290. Monaco CL, Gootenberg DB, Zhao G, Handley SA, Ghebremichael MS, Lim ES, et al. Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host Microbe. 2016;19:311–22.

291. Dinh DM, Volpe GE, Duffalo C, Bhalchandra S, Tai AK, Kane A V., et al. Intestinal Microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. J. Infect. Dis. 2015;211:19–27.

292. Lecerf J-M, Depeint F, Clerc E, Dugenet Y, Niamba CN, Rhazi L, et al. Xylo-oligosaccharide (XOS) in combination with inulin modulates both the intestinal environment and immune status in healthy subjects, while XOS alone only shows prebiotic properties. Br. J. Nutr. England; 2012;108:1847–58.

293. Majid HA, Emery PW, Whelan K. Faecal microbiota and short-chain fatty acids in patients receiving enteral nutrition with standard or fructo-oligosaccharides and fibre-enriched formulas. J. Hum. Nutr. Diet. England; 2011;24:260–8.

294. Halmos EP, Christophersen CT, Bird AR, Shepherd SJ, Gibson PR, Muir JG. Diets that differ in their FODMAP content alter the colonic luminal microenvironment. Gut. England; 2015;64:93–100.

295. Salazar N, Dewulf EM, Neyrinck AM, Bindels LB, Cani PD, Mahillon J, et al. Inulin-type

fructans modulate intestinal Bifidobacterium species populations and decrease fecal short-chain fatty acids in obese women. Clin. Nutr. England; 2015;34:501–7.

296. Cahn P, Ruxrungtham K, Gazzard B, Diaz RS, Gori A, Kotler DP, et al. The immunomodulatory nutritional intervention NR100157 reduced CD4+ T-cell decline and immune activation: a 1-year multicenter randomized controlled double-blind trial in HIV-infected persons not receiving antiretroviral therapy (The BITE Study). Clin. Infect. Dis. United States; 2013;57:139–46.

297. Gonzalez-Hernandez LA, Jave-Suarez LF, Fafutis-Morris M, Montes-Salcedo KE, Valle-Gutierrez LG, Campos-Loza AE, et al. Synbiotic therapy decreases microbial translocation and inflammation and improves immunological status in HIV-infected patients: a double-blind randomized controlled pilot trial. Nutr. J. England; 2012;11:90.

298. Hummelen R, Changalucha J, Butamanya NL, Koyama TE, Cook A, Habbema JDF, et al. Effect of 25 weeks probiotic supplementation on immune function of HIV patients. Gut Microbes. United States; 2011;2:80–5.

299. Gori A, Rizzardini G, Van't Land B, Amor KB, van Schaik J, Torti C, et al. Specific prebiotics modulate gut microbiota and immune activation in HAART-naive HIV-infected adults: results of the "COPA" pilot randomized trial. Mucosal Immunol. 2011;4:554–63.

300. Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ, Pot B, et al. Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat. Rev. Gastroenterol. Hepatol. England; 2014;11:506–14.

301. McFarland L V. Use of probiotics to correct dysbiosis of normal microbiota following disease or disruptive events: a systematic review. BMJ Open. England; 2014;4:e005047.

302. Sanders ME. Considerations for use of probiotic bacteria to modulate human health. J. Nutr. United States; 2000;130:384S–390S.

303. Kumar M, Nagpal R, Kumar R, Hemalatha R, Verma V, Kumar A, et al. Cholesterol-lowering probiotics as potential biotherapeutics for metabolic diseases. Exp. Diabetes Res. United States; 2012;2012:902917.

304. Cani PD. Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced

diabetes in mice through a mechanism associated with endotoxaemia. Diabetologia. ; 2007;50:2374–83.

305. Stiksrud B, Nowak P, Nwosu FC, Kvale D, Thalme A, Sonnerborg A, et al. Reduced Levels of D-dimer and Changes in Gut Microbiota Composition After Probiotic Intervention in HIV-Infected Individuals on Stable ART. J. Acquir. Immune Defic. Syndr. United States; 2015;70:329–37.

306. Villar-Garcia J, Hernandez JJ, Guerri-Fernandez R, Gonzalez A, Lerma E, Guelar A, et al. Effect of probiotics (Saccharomyces boulardii) on microbial translocation and inflammation in HIV-treated patients: a double-blind, randomized, placebo-controlled trial. J. Acquir. Immune Defic. Syndr. United States; 2015;68:256–63.

307. Klatt NR, Canary LA, Sun X, Vinton CL, Funderburg NT, Morcock DR, et al. Probiotic/prebiotic supplementation of antiretrovirals improves gastrointestinal immunity in SIV-infected macaques. J. Clin. Invest. United States; 2013;123:903–7.

308. EISEMAN B, SILEN W, BASCOM GS, KAUVAR AJ. Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. Surgery. United States; 1958;44:854–9.

309. Wyatt JP. Oxford handbook of emergency medicine. Oxford University Press, USA; 2006.

310. Vrieze A, de Groot PF, Kootte RS, Knaapen M, van Nood E, Nieuwdorp M. Fecal transplant: a safe and sustainable clinical therapy for restoring intestinal microbial balance in human disease? Best Pract. Res. Clin. Gastroenterol. Netherlands; 2013;27:127–37.

311. Smith MB, Kelly C, Alm EJ. How to regulate faecal transplants. Nature. 2014;506:290–1.

312. Kunde S, Pham A, Bonczyk S, Crumb T, Duba M, Conrad HJ, et al. Safety, tolerability, and clinical response after fecal transplantation in children and young adults with ulcerative colitis. J. Pediatr. Gastroenterol. Nutr. United States; 2013;56:597–601.

313. Borody TJ, Warren EF, Leis S, Surace R, Ashman O. Treatment of ulcerative colitis using fecal bacteriotherapy. J. Clin. Gastroenterol. United States; 2003;37:42–7.

314. Vermeire S, Joossens M, Verbeke K, Hildebrand F, Machiels K, Van den Broeck K, et al. Sa1922 Pilot Study on the Safety and Efficacy of Faecal Microbiota Transplantation in Refractory Crohn’s Disease. Gastroenterology. AGA; 2012;142:S-360.

315. Serrano-Villar S, Vázquez-Castellanos JF, Vallejo A, Latorre A, Sainz T, Ferrando-Martínez

S, et al. The Effects of Prebiotics on Microbial Dysbiosis, Butyrate Production and Immunity in HIV-Infected Subjects. Mucosal Immunol. 2016;(in press).

316. Yeboah J, Folsom AR, Burke GL, Johnson C, Polak JF, Post W, et al. Predictive value of brachial flow-mediated dilation for incident cardiovascular events in a population-based study: the multi-ethnic study of atherosclerosis. Circulation. United States; 2009;120:502–9.

317. Koeth R a, Wang Z, Levison BS, Buffa J a, Org E, Sheehy BT, et al. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat. Med. Nature Publishing Group; 2013;19:576–85.

318. Elsbach P. The bactericidal/permeability-increasing protein (BPI) in antibacterial host defense. J. Leukoc. Biol. United States; 1998;64:14–8.

319. Paiardini M, Muller-Trutwin M. HIV-associated chronic immune activation. Immunol. Rev. England; 2013;254:78–101.

320. Lu W, Mehraj V, Vyboh K, Cao W, Li T, Routy J-P. CD4:CD8 ratio as a frontier marker for clinical outcome, immune dysfunction and viral reservoir size in virologically suppressed HIV-positive patients. J. Int. AIDS Soc. Switzerland; 2015;18:20052.

321. Okoye AA, Picker LJ. CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure. Immunol. Rev. England; 2013;254:54–64.

322. Wang L, Xie Y, Zhu L-J, Chang T-T, Mao Y-Q, Li J. An association between immunosenescence and CD4(+)CD25(+) regulatory T cells: a systematic review. Biomed. Environ. Sci. Netherlands; 2010;23:327–32.

323. Marchetti G, Tincati C, Silvestri G. Microbial translocation in the pathogenesis of HIV infection and AIDS. Clin. Microbiol. Rev. United States; 2013;26:2–18.

324. Ferrando-Martínez S, Franco JM, Ruiz-Mateos E, Hernández A, Ordoñez A, Gutierrez E, et al. A reliable and simplified sj/beta-TREC ratio quantification method for human thymic output measurement. J. Immunol. Methods. 2010;352:111–7.

325. Tang WHW, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. N. Engl. J. Med. United States; 2013;368:1575–84.

326. Wang Z, Roberts AB, Buffa JA, Levison BS, Zhu W, Org E, et al. Non-lethal Inhibition of Gut Microbial Trimethylamine Production for the Treatment of Atherosclerosis. Cell. United States; 2015;163:1585–95.

327. Pérez-Cobas AE, Gosalbes MJ, Friedrichs A, Knecht H, Artacho A, Eismann K, et al. Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. Gut. 2013;62:1591–601.

328. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863–4.

329. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. Nat. Publ. Gr. Nature Publishing Group; 2010;7:335–6.

330. Sokal RR, Sneath PHA. Principles of Numerical Taxonomy. W. H. Freeman; 1963.

331. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.

332. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, et al. The Ribosomal Database Project (RDP-II): Sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res. 2005;33:294–6.

333. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 2006;72:5069–72.

334. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26:266–7.

335. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490.

336. Koren O, Knights D, Gonzalez A, Waldron L, Segata N, Knight R, et al. A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. PLoS Comput. Biol. 2013;9:e1002863.

337. Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. Appl. Environ. Microbiol. 2005;71:8228–35.

338. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. BioMed Central Ltd; 2011;12:R60.

339. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. 2010;33:1–22.

340. Hastie T, Qian J, Start Q, Regression L, Regression L, Models P, et al. Glmnet Vignette. 2016.

341. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan MB, editor. Probabilistic Reason. Intell. Syst. networks plausible inference. San Francisco California USA: Morgan Kaufmann Publishers; 1988.

342. Faltin F, Kenett R. Bayesian Networks. Encycl. Stat. Qual. Reliab. 2007;1:4.

343. Markov A. Teoriia algorifmov. Moscow: Tr. Matematicheskogo instituta AN SSSR, vol. 42.; 1954.

344. McGeachie MJ, Sordillo JE, Gibson T, Weinstock GM, Liu Y-Y, Gold DR, et al. Longitudinal Prediction of the Infant Gut Microbiome with Dynamic Bayesian Networks. Sci. Rep. Nature Publishing Group; 2016;6:20359.

345. Pérez-Cobas AE, Artacho A, Ott SJ, Andrés, Moya A, Gosalbes MJ, Latorre A. Structural and functional changes in the gut microbiota associated to Clostridium difficile infection. Front. Microbiol. 2014;5:1–15.

346. Džunková M, D'Auria G, Xu H, Huang J, Duan Y, Moya A, et al. The Monoclonal Antitoxin Antibodies (Actoxumab-Bezlotoxumab) Treatment Facilitates Normalization of the Gut Microbiota of Mice with Clostridium difficile Infection. Front. Cell. Infect. Microbiol. 2016;6:119.

347. Džunková M, Moya A, Vázquez-Castellanos JF, van de Guchte M, Chen X, Kelly C, et al. Active and Secretory IgA-Coated Bacterial Fractions Elucidate Dysbiosis in Clostridium difficile Infection. mSphere. 2016;1:1–11.

348. Vazquez-Castellanos J, Serrano-Villar S, Latorre A, Artacho A, Madrid N, Vera M, et al. Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. Mucosal Immunol. 2014;8:760–72.

349. Jiang W, Lederman MM, Hunt P, Sieg SF, Haley K, Rodriguez B, et al. Plasma levels of bacterial DNA correlate with immune activation and the magnitude of immune restoration in

persons with antiretroviral-treated HIV infection. J. Infect. Dis. 2009;199:1177–85.

350. Mavigner M, Cazabat M, Dubois M, L'Faqihi F-E, Requena M, Pasquier C, et al. Altered CD4+ T cell homing to the gut impairs mucosal immune reconstitution in treated HIV-infected individuals. J. Clin. Invest. 2012;122:62–9.

351. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat. Rev. Genet. England; 2012;13:260–70.

352. Presley LL, Ye J, Li X, Leblanc J, Zhang Z, Ruegger PM, et al. Host-microbe relationships in inflammatory bowel disease detected by bacterial and metaproteomic analysis of the mucosal-luminal interface. Inflamm. Bowel Dis. 2012;18:409–17.

353. McHardy IH, Li X, Tong M, Ruegger P, Jacobs J, Borneman J, et al. HIV Infection is associated with compositional and functional shifts in the rectal mucosal microbiota. Microbiome. 2013;1:26.

354. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008;9:386.

355. Scutari M. Learning Bayesian Networks with the bnlearn R Package. J. Stat. Softw. 2010;35:1–22.

356. Su C, Andrew A, Karagas MR, Borsuk ME. Using Bayesian networks to discover relations between genes, environment, and disease. BioData Min. BioData Mining; 2013;6:6.

357. Deeks SG, Tracy R, Douek DC. Systemic Effects of Inflammation on Health during Chronic HIV Infection. Immunity. Elsevier Inc.; 2013;39:633–45.

358. Yu G, Fadrosh D, Ma B, Ravel J, Goedert JJ. Anal microbiota profiles in HIV-positive and HIV-negative MSM. AIDS. 2013;28:753–60.

359. Durbán A, Abellán JJ, Jiménez-Hernández N, Ponce M, Ponce J, Sala T, et al. Assessing gut microbial diversity from feces and rectal mucosa. Microb. Ecol. 2011;61:123–33.

360. Winter SE, Winter MG, Xavier MN, Thiennimitr P, Poon V, Keestra a M, et al. Host-derived nitrate boosts growth of E. coli in the inflamed gut. Science (80-.). 2013;339:708–11.

361. Mukhopadhya I, Hansen R, El-Omar EM, Hold GL. IBD-what role do Proteobacteria play?

Nat Rev Gastroenterol Hepatol. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2012;9:219–30.

362. Serrano-Villar S, Sainz T, Lee S a, Hunt PW, Sinclair E, Shacklett BL, et al. HIV-Infected Individuals with Low CD4/CD8 Ratio despite Effective Antiretroviral Therapy Exhibit Altered T Cell Subsets, Heightened CD8+ T Cell Activation, and Increased Risk of Non-AIDS Morbidity and Mortality. PLoS Pathog. 2014;10:e1004078.

363. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science (80-.). 2011;334:105–8.

364. Vinolo M a R, Rodrigues HG, Nachbar RT, Curi R. Regulation of inflammation by short chain fatty acids. Nutrients. 2011;3:858–76.

365. Carvalho BM, Saad MJA. Influence of gut microbiota on subclinical inflammation and insulin resistance. Mediators Inflamm. 2013;2013:13.

366. Rook G a. Regulation of the immune system by biodiversity from the natural environment: An ecosystem service essential to health. Proc. Natl. Acad. Sci. U. S. A. 2013;110:18360–7.

367. Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E, et al. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. PLoS One. 2010;5:e10667.

368. Ivanov II, Frutos RDL, Manel N, Yoshinaga K, Rifkin DB, Sartor RB, et al. Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. Cell Host Microbe. 2008;4:337–49.

369. Deeks SG, Verdin E, McCune JM. Immunosenescence and HIV. Curr. Opin Immunol. 2012;24:501–6.

370. Verhasselt V, Buelens C, Willem F, Groote D De, Haeffner-cavaillon N, Goldman M. Bacterial Lipopolysaccharide Stimulates the Production of Cytokines and the Expression of Costimulatory Molecules by Human Peripheral Blood Dendritic Cells. J Immunol. 1997;158:2919–25.

371. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat. Commun. 2012;3:1245.

372. Naik SR, Rupawalla EN, Sheth UK. Anti-inflammatory activity of thiamine and nicotinic acid.

Biochem. Pharmacol. 1970;19:2867–73.

373. Moallem S, Hosseinzadeh H, Farahi S. A study of acute and chronic anti-nociceptive and anti-inflammatory effects of thiamine in mice. Iran Biomed J. 2008;12:173–8.

374. Kuroki F, Iida M, Tominaga M, Matsumoto T, Hirakawa K, Sugiyama S, et al. Multiple vitamin status in Crohn's disease. Dig. Dis. Sci. Kluwer Academic Publishers-Plenum Publishers; 1993;38:1614–8.

375. Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol. Lett. England; 2009;294:1–8.

376. Serrano-villar S, Rojo D, Martinez M, Deusch S, Vazquez-Castellanos J, Bargiela R, et al. Gut Bacteria Metabolism Impacts Immune Recovery in HIV-infected Individuals. EBioMedicine. 2016;8:203–216.

377. Bik EM. Composition and function of the human-associated microbiota. Nutr. Rev. United States; 2009;67 Suppl 2:S164-71.

378. Hummelen R, Vos AP, van't Land B, van Norren K, Reid G. Altered host-microbe interaction in HIV: a target for intervention with pro- and prebiotics. Int. Rev. Immunol. England; 2010;29:485–513.

379. Schunter M, Chu H, Hayes TL, McConnell D, Crawford SS, Luciw PA, et al. Randomized pilot trial of a synbiotic dietary supplement in chronic HIV-1 infection. BMC Complement. Altern. Med. England; 2012;12:84.

380. Wang L, Hou Y, Yi D, Li Y, Ding B, Zhu H, et al. Dietary supplementation with glutamate precursor alpha-ketoglutarate attenuates lipopolysaccharide-induced liver injury in young pigs. Amino Acids. Austria; 2015;47:1309–18.

381. Pai M-H, Shih Y-M, Shih J-M, Yeh C-L. Glutamine Modulates Changes in Intestinal Intraepithelial gammadeltaT-Lymphocyte Expressions in Mice With Ischemia/Reperfusion Injury. Shock. United States; 2015;44:77–82.

382. Moore SR, Guedes MM, Costa TB, Vallance J, Maier EA, Betz KJ, et al. Glutamine and alanyl-glutamine promote crypt expansion and mTOR signaling in murine enteroids. Am. J. Physiol. Gastrointest. Liver Physiol. United States; 2015;308:G831-9.

383. Chauvin J, Jarry A, Meurette G, Sarrabayrouse G, Bridonneau C, Preisser L, et al. CD4CD8 aa Lymphocytes, A Novel Human Regulatory T Cell Subset Induced by Colonic Bacteria and Deficient in Patients with Inflammatory Bowel Disease ´. PLoS Biol. 2014;12.

384. Somsouk M, Estes JD, Deleage C, Dunham RM, Albright R, Inadomi JM, et al. Gut epithelial barrier and systemic inflammation during chronic HIV infection. AIDS. England; 2015;29:43–51.

385. Flint HJ, Scott KP, Louis P, Duncan SH. The role of the gut microbiota in nutrition and health. Nat. Rev. Gastroenterol. Hepatol. Nature Publishing Group; 2012;9:577–89.

386. Hooper L V, Midtvedt T, Gordon JI. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu. Rev. Nutr. United States; 2002;22:283–307.

387. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic analysis of the human distal gut microbiome. Science. United States; 2006;312:1355–9.

388. Zhang C, Yin A, Li H, Wang R, Wu G, Shen J, et al. Dietary Modulation of Gut Microbiota Contributes to Alleviation of Both Genetic and Simple Obesity in Children. EBioMedicine. Netherlands; 2015;2:968–84.

389. Sun Y, Ma Y, Lin P, Tang Y-W, Yang L, Shen Y, et al. Fecal bacterial microbiome diversity in chronic HIV-infected patients in China. Emerg. Microbes Infect. United States; 2016;5:e31.

390. Kim M-S, Hwang S-S, Park E-J, Bae J-W. Strict vegetarian diet improves the risk factors associated with metabolic diseases by modulating gut microbiota and reducing intestinal inflammation. Environ. Microbiol. Rep. United States; 2013;5:765–75.

391. Lozupone CA, Rhodes ME, Neff CP, Fontenot AP, Campbell TB, Palmer BE. HIV-induced alteration in gut microbiota: driving factors, consequences, and effects of antiretroviral therapy. Gut Microbes. United States; 2014;5:562–70.

392. Williams B, Landay A, Presti RM. Microbiome alterations in HIV infection a review. Cell. Microbiol. 2016;18:645–51.

393. Klatt NR, Funderburg NT, Brenchley JM. Microbial translocation, immune activation, and HIV disease. Trends Microbiol. England; 2013;21:6–13.

394. Handley SA, Thackray LB, Zhao G, Presti R, Miller AD, Droit L, et al. Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. Cell. United

States; 2012;151:253–66.

395. Barbian HJ, Li Y, Ramirez M, Klase Z, Lipende I, Mjungu D, et al. Destabilization of the gut microbiome marks the end-stage of simian immunodeficiency virus infection in wild chimpanzees. Am. J. Primatol. United States; 2015;

396. Vujkovic-Cvijin I, Swainson LA, Chu SN, Ortiz AM, Santee CA, Petriello A, et al. Gut-Resident Lactobacillus Abundance Associates with IDO1 Inhibition and Th17 Dynamics in SIV-Infected Macaques. Cell Rep. United States; 2015;13:1589–97.

397. Seekatz AM, Aas J, Gessert CE, Rubin TA, Saman DM, Bakken JS, et al. Recovery of the gut microbiome following fecal microbiota transplantation. MBio. United States; 2014;5:e00893-14.

398. Laval L, Martin R, Natividad JN, Chain F, Miquel S, Desclee de Maredsous C, et al. Lactobacillus rhamnosus CNCM I-3690 and the commensal bacterium Faecalibacterium prausnitzii A2-165 exhibit similar protective effects to induced barrier hyper-permeability in mice. Gut Microbes. United States; 2015;6:1–9.

399. Schwiertz A, Lehmann U, Jacobasch G, Blaut M. Influence of resistant starch on the SCFA production and cell counts of butyrate-producing Eubacterium in the human intestine verursacht. J. Appl. Microbiol. 2002;157–62.

400. Zhang C, Zhang M, Wang S, Han R, Cao Y, Hua W, et al. Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. ISME J. England; 2010;4:232–41.

401. Moya A, Ferrer M. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. Trends Microbiol. England; 2016;24:402–13.

402. Kageyama A, Benno Y. Catenibacterium mitsuokai gen. nov., sp. nov., a gram-positive anaerobic bacterium isolated from human faeces. Int. J. Syst. Evol. Microbiol. England; 2000;50 Pt 4:1595–9.

403. den Besten G, Lange K, Havinga R, van Dijk TH, Gerding A, van Eunen K, et al. Gut-derived short-chain fatty acids are vividly assimilated into host carbohydrates and lipids. Am. J. Physiol. Gastrointest. Liver Physiol. United States; 2013;305:G900-10.

404. Holling CS. Resilience and stability of ecological systems. Annu. Rev. Ecol. Syst. Annual

Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA; 1973;4:1–23.

405. Serrano-villar S, Rojo D, Martinez M, Deusch S, Vazquez-Castellanos J, Bargiela R, et al. HIV infection results in metabolic alterations in the gut microbiota different from those induced by other diseases. Sci. Rep. 2016;18:1–11.

406. Serrano-Villar S, Rojo D, Martinez-Martinez M, Deusch S, Vazquez-Castellanos JF, Bargiela R, et al. Gut Bacteria Metabolism Impacts Immune Recovery in HIV-infected Individuals. EBioMedicine. Netherlands; 2016;8:203–16.

407. EA F, XC M, Segata N, Waldron L, Reyes J, AM E. Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci USA. 2014;111:E2329–38.

408. Bashiardes S, Zilberman-Schapira G, Elinav E. Use of Metatranscriptomics in Microbiome Research. Bioinform. Biol. Insights. United States; 2016;10:19–25.

409. Gosalbes MJ, Durban A, Pignatelli M, Abellan JJ, Jimenez-Hernandez N, Perez-Cobas AE, et al. Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS One. United States; 2011;6:e17447.

410. Reck M, Tomasch J, Deng Z, Jarek M, Husemann P, Wagner-Döbler I. Stool metatranscriptomics: A technical guideline for mRNA stabilisation and isolation. BMC Genomics. 2015;16:494.

411. Bikel S, Valdez-Lara A, Cornejo-Granados F, Rico K, Canizales-Quinteros S, Soberon X, et al. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. Comput. Struct. Biotechnol. J. Netherlands; 2015;13:390–401.

412. Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. Metatranscriptomics of the human oral microbiome during health and disease. MBio. United States; 2014;5:e01012-14.

413. Booijink CCGM, Boekhorst J, Zoetendal EG, Smidt H, Kleerebezem M, de Vos WM. Metatranscriptome analysis of the human fecal microbiota reveals subject-specific expression profiles, with genes encoding proteins involved in carbohydrate metabolism being dominantly expressed. Appl. Environ. Microbiol. United States; 2010;76:5533–40.

414. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-

occurrence relationships in the human microbiome. PLoS Comput. Biol. 2012;8:e1002606.

415. Gaedke U. Ecological Networks. Anal. Biol. Networks. 2007. p. 283–304.

416. Pimm SL, Sole R V. Ecological networks and their fragility. Nature. 2006;442:259–64.

417. Widder S, Besemer K, Singer GA, Ceola S, Bertuzzo E, Quince C, et al. Fluvial network organization imprints on microbial co-occurrence networks. PNAS. 2014;111:12799–804.

418. Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and in fl ammatory bowel disease. PNAS. 2011;109.

419. Eiler A, Heinrich F, Bertilsson S. Coherent dynamics and association networks among lake bacterioplankton taxa. ISME J. England; 2012;6:330–42.

420. Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI, et al. Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. PLoS One. United States; 2013;8:e47879.

421. Zimmerman AE, Martiny AC, Allison SD. Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. ISME J. England; 2013;7:1187–99.

422. Bauer B, Jordán F, Podani J. Node centrality indices in food webs: Rank orders versus distributions. Ecol. Complex. 2010;7:471–7.

423. Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J. England; 2011;5:1414–25.

424. Faust K, Raes J. Microbial interactions: from networks to models. Nat. Rev. Microbiol. Nature Publishing Group; 2012;10:538–50.

425. Chaffron S, Rehrauer H, Pernthaler J, von Mering C. A global network of coexisting microbes from environmental and whole-genome sequence data. Genome Res. United States; 2010;20:947–59.

426. Oakley BB, Morales CA, Line J, Berrang ME, Meinersmann RJ, Tillman GE, et al. The poultry-associated microbiome: network analysis and farm-to-fork characterizations. PLoS One. United States; 2013;8:e57190.

427. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2.

428. Hannon Lab. FASTX-Toolkit FASTQ/A short-reads pre-processing tools [Internet]. Available from: http://hannonlab.cshl.edu/fastx_toolkit/

429. Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One. 2011;6.

430. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics. England; 2012;13:31.

431. Langille M, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes J, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. Nature Publishing Group; 2013;31:814–21.

432. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat. Biotechnol. 2014;32:822–8.

433. Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biol. BioMed Central Ltd; 2012;13:R122.

434. García-López R, Vazquez-Castellanos J, Moya A. Fragmentation and coverage variation in viral metagenome assemblies, and their effect in diversity calculations. Front. Bioeng. Biotechnol. 2015;3:1–15.

435. Peng Y, Leung HCM, Yiu SM, Chin FYL. Meta-IDBA: a de Novo assembler for metagenomic data. Bioinformatics. 2011;27:i94--101.

436. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38:1–15.

437. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28.

438. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob. Agents Chemother. 2013;57:3348–57.

439. Ye Y, Choi J, Tang H. RAPSearch: a fast protein similarity search tool for short reads. BMC

Bioinformatics. BioMed Central Ltd; 2011;12:159.

440. Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput. Biol. 2012;8.

441. Gu S, Fang L, Xu X. Using SOAPaligner for short reads alignment. Curr. Protoc. Bioinforma. 2013;1–17.

442. Eddy SR. Multiple alignment using hidden Markov models. Proc. Int. Conf. Intell. Syst. Mol. Biol. 1995;3:114–20.

443. McGrath KC, Thomas-Hall SR, Cheng CT, Leo L, Alexa A, Schmidt S, et al. Isolation and analysis of mRNA from environmental microbial communities. J. Microbiol. Methods. Netherlands; 2008;75:172–6.

444. Embree M, Nagarajan H, Movahedi N, Chitsaz H, Zengler K. Single-cell genome and metatranscriptome sequencing reveal metabolic interactions of an alkane-degrading methanogenic community. ISME J. England; 2014;8:757–67.

445. Dimitrov D V. The human gutome: nutrigenomics of the host-microbiome interactions. OMICS. United States; 2011;15:419–30.

446. Martinez X, Pozuelo M, Pascal V, Campos D, Gut I, Gut M, et al. MetaTrans: an open-source pipeline for metatranscriptomics. Sci. Rep. England; 2016;6:26447.

447. Westreich ST, Korf I, Mills DA, Lemay DG. SAMSA: a comprehensive metatranscriptome analysis pipeline. BMC Bioinformatics. England; 2016;17:399.

448. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 2013;8:1494–512.

449. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. United States; 2012;19:455–77.

450. Celaj A, Markle J, Danska J, Parkinson J. Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. MIcrobiome. 2014;2:1–13.

451. Toseland A, Moxon S, Mock T, Moulton V. Metatranscriptomes from diverse microbial communities: assessment of data reduction techniques for rigorous annotation. BMC Genomics. 2014;15:1–7.

452. Davids M, Hugenholtz F, Dos Santos VM, Smidt H, Kleerebezem M, Schaap PJ. Functional profiling of unfamiliar microbial communities using a validated de novo assembly metatranscriptome pipeline. PLoS One. 2016;11:1–18.

453. Hass B. TransDecoder (Find Coding Regions Within Transcripts) [Internet]. 2015. Available from: http://transdecoder.github.io/

454. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:e323.

455. Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, et al. Relating the metatranscriptome and metagenome of the human gut. Proc. Natl. Acad. Sci. U. S. A. 2014;111:E2329-38.

456. Kohn AJ. Pielou, E. C. 1975. Ecological diversity. John Wiley & Sons, New York, viii + 165 p. $14.95. Limnol. Oceanogr. 1977;22:174.

457. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput. Biol. 2012;8:e1002687.

458. The Huttenhower Lab. MaAsLin: Multivariate Association with Linear Models [Internet]. Available from: https://huttenhower.sph.harvard.edu/maaslin

459. Srinivasa S, Fitch K V, Lo J, Kadar H, Knight R, Wong K, et al. Plaque burden in HIV-infected patients is associated with serum intestinal microbiota-generated trimethylamine. AIDS. England; 2015;29:443–52.

460. Aitchison J. A New Approach to Null Correlations of Proportions. Math Geol. 1981;12:175–89.

461. TR E-Y, Dhiman N, Benavides R, CW S. Myroides odoratimimus bacteremia in a diabetic patient. Proc (Bayl Univ Med Cent). 2015;28:342–3.

462. Benedetti P, Rassu M, Pavan G, Sefton A, Pellizzer G. Septic shock, pneumonia, and soft tissue infection due to Myroides odoratimimus: Report of a case and review of Myroides infections.

Infection. 2011;39:161–5.

463. Maraki S, Sarchianaki E, Barbagadakis S. Myroides odoratimimus soft tissue infection in an immunocompetent child following a pig bite: Case report and literature review. Brazilian J. Infect. Dis. Elsevier Editora Ltda; 2012;16:390–2.

464. Zilberman-Schapira G, Zmora N, Itav S, Bashiardes S, Elinav H, Elinav E. The gut microbiome in human immunodeficiency virus infection. BMC Med. BMC Medicine; 2016;14:4–11.

465. Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ, et al. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. Cell. Elsevier Inc.; 2011;145:745–57.

466. Russell WR, Hoyles L, Flint HJ, Dumas ME. Colonic bacterial metabolites and human health. Curr. Opin. Microbiol. Elsevier Ltd; 2013;16:246–54.

467. Vujkovic-Cvijin I, Swainson LA, Chu SN, Ortiz AM, Santee CA, Petriello A, et al. Gut-Resident Lactobacillus Abundance Associates with IDO1 Inhibition and Th17 Dynamics in SIV-Infected Macaques. Cell Rep. The Authors; 2015;13:1589–97.

468. Huang L, Li L, Klonowski KD, Tompkins SM, Tripp RA, Mellor AL. Induction and role of indoleamine 2,3 dioxygenase in mouse models of influenza a virus infection. PLoS One. United States; 2013;8:e66546.

469. Lamas B, Richard ML, Leducq V, Pham H-P, Michel M-L, Da Costa G, et al. CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. Nat. Med. United States; 2016;22:598–605.

470. Craciun S, Balskus EP. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. Proc. Natl. Acad. Sci. U. S. A. United States; 2012;109:21307–12.

471. Martinez-del Campo A, Bodea S, Hamer HA, Marks JA, Haiser HJ, Turnbaugh PJ, et al. Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. MBio. United States; 2015;6.

472. Liu J, Williams B, Frank D, Dillon SM, Wilson CC, Landay AL. Inside Out: HIV, the Gut Microbiome, and the Mucosal Immune System. J. Immunol. 2017;198:605–14.

473. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure, function and diversity of the healthy human microbiome. Nature. Nature Publishing Group; 2012;486:207–14.

474. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol. 2014;32:834–41.

475. Velazquez OC, Lederer HM, Rombeau JL. Butyrate and the colonocyte. Production, absorption, metabolism, and therapeutic implications. Adv. Exp. Med. Biol. United States; 1997;427:123–34.

476. Peng L, He Z, Chen W, Holzman IR, Lin J. Effects of butyrate on intestinal barrier function in a Caco-2 cell monolayer model of intestinal barrier. Pediatr. Res. United States; 2007;61:37–41.

477. Bhogal HK, Sanyal AJ. The molecular pathogenesis of cholestasis in sepsis. Front. Biosci. (Elite Ed). United States; 2013;5:87–96.

478. Rahman AH, Aloman C. Dendritic cells and liver fibrosis. Biochim. Biophys. Acta. Netherlands; 2013;1832:998–1004.

479. Carambia A, Freund B, Schwinge D, Heine M, Laschtowitz A, Huber S, et al. TGF-beta-dependent induction of CD4(+)CD25(+)Foxp3(+) Tregs by liver sinusoidal endothelial cells. J. Hepatol. Netherlands; 2014;61:594–9.

480. Licata LA, Nguyen CT, Burga RA, Falanga V, Espat NJ, Ayala A, et al. Biliary obstruction results in PD-1-dependent liver T cell dysfunction and acute inflammation mediated by Th17 cells and neutrophils. J. Leukoc. Biol. United States; 2013;94:813–23.

481. O'Brien KM, Allen KM, Rockwell CE, Towery K, Luyendyk JP, Copple BL. IL-17A synergistically enhances bile acid-induced inflammation during obstructive cholestasis. Am. J. Pathol. United States; 2013;183:1498–507.

482. Sayin SI, Wahlstrom A, Felin J, Jantti S, Marschall H-U, Bamberg K, et al. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. Cell Metab. United States; 2013;17:225–35.

483. Zhu C, Fuchs CD, Halilbasic E, Trauner M. Bile acids in regulation of inflammation and immunity: Friend or foe? Clin. Exp. Rheumatol. 2016;34:S25–31.

484. Pols TWH, Noriega LG, Nomura M, Auwerx J, Schoonjans K. The bile acid membrane receptor TGR5 as an emerging target in metabolism and inflammation. J. Hepatol. Netherlands; 2011;54:1263–72.

485. Trauner M, Fickert P, Halilbasic E, Moustafa T. Lessons from the toxic bile concept for the pathogenesis and treatment of cholestatic liver diseases. Wien. Med. Wochenschr. Austria; 2008;158:542–8.

486. Smit JJ, Schinkel AH, Oude Elferink RP, Groen AK, Wagenaar E, van Deemter L, et al. Homozygous disruption of the murine mdr2 P-glycoprotein gene leads to a complete absence of phospholipid from bile and to liver disease. Cell. United States; 1993;75:451–62.

487. Schaap FG, Trauner M, Jansen PLM. Bile acid receptors as targets for drug development. Nat. Rev. Gastroenterol. Hepatol. England; 2014;11:55–67.

488. Sharpstone D, Neild P, Crane R, Taylor C, Hodgson C, Sherwood R, et al. Small intestinal transit, absorption, and permeability in patients with AIDS with and without diarrhoea. Gut. England; 1999;45:70–6.

12. APPENDIX

12.1. Chapter 1 supplementary information

Supplementary Figures

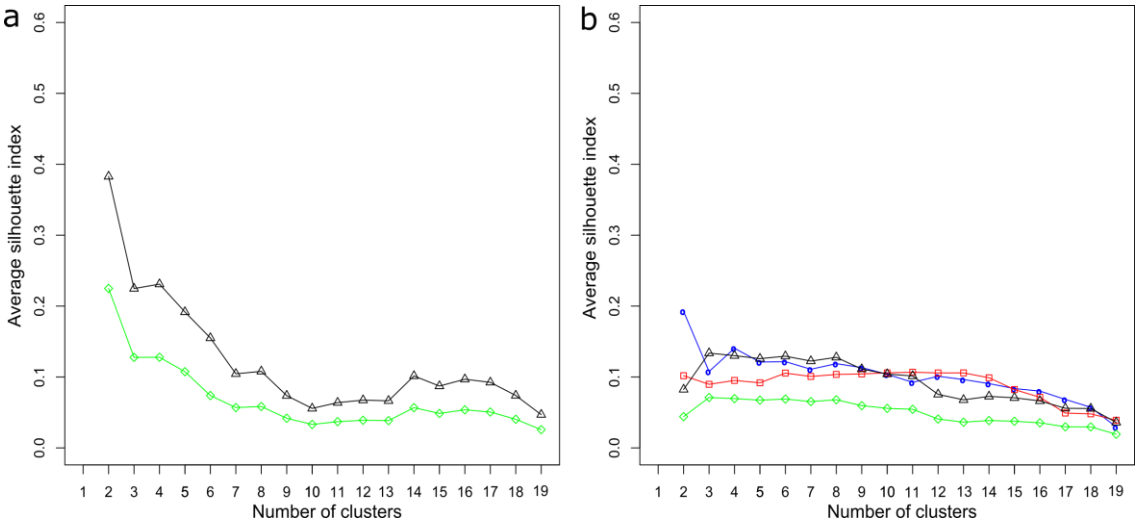


Figure S12.1.1 Average silhouette index. Average silhouette index from all the possible numbers of cluster configurations within the genus (**a**) and OTUs (**b**). The Bray-Curtis index (red squares), the Jensen−Shannon distance (green diamonds) and the Jensen−Shannon divergence (black triangles) were tested for both taxonomical levels in order to ascertain the distance that maximizes the average silhouette index. Since the weighted UniFrac distance (blue circles) could only be computed by estimating a phylogenetic tree and was incompatible for use at higher levels, it was analyzed only at the OTU level. Modified from Vazquez-Castellanos et al. (2014) [348] with permission from © 2017 Society for Mucosal Immunology.

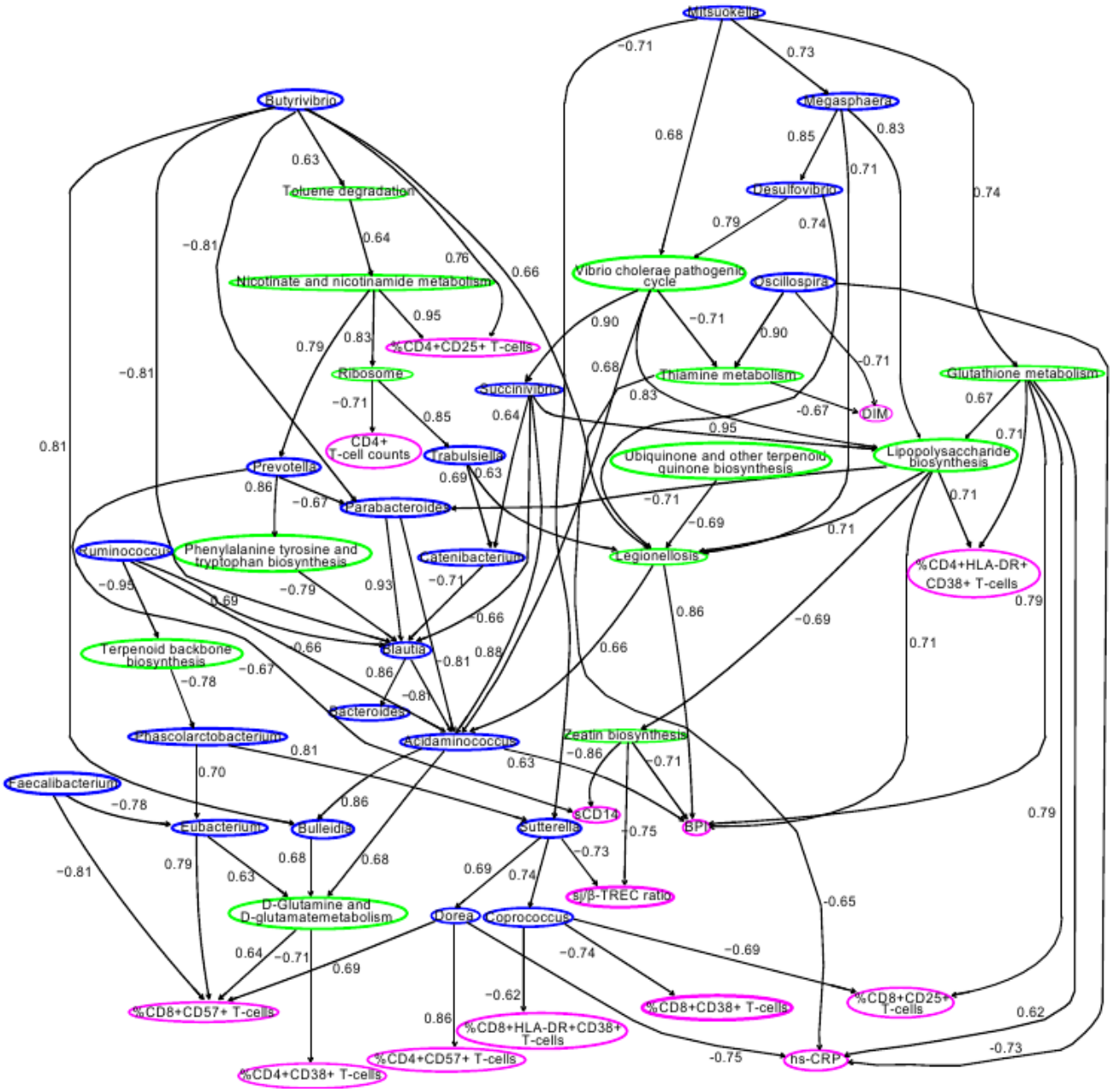


Figure S12.1.2 Total Bayesian network. The network represents the relationships between genus abundance (blue ellipses), pathway abundance (green ellipses) and markers of adaptive immunity, thymic function, and bacterial translocation (pink ellipses). Arrows indicate conditional dependencies between variables. The Spearman correlation coefficient is indicated next to the lines. Modified from Vazquez-Castellanos et al. (2014) [348] with permission from © 2017 Society for Mucosal Immunology.

Supplementary Tables

Table S12.1.1 Clinical variables of participants.

	Cases N = 15	Controls N = 15	p-value[a]	q-value[f]
Clinical characteristics				
Age	43 (34-48)	48.5 (31-54)	0.76	0.97
Sex ratio (F/M)	3/12	7/8	0.13	0.33
Hypertensive (Y/N)	1/14	2/11	0.99	0.99
Smoker (Y/N)	7/8	2/12	0.99	0.99
Body mass index (kg/m^2)	24.5 (23.2-24.7)	23.5 (21.2-28.3)	0.63	0.92
Framingham risk score (%)	4.5 (1-7)	2 (1-6)	0.12	0.33
Time from HIV diagnosis to initiation of ART (months)	14 (3-25)	NA	-	-
Time on HIV suppression (months)	74 (52-113)	NA	-	-
Nadir CD4+ T-cell count (cells/µL)	203 (127-284)	NA	-	-
CD4+ T-cell count (cells/µL)	584 (466-794)	762 (645-927)	-	-
CD4/CD8 ratio	1.2 (0.9-1.3)	1.5 (1.2-1.9)	-	-
Nadir CD4+ T-cell count (cells/µL)	203 (127-284)	NA	-	-
Metabolic profile in plasma				
Glucose (mg/dL)	91 (81-96)	89 (86-95)	0.84	0.97
Creatinine (mg/dL)	1.0 (0.9-1.1)	0.9 (0.7-1.0)	0.1	0.31
Total cholesterol (mg/dL)	190 (169-214)	201 (157-230)	0.62	0.92
LDL cholesterol (mg/dL)	106 (97-124)	114 (81-136)	0.87	0.97
HDL cholesterol (mg/dL)	55 (50-63)	56 (49-75)	0.56	0.92
Triglycerides (mg/dL)	106 (78-155)	75 (61-176)	0.38	0.71
25-hidroxy-vitamin D (mg/dL)	28.2 (21.7-36)	28.4 (21.9-33.9)	0.93	0.99
Markers of innate immunity				
Inflammation				
hs-CRP[b] (mg/L)	0.18 (0.06-0.47)	0.08 (0.04-0.29)	0.24	0.54
IL6[c] (pg/mL)	2 (2-2)	2 (1-2.6)	0.51	0.89
Thrombosis				

Dimers-D (ng/mL)	199 (168-301)	212 (122-304)	0.87	0.97
Bacterial translocation				
sCD14[d] (ng/mL)	1663 (1483-1958)	1439.5 (1263-1516)	0.05	0.20
BPI[e] (ng/mL)	28.3 (3.0-113.6)	10.0 (4.9-12.3)	0.66	0.92
Endothelial function				
ADMA (µM/L)	0.97 (0.89-1.12)	1.10 (1.05-1.10)	0.25	0.54
Markers of adaptive immunity				
T cell markers				
CD4+ T-cells				
%HLADR+CD38+	2.2 (1.8-2.6)	1.1 (0.6-1.2)	<0.001	0.01
%CD38+	16.1 (14.3-22.8)	11.6 (10.7-13.2)	<0.001	0.01
%CD25+	4.3 (3.7-6.6)	2.8 (1.9-4.1)	0.01	0.06
%CD57+	5.7 (4.0-9.5)	2.5 (1.6-5.7)	0.04	0.19
CD8+ T-cells				
%HLADR+CD38+	3.6 (2.5-7.1)	1.5 (1.1-1.7)	<0.001	0.01
%CD38+	7.3 (6.1-12.9)	5.4 (4.0-8.4)	0.01	0.06
%CD25+	0.4 (0.3-0.7)	5.4 (4.0-8.4)	0.29	0.58
%CD57+	26.5 (17.5-41.8)	23.1 (15.8-43.8)	0.77	0.97
Thymic function				
sj/β-TREC ratio	5.7 (0-13.6)	18-5 (3.2-57.8)	0.06	0.21

All values are expressed as median (P25-P75).

[a]Analysis was performed using a Wilcoxon rank-sum test. P is probability at α=0.05.

[b]High-sensitivity C reactive protein.

[c] Interleukin-6.

[d] Soluble CD14.

[e] Bactericidal-permeability increasing protein.

[f] p-value adjusted according to the Benjamini-Hochberg method.

Table S12.1.2 LEfSe biomarker statistics for KEGG pathways.

Condition	Biomarker pathway	Log LDA	p-value[a]	%Control coverage[b]	%Case coverage[c]
Control	Starch and sucrose metabolism [PATH:ko00500]	3.08	0.03	51.04	47.92
Control	Glycolysis / Gluconeogenesis [PATH:ko00010]	2.96	0.01	60.44	50.55
Control	Valine, leucine, and isoleucine degradation [PATH:ko00280]	2.94	0.00	44.26	29.51
Control	Lysosome [PATH:ko04142]	2.89	0.04	18.18	8.08
Control	Pyruvate metabolism [PATH:ko00620]	2.87	0.04	78.95	65.79
Control	Glycine, serine, and threonine metabolism [PATH:ko00260]	2.87	0.04	45.88	42.35
Control	Fatty acid metabolism [PATH:ko00071]	2.75	0.01	34.69	24.49
Control	Histidine metabolism [PATH:ko00340]	2.66	0.04	51.28	56.41
Control	PPAR signaling pathway [PATH:ko03320]	2.63	0.01	7.14	5.36
Control	Ascorbate and aldarate metabolism [PATH:ko00053]	2.59	0.01	32.43	24.32
Control	Tryptophan metabolism [PATH:ko00380]	2.55	0.00	16.92	15.38
Control	Polycyclic aromatic hydrocarbon degradation [PATH:ko00624]	2.55	0.01	9.68	6.45
Control	Lysine degradation [PATH:ko00310]	2.50	0.01	23.33	20.00
Control	Caprolactam degradation [PATH:ko00930]	2.49	0.01	28.57	14.29

Control	Dioxin degradation [PATH:ko00621]	2.47	0.01	24.00	8.00
Control	Xylene degradation [PATH:ko00622]	2.47	0.01	18.75	6.25
Control	Benzoate degradation [PATH:ko00362]	2.47	0.03	17.24	11.49
Control	Steroid hormone biosynthesis [PATH:ko00140]	2.44	0.00	10.53	10.53
Control	MAPK signaling pathway - yeast [PATH:ko04011]	2.40	0.02	2.44	2.44
Control	Naphthalene degradation [PATH:ko00626]	2.30	0.02	18.75	15.63
Control	Phosphonate and phosphinate metabolism [PATH:ko00440]	2.30	0.03	17.65	14.71
Control	Proximal tubule bicarbonate reclamation [PATH:ko04964]	2.29	0.03	25.00	25.00
Control	Geraniol degradation [PATH:ko00281]	2.21	0.04	37.50	37.50
Case	Ribosome [PATH:ko03010]	3.19	0.01	52.78	40.28
Case	Lipopolysaccharide biosynthesis [PATH:ko00540]	3.15	0.00	58.82	44.12
Case	Phenylalanine, tyrosine, and tryptophan biosynthesis [PATH:ko00400]	2.87	0.00	53.73	53.73
Case	Vibrio cholerae pathogenic cycle [PATH:ko05111]	2.87	0.01	13.95	20.93
Case	Legionellosis [PATH:ko05134]	2.87	0.00	7.35	7.35
Case	Terpenoid backbone biosynthesis [PATH:ko00900]	2.79	0.04	41.67	33.33

Case	Fatty acid biosynthesis [PATH:ko00061]	2.68	0.04	53.33	46.67
Case	Nicotinate and nicotinamide metabolism [PATH:ko00760]	2.68	0.03	53.66	53.66
Case	Thiamine metabolism [PATH:ko00730]	2.59	0.01	59.09	63.64
Case	Ubiquinone and other terpenoid-quinone biosynthesis [PATH:ko00130]	2.52	0.04	36.36	25.00
Case	Zeatin biosynthesis [PATH:ko00908]	2.47	0.01	12.50	12.50
Case	Toluene degradation [PATH:ko00623]	2.47	0.01	17.65	11.76

[a] Analysis was performed using a Wilcoxon rank-sum test. P is probability at α=0.05.

[b] The percentage of control coverage was calculated as the observed number of KOs per pathway divided by the total number of KOs for each condition.

[c] The percentage of case coverage was calculated as the observed number of KOs per pathway divided by the total number of KOs for each condition.

12.2. Chapter 2 supplementary information

Supplementary Figures

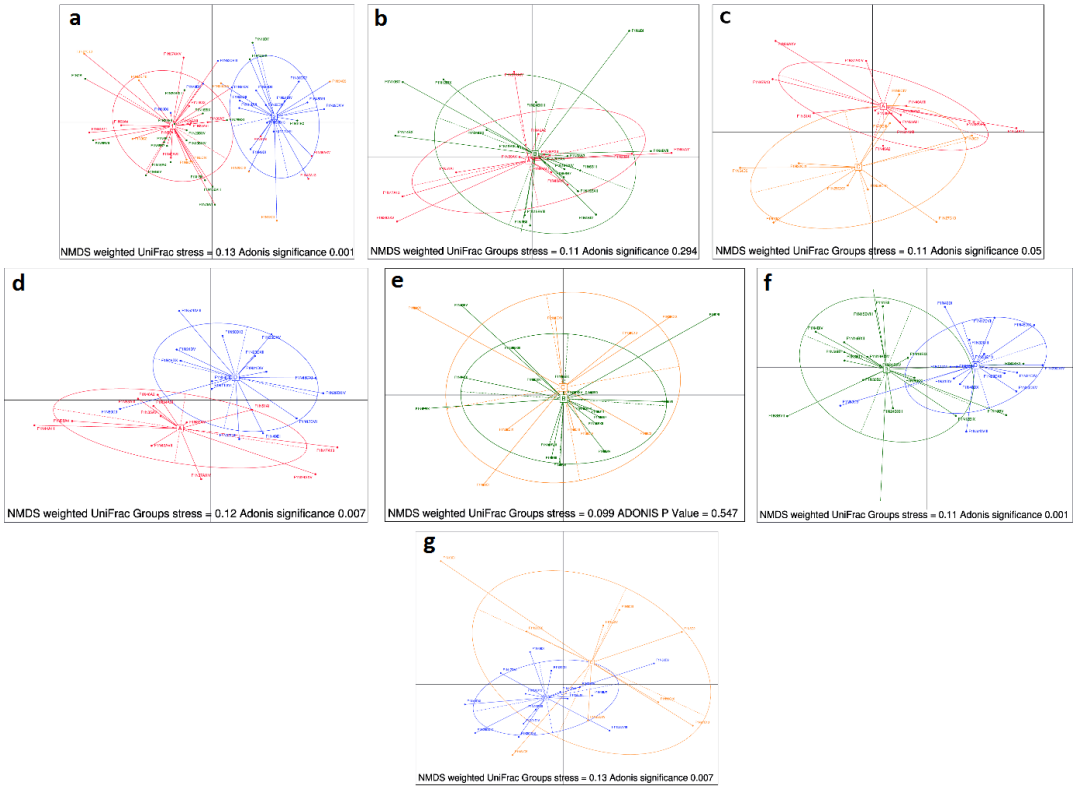


Figure S12.2.1 Clustering analysis of the microbiota composition at baseline. Genera distribution composition based on a Nonmetric Multidimensional Scaling (NMDS) at baseline. (**a**) Clustering between HIV$^-$ (in blue) and HIV$^+$ (VU, red; INR, orange, IR, green) (**b**) Clustering between INR (in orange) and VU (in red) (**c**) Clustering between IR (in green) and VU (in red). (**d**) Clustering between HIV$^-$ (in blue) and VU (in red). (**e**) Clustering between INR (in orange) and IR (in green). (**f**) Clustering between HIV$^-$ (in blue) and INR (in orange). (**g**) Clustering between HIV$^-$ (in blue) and IR (in green). Abbreviations: VU, viremic untreated; INR, immunological non-responder, IR, immunological responder. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

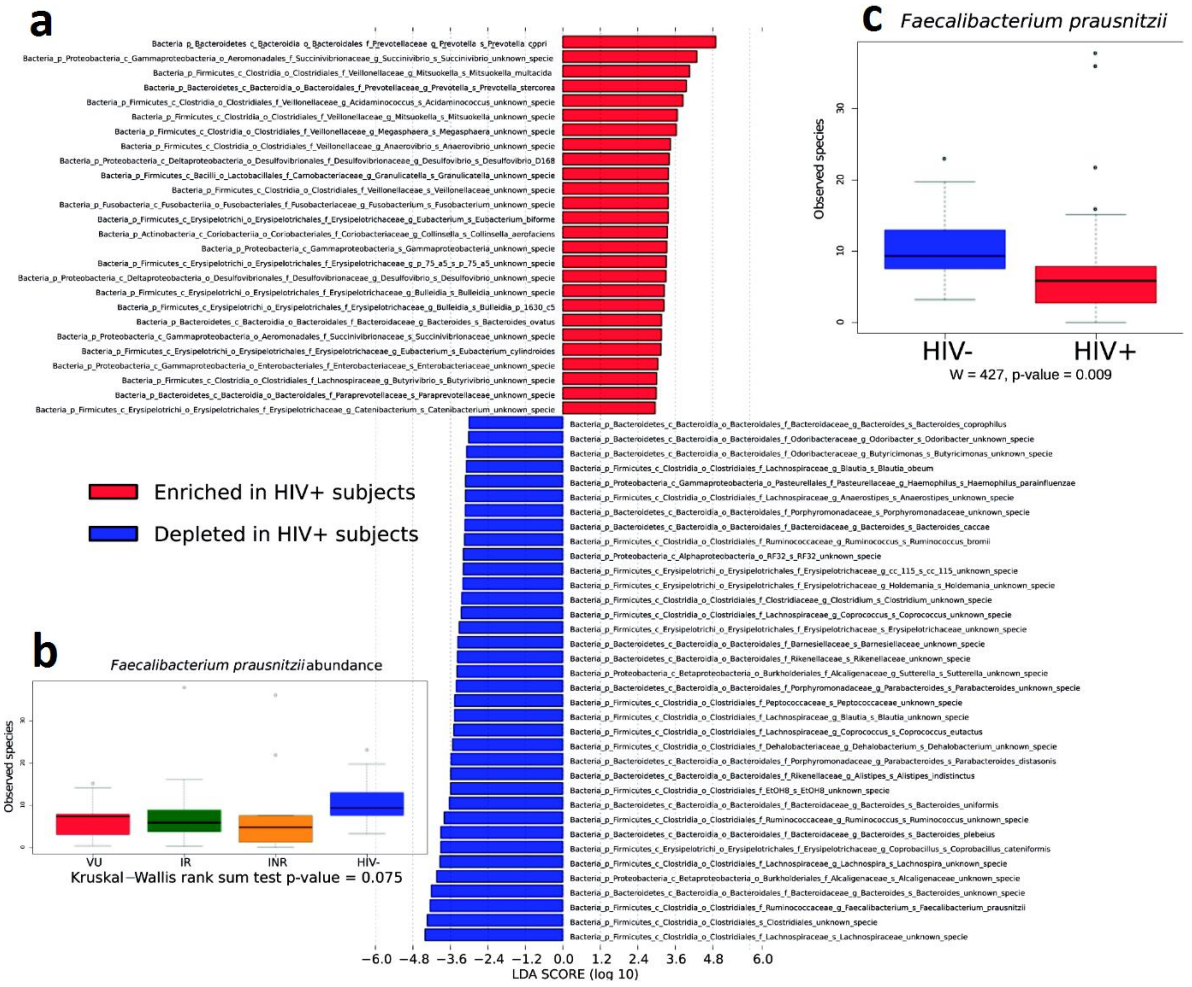


Figure S12.2.2 Bacterial taxa driving HIV-associated dysbiosis using LEfSe analysis. (**a**) All HIV+ patients vs. HIV- individuals. (**b**) and (**c**) abundance of *Faecalibacterium prausnitzii* in different groups. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

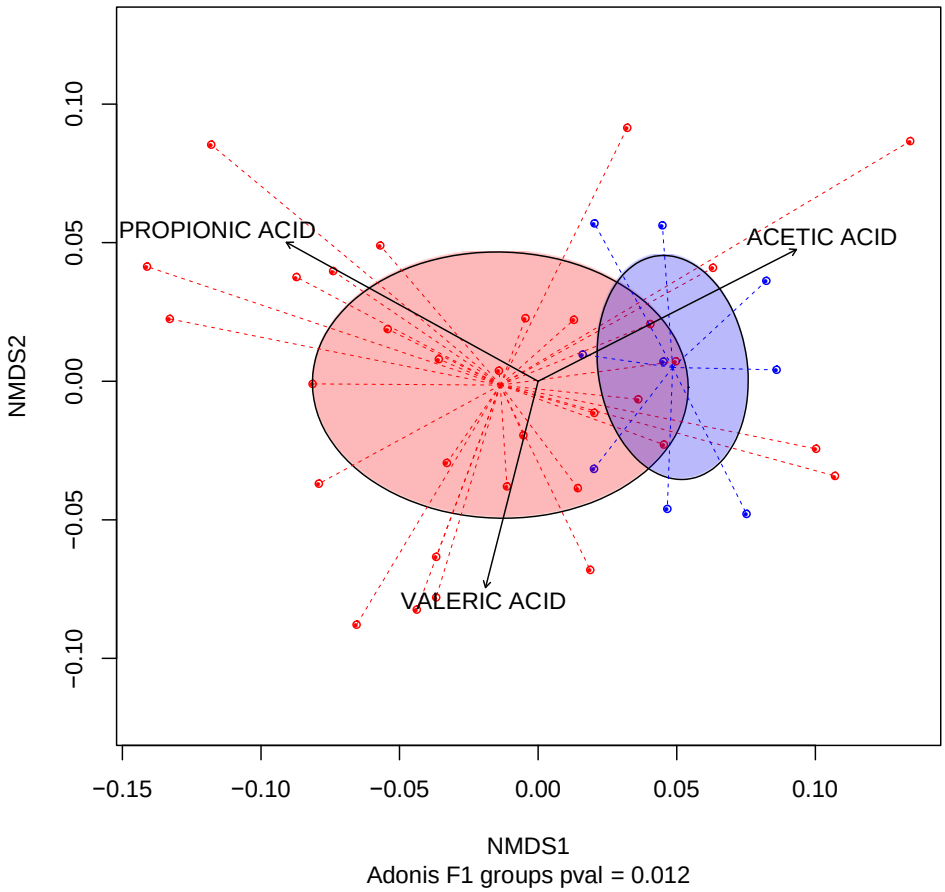


Figure S12.2.3 SCFA profile characterizing HIV-infected individuals at baseline. NMDS analysis was used to determine the clustering of the samples based on the relative abundances of SCFAs. The input distance matrix was calculated using the Bray-Curtis dissimilarity index and the statistical difference between the HIV^{+} subjects and the HIV^{-} subjects were calculated by the ADONIS test. Reproduced from Serrano-Villar et al. (2016) [315] with permission from © 2017 Society for Mucosal Immunology.

Supplementary Tables

Table S12.2.1 Between-group comparison of alpha diversity at baseline.

Active arm (n=34)				
Reference group	**Mean± SD**	**HIV$^+$ group**	**Mean± SD**	**p value**
HIV-	980.64±118.46	VU	1127.69±295.94	0.07
	980.64±118.46	IR	833.47±261.79	0.19
	980.64±118.46	INR	689.46±259.23	0.03
ACE				
HIV-	941.13±161.74	VU	1079.55±275.21	0.3
	941.13±161.74	IR	805.06±254.08	0.27
	941.13±161.74	INR	651.88±224.48	0.01
Chao1				
HIV-	548.71±78.1	VU	590±126.57	0.37
	548.71±78.1	IR	458.6±140.76	0.16
	548.71±78.1	INR	379.88±120.66	0.02
Observed species				
HIV-	7.78±0.46	VU	7.97±0.59	0.41
	7.78±0.46	IR	7.17±0.92	0.13
Shannon	7.78±0.46	INR	6.72±0.87	0.01

Table S12.2.2 Comparison of changes in alpha diversity.

Group	Index	Baseline statistics (mean±SD)	Post-intervention statistics (mean±SD)	p-value
VU	Observed species	590±126.57	615.11±68.84	0.43
	Chao1	1079.55±275.21	1101.42±161.58	0.57
	ACE	1127.69±295.94	1165.79±177.6	0.5
	Shannon	7.97±0.59	8.11±0.28	0.43
IR	Observed species	458.6±140.76	473.7±184.06	0.85
	Chao1	805.06±254.08	894.33±385.37	0.43
	ACE	833.47±261.79	914.95±408.31	0.43
	Shannon	7.17±0.92	7.22±1.06	0.49
INR	Observed species	379.88±120.66	423.12±86.48	0.31
	Chao1	651.88±224.48	766.01±235.06	0.11
	ACE	689.46±259.23	784.69±220.16	0.11
	Shannon	6.72±0.87	7.09±0.56	0.74
HIV$^-$	Observed species	548.71±78.1	538.71±117.17	0.3
	Chao1	941.13±161.74	967.59±245.11	0.94
	ACE	980.64±118.46	1001.58±246.79	1.0
	Shannon	7.78±0.46	7.72±0.58	0.38

Table S12.2.3 Between-group comparison of beta diversity at baseline.

Group Distances

Pair1	Pair2	p-value
VU	IR	0.328
VU	INR	1.0
VU	HIV$^-$	0.192
IR	INR	0.1
IR	HIV$^-$	0.1
INR	HIV$^-$	<0.001

Table S12.2.4 Comparison of changes in beta diversity.

Pair 1	Pair 2	statistic	p-value
Pre-VU vs. Pre-VU	Post-VU vs. Post-VU	834	0.0361
Pre-IR vs. Pre-IR	Post-IR vs. Post-IR	1401	0.0015
Pre-INR vs. Pre-INR	Post-INR vs. Post-INR	579	0.0018
Pre-HIV$^-$ vs. Pre-HIV$^-$	Post-VIH$^-$ vs. Post-HIV$^-$	387	<0.0001

Abbreviations: Pre, pre-intervention; Post, post-intervention; VU, viremic untreated; INR, immunological non-responder, IR, munological responder, HIV$^-$, HIV-uninfected.

Table S12.2.5 Effect of HIV serostatus on gut microbiota composition after adjustment by sexual orientation.

Independent variable	Sum of squares	F model	R^2	p-value
HIV serostatus	1.24	7.77	0.12	0.0002
Study group	0.72	2.26	0.07	0.0154
Sexual orientation	0.89	5.58	0.09	0.0004
Study group*Sexual orientation$^-$	0.39	1.23	0.03	0.2493

Table S12.2.6 Correlations between shifts in species abundance and changes in SCFA abundance.

SCFA	Taxa	GLM coefficient	Spearman index	p-value	q-value[a]
Acetic acid	Bacteria_p_Bacteroidetes_c_Bacteroidia_o_Bacteroidales_f_Barnesiellaceae_s_Barnesiellaceae_unknown_specie	-3,42E-044	-0,4192	0,0064	0,0099
Acetic acid	Bacteria_p_Bacteroidetes_c_Bacteroidia_o_Bacteroidales_f_Porphyromonadaceae_g_Parabacteroides_s_Parabacteroides_distasonis	6,63E-044	0,3870	0,0124	0,0139
Butyric acid	Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Lachnospiraceae_g_Roseburia_s_Roseburia_faecis	1,29E-044	0,5363	0,0003	0,0016
Butyric acid	Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Lachnospiraceae_g_Lachnospira_s_Lachnospira_unknown_specie	6,43E-045	0,4821	0,0016	0,0038
Butyric acid	Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Ruminococcaceae_g_Ruminococcus_s_Ruminococcus_torques	6,97E-044	0,4013	0,0093	0,0128
Butyric acid	Bacteria_p_Bacteroidetes_c_Bacteroidia_o_Bacteroidales_f_Rikenellaceae_g_Alistipes_s_Alistipes_indistinctus	-3,29E-045	-0,3892	0,0119	0,0139
Butyric acid	Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Ruminococcaceae_g_Faecalibacterium_s_Faecalibacterium_prausnitzii	5,12E-038	0,3902	0,0121	0,0139
Propionic acid	Bacteria_p_Lentisphaerae_c_Lentisphaeria_o_Z20_f_R4-45B_s_R4-45B_unknown_specie	-1,65E-044	-0,5081	0,0007	0,0023
Propionic acid	Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Mogibacteriaceae_s_Mogibacteriaceae_unknown_specie	-3,09E-044	-0,4863	0,0013	0,0032
Propionic acid	Bacteria_p_Bacteroidetes_c_Bacteroidia_o_Bacteroidales_f_Bacteroidaceae_g_Bacteroides_s_Bacteroides_ovatus	-1,28E-044	-0,4395	0,0040	0,0068
Propionic acid	Bacteria_p_Bacteroidetes_c_Bacteroidia_o_Bacteroidales_s_Bacteroidales_unknown_specie	7,01E-044	-0,4388	0,0041	0,0068
Propionic acid	Bacteria_p_Bacteroidetes_c_Bacteroidia_o_Bacteroidales_f_Bacteroidaceae_g_Bacteroides_s_Bacteroides_unknown_specie	-6,90E-039	-0,4145	0,0075	0,0106
Propionic acid	Bacteria_p_Verrucomicrobia_c_Verruco-5_o_WCHB1-41_f_RFP12_s_RFP12_unknown_specie	-1,02E-043	-0,3974	0,0101	0,0135
Propionic acid	Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Veillonellaceae_g_Mitsuokella_s_Mitsuokella_unknown_specie	1,37E-045	0,3876	0,0123	0,0139
Propionic acid	Bacteria_p_Proteobacteria_c_Betaproteobacteria_o_Burkholderiales_f_Oxalobacteraceae_g_Oxalobacter_s_Oxalobacter_formigenes	-3,81E-045	-0,3777	0,0149	0,0149
Valeric acid	Bacteria_p_Firmicutes_c_Erysipelotrichi_o_Erysipelotrichales_f_Erysipelotrichaceae_g_Catenibacterium_s_Catenibacterium_unknown_specie	9,26E-045	0,5225	0,0005	0,0020

Valeric acid	Bacteria_p_Firmicutes_c_Erysipelotrichi_o_Erysipelotrichales_f_Erysipelotrichaceae_g_Bulleidia_s_Bulleidia_p-1630-c5	3,01E-044	0,4611	0,0024	0,0050
Valeric acid	Bacteria_p_Proteobacteria_c_Betaproteobacteria_s_Betaproteobacteria_unknown_specie	-1,50E-043	-0,3841	0,0132	0,0139
Valeric acid	Bacteria_p_Firmicutes_c_Clostridia_o_Clostridiales_f_Veillonellaceae_g_Mitsuokella_s_Mitsuokella_multacida	6,72E-044	0,3825	0,0136	0,0139

[a]Benjamini-Hodge adjusted P-value.

Table S12.2.7 Associations between changes in species contributing to the HIV-associated dysbiosis, SCFA abundance and peripheral biomarkers using generalized linear models (GLM).

GLM Models in the viremic untreated group.

Response variable	Predictor	GLM Coef	Spearman Correlation	Spearman p-value	Benjamini & Hochberg correction
BUTYRIC ACID	Lachnospira sp.	0.22	0.767	0.021	0.087
BUTYRIC ACID	Faecalibacterium prausnitzii	0.016	0.633	0.076	0.097
CD4/CD8ratio	Faecalibacterium prausnitzii	-0.001	-0.617	0.086	0.097
hs-CRP	BUTYRIC ACID	-0.002	-0.717	0.037	0.087
hs-CRP	Collinsella aerofaciens	-0.009	-0.7	0.043	0.087
hs-CRP	Roseburia faecis	-0.029	-0.617	0.086	0.097
hs-CRP	Faecalibacterium prausnitzii	-0.001	-0.6	0.097	0.097
sCD14	BUTYRIC ACID	-0.275	-0.7	0.043	0.087

GLM models in the HIV+ group

Response variable	Predictor	GLM Coef	Spearman Cor	Spearman p-value	Benjamini & Hochberg correction
BUTYRIC ACID	Faecalibacterium prausnitzii	0.016	0.499	0.008	0.033
BUTYRIC ACID	Roseburia faecis	0.317	0.464	0.015	0.044
BUTYRIC ACID	VALERIC ACID	0.041	0.451	0.018	0.046
BUTYRIC ACID	Veillonellaceae sp.	-0.024	-0.422	0.028	0.059
BUTYRIC ACID	Coprococcus eutactus	0.202	0.395	0.041	0.075
BUTYRIC ACID	Parabacteroides sp.	-0.064	-0.327	0.096	0.099
%CD4+HLA-DR+CD38+T-cells	Paraprevotellaceae sp.	-0.003	-0.479	0.011	0.04

%CD4+HLA-DR+CD38+T-cells	Coprococcus eutactus	-0.011	-0.444	0.02	0.049
%CD4+HLA-DR+CD38+T-cells	Bulleidia p 1630 c5	-0.007	-0.391	0.044	0.075
%CD4+HLA-DR+CD38+T-cells	Bulleidia p 1630 c5.1	-0.007	-0.391	0.044	0.075
%CD4+HLA-DR+CD38+T-cells	Acidaminococcus sp.	0.003	0.366	0.061	0.089
%CD4+HLA-DR+CD38+T-cells	Prevotella copri	0	0.361	0.064	0.091
%CD4+HLA-DR+CD38+T-cells	Veillonellaceae sp.	0.001	0.356	0.068	0.094
%CD4+HLA-DR+CD38+T-cells	Butyrivibrio sp.	-0.017	-0.352	0.072	0.096
%CD4+HLA-DR+CD38+T-cells	Bacteroides sp.	-0.001	-0.346	0.078	0.097
%CD4+HLA-DR+CD38+T-cells	Collinsella aerofaciens	0.004	0.333	0.09	0.098
CD4/CD8ratio	Enterobacteriaceae sp.	0	-0.479	0.012	0.04
CD4/CD8ratio	Veillonellaceae sp.	0	0.418	0.03	0.06
CD4/CD8ratio	Sutterella sp.	0	0.366	0.061	0.089
CD4/CD8ratio	Erysipelotrichaceae sp.	0	0.337	0.086	0.098
IL-6	Butyricimonas sp.	1.594	0.43	0.025	0.055
IL-6	Coprococcus sp.	0.014	0.386	0.047	0.076
IL-6	Bacteroides ovatus	0.013	-0.374	0.054	0.084
IL-6	BUTYRIC ACID	0.036	0.333	0.09	0.098
hs-CRP	Sutterella sp.	-0.008	-0.453	0.018	0.046
hs-CRP	Collinsella aerofaciens	-0.004	-0.348	0.075	0.097
hs-CRP	Mitsuokella sp.	0.002	0.344	0.079	0.097
PROPIONIC ACID	Veillonellaceae sp.	0.089	0.596	0.001	0.012

PROPIONIC ACID	VALERIC ACID	0.182	0.536	0.004	0.031
PROPIONIC ACID	Bacteroidales sp.	-0.057	-0.529	0.005	0.031
PROPIONIC ACID	Megasphaera sp.	0.153	0.52	0.005	0.033
PROPIONIC ACID	Mitsuokella sp.	0.071	0.5	0.008	0.033
PROPIONIC ACID	Mogibacteriaceae sp.	-1.594	-0.497	0.008	0.033
PROPIONIC ACID	Prevotella copri	0.012	0.43	0.025	0.055
PROPIONIC ACID	Bacteroides ovatus	-0.177	-0.385	0.047	0.076
PROPIONIC ACID	Parabacteroides distasonis	-0.08	-0.338	0.085	0.098
PROPIONIC ACID	Enterobacteriaceae sp.	-0.173	-0.328	0.095	0.099
PROPIONIC ACID	Paraprevotellaceae sp.	-0.09	-0.326	0.097	0.099
VALERIC ACID	Bulleidia p 1630 c5	0.06	0.716	0	0.001
VALERIC ACID	Bulleidia p 1630 c5.1	0.06	0.716	0	0.001
VALERIC ACID	Megasphaera sp.	0.032	0.673	0	0.002
VALERIC ACID	Coriobacteriaceae sp.	0.207	0.529	0.005	0.031
VALERIC ACID	Catenibacterium sp.	0.027	0.504	0.007	0.033
VALERIC ACID	Mitsuokella sp.	0.021	0.468	0.014	0.044
VALERIC ACID	BUTYRIC ACID	0.014	0.451	0.018	0.046
VALERIC ACID	Bacteroides ovatus	-0.046	-0.391	0.044	0.075
VALERIC ACID	Bacteroides uniformis	-0.051	-0.337	0.086	0.098
VALERIC ACID	Blautia obeum	0.286	0.325	0.099	0.099

Table S12.2.8 The power of the study to detect differences in SCFA abundance.

To detect differences in baseline vs. post-prebiotic intervention levels of systemic markers.				
Group	**Systemic marker**	**Power**	**W-T p-value**	**W-T Permutation Test p-value**
VU	CD4+ T-cell counts	1	0.82	0.8189819
VU	CD8+ T-cell counts	1	0.652	0.64546455
VU	CD4/CD8 ratio	0.02587823	0.91	0.90849085
VU	Flow mediated dilation (%)	0.24333059	0.91	0.90079008
VU	Peak hyperemic flow (m/s)	0.98539999	0.652	0.65236524
VU	%CD4+HLA-DR+CD38+ T-cells	0.99258252	0.027	0.02490249
VU	%CD4+ CD25+	0.63983317	0.203	0.20942094
VU	%CD4+ CD57+	0.99737786	0.301	0.30453045
VU	%CD8+HLA-DR+CD38+ T-cells	1	0.055	0.05360536
VU	%CD8+ CD25+	0.12006346	0.301	0.30323032
VU	%CD8+ CD57+	0.3072068	0.91	0.90129013
VU	sj/β-TREC	0.6706314	0.203	0.20492049
VU	sj/β-TREC ratio	0.25428828	0.496	0.49824982
VU	IL6 (pg/ml)	0.70542533	0.201	0.2520252
VU	Hs-CRP (mg/L)	0.02701124	0.726	0.77337734
VU	Dimers-D (ng/ml)	1	0.859	0.84318432
VU	ADMA (uM/L)	0.04359359	0.203	0.20052005
VU	BPI (ng/ml)	1	0.294	0.38843884
VU	sCD14 (ng/ml)	1	0.91	0.90969097
VU	TP53	0.09040781	0.012	0.00650065
VU	CCR2	0.37117613	0.012	0.0070007
VU	APOBEC3G	0.10311894	0.023	0.01830183
VU	CCL2	1	0.023	0.02050205
VU	BCL2	0.03491969	0.151	0.13011301
VU	STAT1	0.10169469	0.073	0.06630663
VU	SCARB1	0.07616985	0.012	0.00770077

VU	CXCL10	0.99999967	0.009	0.00490049
VU	IFNA1	0.54288698	0.009	0.00450045
VU	PPARG	0.03951369	0.403	0.3960396
VU	PDCD1	0.10329375	0.282	0.28972897
IR	CD4+ T-cell counts	1	0.126	0.1260126
IR	CD8+ T-cell counts	1	0.432	0.43544354
IR	CD4/CD8 ratio	0.03137404	0.232	0.23422342
IR	Flow mediated dilation (%)	0.99999045	0.16	0.16361636
IR	Peak hyperemic flow (m/s)	0.78605199	1	1
IR	%CD4+HLA-DR+CD38+ T-cells	0.19686365	0.037	0.02910291
IR	%CD4+ CD25+	0.16957274	0.625	0.6229623
IR	%CD4+ CD57+	0.05029652	0.695	0.69866987
IR	%CD8+HLA-DR+CD38+ T-cells	0.03567946	0.432	0.42884288
IR	%CD8+ CD25+	0.03202526	1	1
IR	%CD8+ CD57+	0.04078954	0.846	0.85028503
IR	sj/β-TREC	0.0519716	0.922	0.92289229
IR	sj/β-TREC ratio	0.07649536	0.557	0.55455546
IR	IL6 (pg/ml)	0.09619101	0.832	0.90169017
IR	Hs-CRP (mg/L)	0.04229619	0.878	0.8719872
IR	Dimers-D (ng/ml)	1	0.636	0.59145915
IR	ADMA (uM/L)	0.02742676	0.557	0.56265627
IR	BPI (ng/ml)	1	0.058	0.0560056
IR	sCD14 (ng/ml)	1	0.16	0.16251625
IR	TP53	0.04040568	0.05	0.04890489
IR	CCR2	0.10919866	0.148	0.16531653
IR	APOBEC3G	0.05058716	0.079	0.07720772
IR	CCL2	0.3731573	0.03	0.01980198
IR	BCL2	0.04444139	0.039	0.02630263
IR	STAT1	0.06908158	0.039	0.02280228

IR	SCARB1	0.04135833	0.079	0.07930793
IR	CXCL10	1	0.013	0.01050105
IR	IFNA1	0.0603315	0.121	0.12651265
IR	PPARG	0.7566191	0.098	0.08950895
IR	PDCD1	0.19951826	0.05	0.04960496
INR	CD4+ T-cell counts	0.56873373	0.674	0.66986699
INR	CD8+ T-cell counts	1	0.461	0.46554655
INR	CD4/CD8 ratio	0.70714118	0.016	0.01610161
INR	Flow mediated dilation (%)	0.99763724	0.107	0.10581058
INR	Peak hyperemic flow (m/s)	0.25784604	0.233	0.23212321
INR	%CD4+HLA-DR+CD38+ T-cells	0.06850644	0.547	0.5379538
INR	%CD4+ CD25+	0.0483401	0.844	0.84838484
INR	%CD4+ CD57+	0.53326776	0.195	0.19571957
INR	%CD8+HLA-DR+CD38+ T-cells	0.05283075	0.547	0.54265427
INR	%CD8+ CD25+	0.114252	0.383	0.38313831
INR	%CD8+ CD57+	0.99246144	0.383	0.37783778
INR	sj/β-TREC	0.77408521	0.041	0.03960396
INR	sj/β-TREC ratio	0.12634462	0.383	0.38433843
INR	IL6 (pg/ml)	0.99999999	0.052	0.04550455
INR	Hs-CRP (mg/L)	0.22374038	0.035	0.03230323
INR	Dimers-D (ng/ml)	0.99205549	0.547	0.54585459
INR	ADMA (uM/L)	0.05875855	0.25	0.25112511
INR	BPI (ng/ml)	1	0.313	0.31333133
INR	sCD14 (ng/ml)	1	0.25	0.24562456
INR	TP53	0.08336272	0.042	0.04040404
INR	CCR2	0.62923391	0.014	0.00850085
INR	APOBEC3G	0.03503434	0.624	0.61236124
INR	CCL2	0.15236113	0.944	0.90409041
INR	BCL2	0.04108921	0.528	0.52565257

INR	STAT1	0.05534426	0.528	0.53265327
INR	SCARB1	0.06314945	0.058	0.06030603
INR	CXCL10	1	0.528	0.52865287
INR	IFNA1	0.03047288	0.183	0.17981798
INR	PPARG	0.99938577	0.233	0.24182418
INR	PDCD1	0.98568278	0.107	0.10641064
HIV-	CD4+ T-cell counts	1	0.813	0.81138114
HIV-	CD8+ T-cell counts	1	0.937	0.94089409
HIV-	CD4/CD8 ratio	0.10848698	0.578	0.57325733
HIV-	Flow mediated dilation (%)	1	0.093	0.09470947
HIV-	Peak hyperemic flow (m/s)	1	0.375	0.36723672
HIV-	%CD4+HLA-DR+CD38+ T-cells	0.0533615	0.469	0.47824782
HIV-	%CD4+ CD25+	0.03898973	0.937	0.93989399
HIV-	%CD4+ CD57+	0.09819543	0.469	0.47344734
HIV-	%CD8+HLA-DR+CD38+ T-cells	0.12671403	0.078	0.07980798
HIV-	%CD8+ CD25+	0.03735285	0.938	0.93989399
HIV-	%CD8+ CD57+	0.97523529	0.578	0.58035804
HIV-	sj/β-TREC	0.11089219	0.937	0.93829383
HIV-	sj/β-TREC ratio	0.4479971	0.156	0.16281628
HIV-	IL6 (pg/ml)	0.13566428	1	1
HIV-	Hs-CRP (mg/L)	0.03118241	1	1
HIV-	Dimers-D (ng/ml)	1	0.578	0.5719572
HIV-	ADMA (uM/L)	0.03284584	0.207	0.19331933
HIV-	BPI (ng/ml)	1	0.022	0.01490149
HIV-	sCD14 (ng/ml)	1	0.156	0.14761476
HIV-	TP53	0.05411772	0.034	0.03020302
HIV-	CCR2	0.11284618	0.271	0.30383038
HIV-	APOBEC3G	0.03661191	0.352	0.36473647
HIV-	CCL2	0.13959823	0.034	0.03240324

HIV-	BCL2	0.099222	0.022	0.01580158
HIV-	STAT1	0.11531544	0.034	0.03090309
HIV-	SCARB1	0.07914949	0.022	0.01680168
HIV-	CXCL10	0.02875013	0.672	0.67166717
HIV-	IFNA1	0.25331006	0.022	0.0160016
HIV-	PPARG	0.02791027	0.799	0.77867787
HIV-	PDCD1	0.10170813	0.034	0.03310331
	AVERAGE	0.42245085	0.36866667	0.36874521
	SD	0.41716589	0.32624655	0.32700899

To detect differences in baseline SCFA abundance.

			One-Way Permutation
SCFA	**Power**	**K-W p-value**	**Test p-value**
ISOB.ACID	0.124	0.031	0.031
BUT.ACID	0.122	0.029	0.045
ISOVAL.ACID	0.121	0.108	0.076
VALERIC.ACID	0.115	0.185	0.54
AVERAGE	0.1205	0.08825	0.173
SD	0.003872983	0.07424902	0.24538813

To detect differences in baseline vs. post-prebiotic intervention in SCFA abundance.

				W-T Permutation
Group	**SCFA**	**Power**	**W-T p-value**	**Test p-value**
VU	ISOB.ACID	0.18333219	0.57	0.57185719
VU	BUT.ACID	0.97371978	0.039	0.0410041
VU	ISOVAL.ACID	0.38722395	0.496	0.49894989
VU	VALERIC.ACID	0.51066867	0.027	0.02840284

IR	ISOB.ACID	0.06385258	0.77	0.77317732
IR	BUT.ACID	0.53283198	0.492	0.48824882
IR	ISOVAL.ACID	0.17098553	0.77	0.77247725
IR	VALERIC.ACID	0.14629355	0.695	0.69716972
INR	ISOB.ACID	0.0708002	0.547	0.54455446
INR	BUT.ACID	0.03347488	0.547	0.55065507
INR	ISOVAL.ACID	0.09032126	0.461	0.46674667
INR	VALERIC.ACID	0.9839282	0.039	0.03940394
HIV-	ISOB.ACID	0.06435771	0.469	0.47784778
HIV-	BUT.ACID	0.46503332	0.375	0.36923692
HIV-	ISOVAL.ACID	0.03811596	0.688	0.68806881
HIV-	VALERIC.ACID	0.02735107	0.938	0.93509351
	AVERAGE	0.29639318	0.4951875	0.49643089
	SD	0.3194279	0.26918624	0.26907211

12.3. Chapter 3 supplementary information

Supplementary Results

S12.3.1. Functional metagenomic pipeline comparison

At the time of the present thesis, there was no a consensus methodology for the study of the metagenomic functional characterization. The different bioinformatic strategies mainly varied between the use of unassembled reads vs contigs and the previous ORF prediction before the read mapping against the functional databases. Taking this into account, eight different pipelines were performed in order to compare the effects of the bioinformatic strategies over the functional annotations. The variation between them was the use of an assembly software (Ray-meta and Meta-IDBA) vs the unassembled reads and the previous use of an ORF prediction software (MetaGeneMark) vs the direct read mapping against the database (coverage approximation). Additionally, in the comparison, we also included the *in-silico* metagenome predictions retrieved by PICRUST as well as a modified version of the Nielsen et al., 2014 [432] pipeline approximation (see in the Chapter 3 methods section "6.2.3. Metagenomic functional annotation strategies"). This last, proposed by the Metahit consortium, allowed the discovery of new microbial organisms by means of its ORF relative abundance, thus validating its accuracy in functional characterization. The summary of these approximations can be seen in **Table 6.1**.

The results retrieved from the ADONIS test and the NMDS analysis (**Figure S12.3.1**) shows that all the eight functional annotations present significant differences among the gene composition between the four groups and between the HIV+ and HIV- conditions. This indicates that the effect of the sample segregation is given by the effect of the drastic change in the bacterial gene composition rather than biases given the bioinformatics analysis. The results obtained from the Mantel test (**Table S12.3.1**) showed that for almost all the methods exist correlations between their distance matrices, with the exception of the methods that were based on the coverage approximation. This difference was also highlighted when comparing with the Nielsen et al. method [432] from which all the coverage methods differ (**Table S12.3.1**). The coverage annotations are based on translating the query sequencing into an amino acid sequence (using rapsearch2) prior to the local alignment, and then the algorithm assigns as putative genes all the no-overlapping high-quality hits. This step could be prone to assign long multi-domain proteins as different gene and generate false positive annotations. Taking these biases in mind we select the custom Nielsen et al.,

2014 [432] pipeline, in addition, similar approximation have been widely used in metagenomics [354,432] and it contains the advantage of the increased coverage of the 454 reads.

Supplementary Supplemental Figures

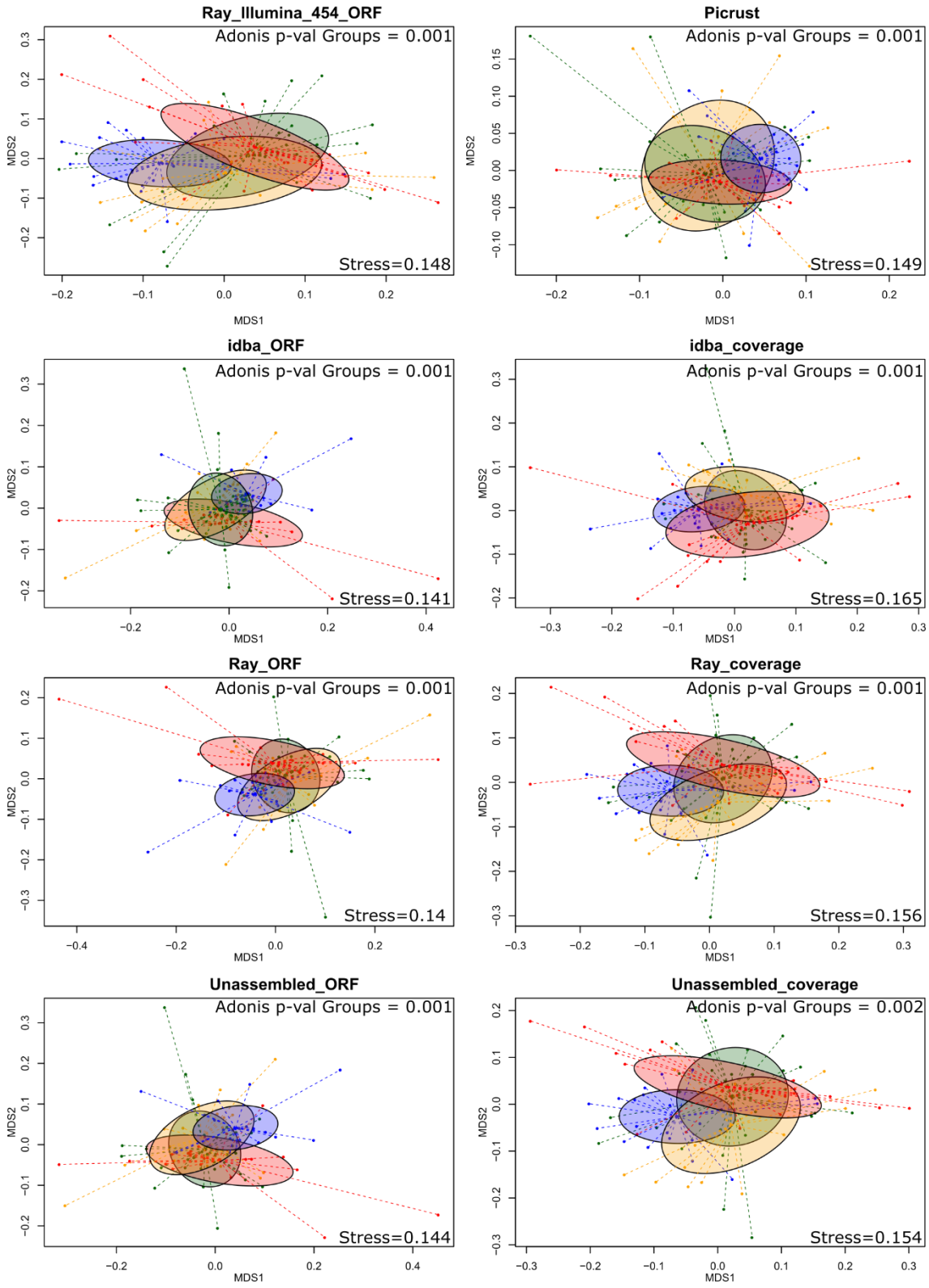


Figure S12.3.1 Comparison of the microbiota KO gene composition between the different bioinformatic strategies. NMDS analysis of the KO gene composition retrieved from the bioinformatic strategies. Ray_illumina_454_ORF (Nielsen et al.(ref)

modified pipeline), PICRUST (*in-silico* PICRUST gene prediction), idba_ORF (Meta-IDBA assembly and ORF prediction), idba_coverage (Meta-IDBA assembly and read mapping), Ray_ORF (Ray-Meta assembly and ORF prediction), Ray_coverage Ray-Meta assembly and read mapping), Unassembled_ORF (no assembly and ORF prediction), Unassembled_coverage no assembly and read mapping). The VU (red), IR (green), INR (orange) and HIV- (blue) subjects are represented by the ellipses which represent 70% of the samples; the groups cluster configuration was validated using the ADONIS test (p-value < 0.01) by the four groups of the cohort and between HIV+ vs the HIV- subjects. INR, immunological non-responder; IR, immunological responder; VU, viremic untreated.

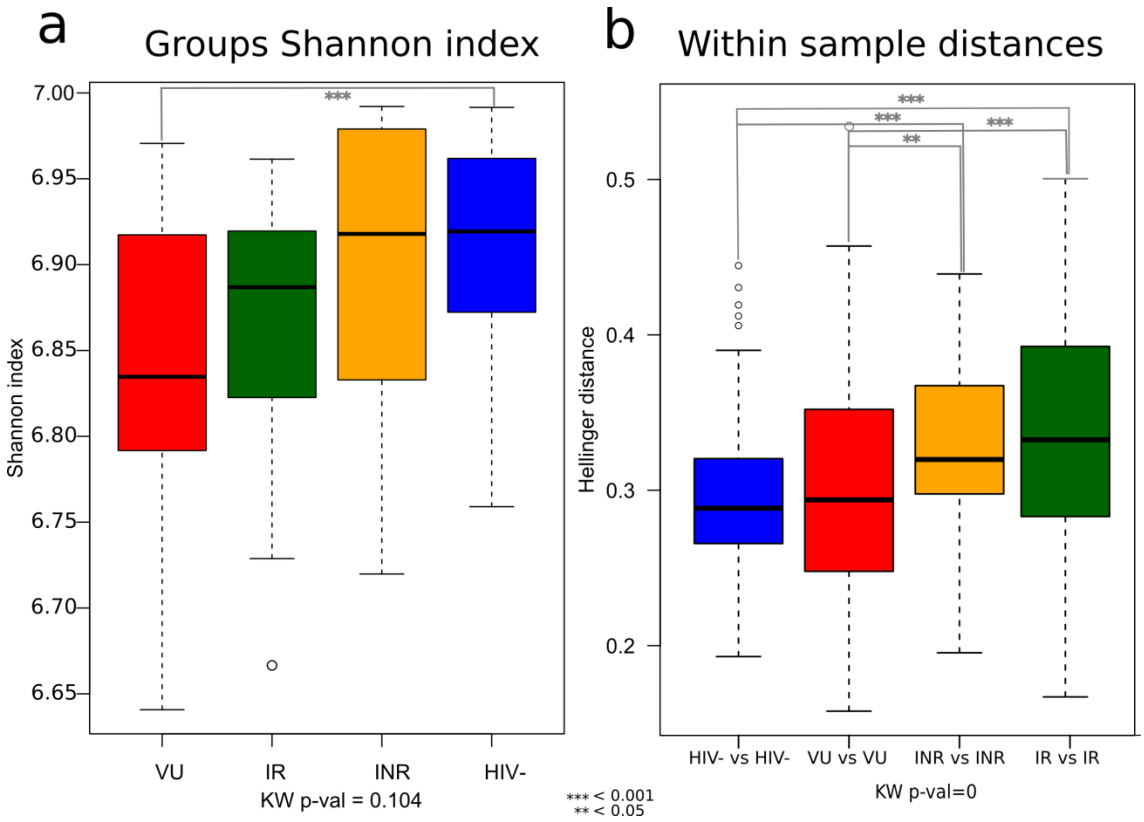


Figure S12.3.2 Diversity parameters of the metagenomic functional annotations. (**a**) Shannon diversity index and (**b**) beta diversity pairwise Hellinger distances from the groups calculated with the KO abundance matrix. VU (red), IR (green), INR (orange) and HIV- (blue) groups. The group differences were compared using the Kruskal-Wallis test while the group pairwise comparison was set by the Wilcoxon signed-rank test. INR, immunological non-responder; IR, immunological responder; VU, viremic untreated.

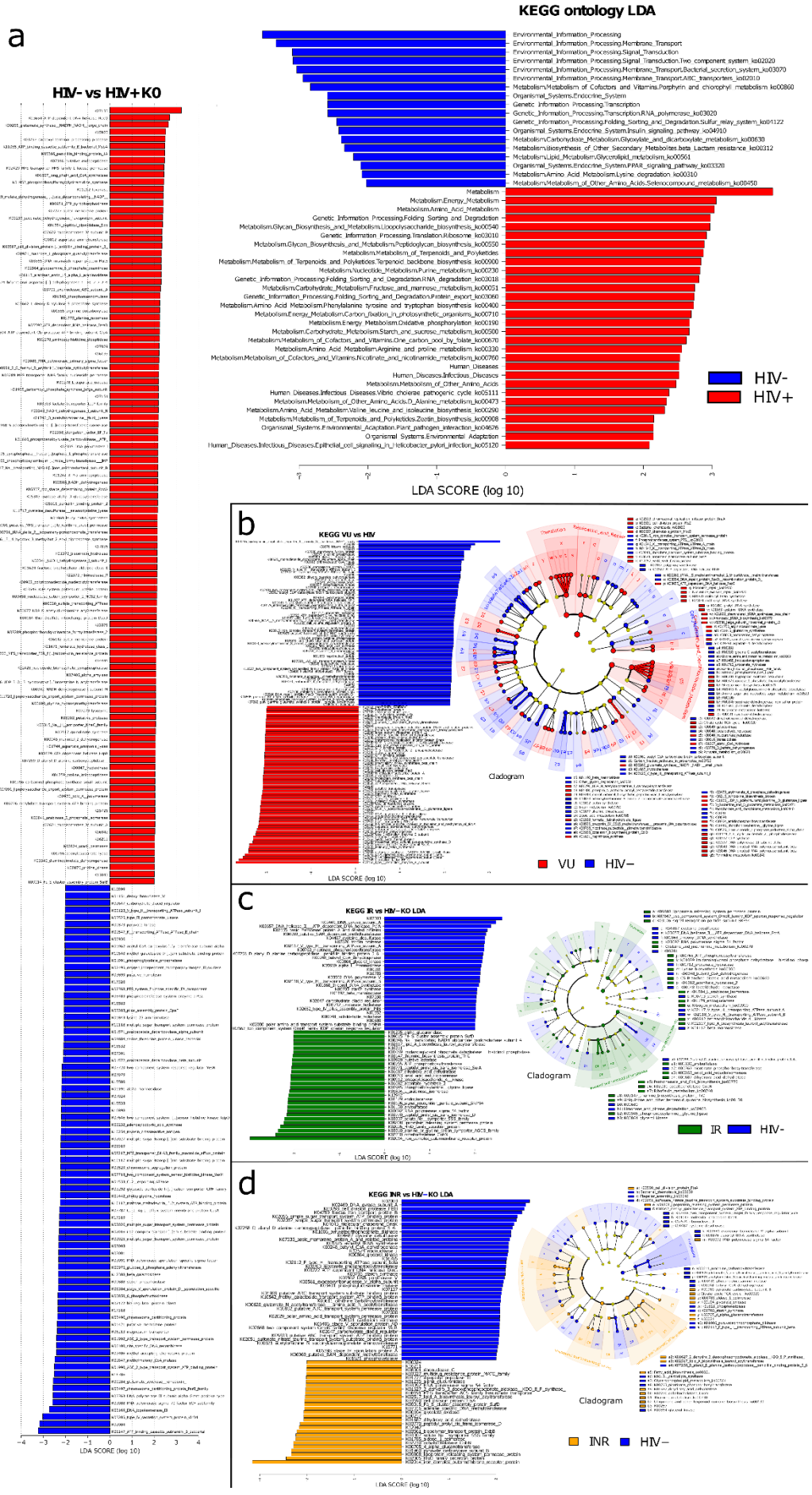

a
KEGG ontology LDA
HIV- vs HIV+KO
HIV-
HIV+
LDA SCORE (log 10)
b
KEGG VU vs HIV
Cladogram
VU
HIV−
LDA SCORE (log 10)
c
KEGG IR vs HIV−KO LDA
Cladogram
IR
HIV−
LDA SCORE (log 10)
d
KEGG INR vs HIV−KO LDA
Cladogram
INR
HIV−
LDA SCORE (log 10)

Figure S12.3.3 Metagenomic KEGG Orthology (KO) biomarkers. (**a**) The linear discriminative analysis (LDA) effect size LEfSe) analysis between the HIV+ (in red) and HIV- (in blue) subjects. (**b**) VU (red) and HIV- (blue), (**c**) IR (green) and HIV- (blue) and (**d**) INR (orange) and HIV- (blue). LDA scores (log 10) for the most discriminative KO from each of the cohort groups. The cladogram represents the biomarkers of the upper hierarchical classes within the KEGG database. INR, immunological non-responder; IR, immunological responder; VU, viremic untreated.

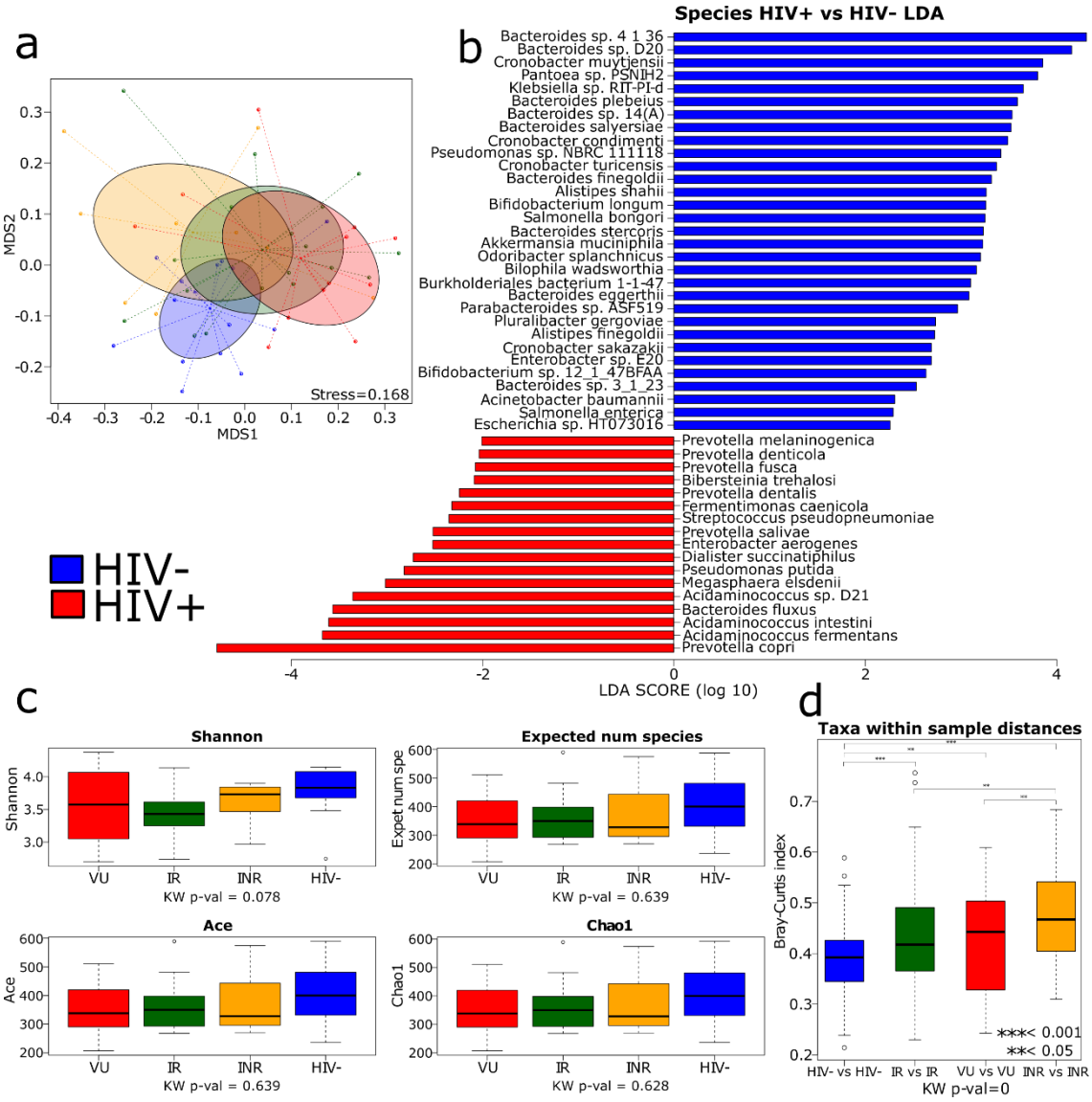


Figure S12.3.4 Species diversity from the metagenomic data set (**a**) NMDS analysis of the taxonomic species composition. The VU (red), IR (green), INR (orange) and HIV- (blue) subjects are represented by the ellipses which represent 70% of the samples; the groups cluster configuration was validated using the ADONIS test for the four groups of the cohort and for HIV+ vs the HIV- subjects (p-value=0.002 and p-value=0.003 respectively). (**b**) The linear discriminative analysis (LDA) effect size LEfSe analysis between the HIV+ (in red) and HIV- (in blue) subjects. LDA scores (log 10) for the most discriminative species in controls are represented on the positive scale (blue bars), whereas LDA-negative scores indicate enriched pathways in HIV+ subjects (red bars). (**c**) Alpha diversity estimator for the four groups of the cohort, the differences between the groups were assessed by the Kruskal-Wallis test. (**d**) Beta diversity pairwise Bray-Curtis dissimilarity from the taxonomic composition of the groups. The group differences were compared using the Kruskal-Wallis test while the group pairwise comparisons were set by the Wilcoxon signed-rank test. INR, immunological non-responder; IR, immunological responder; VU, viremic untreated.

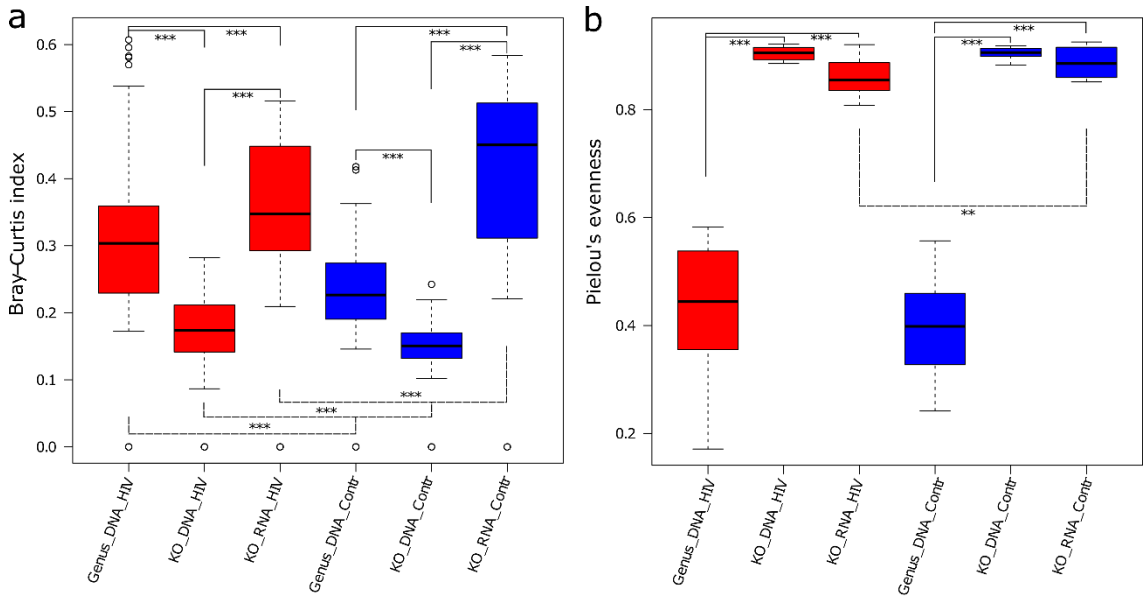


Figure S12.3.5 Diversity comparisons between metagenomic and metatranscriptomic data. (**a**) Bray-Curtis pairwise differences and (**b**) Pielou's evenness index differences between the genus and KO gene composition from HIV+ (red) and HIV- (blue) subjects. The group pairwise comparison was set by the Wilcoxon signed-rank test.

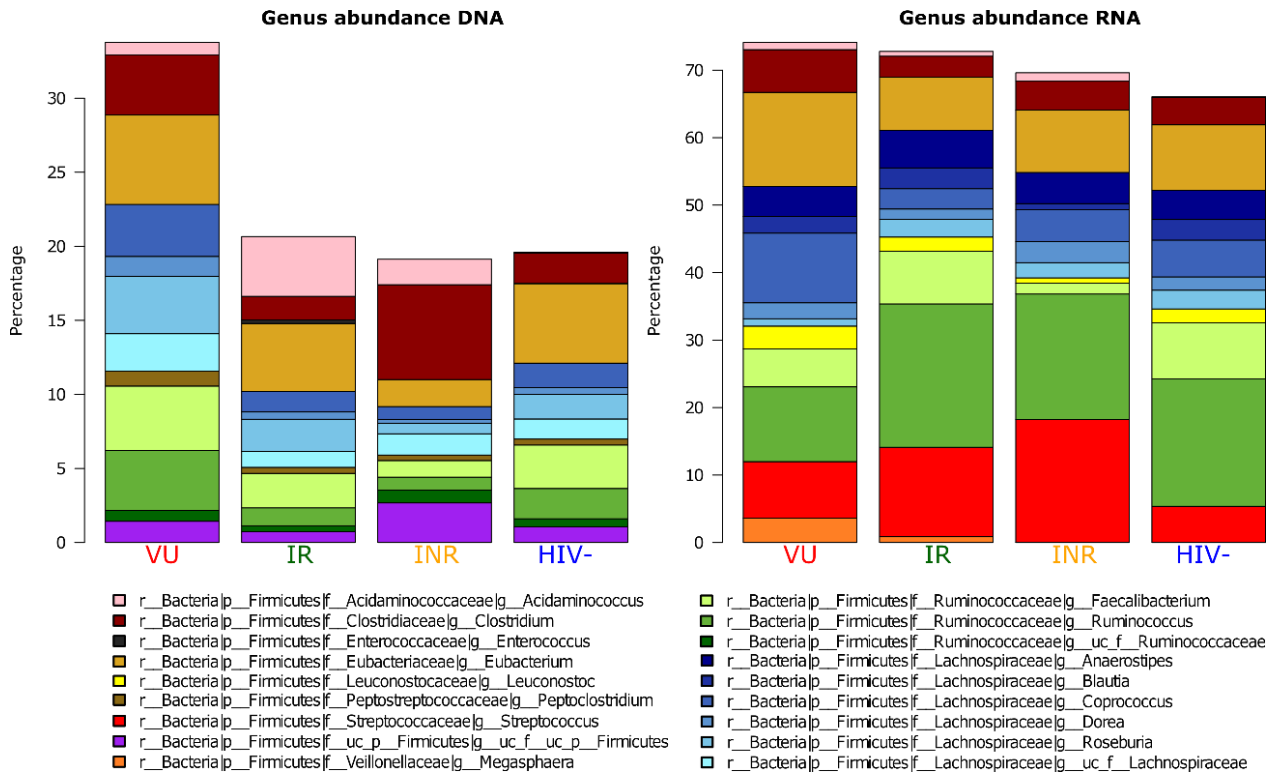


Figure S12.3.6 Metagenomic and metatranscriptomic genus relative abundance. Barplot representation of the most abundant genus (relative abundance mean > 0.5 and be present in at least 70% of the samples).

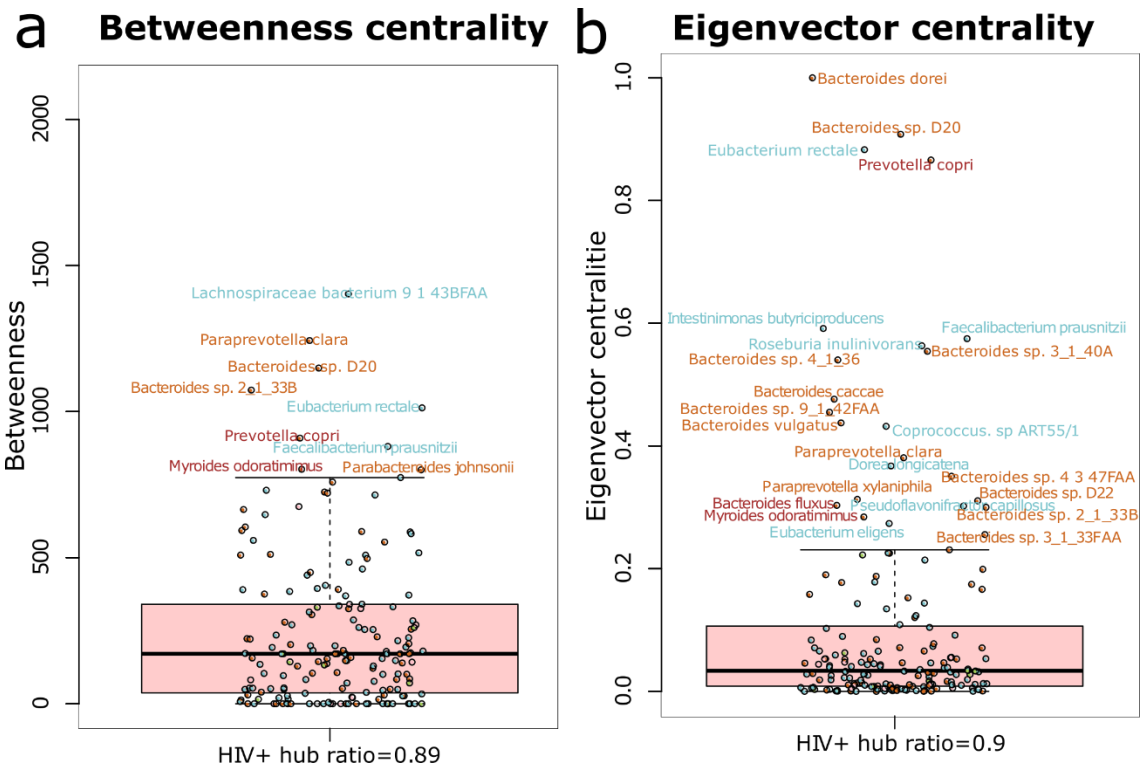


Figure S12.3.7 HIV+ species centrality. (**a**) Betweenness centrality which quantifies the number of times that a node act as a bridge along the shortest path between two other nodes. (**b**) eigenvector centrality which measures the influence of a node in a network by assigning a weighted score that takes into account its own degree centrality and the centrality of the nodes from which it is connected. The nodes represent species belonging to Firmicutes (cyan), Bacteroidetes (orange), Actinobacteria (pink) and Proteobacteria (green) phyla. Only the outlier's species (quantile 95) were label within the boxplot, highlighting those that also were set as HIV+ LEfSe biomarkers (*Prevotella copri*, *Myroides odoratimimus* and *Bacteroides fluxus*).

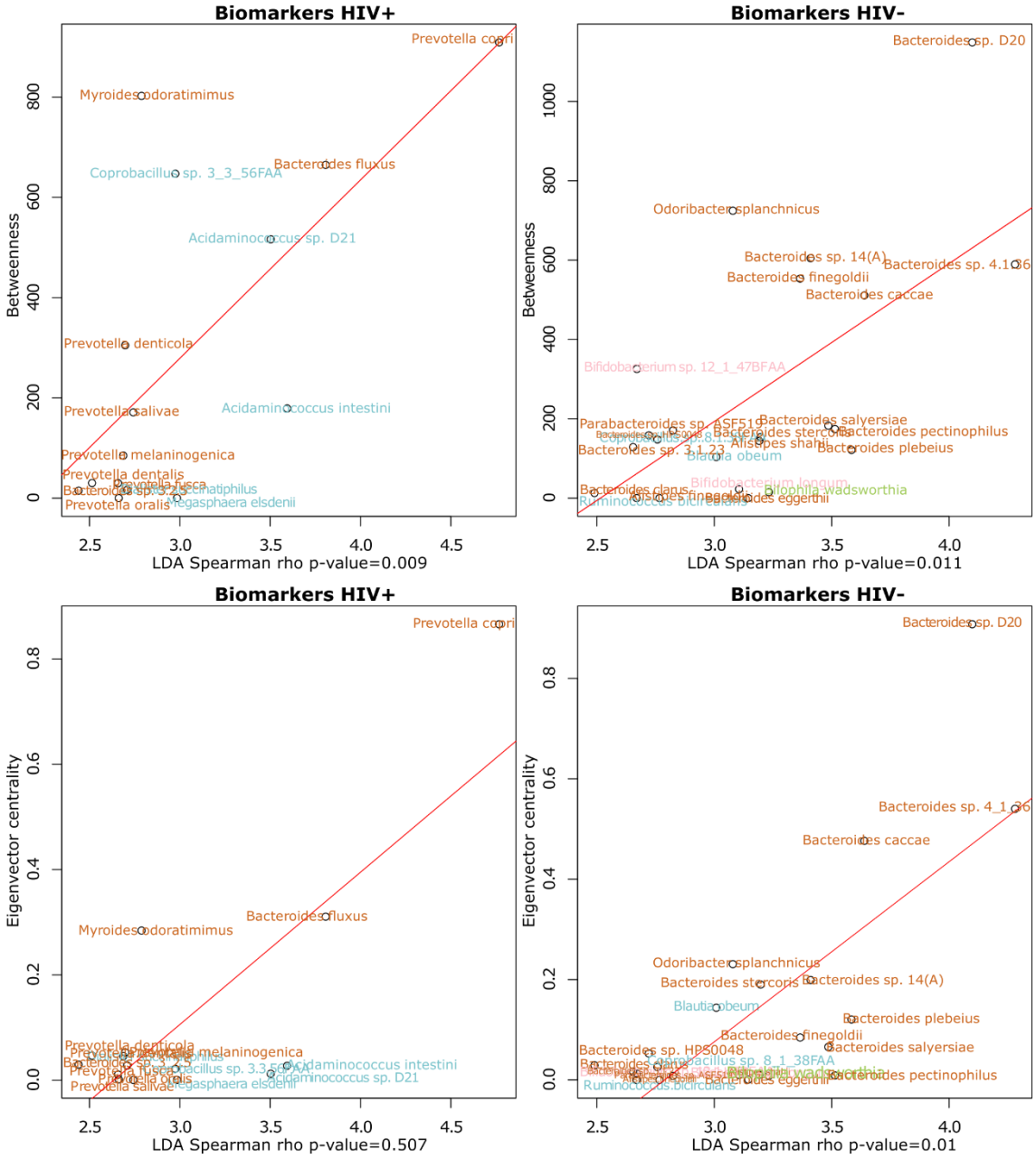


Figure S12.3.8 Taxonomic biomarkers centrality correlation. Spearman correlation between the LDA scores (log 10) of the taxonomic LEfSe biomarkers and its corresponding centrality score, the betweenness and the eigenvector centralities: The regression line is represented in red. Correlations were set for the HIV+ and the HIV- biomarkers independently. Species belong to Firmicutes (cyan), Bacteroides (orange), Actinobacteria (pink) and Proteobacteria (green) phyla.

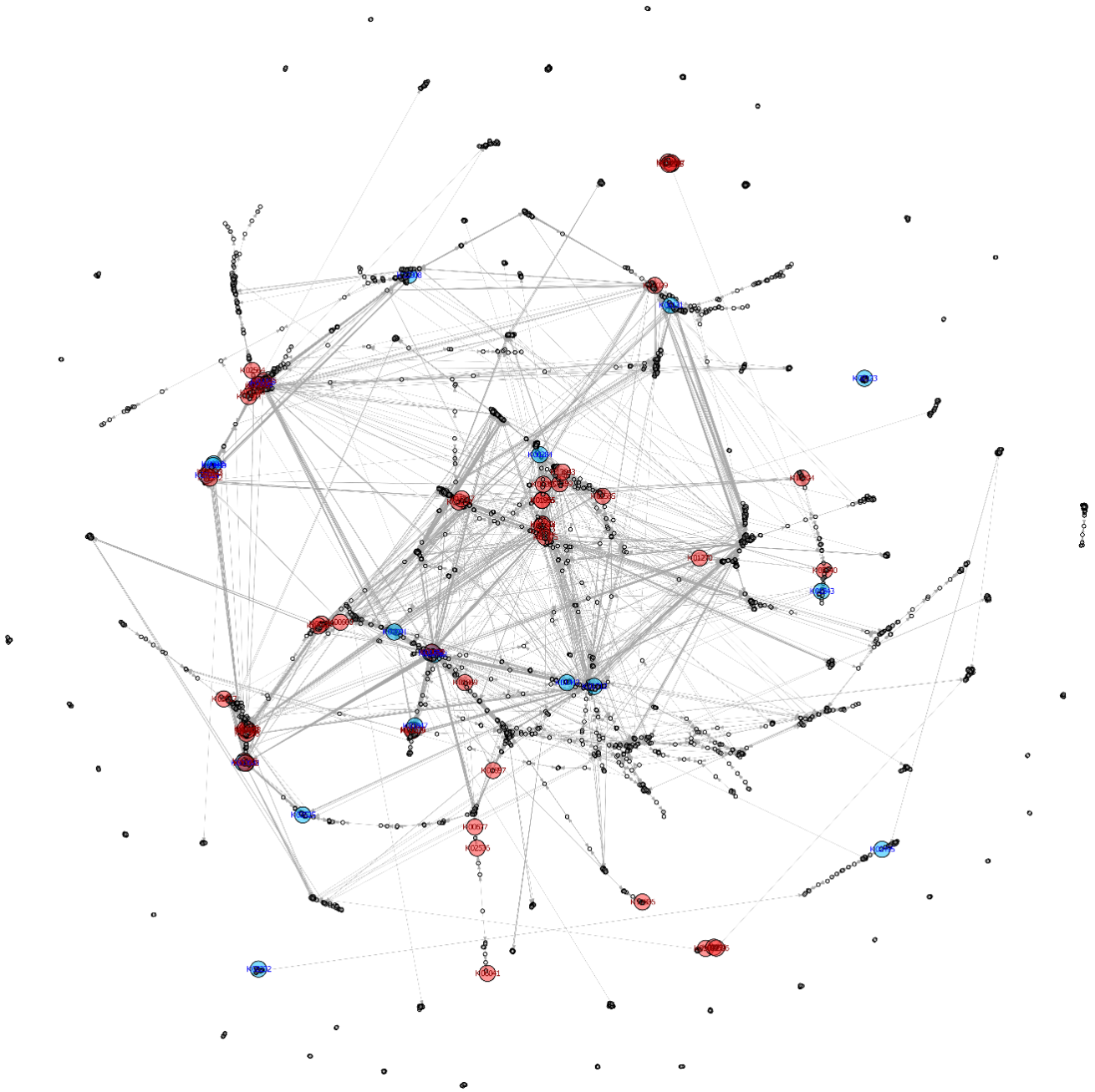


Figure S12.3.9 Metabolic network. Metabolic network retrieved from the pathways inferred by the ORF KEGG-annotations enzymes. Each enzyme of the network can be associated with several reactions, and each reaction can be involved with different enzymes. The nodes represent the KEGG orthologous groups (KO) meanwhile the direct edges indicate that the product metabolite of one of the enzymes y the substrate metabolite from the enzyme which the edge is pointing. In blue are represented all the KO LEfSe biomarkers related to the HIV- subjects, while in red all the KO LEfSe biomarkers related to the HIV condition.

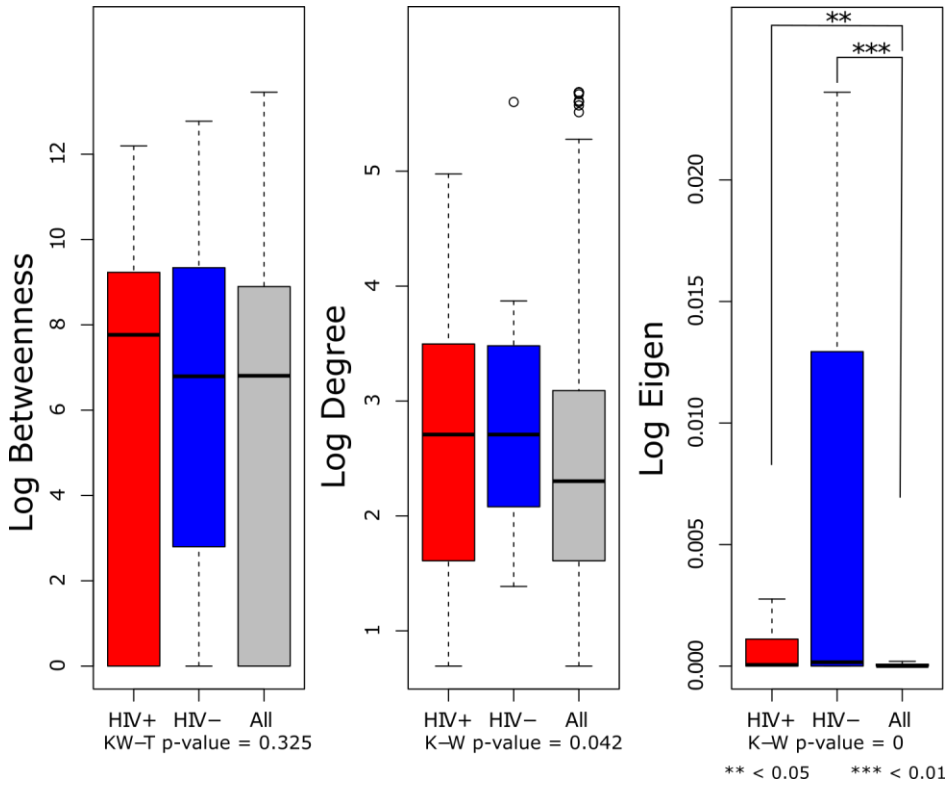


Figure S12.3.10 Biomarker KO centrality. Betweenness degree and eigenvector centrality for the HIV- (blue), HIV+ (red) KO biomarkers and the rest of the enzymes into the network (gray). The differences between the groups were asses by the Kruskal-Wallis test while the group pairwise comparisons were set by the Wilcoxon signed-rank test. The centralities values were expressed in the log scale.

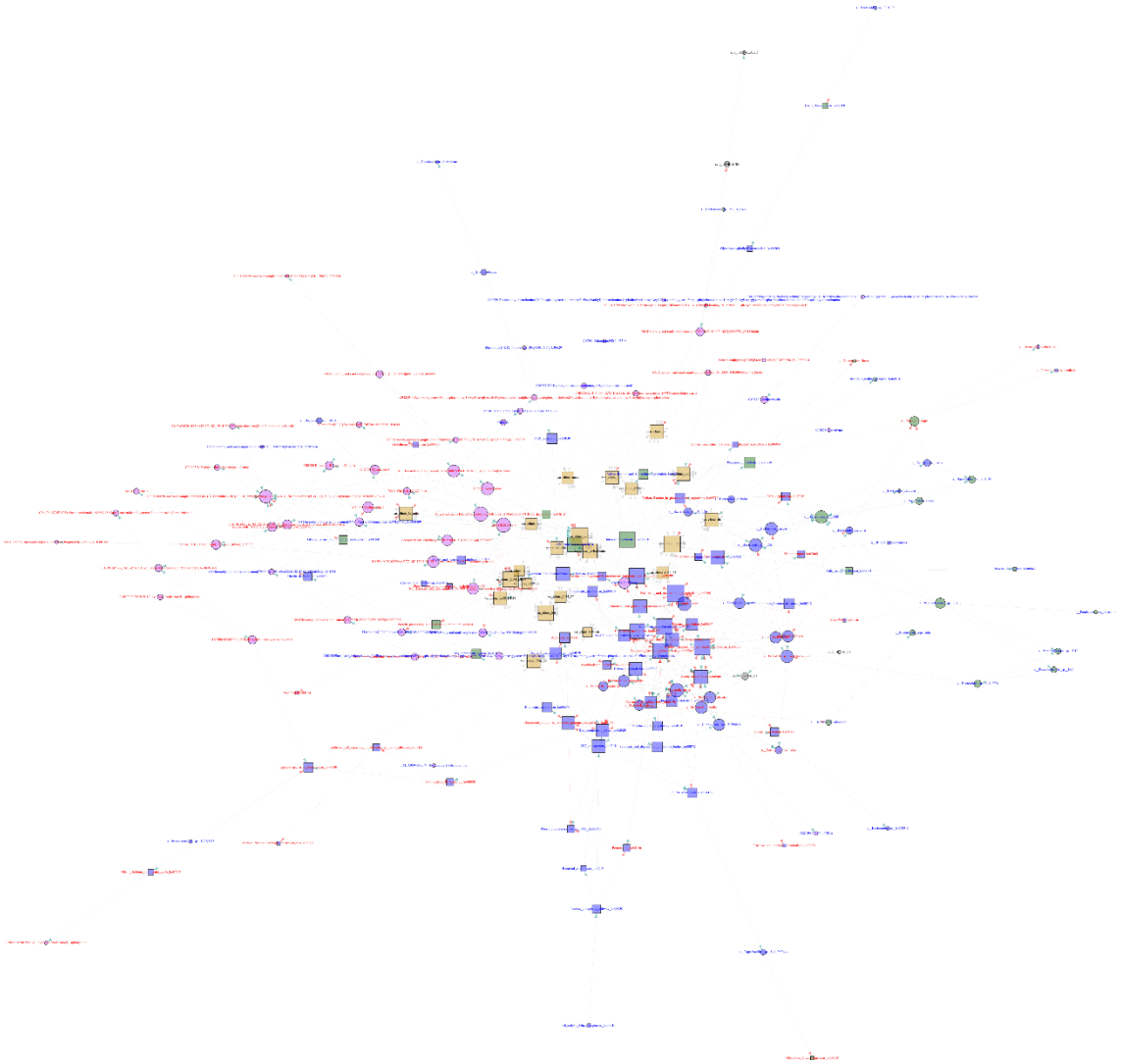


Figure S12.3.11 HIV+ Bayesian network. “Multiomic” Bayesian networks composed of the metagenomic (blue nodes), metatranscriptomic (green nodes), metabolomic (pink circle nodes) and the clinical variables (gold square nodes) from the HIV+ subjects. The data from the metagenomic and metatranscriptomic include the information of the species relative abundance (circles) and pathway relative abundance (squares). The labels of the nodes represent overrepresentation in the HIV- subjects (blue labels) or in HIV+ subjects (red label). Arrows indicate conditional dependencies between variables. The Spearman correlation coefficient is represented by the arrow's color, blue if it is significantly positive (PH adjusted p-value < 0.1), red if it has a significantly negative correlation (PH adjusted p-value < 0.1) or nonsignificant (gray).

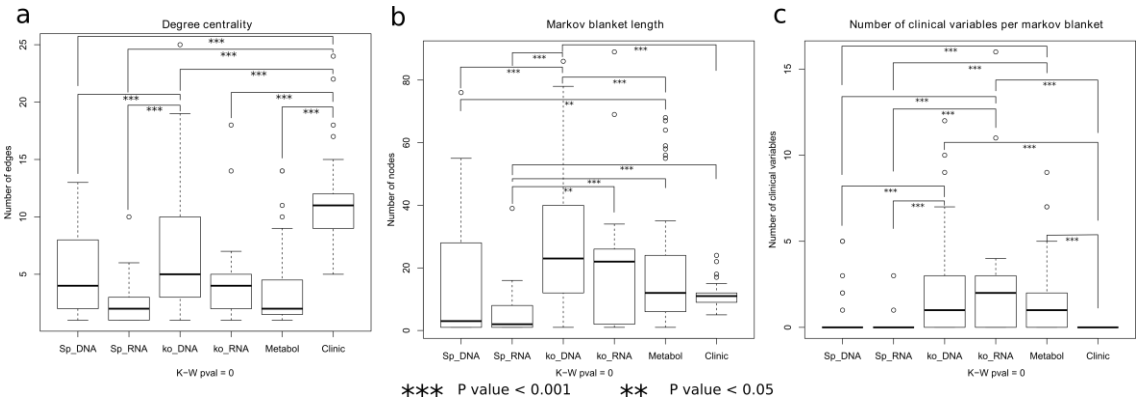


Figure S12.3.12 Bayesian network statistics. (**a**) Degree centrality, (**b**) a number of nodes per MB and (**c**) number of clinical variables within each MB for each component of the BN. The metagenomic and metatranscriptomic species (Sp_DNA and Sp_RNA respectively), the metagenomic and metatranscriptomic pathways (ko_DNA and ko_RNA respectively), the metabolites (Metabol) and the clinical variables (Clinic). The differences between the groups were asses by the Kruskal-Wallis test while the group pairwise comparison was set by the Wilcoxon signed-rank test. The centralities values were expressed in the log scale.

Supplementary Tables

Table S12.3.1 Mantel test pairwise comparison.

Analysis method 1	Analysis method 2	Mantel test p-value	BH p-value adjust
Ray_Illumina_454_ORF	Picrust	0.093	0.326
Ray_Illumina_454_ORF	idba_ORF	0.672	0.918
Ray_Illumina_454_ORF	idba_coverage	0.001	0.005
Ray_Illumina_454_ORF	ray_ORF	0.993	0.993
Ray_Illumina_454_ORF	ray_coverage	0.001	0.005
Ray_Illumina_454_ORF	Unassembled_ORF	0.746	0.918
Ray_Illumina_454_ORF	Unassembled_coverage	0.001	0.005
Picrust	idba_ORF	0.174	0.443
Picrust	idba_coverage	0.66	0.918
Picrust	ray_ORF	0.077	0.308
Picrust	ray_coverage	0.294	0.686
Picrust	Unassembled_ORF	0.507	0.918
Picrust	Unassembled_coverage	0.15	0.42
idba_ORF	idba_coverage	0.33	0.711
idba_ORF	ray_ORF	0.138	0.42
idba_ORF	ray_coverage	0.537	0.918
idba_ORF	Unassembled_ORF	0.501	0.918
idba_ORF	Unassembled_coverage	0.558	0.918
idba_coverage	ray_ORF	0.914	0.993

idba_coverage	ray_coverage	0.001	0.005
idba_coverage	Unassembled_ORF	0.613	0.918
idba_coverage	Unassembled_coverage	0.001	0.005
ray_ORF	ray_coverage	0.942	0.993
ray_ORF	Unassembled_ORF	0.906	0.993
ray_ORF	Unassembled_coverage	0.98	0.993
ray_coverage	Unassembled_ORF	0.754	0.918
ray_coverage	Unassembled_coverage	0.001	0.005
Unassembled_ORF	Unassembled_coverage	0.738	0.918

Table S12.3.2 Network bootstrap p-value.

	Mean Degree	Mean Betweenness	Mean Closeness	Mean Eigenvector centrality	%Pos	%Neg	Ratio	Fragmentation	Frag Modularity	Modularity	Number of clusters above 1	Transitivity	Diameter	Mean-short pathways	K-Smirnov p-value
Co-ocuurence network	*	*	*	*	0	0	0	*	0	0	0	0	0	0	*
Enzimatic network	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

*P-value below 0.001

Table S12.3.3 GLM between the taxonomic and functional biomarkers.

Class Biomarker	ko_pathway	Species	GLM Coefficient	Spearman Correlation	Spearman p-value	BH adjusted p-value
Control	ABC_transporters_ko02010	Prevotella_dentalis	-0.404610244	-0.573105613	5.92E-006	1.84E-005
Control	ABC_transporters_ko02010	Prevotella_copri	-0.001258591	-0.522622451	6.48E-005	0.000154449
Control	ABC_transporters_ko02010	Alistipes_shahii	0.00448976	0.462626263	0.000492975	0.000893517
Control	ABC_transporters_ko02010	Blautia_obeum	0.015277173	0.411092053	0.00218239	0.003360494
Control	ABC_transporters_ko02010	Coprococcus_sp__HPP0048	0.05631928	0.344602784	0.010718795	0.014237178
Control	ABC_transporters_ko02010	Faecalibacterium_prausnitzii	0.004275731	0.275242996	0.044299318	0.053159181
Control	MAPK_signaling_pathway_yeast_ko04011	Bacteroides_fluxus	0.000776101	-0.5687503	7.22E-006	2.19E-005
Control	MAPK_signaling_pathway_yeast_ko04011	Prevotella_melaninogenica	0.006653619	-0.509011715	8.48E-005	0.000195418
Control	MAPK_signaling_pathway_yeast_ko04011	Myroides_odoratimimus	-0.049419572	-0.496027079	0.000136589	0.000295235
Control	MAPK_signaling_pathway_yeast_ko04011	Prevotella_ruminicola	-0.145535943	-0.478375204	0.000253457	0.000481983
Control	MAPK_signaling_pathway_yeast_ko04011	Bacteroides_sp__4_1_36	0.000590377	0.470259115	0.000333084	0.000619857
Control	MAPK_signaling_pathway_yeast_ko04011	Megasphaera_elsdenii	0.000130664	-0.39438499	0.003168073	0.004731714
Control	MAPK_signaling_pathway_yeast_ko04011	Bacteroides_stercoris	3.47E-005	0.383382096	0.004215394	0.006112322
Control	MAPK_signaling_pathway_yeast_ko04011	Bacteroides_sp__HPS0048	0.002223169	0.380784191	0.004503223	0.006502579
Control	MAPK_signaling_pathway_yeast_ko04011	Bilophila_wadsworthia	0.000875843	0.379509246	0.004650696	0.006632959
Control	MAPK_signaling_pathway_yeast_ko04011	Bacteroides_sp__14_A_	0.002598421	0.361063083	0.007311308	0.009977785
Control	MAPK_signaling_pathway_yeast_ko04011	Alistipes_finegoldii	0.002268105	0.305139293	0.024855275	0.031226121
Control	MAPK_signaling_pathway_yeast_ko04011	Parabacteroides_sp__ASF519	0.008180161	0.297841171	0.028715399	0.035689139
Control	MAPK_signaling_pathway_yeast_ko04011	Bacteroides_salyersiae	0.000879788	0.297116349	0.029124735	0.035941162
Control	MAPK_signaling_pathway_yeast_ko04011	Faecalibacterium_prausnitzii	-0.000336933	-0.283974003	0.037433949	0.045549001
Control	MAPK_signaling_pathway_yeast_ko04011	Acidaminococcus_sp__D21	-0.001417503	-0.276562331	0.042925921	0.051689344
Control	MAPK_signaling_pathway_yeast_ko04011	Coprococcus_sp__HPP0048	0.105402087	0.27196257	0.046655083	0.055602633
Control	MAPK_signaling_pathway_yeast_ko04011	Bacteroides_clarus	-0.014087447	0.27060243	0.047807141	0.056588044
Control	MAPK_signaling_pathway_yeast_ko04011	Bacteroides_sp__3_1_23	0.002201815	0.268768614	0.049397044	0.058253754
Control	MAPK_signaling_pathway_yeast_ko04011	Parabacteroides_sp__HGS0025	-0.087663623	0.254725231	0.063050669	0.073138776
Control	MAPK_signaling_pathway_yeast_ko04011	Bifidobacterium_longum	0.001449896	0.229188759	0.095504453	0.108259119
Control	Two_component_system_ko02020	Prevotella_fusca	-0.159876866	-0.590540216	2.60E-006	8.69E-006
Control	Two_component_system_ko02020	Prevotella_copri	-0.000754814	-0.580331618	6.47E-006	1.97E-005

Control	Two_component_system_ko02020	Prevotella_sp__oral_taxon_299	-0.309201649	-0.564406917	8.78E-006	2.63E-005
Control	Two_component_system_ko02020	Bacteroides_fluxus	-0.008426397	-0.491747665	0.000192224	0.000391193
Control	Two_component_system_ko02020	Coprobacillus_sp__8_1_38FAA	0.012008485	0.484237412	0.000207181	0.000411994
Control	Two_component_system_ko02020	Bacteroides_sp__14_A_	0.002761163	0.464773169	0.000399119	0.00072719
Control	Two_component_system_ko02020	Bacteroides_caccae	0.001666558	0.451877263	0.000684655	0.001185373
Control	Two_component_system_ko02020	Bacteroides_pectinophilus	0.007637343	0.447226987	0.00078678	0.001342153
Control	Two_component_system_ko02020	Alistipes_shahii	0.003205701	0.427101201	0.001406729	0.002276938
Control	Two_component_system_ko02020	Coprococcus_sp__HPP0048	0.400294857	0.391231574	0.003441693	0.005096634
Control	Two_component_system_ko02020	Blautia_obeum	0.023382653	0.362378502	0.007388265	0.010043423
Control	Two_component_system_ko02020	Bacteroides_plebeius	0.000375097	0.298036974	0.028990552	0.035902889
Control	Sulfur_relay_system_ko04122	Prevotella_copri	-0.003366469	-0.631484658	6.21E-007	2.40E-006
Control	Sulfur_relay_system_ko04122	Prevotella_dentalis	-0.810255163	-0.589380727	2.75E-006	9.10E-006
Control	Sulfur_relay_system_ko04122	Prevotella_sp__oral_taxon_299	-0.933634956	-0.520157721	5.54E-005	0.000133984
Control	Sulfur_relay_system_ko04122	Parabacteroides_sp__HGS0025	4.648204165	0.466637111	0.000375458	0.00069132
Control	Sulfur_relay_system_ko04122	Alistipes_shahii	0.013719666	0.441738136	0.000924929	0.001570123
Control	Sulfur_relay_system_ko04122	Fermentimonas_caenicola	-2.096320995	-0.431582773	0.001120289	0.001838965
Control	Sulfur_relay_system_ko04122	Blautia_obeum	0.020734736	0.412388031	0.002107719	0.003259939
Control	Sulfur_relay_system_ko04122	Coprococcus_sp__HPP0048	2.60112256	0.406723627	0.002273316	0.003485083
Control	Sulfur_relay_system_ko04122	Bacteroides_pectinophilus	0.009308186	0.356203545	0.008521789	0.011539232
Control	Sulfur_relay_system_ko04122	Faecalibacterium_prausnitzii	0.010725168	0.289879931	0.033857264	0.04148707
Control	Sulfur_relay_system_ko04122	Ruminococcus_bicirculans	0.014567625	0.268595682	0.04954917	0.058253754
Control	Amoebiasis_ko05146	Prevotella_denticola	-0.016405769	-0.6371651	2.21E-007	9.29E-007
Control	Amoebiasis_ko05146	Prevotella_sp__oral_taxon_299	-0.063714374	-0.63168952	3.02E-007	1.22E-006
Control	Amoebiasis_ko05146	Prevotella_copri	-1.12E-005	-0.550445405	0.00001616	4.46E-005
Control	Amoebiasis_ko05146	Bacteroides_clarus	0.003226255	0.483381457	0.000213418	0.000421986
Control	Amoebiasis_ko05146	Myroides_odoratimimus	-0.002572225	-0.481637783	0.000226658	0.000440653
Control	Amoebiasis_ko05146	Parabacteroides_sp__ASF519	0.005395684	0.39558732	0.003068928	0.004603392
Control	Amoebiasis_ko05146	Streptococcus_pneumoniae	-0.008861073	-0.367274672	0.006296089	0.008764156
Control	Amoebiasis_ko05146	Odoribacter_splanchnicus	0.000609267	0.366108363	0.006476732	0.008979692
Control	Amoebiasis_ko05146	Streptococcus_pseudopneumoniae	-0.103425005	-0.365472978	0.006577035	0.009082572
Control	Amoebiasis_ko05146	Bacteroides_sp__14_A_	0.000365933	0.333725496	0.01365775	0.017935

					2	463
Control	Lysine_degradation_ko00310	Prevotella_copri	-0.000894318	-0.799352011	0	0
Control	Valine_leucine_and_isoleucine_degradation_ko00280	Prevotella_salivae	-0.176130143	-0.755517439	0	0
Control	Valine_leucine_and_isoleucine_degradation_ko00280	Prevotella_copri	-0.002328149	-0.792872117	0	0
Control	Valine_leucine_and_isoleucine_degradation_ko00280	Prevotella_fusca	-0.312005607	-0.833620741	5.15E-015	4.27E-014
Control	Valine_leucine_and_isoleucine_degradation_ko00280	Prevotella_dentalis	-0.466903929	-0.821129187	2.87E-014	2.17E-013
Control	Valine_leucine_and_isoleucine_degradation_ko00280	Bacteroides_clarus	0.078789921	0.610520297	0.000001674	5.83E-006
Control	Valine_leucine_and_isoleucine_degradation_ko00280	Bacteroides_eggerthii	0.003424819	0.592605298	3.78E-006	1.22E-005
Control	Valine_leucine_and_isoleucine_degradation_ko00280	Bacteroides_caccae	0.003146063	0.559824662	1.53E-005	4.26E-005
Control	Valine_leucine_and_isoleucine_degradation_ko00280	Dialister_succinatiphilus	-0.009974321	-0.540689918	3.28E-005	8.50E-005
Control	beta_Lactam_resistance_ko00312	Prevotella_salivae	-0.783418532	-0.836859158	0	0
Control	beta_Lactam_resistance_ko00312	Prevotella_copri	-0.005526572	-0.861406518	0	0
Control	beta_Lactam_resistance_ko00312	Prevotella_fusca	-0.412349027	-0.862712521	5.08E-017	5.19E-016
Control	beta_Lactam_resistance_ko00312	Prevotella_sp__oral_taxon_299	-3.206797393	-0.853710917	2.36E-016	2.34E-015
Control	beta_Lactam_resistance_ko00312	Bacteroides_sp__4_1_36	0.005671355	0.655422146	1.91E-007	8.23E-007
Control	beta_Lactam_resistance_ko00312	Dialister_succinatiphilus	-0.192105116	-0.650543168	2.44E-007	1.00E-006
Control	beta_Lactam_resistance_ko00312	Bacteroides_sp__14_A_	0.017023445	0.552954642	0.000014512	4.11E-005
Control	beta_Lactam_resistance_ko00312	Bacteroides_eggerthii	0.00756203	0.532304174	4.52E-005	0.000112246
Control	beta_Lactam_resistance_ko00312	Alistipes_finegoldii	0.014018659	0.517819707	7.72E-005	0.000181589
Control	beta_Lactam_resistance_ko00312	Streptococcus_pneumoniae	-0.589259145	-0.457006706	0.000512954	0.000915426
Control	beta_Lactam_resistance_ko00312	Bifidobacterium_sp__12_1_47BFAA	0.115633289	0.444550495	0.000757822	0.001299124
Control	beta_Lactam_resistance_ko00312	Odoribacter_splanchnicus	0.019532226	0.438841243	0.001006353	0.001683705
Control	beta_Lactam_resistance_ko00312	Bacteroides_sp__3_1_23	0.075550068	0.42915952	0.001327564	0.002158843
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Prevotella_copri	-0.001662197	-0.758490566	0	0
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Prevotella_fusca	-0.831174904	-0.772764709	7.62E-012	5.10E-011
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Prevotella_dentalis	-0.018170057	-0.727910222	4.48E-010	2.73E-009
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Prevotella_salivae	-0.060872961	-0.708709739	5.81E-009	3.17E-008
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Bacteroides_clarus	0.015328671	0.601143511	2.57E-006	8.69E-006
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Bacteroides_stercoris	0.005539381	0.588259958	4.58E-006	1.45E-005
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Bacteroides_finegoldii	0.024988801	0.504631218	0.000123532	0.000273815
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Bacteroides_sp__3_1_23	0.00544084	0.35460263	0.008839365	0.011922864
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Bacteroides_plebeius	0.000470963	0.327844483	0.015895426	0.020673515
Control	Glyoxylate_and_dicarboxylate_metabolism_ko00630	Clostridiales_bacterium_V	-0.386822062	-0.312064037	0.02199682	0.027937

		E202_14			8	578
Control	Lipoic_acid_metabolism_ko00785	Prevotella_copri	-0.004775056	-0.870418724	1.25E-017	1.40E-016
Control	Lipoic_acid_metabolism_ko00785	Prevotella_fusca	-1.173112219	-0.853451486	2.46E-016	2.38E-015
Control	Lipoic_acid_metabolism_ko00785	Prevotella_ruminicola	-2.790519067	-0.835840226	3.74E-015	3.18E-014
Control	Lipoic_acid_metabolism_ko00785	Prevotella_salivae	-0.359873991	-0.802416665	2.98E-013	2.21E-012
Control	Lipoic_acid_metabolism_ko00785	Bacteroides_sp__D20	0.00071518	0.730602832	3.59E-010	2.23E-009
Control	Lipoic_acid_metabolism_ko00785	Bacteroides_sp__4_1_36	0.000336785	0.707541596	2.21E-009	1.26E-008
Control	Lipoic_acid_metabolism_ko00785	Bacteroides_eggerthii	0.013911565	0.687453544	9.37E-009	4.94E-008
Control	Lipoic_acid_metabolism_ko00785	Dialister_succinatiphilus	-0.124798646	-0.638739065	2.02E-007	8.59E-007
Control	Lipoic_acid_metabolism_ko00785	Fermentimonas_caenicola	-11.16497396	-0.609816071	9.84E-007	3.64E-006
Control	Lipoic_acid_metabolism_ko00785	Bacteroides_stercoris	0.019979627	0.602412853	1.44E-006	5.06E-006
Control	Lipoic_acid_metabolism_ko00785	Bacteroides_sp__14_A_	0.04164849	0.587658942	0.000002985	9.71E-006
Control	Lipoic_acid_metabolism_ko00785	Streptococcus_pneumoniae	-3.80967269	-0.534309555	3.16E-005	0.000082754
Control	Lipoic_acid_metabolism_ko00785	Alistipes_finegoldii	0.072706101	0.523127942	0.000049382	0.000121021
Control	Lipoic_acid_metabolism_ko00785	Clostridiales_bacterium_VE202_14	-0.198114395	-0.472812518	0.000305874	0.000575374
Control	Lipoic_acid_metabolism_ko00785	Bacteroides_plebeius	0.002252228	0.456764947	0.000516928	0.000917781
Control	Lipoic_acid_metabolism_ko00785	Streptococcus_pseudopneumoniae	-2.521267334	-0.435954757	0.000983651	0.001653674
Control	Lipoic_acid_metabolism_ko00785	Parabacteroides_sp__ASF519	-0.041601876	0.41232438	0.001947281	0.003052494
Control	Lipoic_acid_metabolism_ko00785	Faecalibacterium_prausnitzii	-0.002882178	-0.396843851	0.002968258	0.004510716
Control	Lipoic_acid_metabolism_ko00785	Bacteroides_pectinophilus	-0.005368301	-0.390554423	0.003503101	0.005143794
Control	Lipoic_acid_metabolism_ko00785	Bibersteinia_trehalosi	10.43115112	-0.337387411	0.012599375	0.016671417
Control	Porphyrin_and_chlorophyll_metabolism_ko00860	Prevotella_salivae	-0.292222324	-0.805831904	0	0
Control	Porphyrin_and_chlorophyll_metabolism_ko00860	Prevotella_copri	-0.001858476	-0.853554412	0	0
Control	Porphyrin_and_chlorophyll_metabolism_ko00860	Bacteroides_sp__4_1_36	0.002331086	0.67890223	5.58E-008	2.66E-007
Control	Limonene_and_pinene_degradation_ko00903	Prevotella_dentalis	-0.522900256	-0.704223436	2.83E-009	1.59E-008
Control	Limonene_and_pinene_degradation_ko00903	Prevotella_copri	-0.000924568	-0.67608157	6.52E-008	3.07E-007
Control	Limonene_and_pinene_degradation_ko00903	Bacteroides_fluxus	-0.002261777	-0.608995617	1.80E-006	6.19E-006
Control	Limonene_and_pinene_degradation_ko00903	Bacteroides_salyersiae	0.004305154	0.555250619	1.84E-005	5.05E-005
Control	Limonene_and_pinene_degradation_ko00903	Bacteroides_eggerthii	0.002080767	0.516066324	8.23E-005	0.000190909
Control	Limonene_and_pinene_degradation_ko00903	Bacteroides_sp__HPS0048	0.143203385	0.39028016	0.00375193	0.005463061
Control	Aminobenzoate_degradation_ko00627	Prevotella_copri	-0.000556968	-0.786163522	0	0
Control	Aminobenzoate_degradation_ko00627	Prevotella_dentalis	-0.193123698	-0.824414773	1.85E-014	1.43E-

						013
Control	Aminobenzoate_degradation_ko00627	Prevotella_sp__oral_taxon_299	-0.212768645	-0.789251015	1.34E-012	9.52E-012
Control	Aminobenzoate_degradation_ko00627	Bacteroides_sp__D20	0.002564025	0.717629121	2.54E-010	1.61E-009
Control	Aminobenzoate_degradation_ko00627	Bacteroides_sp__4_1_36	0.001491365	0.659767486	1.54E-007	6.86E-007
Control	Aminobenzoate_degradation_ko00627	Bacteroides_fluxus	-0.000375147	-0.64436821	3.32E-007	1.31E-006
Control	Aminobenzoate_degradation_ko00627	Bacteroides_sp__3_1_23	0.03676153	0.531999238	4.57E-005	0.000112742
Control	Aminobenzoate_degradation_ko00627	Fermentimonas_caenicola	-0.252350844	-0.521382612	5.29E-005	0.000128649
Control	Aminobenzoate_degradation_ko00627	Bacteroides_stercoris	0.000594911	0.508976558	0.000106025	0.000239588
Control	Aminobenzoate_degradation_ko00627	Bacteroides_eggerthii	0.000368549	0.497160282	0.000159952	0.000335321
Control	Aminobenzoate_degradation_ko00627	Bacteroides_sp__14_A_	0.005881099	0.478802906	0.000249787	0.000481084
Control	Aminobenzoate_degradation_ko00627	Bacteroides_sp__HPS0048	0.085682041	0.461558986	0.000509546	0.000914031
Control	Aminobenzoate_degradation_ko00627	Dialister_succinatiphilus	0.029362779	-0.458433391	0.000561027	0.000986048
Control	Aminobenzoate_degradation_ko00627	Bacteroides_salyersiae	0.001699204	0.398818372	0.003015754	0.004547118
Control	Aminobenzoate_degradation_ko00627	Bacteroides_plebeius	0.002784568	0.352086907	0.009359276	0.012575398
Control	Aminobenzoate_degradation_ko00627	Alistipes_finegoldii	0.001583667	0.335010482	0.013645164	0.017935463
Control	Aminobenzoate_degradation_ko00627	Faecalibacterium_prausnitzii	-0.001057403	-0.327768249	0.015920983	0.020673515
Control	Aminobenzoate_degradation_ko00627	Odoribacter_splanchnicus	-0.004823793	0.246426529	0.072614389	0.083124367
Control	Benzoate_degradation_ko00362	Prevotella_copri	-0.001117329	-0.628435296	7.19E-007	2.75E-006
Control	Benzoate_degradation_ko00362	Prevotella_dentalis	-0.048555287	-0.610240381	9.63E-007	0.000003603
Control	Benzoate_degradation_ko00362	Prevotella_melaninogenica	-0.141158893	-0.578399552	4.63E-006	1.45E-005
Control	Benzoate_degradation_ko00362	Bacteroides_fluxus	-0.004750316	-0.563026491	1.34E-005	0.000038279
Control	Benzoate_degradation_ko00362	Bilophila_wadsworthia	0.008478258	0.532735887	3.37E-005	8.62E-005
Control	Benzoate_degradation_ko00362	Ruminococcus_bicirculans	0.006745017	0.419257305	0.001601794	0.002580668
Control	Benzoate_degradation_ko00362	Fermentimonas_caenicola	-0.387595443	-0.418797971	0.001622863	0.002602563
Control	Benzoate_degradation_ko00362	Blautia_obeum	0.01696103	0.417724414	0.001823758	0.002898027
Control	Benzoate_degradation_ko00362	Bacteroides_plebeius	8.40E-005	0.415589861	0.001932946	0.003043734
Control	Benzoate_degradation_ko00362	Parabacteroides_sp__ASF519	0.032656835	0.379904699	0.004604506	0.006594108
Control	Benzoate_degradation_ko00362	Bacteroides_stercoris	-0.000450685	0.372441395	0.005822631	0.008170467
Control	Benzoate_degradation_ko00362	Bacteroides_sp__HPS0048	0.038100793	0.350181056	0.009770691	0.013077694
Control	Benzoate_degradation_ko00362	Bacteroides_pectinophilus	0.002478532	0.28362874	0.038033283	0.046117012
Control	Benzoate_degradation_ko00362	Bacteroides_sp__3_2_5	-0.003563527	-0.283400038	0.038193836	0.046150885

Control	Benzoate_degradation_ko00362	Acidaminococcus_sp__D21	-0.005591464	-0.250085763	0.068369106	0.078522933
Control	Benzoate_degradation_ko00362	Megasphaera_elsdenii	0.015360524	-0.232704403	0.09044369	0.103194767
Control	Insulin_signaling_pathway_ko04910	Prevotella_copri	-0.000811217	-0.607623404	1.91E-006	6.53E-006
Control	Insulin_signaling_pathway_ko04910	Prevotella_ruminicola	-0.253657875	-0.556040851	1.27E-005	0.000036521
Control	Insulin_signaling_pathway_ko04910	Prevotella_dentalis	-0.089897093	-0.536429299	2.90E-005	7.71E-005
Control	Insulin_signaling_pathway_ko04910	Prevotella_salivae	-0.008635522	-0.50806175	0.000109508	0.000245864
Control	Insulin_signaling_pathway_ko04910	Bacteroides_salyersiae	0.001833937	0.501353154	0.000138455	0.000295596
Control	Insulin_signaling_pathway_ko04910	Coprococcus_sp__HPP0048	1.098984152	0.48761809	0.000184135	0.000379165
Control	Insulin_signaling_pathway_ko04910	Parabacteroides_sp__HGS0025	1.558672848	0.486827217	0.000189306	0.000387521
Control	Insulin_signaling_pathway_ko04910	Fermentimonas_caenicola	-0.636398493	-0.43127655	0.001130469	0.001846963
Control	Insulin_signaling_pathway_ko04910	Blautia_obeum	0.011370429	0.395159139	0.003313849	0.004928288
Control	Insulin_signaling_pathway_ko04910	Bifidobacterium_sp__12_1_47BFAA	0.032675009	0.390273759	0.003528836	0.00515981
Control	Insulin_signaling_pathway_ko04910	Bacteroides_plebeius	0.00065292	0.382199352	0.004591191	0.006594108
Control	Insulin_signaling_pathway_ko04910	Parabacteroides_sp__ASF519	0.020298779	0.325224047	0.016412505	0.021232535
Control	Insulin_signaling_pathway_ko04910	Bacteroides_pectinophilus	0.005530938	0.308023633	0.023846277	0.030067045
Control	Insulin_signaling_pathway_ko04910	Alistipes_finegoldii	-0.003297811	0.289117591	0.034345067	0.041937134
HIV+	Protein_export_ko03060	Prevotella_salivae	0.630327914	0.708557271	5.93E-009	3.17E-008
HIV+	Protein_export_ko03060	Bacteroides_sp__D20	-0.000344256	-0.544730322	2.80E-005	0.00007493
HIV+	Protein_export_ko03060	Alistipes_finegoldii	-0.002922092	-0.364665523	0.007003267	0.009595027
HIV+	RNA_degradation_ko03018	Prevotella_salivae	0.489633859	0.748580141	0	0
HIV+	RNA_degradation_ko03018	Prevotella_copri	0.000982003	0.737145035	0	0
HIV+	RNA_degradation_ko03018	Bacteroides_pectinophilus	0.00491378	0.312826377	0.02166191	0.027612985
HIV+	RNA_degradation_ko03018	Bibersteinia_trehalosi	-0.331645263	0.293629397	0.031162701	0.038320212
HIV+	Epithelial_cell_signaling_in_Helicobacter_pylori_infection_ko05120	Fermentimonas_caenicola	1.023448054	0.448348472	0.000673857	0.001178404
HIV+	Epithelial_cell_signaling_in_Helicobacter_pylori_infection_ko05120	Streptococcus_pseudopneumoniae	2.067914861	0.396214039	0.003018346	0.004547118
HIV+	Epithelial_cell_signaling_in_Helicobacter_pylori_infection_ko05120	Odoribacter_splanchnicus	-0.00151189	-0.254507338	0.063507989	0.073424518
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Prevotella_copri	0.000224926	0.686601868	3.56E-008	1.72E-007
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Prevotella_salivae	0.057339171	0.662206975	1.36E-007	6.13E-007
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Prevotella_ruminicola	2.115990035	0.640627435	1.81E-007	7.89E-007
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Prevotella_melaninogenica	0.276605282	0.6355523	2.43E-007	1.00E-006
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Acidaminococcus_intestini	0.012038269	0.560434534	1.49E-005	4.19E-005

HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Dialister_succinatiphilus	0.098358216	0.554640747	1.89E-005	5.13E-005
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Bilophila_wadsworthia	-0.019762194	-0.539944814	0.000025114	6.77E-005
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Bacteroides_eggerthii	-0.004636945	-0.50409758	0.000125856	0.000277201
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Bifidobacterium_longum	-0.002182973	-0.490909091	0.000197727	0.000400051
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Bacteroides_finegoldii	-0.000509064	-0.413607776	0.002039538	0.003168568
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Ruminococcus_bicirculans	-0.022716714	-0.375314332	0.005166308	0.007338267
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Coprobacillus_sp__8_1_38FAA	-0.063941187	-0.304921283	0.024963915	0.031249793
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Blautia_obeum	-0.005948943	-0.262511912	0.055419755	0.064718372
HIV+	Vibrio_cholerae_pathogenic_cycle_ko05111	Bacteroides_sp__3_2_5	0.008743212	0.252144082	0.066070226	0.076133903
HIV+	Alanine_aspartate_and_glutamate_metabolism_ko00250	Prevotella_salivae	0.597740174	0.806899181	0	0
HIV+	Alanine_aspartate_and_glutamate_metabolism_ko00250	Prevotella_copri	0.006633203	0.890909091	0	0
HIV+	Alanine_aspartate_and_glutamate_metabolism_ko00250	Prevotella_fusca	0.827176859	0.870195445	1.30E-017	1.41E-016
HIV+	Alanine_aspartate_and_glutamate_metabolism_ko00250	Prevotella_sp__oral_taxon_299	0.787709191	0.845978791	8.13E-016	7.45E-015
HIV+	Alanine_aspartate_and_glutamate_metabolism_ko00250	Prevotella_intermedia	0.882819982	0.779715516	3.73E-012	2.60E-011
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Prevotella_salivae	0.122106449	0.752391843	0	0
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Prevotella_copri	0.002417975	0.828092243	0	0
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Prevotella_fusca	0.358790813	0.841638161	1.58E-015	1.41E-014
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Prevotella_dentalis	0.155941444	0.836258167	3.52E-015	3.06E-014
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Prevotella_ruminicola	2.445823921	0.827009201	1.30E-014	1.03E-013
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Prevotella_intermedia	2.023986526	0.77558943	5.72E-012	3.90E-011
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Dialister_succinatiphilus	0.029707247	0.61982085	1.08E-006	3.97E-006
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Bacteroides_eggerthii	-0.001618533	-0.566152087	1.18E-005	3.45E-005
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Bacteroides_sp__HPS0048	-0.271486108	-0.565770917	1.20E-005	3.47E-005
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Bacteroides_salyersiae	-0.002680836	-0.538860301	3.52E-005	8.93E-005
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Odoribacter_splanchnicus	-0.021115026	-0.502496665	0.000133071	0.000289429
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Bifidobacterium_sp__12_1_47BFAA	-0.006566333	-0.485248511	0.000200027	0.000400054
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Bilophila_wadsworthia	0.005086793	-0.370744814	0.005784391	0.008149668
HIV+	Phenylalanine_tyrosine_and_tryptophan_biosynthesis_ko00400	Ruminococcus_bicirculans	-0.029883963	-0.229935053	0.094400455	0.10735738
HIV+	Citrate_cycle_TCA_cycle__ko00020	Prevotella_salivae	0.642743553	0.756279779	0	0
HIV+	Citrate_cycle_TCA_cycle__ko00020	Acidaminococcus_intestini	4.92E-005	0.686983038	3.48E-008	1.71E-007
HIV+	Fructose_and_mannose_metabolism_ko00051	Prevotella_denticola	0.536892355	0.482560962	0.000219558	0.000431673

HIV+	Fructose_and_mannose_metabolism_ko00051	Prevotella_dentalis	0.085864011	0.47820565	0.000254925	0.000482141
HIV+	Fructose_and_mannose_metabolism_ko00051	Prevotella_copri	0.000296255	0.437926434	0.001033378	0.001720648
HIV+	Carbon_fixation_in_photosynthetic_organisms_ko00710	Prevotella_salivae	0.836564665	0.770688012	0	0
HIV+	Carbon_fixation_in_photosynthetic_organisms_ko00710	Prevotella_copri	0.003241563	0.765885268	0	0
HIV+	Carbon_fixation_in_photosynthetic_organisms_ko00710	Prevotella_melaninogenica	0.066786524	0.793648178	8.21E-013	5.96E-012
HIV+	Carbon_fixation_in_photosynthetic_organisms_ko00710	Bacteroides_stercoris	-0.009513908	-0.506155899	0.000117107	0.000261239
HIV+	Carbon_fixation_in_photosynthetic_organisms_ko00710	Streptococcus_pneumoniae	2.91801112	0.495672105	0.000138343	0.000295596
HIV+	Carbon_fixation_in_photosynthetic_organisms_ko00710	Clostridiales_bacterium_VE202_14	1.224250705	0.493119878	0.000183522	0.000379165
HIV+	Carbon_fixation_in_photosynthetic_organisms_ko00710	Alistipes_finegoldii	-0.006455168	-0.437392796	0.001049443	0.001730835
HIV+	Carbon_fixation_in_photosynthetic_organisms_ko00710	Bibersteinia_trehalosi	-1.310449328	0.270662686	0.047755616	0.056588044
HIV+	Oxidative_phosphorylation_ko00190	Prevotella_copri	0.000435808	0.542443301	3.06E-005	8.07E-005
HIV+	Oxidative_phosphorylation_ko00190	Prevotella_dentalis	0.01313649	0.531386305	3.56E-005	8.97E-005
HIV+	Oxidative_phosphorylation_ko00190	Bacteroides_fluxus	0.00358923	0.483666857	0.000251601	0.000481084
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Prevotella_fusca	0.150236935	0.673539598	2.39E-008	1.22E-007
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Prevotella_ruminicola	9.183981911	0.66860578	3.29E-008	0.000000166
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Prevotella_copri	0.004909974	0.687059272	3.46E-008	1.71E-007
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Prevotella_salivae	0.364890392	0.671964932	8.15E-008	3.73E-007
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Prevotella_melaninogenica	0.851269653	0.641698989	1.71E-007	7.51E-007
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bacteroides_fluxus	0.005905104	0.624623594	0.000000863	3.26E-006
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Prevotella_oralis	-1.385958208	0.606900116	1.14E-006	4.15E-006
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Acidaminococcus_intestini	0.040889744	0.61661902	1.26E-006	4.52E-006
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bifidobacterium_longum	-0.043668026	-0.581551363	6.14E-006	1.89E-005
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Riemerella_anatipestifer	0.158894854	0.533289653	3.30E-005	8.50E-005
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bacteroides_stercoris	0.022152402	-0.521250238	6.81E-005	0.000161325
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bibersteinia_trehalosi	6.667541297	0.50484658	9.90E-005	0.000226686
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Fermentimonas_caenicola	2.671034976	0.49719101	0.000130976	0.000286665
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Parabacteroides_sp__HGS0025	-1.983556429	-0.493045463	0.000151983	0.000322501
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bacteroides_plebeius	-0.006023894	-0.494034687	0.000177923	0.000370761
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Odoribacter_splanchnicus	0.006868554	-0.490680389	0.000199252	0.000400054
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Dialister_succinatiphilus	0.34678993	0.483666857	0.000251601	0.000481084
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bacteroides_eggerthii	0.012927647	-0.475967219	0.00032332	0.000604

					2	926
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bacteroides_sp__14_A_	-0.008273623	-0.465878773	0.00038492 6	0.000705 023
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Coprobacillus_sp__8_1_38 FAA	-0.044760255	-0.45458011	0.00055411 8	0.000978 849
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bacteroides_caccae	0.010611613	-0.450962455	0.00070374 1	0.001212 386
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Parabacteroides_sp__ASF5 19	-0.114727764	-0.437712139	0.00093308 7	0.001576 283
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Alistipes_finegoldii	0.025365913	-0.420163903	0.00170576 7	0.002722 968
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Blautia_obeum	-0.032163756	-0.414751286	0.00197741 8	0.003085 836
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bacteroides_sp__HPS0048	-0.552418218	-0.36611397	0.00676863 5	0.009310 218
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Coprococcus_sp__HPP004 8	-2.033517547	-0.331400197	0.01436865 8	0.018798 093
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bacteroides_sp__3_2_5	0.156426259	0.303602058	0.02601956	0.032454 505
HIV+	Lipopolysaccharide_biosynthesis_ko00540	Bacteroides_pectinophilus	-0.015602444	-0.260377359	0.05748977 8	0.066911 18
HIV+	N_Glycan_biosynthesis_ko00510	Prevotella_copri	0.000175869	0.516142558	0.00008206 2	0.000190 909
HIV+	N_Glycan_biosynthesis_ko00510	Fermentimonas_caenicola	0.781727328	0.491985222	0.00015782 9	0.000332 875
HIV+	N_Glycan_biosynthesis_ko00510	Prevotella_dentalis	0.008432226	0.41325801	0.00189718 5	0.003001 002
HIV+	N_Glycan_biosynthesis_ko00510	Bacteroides_eggerthii	-0.000986601	-0.399504479	0.00296259 7	0.004510 716
HIV+	N_Glycan_biosynthesis_ko00510	Parabacteroides_sp__HGS 0025	-0.314962578	-0.373900806	0.00535100 4	0.007569 713
HIV+	N_Glycan_biosynthesis_ko00510	Bacteroides_stercoris	-0.001063969	-0.314503526	0.02094032 7	0.026791 3
HIV+	N_Glycan_biosynthesis_ko00510	Coprococcus_sp__HPP004 8	-0.072767395	-0.265921471	0.05195059	0.060871 398
HIV+	N_Glycan_biosynthesis_ko00510	Faecalibacterium_prausnitz ii	-0.000823985	-0.227901658	0.09744031 6	0.110094 903
HIV+	Peptidoglycan_biosynthesis_ko00550	Bacteroides_sp__D20	-0.002507641	-0.728606823	0	0
HIV+	Peptidoglycan_biosynthesis_ko00550	Prevotella_salivae	0.771067843	0.812388031	0	0
HIV+	Peptidoglycan_biosynthesis_ko00550	Prevotella_copri	0.00519608	0.837164094	0	0
HIV+	Peptidoglycan_biosynthesis_ko00550	Prevotella_dentalis	1.382983757	0.872323209	8.70E-018	1.01E-016
HIV+	Peptidoglycan_biosynthesis_ko00550	Prevotella_sp__oral_taxon _299	3.20351697	0.846514681	7.48E-016	7.04E-015
HIV+	Peptidoglycan_biosynthesis_ko00550	Faecalibacterium_prausnitz ii	0.016487105	0.371755289	0.00591925 1	0.008272 687
HIV+	Nicotinate_and_nicotinamide_metabolism_ko00760	Prevotella_copri	0.002565147	0.765199161	0	0
HIV+	Nicotinate_and_nicotinamide_metabolism_ko00760	Prevotella_sp__oral_taxon _299	1.188732661	0.727317452	4.70E-010	2.82E-009
HIV+	Nicotinate_and_nicotinamide_metabolism_ko00760	Prevotella_denticola	0.81085705	0.725235907	5.56E-010	3.28E-009
HIV+	Nicotinate_and_nicotinamide_metabolism_ko00760	Prevotella_fusca	0.120275338	0.716642772	1.10E-009	6.38E-009
HIV+	Nicotinate_and_nicotinamide_metabolism_ko00760	Bacteroides_sp__D20	-0.000525527	-0.598170383	2.94E-006	9.66E-006
HIV+	Nicotinate_and_nicotinamide_metabolism_ko00760	Bacteroides_sp__HPS0048	-0.072662376	-0.462321327	0.00049765 9	0.000897 333
HIV+	One_carbon_pool_by_folate_ko00670	Prevotella_salivae	0.81149853	0.591614256	3.95E-006	0.000012 602

HIV+	One_carbon_pool_by_folate_ko00670	Bacteroides_sp__3_1_23	-0.033884814	-0.487173623	0.000224032	0.000437995
HIV+	One_carbon_pool_by_folate_ko00670	Prevotella_oralis	0.785144755	0.468924719	0.000348162	0.00064447
HIV+	One_carbon_pool_by_folate_ko00670	Bacteroides_caccae	-0.001178167	-0.39310082	0.003492791	0.005143794
HIV+	Metabolism|Metabolism_of_Other_Amino_Acids|D_Alanine_metabolism_ko00473	Prevotella_salivae	0.361960852	0.755441205	0	0
HIV+	Metabolism|Metabolism_of_Other_Amino_Acids|D_Alanine_metabolism_ko00473	Prevotella_copri	0.012571451	0.839146179	0	0
HIV+	Metabolism|Metabolism_of_Other_Amino_Acids|D_Alanine_metabolism_ko00473	Prevotella_dentalis	0.493603635	0.827394724	1.23E-014	9.99E-014
HIV+	Metabolism|Metabolism_of_Other_Amino_Acids|D_Alanine_metabolism_ko00473	Bacteroides_fluxus	0.047915676	0.673413379	7.54E-008	3.50E-007
HIV+	Metabolism|Metabolism_of_Other_Amino_Acids|D_Alanine_metabolism_ko00473	Parabacteroides_sp__ASF519	-0.166124553	-0.433898983	0.001045915	0.001730835
HIV+	Terpenoid_backbone_biosynthesis_ko00900	Bacteroides_sp__D20	-0.000990142	-0.760853821	0	0
HIV+	Terpenoid_backbone_biosynthesis_ko00900	Prevotella_copri	0.004773795	0.897846388	0	0
HIV+	Terpenoid_backbone_biosynthesis_ko00900	Prevotella_dentalis	1.547405825	0.90693648	3.63E-021	4.35E-020
HIV+	Terpenoid_backbone_biosynthesis_ko00900	Prevotella_sp__oral_taxon_299	0.895389824	0.863663158	4.29E-017	4.52E-016
HIV+	Terpenoid_backbone_biosynthesis_ko00900	Bacteroides_sp__4_1_36	-0.000347414	-0.708557271	5.93E-009	3.17E-008
HIV+	Terpenoid_backbone_biosynthesis_ko00900	Fermentimonas_caenicola	6.137082373	0.631086934	3.12E-007	1.25E-006
HIV+	Terpenoid_backbone_biosynthesis_ko00900	Bacteroides_clarus	-0.04616053	-0.61623785	1.28E-006	4.55E-006
HIV+	Terpenoid_backbone_biosynthesis_ko00900	Bacteroides_sp__14_A_	-0.007914154	-0.560541375	0.000010427	3.10E-005
HIV+	Terpenoid_backbone_biosynthesis_ko00900	Bacteroides_salyersiae	-0.011945584	-0.509738898	0.000103199	0.000234728
HIV+	Zeatin_biosynthesis_ko00908	Prevotella_copri	0.001080912	0.77442348	0	0
HIV+	Zeatin_biosynthesis_ko00908	Prevotella_dentalis	2.314246958	0.739447979	1.70E-010	1.10E-009
HIV+	Zeatin_biosynthesis_ko00908	Fermentimonas_caenicola	0.744041474	0.559966686	1.07E-005	3.15E-005
HIV+	Zeatin_biosynthesis_ko00908	Parabacteroides_sp__HGS0025	-3.340567	-0.519453814	5.70E-005	0.000136739
HIV+	Zeatin_biosynthesis_ko00908	Bacteroides_sp__HPS0048	-0.033429809	-0.452029731	0.00068152	0.001185373
HIV+	Zeatin_biosynthesis_ko00908	Megasphaera_elsdenii	-0.036514968	0.347131694	0.010461707	0.013948943
HIV+	Zeatin_biosynthesis_ko00908	Acidaminococcus_sp__D21	0.013281067	0.325404993	0.016730934	0.021564315
HIV+	Zeatin_biosynthesis_ko00908	Coprobacillus_sp__8_1_38FAA	-0.174983397	-0.32215149	0.01751841	0.022495966
HIV+	Toluene_degradation_ko00623	Prevotella_copri	0.001917409	0.745683248	0	0
HIV+	Toluene_degradation_ko00623	Prevotella_dentalis	0.820839323	0.752437507	5.37E-011	3.53E-010
HIV+	Toluene_degradation_ko00623	Prevotella_salivae	0.195595019	0.695826186	1.94E-008	1.01E-007
HIV+	Toluene_degradation_ko00623	Bacteroides_fluxus	0.000418738	0.644215742	3.34E-007	1.31E-006
HIV+	Toluene_degradation_ko00623	Bacteroides_sp__4_1_36	-0.001583292	-0.537869259	0.000036526	9.14E-005
HIV+	Toluene_degradation_ko00623	Bacteroides_eggerthii	0.009439637	-0.308862207	0.023451956	0.029677384
HIV+	Toluene_degradation_ko00623	Ruminococcus_bicirculans	-0.014396793	-0.272576579	0.04614249	0.055180

						1	711
HIV+	Toluene_degradation_ko00623	Blautia_obeum	-0.010927638	-0.227215552	0.098473454	0.110902142	

Table S12.3.4 Biomarker species gene mapping

Health status[a]	Species[b]	ko pathways[c]	KO in ko pathways[d]	KO present[e]	Percentage %[f]	KO list[g]
Control	Bacteroides_caccae	ko00280	61	10	16.39	K00186 K00187 K00382 K00826 K01847 K01965 K01966 K01968 K05606 K11381
Control	Bacteroides_clarus	ko00280	61	10	16.39	K00186 K00187 K00382 K00826 K01847 K01965 K01966 K01968 K05606 K11381
Control	Bacteroides_eggerthii	ko00280	61	8	13.11	K00186 K00187 K00382 K00826 K01847 K01966 K05606 K11381
Control	Alistipes_finegoldii	ko00312	7	2	28.57	K02171 K02172
Control	Bacteroides_eggerthii	ko00312	7	3	42.86	K02171 K02172 K17836
Control	Bacteroides_sp__14_A_	ko00312	7	3	42.86	K02171 K02172 K17836
Control	Bacteroides_sp__3_1_23	ko00312	7	3	42.86	K02171 K02172 K17836
Control	Bacteroides_sp__4_1_36	ko00312	7	3	42.86	K02171 K02172 K17836
Control	Odoribacter_splanchnicus	ko00312	7	2	28.57	K02171 K02172
Control	Bacteroides_pectinophilus	ko00362	87	4	4.6	K00680 K01615 K01726 K04072
Control	Bacteroides_plebeius	ko00362	87	4	4.6	K00680 K01607 K01615 K01726
Control	Bacteroides_sp__HPS0048	ko00362	87	3	3.45	K00680 K01615 K01726
Control	Bacteroides_stercoris	ko00362	87	4	4.6	K00680 K01607 K01615 K01726
Control	Bilophila_wadsworthia	ko00362	87	5	5.75	K00074 K00626 K00680 K01607 K04021
Control	Blautia_obeum	ko00362	87	4	4.6	K01607 K01615 K01726 K04072
Control	Parabacteroides_sp__ASF519	ko00362	87	5	5.75	K00680 K01607 K01615 K01666 K01726
Control	Ruminococcus_bicirculans	ko00362	87	4	4.6	K01607 K01615 K01666 K04072
Control	Alistipes_finegoldii	ko00627	67	1	1.49	K09461
Control	Bacteroides_eggerthii	ko00627	67	1	1.49	K09461
Control	Bacteroides_plebeius	ko00627	67	1	1.49	K09461
Control	Bacteroides_salyersiae	ko00627	67	1	1.49	K09461
Control	Bacteroides_sp__14_A_	ko00627	67	1	1.49	K09461
Control	Bacteroides_sp__3_1_23	ko00627	67	1	1.49	K09461
Control	Bacteroides_sp__4_1_36	ko00627	67	1	1.49	K09461
Control	Bacteroides_sp__D20	ko00627	67	1	1.49	K09461
Control	Bacteroides_stercoris	ko00627	67	1	1.49	K09461
Control	Bacteroides_clarus	ko00630	73	9	12.33	K00018 K00048 K00284 K00600 K00865 K01091 K01433 K01625 K02437
Control	Bacteroides_finegoldii	ko00630	73	8	10.96	K00018 K00048 K00284 K00600 K01091 K01433 K01625 K02437
Control	Bacteroides_plebeius	ko00630	73	9	12.33	K00018 K00048 K00284 K00600 K00865 K01091 K01625 K02437 K03781
Control	Bacteroides_sp__3_1_23	ko00630	73	10	13.7	K00018 K00048 K00284 K00600 K00865 K01091 K01433 K01625 K02437 K03781
Control	Bacteroides_stercoris	ko00630	73	9	12.33	K00018 K00048 K00284 K00600 K00865 K01091

						K01433 K01625 K02437
Control	Alistipes_finegoldii	ko00785	4	2	50	K03644 K03801
Control	Bacteroides_eggerthii	ko00785	4	3	75	K03644 K03800 K03801
Control	Bacteroides_plebeius	ko00785	4	2	50	K03644 K03801
Control	Bacteroides_sp__14_A_	ko00785	4	3	75	K03644 K03800 K03801
Control	Bacteroides_sp__4_1_36	ko00785	4	2	50	K03644 K03801
Control	Bacteroides_sp__D20	ko00785	4	2	50	K03644 K03801
Control	Bacteroides_stercoris	ko00785	4	2	50	K03644 K03801
Control	Parabacteroides_sp__ASF519	ko00785	4	3	75	K03644 K03800 K03801
Control	Bacteroides_sp__4_1_36	ko00860	105	19	18.1	K00595 K00768 K01195 K01719 K01885 K02188 K02190 K02217 K02224 K02226 K02227 K02231 K02232 K02233 K02495 K03394 K04720 K05934 K13541
Control	Alistipes_shahii	ko02010	426	16	3.76	K02013 K02015 K02016 K05655 K06148 K06861 K07091 K09808 K09810 K09811 K09812 K09815 K09816 K09817 K11085 K11720
Control	Blautia_obeum	ko02010	426	80	18.78	K01995 K01996 K01997 K01998 K01999 K02007 K02008 K02010 K02013 K02015 K02016 K02017 K02018 K02020 K02036 K02037 K02038 K02040 K02045 K02046 K02048 K02062 K02071 K02072 K02073 K03523 K05655 K05658 K05816 K06148 K09690 K09691 K09692 K09812 K09815 K09816 K09972 K10004 K10008 K10010 K10038 K10041 K10111 K10112 K10117 K10118 K10119 K10188 K10189 K10190 K10439 K10440 K10441 K10542 K10543 K10545 K10546 K10547 K10548 K10559 K10560 K10561 K10562 K10823 K11069 K11070 K11071 K11072 K11085 K15583 K16785 K16786 K16787 K16960 K16963 K17073 K17074 K17076 K17215 K17318
Control	Coprococcus_sp__HPP0048	ko02010	426	66	15.49	K02006 K02007 K02010 K02017 K02018 K02020 K02036 K02037 K02038 K02040 K02071 K02072 K02073 K03523 K05655 K05685 K05816 K06148 K09690 K09691 K09697 K09811 K09812 K09972 K10004 K10008 K10010 K10038 K10041 K10111 K10112 K10117 K10118 K10119 K10191 K10200 K10201 K10202 K10439 K10441 K10542 K10548 K10823 K11069 K11070 K11071 K11072 K11085 K12371 K13892 K15580 K15581 K15582 K15583 K15770 K15772 K16202 K16785 K16786 K16787 K16960 K17076 K17215 K17318 K17319 K17320
Control	Faecalibacterium_prausnitzii	ko02010	426	72	16.9	K01995 K01996 K01997 K01998 K01999 K02010 K02013 K02015 K02016 K02017 K02018 K02020 K02036 K02037 K02038 K02040 K02071 K02072 K02073 K02424 K03523 K05685 K05816 K06148 K09690 K09691 K09810 K09811 K09812 K09972 K10004 K10008 K10009 K10010 K10038 K10041 K10111 K10112 K10117 K10119 K10192 K10193 K10194 K10200 K10201 K10202 K10441 K10540 K10541 K10542 K10548 K10562 K10823 K11069 K11070 K11071 K11072 K11085 K12371 K13892 K15580 K15581 K15582 K15583 K16202 K16785 K16786 K16787 K16960 K16963 K17076 K17215
Control	Alistipes_shahii	ko02020	384	28	7.29	K00027 K00426 K01546 K01547 K01548 K01915 K02313 K02584 K02667 K03092 K07636 K07646 K07652 K07657 K07658 K07662 K07665 K07679 K07712 K07713 K07714 K07715 K07783 K10941 K10943 K11626 K13599 K13924
Control	Bacteroides_caccae	ko02020	384	33	8.59	K00027 K00425 K00426 K01077 K01546 K01547 K01548 K01915 K02313 K02488 K02584 K02667 K03092 K07636 K07646 K07651 K07657 K07659 K07662 K07664 K07665 K07679 K07684 K07712 K07713 K07714 K07715 K07787 K07792 K10941 K10943 K13599 K13924
Control	Bacteroides_pectinophilus	ko02020	384	39	10.16	K00027 K00575 K01915 K02405 K02406 K02556 K02660 K03406 K03407 K03408 K03412 K03413 K03415 K03563 K03776 K07636 K07646 K07651 K07652 K07657 K07658 K07662 K07664 K07665 K07667 K07668 K07669 K07670 K07768 K07770 K07774 K07775 K07776 K07813 K08372 K11329 K11521 K11618 K11690
Control	Bacteroides_plebeius	ko02020	384	27	7.03	K00027 K00425 K00426 K01077 K01915 K02313 K02584 K02667 K03092 K04751 K07636 K07657 K07659 K07662 K07665 K07668 K07684 K07712 K07713 K07714 K07715 K07787 K07792 K10941

						K10943 K11626 K13599
Control	Bacteroides_sp__14_A_	ko02020	384	33	8.59	K00027 K00425 K00426 K01077 K01546 K01547 K01548 K01915 K02313 K02584 K02667 K03092 K04751 K07636 K07646 K07657 K07659 K07662 K07665 K07678 K07679 K07684 K07712 K07713 K07714 K07715 K07783 K07787 K10941 K10943 K11626 K13599 K13924
Control	Blautia_obeum	ko02020	384	27	7.03	K00027 K01915 K02313 K03406 K04771 K07636 K07652 K07657 K07658 K07662 K07664 K07665 K07667 K07668 K07670 K07699 K07718 K07720 K07768 K07770 K07774 K07775 K07776 K08372 K11329 K11630 K11690
Control	Coprobacillus_sp__8_1_38 FAA	ko02020	384	9	2.34	K00027 K01915 K07636 K07650 K07658 K07665 K07699 K07770 K11630
Control	Coprococcus_sp__HPP004 8	ko02020	384	31	8.07	K00027 K01546 K01547 K01548 K01915 K02313 K07636 K07646 K07651 K07652 K07657 K07658 K07659 K07662 K07664 K07665 K07667 K07668 K07669 K07670 K07699 K07718 K07720 K07768 K07770 K07774 K07775 K07776 K08372 K11329 K11630
Control	Alistipes_finegoldii	ko04011	41	1	2.44	K01759
Control	Bacteroides_clarus	ko04011	41	1	2.44	K01759
Control	Bacteroides_salyersiae	ko04011	41	1	2.44	K01759
Control	Bacteroides_sp__14_A_	ko04011	41	1	2.44	K01759
Control	Bacteroides_sp__3_1_23	ko04011	41	1	2.44	K01759
Control	Bacteroides_sp__4_1_36	ko04011	41	1	2.44	K01759
Control	Bacteroides_sp__HPS0048	ko04011	41	1	2.44	K01759
Control	Bacteroides_stercoris	ko04011	41	1	2.44	K01759
Control	Coprococcus_sp__HPP004 8	ko04011	41	1	2.44	K01759
Control	Parabacteroides_sp__ASF5 19	ko04011	41	1	2.44	K01759
Control	Alistipes_shahii	ko04122	21	2	9.52	K00566 K04085
Control	Bacteroides_pectinophilus	ko04122	21	4	19.05	K00566 K03151 K04487 K11996
Control	Blautia_obeum	ko04122	21	8	38.1	K00566 K03151 K03637 K03639 K03831 K04085 K04487 K11996
Control	Coprococcus_sp__HPP004 8	ko04122	21	7	33.33	K00566 K03151 K03637 K03639 K03831 K04085 K04487
Control	Faecalibacterium_prausnitzi i	ko04122	21	6	28.57	K00566 K03151 K03637 K03639 K03831 K04487
Control	Parabacteroides_sp__HGS0 025	ko04122	21	2	9.52	K00566 K04085
Control	Ruminococcus_bicirculans	ko04122	21	4	19.05	K00566 K03151 K04487 K11996
Control	Alistipes_finegoldii	ko04910	84	1	1.19	K00688
Control	Bacteroides_pectinophilus	ko04910	84	2	2.38	K00688 K07192
Control	Bacteroides_plebeius	ko04910	84	1	1.19	K00688
Control	Bacteroides_salyersiae	ko04910	84	2	2.38	K00688 K07192
Control	Bifidobacterium_sp__12_1 _47BFAA	ko04910	84	1	1.19	K00688
Control	Blautia_obeum	ko04910	84	1	1.19	K00688
Control	Coprococcus_sp__HPP004 8	ko04910	84	1	1.19	K00688
Control	Parabacteroides_sp__ASF5 19	ko04910	84	3	3.57	K00688 K00844 K07192
Control	Parabacteroides_sp__HGS0 025	ko04910	84	3	3.57	K00688 K00844 K07192
Control	Bacteroides_sp__14_A_	ko05146	78	1	1.28	K01476
Control	Parabacteroides_sp__ASF5	ko05146	78	1	1.28	K01476

	19					
HIV+	Acidaminococcus_intestini	ko00020	56	11	19.64	K00030 K00031 K00174 K00175 K00176 K00177 K01610 K01679 K01681 K01958 K01960
HIV+	Prevotella_salivae	ko00020	56	8	14.29	K00024 K00174 K00175 K00176 K00177 K01610 K01676 K01960
HIV+	Prevotella_copri	ko00051	81	19	23.46	K00011 K00100 K00754 K00847 K00848 K00850 K00895 K00966 K00971 K01624 K01629 K01711 K01803 K01805 K01808 K01813 K01840 K02377 K04041
HIV+	Prevotella_dentalis	ko00051	81	20	24.69	K00100 K00754 K00847 K00848 K00850 K00895 K00966 K00971 K01624 K01629 K01711 K01803 K01805 K01808 K01809 K01813 K01840 K02377 K04041 K05305
HIV+	Prevotella_denticola	ko00051	81	16	19.75	K00100 K00754 K00847 K00850 K00895 K00966 K00971 K01218 K01624 K01711 K01803 K01808 K01809 K01840 K02377 K04041
HIV+	Bacteroides_fluxus	ko00190	212	26	12.26	K00330 K00331 K00335 K00337 K00338 K00339 K00340 K00341 K00342 K00343 K02108 K02109 K02111 K02112 K02113 K02114 K02115 K02117 K02118 K02120 K02121 K02123 K02124 K03885 K13378 K15987
HIV+	Prevotella_copri	ko00190	212	20	9.43	K00330 K00331 K00337 K00338 K00339 K00340 K00341 K00342 K00343 K02108 K02109 K02110 K02111 K02112 K02113 K02114 K02115 K02120 K03885 K13378
HIV+	Prevotella_dentalis	ko00190	212	21	9.91	K00330 K00331 K00337 K00338 K00339 K00340 K00341 K00342 K00343 K02108 K02109 K02110 K02111 K02112 K02113 K02114 K02115 K02120 K03885 K13378 K15987
HIV+	Prevotella_copri	ko00250	60	21	35	K00259 K00262 K00264 K00265 K00266 K00278 K00609 K00610 K00764 K00820 K01424 K01755 K01756 K01914 K01939 K01940 K01953 K01955 K01956 K11540 K11541
HIV+	Prevotella_fusca	ko00250	60	16	26.67	K00262 K00266 K00278 K00609 K00610 K00764 K00820 K01424 K01744 K01755 K01756 K01914 K01939 K01955 K01956 K11541
HIV+	Prevotella_intermedia	ko00250	60	16	26.67	K00262 K00266 K00278 K00609 K00610 K00764 K00820 K01424 K01744 K01755 K01756 K01939 K01953 K01955 K01956 K11541
HIV+	Prevotella_salivae	ko00250	60	16	26.67	K00259 K00262 K00266 K00609 K00610 K00764 K00820 K01424 K01744 K01755 K01756 K01914 K01939 K01955 K01956 K11541
HIV+	Prevotella_sp__oral_taxon_299	ko00250	60	14	23.33	K00262 K00609 K00610 K00764 K00820 K01424 K01744 K01755 K01756 K01914 K01939 K01955 K01956 K11541
HIV+	Dialister_succinatiphilus	ko00400	70	21	30	K00014 K00766 K00800 K00812 K00817 K00832 K00891 K01609 K01626 K01657 K01658 K01695 K01696 K01713 K01735 K01736 K03856 K04517 K06001 K13497 K14170
HIV+	Prevotella_copri	ko00400	70	19	27.14	K00014 K00210 K00766 K00800 K00812 K00817 K00891 K01609 K01657 K01658 K01695 K01696 K01735 K01736 K01817 K03786 K04516 K04518 K06001
HIV+	Prevotella_dentalis	ko00400	70	20	28.57	K00014 K00210 K00766 K00800 K00812 K00817 K00891 K01609 K01626 K01657 K01658 K01695 K01696 K01735 K01736 K01817 K03786 K04516 K04518 K06001
HIV+	Prevotella_fusca	ko00400	70	16	22.86	K00014 K00210 K00766 K00800 K00812 K00817 K00891 K01657 K01658 K01735 K01736 K01817 K03786 K04516 K04518 K06001
HIV+	Prevotella_intermedia	ko00400	70	12	17.14	K00014 K00210 K00800 K00812 K00891 K01609 K01657 K01658 K01735 K01736 K03786 K04516
HIV+	Prevotella_ruminicola	ko00400	70	19	27.14	K00014 K00210 K00766 K00800 K00812 K00817 K00891 K01609 K01657 K01658 K01695 K01696 K01735 K01736 K01817 K03786 K04516 K04518 K06001
HIV+	Prevotella_salivae	ko00400	70	14	20	K00014 K00210 K00766 K00800 K00812 K00891 K01609 K01626 K01735 K01736 K03786 K04516 K04518 K06001
HIV+	Bacteroides_fluxus	ko00473	5	1	20	K01775

HIV+	Prevotella_copri	ko00473	5	1	20	K01775
HIV+	Prevotella_dentalis	ko00473	5	1	20	K01775
HIV+	Prevotella_salivae	ko00473	5	1	20	K01775
HIV+	Fermentimonas_caenicola	ko00510	44	1	2.27	K00721
HIV+	Prevotella_copri	ko00510	44	1	2.27	K00721
HIV+	Prevotella_dentalis	ko00510	44	1	2.27	K00721
HIV+	Acidaminococcus_intestini	ko00540	34	15	44.12	K00677 K00748 K00979 K01627 K02517 K02527 K02535 K02536 K02843 K03270 K03271 K03272 K03273 K03274 K16363
HIV+	Bacteroides_fluxus	ko00540	34	11	32.35	K00677 K00748 K00912 K00979 K01627 K02517 K02527 K02536 K03269 K03270 K16363
HIV+	Bacteroides_sp__3_2_5	ko00540	34	13	38.24	K00677 K00748 K00912 K00979 K01627 K02517 K02527 K02536 K03269 K03270 K03271 K03273 K16363
HIV+	Bibersteinia_trehalosi	ko00540	34	15	44.12	K00677 K00748 K00912 K00979 K01627 K02527 K02535 K02536 K02560 K02843 K03270 K03271 K03272 K03273 K03274
HIV+	Dialister_succinatiphilus	ko00540	34	12	35.29	K00677 K00748 K00912 K00979 K01627 K02527 K02535 K02536 K03271 K03272 K07031 K16363
HIV+	Fermentimonas_caenicola	ko00540	34	10	29.41	K00677 K00748 K00912 K00979 K01627 K02527 K02536 K03269 K03270 K16363
HIV+	Prevotella_copri	ko00540	34	14	41.18	K00677 K00748 K00912 K00979 K01627 K02517 K02527 K02536 K03269 K03270 K03271 K03273 K07031 K16363
HIV+	Prevotella_fusca	ko00540	34	11	32.35	K00677 K00748 K00912 K00979 K01627 K02517 K02527 K02536 K03269 K03270 K16363
HIV+	Prevotella_melaninogenica	ko00540	34	11	32.35	K00677 K00748 K00912 K00979 K01627 K02517 K02527 K02536 K03269 K03270 K16363
HIV+	Prevotella_oralis	ko00540	34	13	38.24	K00677 K00748 K00912 K00979 K01627 K02517 K02527 K02536 K03269 K03270 K03271 K03273 K16363
HIV+	Prevotella_ruminicola	ko00540	34	11	32.35	K00677 K00748 K00912 K00979 K01627 K02517 K02527 K02536 K03269 K03270 K16363
HIV+	Prevotella_salivae	ko00540	34	11	32.35	K00677 K00748 K00912 K00979 K01627 K02517 K02527 K02536 K03269 K03270 K16363
HIV+	Riemerella_anatipestifer	ko00540	34	6	17.65	K00677 K00748 K01627 K02517 K02536 K16363
HIV+	Prevotella_copri	ko00550	38	15	39.47	K00075 K00790 K01000 K01921 K01924 K01925 K01928 K01929 K02563 K03587 K03814 K05366 K05515 K06153 K07259
HIV+	Prevotella_dentalis	ko00550	38	15	39.47	K00075 K00790 K01000 K01921 K01924 K01925 K01928 K01929 K02563 K03587 K03814 K05366 K05515 K06153 K07259
HIV+	Prevotella_salivae	ko00550	38	15	39.47	K00075 K00790 K01000 K01921 K01924 K01925 K01928 K01929 K02563 K03587 K03814 K05366 K05515 K06153 K07259
HIV+	Prevotella_sp__oral_taxon_299	ko00550	38	15	39.47	K00075 K00790 K01000 K01921 K01924 K01925 K01928 K01929 K02563 K03587 K03814 K05366 K05515 K06153 K07259
HIV+	Bacteroides_fluxus	ko00623	51	3	5.88	K00239 K00240 K00241
HIV+	Prevotella_copri	ko00623	51	3	5.88	K00239 K00240 K00241
HIV+	Prevotella_dentalis	ko00623	51	3	5.88	K00239 K00240 K00241
HIV+	Prevotella_salivae	ko00623	51	3	5.88	K00239 K00240 K00241
HIV+	Prevotella_oralis	ko00670	27	11	40.74	K00287 K00297 K00548 K00560 K00602 K00604 K00605 K01491 K01934 K01938 K08289
HIV+	Prevotella_salivae	ko00670	27	10	37.04	K00287 K00297 K00548 K00560 K00602 K00604 K01491 K01938 K03465 K08289
HIV+	Bibersteinia_trehalosi	ko00710	37	7	18.92	K00029 K00134 K00615 K00927 K01595 K01783 K01807
HIV+	Clostridiales_bacterium_VE202_14	ko00710	37	6	16.22	K00029 K00134 K00615 K00927 K01006 K01783

HIV+	Prevotella_copri	ko00710	37	6	16.22	K00029 K00134 K00615 K00927 K01006 K01783
HIV+	Prevotella_melaninogenica	ko00710	37	6	16.22	K00029 K00134 K00615 K00927 K01006 K01783
HIV+	Prevotella_salivae	ko00710	37	6	16.22	K00029 K00134 K00615 K00927 K01006 K01783
HIV+	Streptococcus_pneumoniae	ko00710	37	7	18.92	K00134 K00150 K00615 K00927 K01595 K01783 K01807
HIV+	Prevotella_copri	ko00760	44	9	20.45	K00767 K00858 K00969 K01081 K01950 K03426 K03517 K03783 K03787
HIV+	Prevotella_denticola	ko00760	44	10	22.73	K00763 K00767 K00858 K00969 K01081 K01950 K03426 K03517 K03783 K03787
HIV+	Prevotella_fusca	ko00760	44	10	22.73	K00763 K00767 K00858 K00969 K01081 K01950 K03426 K03517 K03783 K03787
HIV+	Prevotella_sp__oral_taxon_299	ko00760	44	9	20.45	K00763 K00767 K00858 K00969 K01081 K01950 K03426 K03783 K03787
HIV+	Fermentimonas_caenicola	ko00900	50	10	20	K00099 K00806 K00919 K00991 K01662 K01770 K02523 K03526 K03527 K13789
HIV+	Prevotella_copri	ko00900	50	11	22	K00099 K00806 K00919 K00991 K01662 K01770 K02523 K03526 K03527 K12506 K13789
HIV+	Prevotella_dentalis	ko00900	50	11	22	K00099 K00806 K00919 K00991 K01662 K01770 K02523 K03526 K03527 K12506 K13789
HIV+	Prevotella_sp__oral_taxon_299	ko00900	50	11	22	K00099 K00806 K00919 K00991 K01662 K01770 K02523 K03526 K03527 K12506 K13789
HIV+	Acidaminococcus_sp__D21	ko00908	8	1	12.5	K00791
HIV+	Fermentimonas_caenicola	ko00908	8	1	12.5	K00791
HIV+	Megasphaera_elsdenii	ko00908	8	1	12.5	K00791
HIV+	Prevotella_copri	ko00908	8	1	12.5	K00791
HIV+	Prevotella_dentalis	ko00908	8	1	12.5	K00791
HIV+	Bibersteinia_trehalosi	ko03018	75	12	16	K00962 K01689 K03628 K03654 K03666 K03732 K04043 K04077 K05592 K08300 K08311 K12573
HIV+	Prevotella_copri	ko03018	75	10	13.33	K00962 K00970 K03628 K03654 K04043 K04077 K05592 K11927 K12573 K12574
HIV+	Prevotella_salivae	ko03018	75	10	13.33	K00962 K00970 K01689 K03628 K03654 K04043 K04077 K05592 K11927 K12573
HIV+	Prevotella_salivae	ko03060	39	10	25.64	K03070 K03075 K03076 K03100 K03101 K03106 K03110 K03210 K03217 K12257
HIV+	Acidaminococcus_intestini	ko05111	43	1	2.33	K07173
HIV+	Prevotella_copri	ko05111	43	1	2.33	K07173
HIV+	Prevotella_melaninogenica	ko05111	43	1	2.33	K07173
HIV+	Fermentimonas_caenicola	ko05120	90	1	1.11	K08303

The genes from the biomarker pathways were mapped into the genome species

[a]Biomarker condition.

[b]Biomarker specie.

[c]Biomarker pathways.

[d]Number of genes in pathways.

[e]Number of genes in pathways that maps into the biomarker specie.

[f]List of genes in pathways that maps into the biomarker specie.

Table S12.3.5 Relative abundance of the species that contain the genes for the tryptophan catabolism via the kynurenine pathway.

DNA mapping species	VU (mean / sd)	IR (mean / sd)	INR (mean / sd)	HIV- (mean / sd)	K/T ratio Spearman cor-index	P-value	Adjusted p-value *
Xanthomonas arboricola pruni	9.41e-05 / 6.5e-05	7.66e-05 / 3.57e-05	0.0001108 / 8.45e-05	7.61e-05 / 6.1e-05	0.280	0.041	0.222
Trichodesmium erythraeum IMS101	8.97e-05 / 6.94e-05	6.91e-05 / 4.56e-05	9.77e-05 / 6.18e-05	9.11e-05 / 6.69e-05	0.167	0.227	0.265
Streptomyces scabiei	0.0001091 / 9.94e-05	8.52e-05 / 4.64e-05	0.000125 / 8.8e-05	0.0001094 / 8.81e-05	0.213	0.122	0.250
Pseudomonas putida	0.0001933 / 0.0001906	0.0001985 / 0.0001281	0.0002391 / 0.000237	0.0001844 / 0.0001553	0.173	0.211	0.265
Pseudomonas fluorescens	0.0001734 / 0.0001575	0.0001706 / 0.0001096	0.0001964 / 0.0001917	0.0001549 / 0.0001238	0.204	0.140	0.250
Pseudomonas aeruginosa	0.0001707 / 0.0001523	0.0001658 / 0.0001028	0.0002052 / 0.0001739	0.0001493 / 0.0001172	0.199	0.149	0.250
Nitrosococcus oceani ATCC19707	0.0002297 / 0.000164	0.0002088 / 0.0001064	0.0002464 / 0.0001768	0.0002198 / 0.000144	0.194	0.161	0.250
Nitrosococcus halophilus Nc4	0.0001412 / 0.0001042	0.0001196 / 6.85e-05	0.0001532 / 0.0001356	0.0001245 / 0.0001001	0.271	0.048	0.222
Neptuniibacter caesariensis MED92	5.38e-05 / 5.39e-05	5.6e-05 / 4.18e-05	5.4e-05 / 5.33e-05	3.99e-05 / 3.99e-05	0.211	0.126	0.250
Methylocystis sp ATCC 49242	0.0001391 / 0.0001011	8.43e-05 / 4.64e-05	0.0001506 / 7.82e-05	0.0001229 / 8.09e-05	0.171	0.216	0.265
Gemmatimonas aurantiaca	8.17e-05 / 6.32e-05	5.42e-05 / 3.16e-05	8.89e-05 / 4.56e-05	7.15e-05 / 6e-05	0.213	0.121	0.250
Erythrobacter sp SD 2	2.47e-05 / 2.95e-05	6e-06 / 1.19e-05	2.23e-05 / 2.76e-05	2.12e-05 / 2.01e-05	0.287	0.035	0.222
Erythrobacter litoralis	0.0001235 / 9.76e-05	7.31e-05 / 4.4e-05	0.0001265 / 6.67e-05	0.0001322 / 8.27e-05	0.134	0.334	0.338
Citromicrobium bathyomarinum	0.0001136 / 9.57e-05	6.87e-05 / 4.21e-05	0.0001276 / 6.37e-05	0.0001259 / 8.17e-05	0.133	0.338	0.338

RNA mapping species	VU (mean / sd)	IR (mean / sd)	INR (mean / sd)	HIV- (mean / sd)	K/T ratio Spearman cor-index	P-value	Adjusted p-value *
Xanthomonas arboricola pruni	9.5e-06 / 1.75e-05	7.4e-06 / 1.36e-05	1.7e-06 / 4.1e-06	1.32e-05 / 1.69e-05	-0.030	0.868	0.935
Trichodesmium erythraeum IMS101	0 / 0	0 / 0	0 / 0	2.1e-06 / 7.7e-06	0.000	1.000	1.000
Streptomyces scabiei	1.2e-06 / 3.6e-06	2.1e-06 / 3e-06	0 / 0	6.1e-06 / 1.53e-05	-0.162	0.367	0.728
Pseudomonas putida	5e-06 / 6e-06	1.08e-05 / 1.55e-05	1.4e-06 / 3.4e-06	9.6e-06 / 1.93e-05	-0.075	0.679	0.792
Pseudomonas fluorescens	4.6e-06 / 7.4e-06	7.4e-06 / 1.24e-05	1.2e-06 / 3e-06	6.6e-06 / 1.27e-05	-0.155	0.389	0.728

Pseudomonas aeruginosa	3.8e-06 5.7e-06	/	9.3e-06 1.38e-05	/	3e-06 4.7e-06	/	6.5e-06 1.06e-05	/	-0.147	0.415	0.728
Nitrosococcus oceani ATCC19707	4.4e-06 8.7e-06	/	1.22e-05 2.21e-05	/	2.4e-06 6e-06	/	1e-05 / 2e-05		-0.119	0.509	0.728
Nitrosococcus halophilus Nc4	3.8e-06 7.5e-06	/	1.04e-05 1.88e-05	/	2.1e-06 5.1e-06	/	7.6e-06 1.43e-05	/	-0.118	0.512	0.728
Neptuniibacter caesariensis MED92	0 / 0		1.04e-05 1.66e-05	/	0 / 0		3e-06 / 8.2e-06		-0.219	0.220	0.728
Methylocystis sp ATCC 49242	9.7e-06 1.01e-05	/	5.9e-06 7.1e-06	/	0 / 0		2.7e-06 / 5.4e-06		0.110	0.543	0.728
Gemmatimonas aurantiaca	3.3e-06 6.6e-06	/	4.2e-06 7e-06	/	1.8e-06 4.5e-06	/	6.5e-06 / 1.1e-05		-0.172	0.339	0.728
Erythrobacter sp SD 2	7.7e-06 1.63e-05	/	1.02e-05 2.04e-05	/	0 / 0		7.8e-06 1.68e-05	/	-0.161	0.372	0.728
Erythrobacter litoralis	6.1e-06 9.3e-06	/	8.6e-06 1.57e-05	/	0 / 0		8.6e-06 1.05e-05	/	-0.151	0.400	0.728
Citromicrobium bathyomarinum	6e-06 9.1e-06	/	8.3e-06 1.84e-05	/	0 / 0		9.6e-06 1.24e-05	/	-0.102	0.572	0.728

*P-value adjusted using the Benjamini and Hochberg method

Table S12.3.6 Network statistics.

	Mean Degree	Mean betweenness centrality	Mean closeness centrality	Mean Eigenvector centrality	%Positve connection	%Negative connection	ratio	Fragmentation	Modularity	Transitivity	Diameter	Mean-short-pathways	K-Smirnov p-value	Number node
Co-correnceEcological network	3.88	262.38	0.001	0.106	0.431	0.569	0.758	0.134	0.428	0.155	10	4.079	0.603	176
Functional network	17.15	10364.49	1.97E-07	0.018				0.433	0.780	0.369	26	6.600	0.702	3700

Table S12.3.7 Bayesian network quantile 95 nodes. The quantile 95 based on the three different statistics: the degree centrality, the number of direct connections with a clinical variable and the MB size.

Node name	Number of direct connections (degree centrality)
DNA seq Terpenoid backbone biosynthesis ko00900	17
BPI (ng/mL)	17
DNA seq Valine leucine and isoleucine degradation ko00280	18
RNA seq Butanoate metabolism ko00650	18
%CD4+HLA−DR+CD38+ T−cells	18
DNA seq Alanine aspartate and glutamate metabolism ko00250	19
DNA seq Aminobenzoate degradation ko00627	19
sj/β−TREC ratio	22
CD4+ T−cell counts (cells/uL)	24
DNA seq Nicotinate and nicotinamide metabolism ko00760	25

Node name	Number of direct connections with a clinical variable
DNA seq Limonene and pinene degradation ko00903	7
Metabolite Diradylglycerols [GL02]|Glycerolipids [GL]|LMGL02010346	7
Metabolite Diradylglycerols [GL02]|Glycerolipids [GL]|582.5218@10.206741	7
DNA seq Glyoxylate and dicarboxylate metabolism ko00630	9
Metabolite Steroid conjugates [ST05]|Sterol Lipids [ST]|LMST05010043	9
DNA seq Valine leucine and isoleucine degradation ko00280	10
RNA seq ABC transporters ko02010	11
DNA seq Nicotinate and nicotinamide metabolism ko00760	12
RNA seq Butanoate metabolism ko00650	16

Node name	Markov blanket size in nodes
DNA seq Aminobenzoate degradation ko00627	67
Metabolite Diradylglycerols [GL02]|Glycerolipids [GL]|582.5218@10.206741	67
Metabolite Steroid conjugates [ST05]|Sterol Lipids [ST]|LMST05010043	68
RNA seq ABC transporters ko02010	69
DNA seq Glyoxylate and dicarboxylate metabolism ko00630	74
DNA seq D Alanine metabolism ko00473	75

12.4. Publications related to the thesis

Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals

Vázquez-Castellanos JF*, Serrano-Villar S*, Latorre A, Artacho A, Ferrús ML, Madrid N, Vallejo A, Sainz T, Martínez-Botas J, Ferrando-Martínez S, Vera M, Dronda F, Leal M, Del Romero J, Moreno S, Estrada V, Gosalbes MJ, Moya A.
* These authors contributed equally to this work.

Mucosal Immunol. 2015 Jul;8(4):760-72. doi: 10.1038/mi.2014.107. Epub 2014 Nov 19.

Abstract

Altered interplay between gut mucosa and microbiota during treated HIV infection may possibly contribute to increased bacterial translocation and chronic immune activation, both of which are predictors of morbidity and mortality. Although a dysbiotic gut microbiota has recently been reported in HIV+ individuals, the metagenome gene pool associated with HIV infection remains unknown. The aim of this study is to characterize the functional gene content of gut microbiota in HIV+ patients and to define the metabolic pathways of this bacterial community, which is potentially associated with immune dysfunction. We determined systemic markers of innate and adaptive immunity in a cohort of HIV-infected individuals on successful antiretroviral therapy without comorbidities and in healthy non-HIV-infected subjects. Metagenome sequencing revealed an altered functional profile, with enrichment of the genes involved in various pathogenic processes, lipopolysaccharide biosynthesis, bacterial translocation, and other inflammatory pathways. In contrast, we observed depletion of genes involved in amino acid metabolism and energy processes. Bayesian networks showed significant interactions between the bacterial community, their altered metabolic pathways, and systemic markers of immune dysfunction. This study reveals altered metabolic activity of microbiota and provides novel insight into the potential host-microbiota interactions driving the sustained inflammatory state in successfully treated HIV-infected patients.

The effects of prebiotics on microbial dysbiosis, butyrate production and immunity in HIV-infected subjects

Serrano-Villar S*, **Vázquez-Castellanos JF***, Vallejo A, Latorre A, Sainz T, Ferrando-Martínez S, Rojo D, Martínez-Botas J, Del Romero J, Madrid N, Leal M, Mosele JI, Motilva MJ, Barbas C, Ferrer M, Moya A, Moreno S, Gosalbes MJ, Estrada V.
* These authors contributed equally to this work.

Mucosal Immunol. 2016 Dec 21. doi: 10.1038/mi.2016.122. [Epub ahead of print].

Abstract
Altered interactions between the gut mucosa and bacteria during HIV infection seem to contribute to chronic immune dysfunction. A deeper understanding of how nutritional interventions could ameliorate gut dysbiosis is needed. Forty-four subjects, including 12 HIV+ viremic untreated (VU) patients, 23 antiretroviral therapy-treated (ART+) virally suppressed patients (15 immunological responders and 8 non-responders) and 9 HIV- controls (HIV-), were blindly randomized to receive either prebiotics (scGOS/lcFOS/glutamine) or placebo (34/10) over 6 weeks in this pilot study. We assessed fecal microbiota composition using deep 16S rRNA gene sequencing and several immunological and genetic markers involved in HIV immunopathogenesis. The short dietary supplementation attenuated HIV-associated dysbiosis, which was most apparent in VU individuals but less so in ART+ subjects, whose gut microbiota was found more resilient. This compositional shift was not observed in the placebo arm. Significantly, declines in indirect markers of bacterial translocation and T-cell activation, improvement of thymic output, and changes in butyrate production were observed. Increases in the abundance of *Faecalibacterium* and *Lachnospira* strongly correlated with moderate but significant increases of butyrate production and amelioration of the inflammatory biomarkers soluble CD14 and high-sensitivity C-reactive protein, especially among VU. Hence, the bacterial butyrate synthesis pathway holds promise as a viable target for interventions.

HIV infection results in metabolic alterations in the gut microbiota different from those induced by other diseases

Serrano-Villar S, Rojo D, Martínez-Martínez M, Deusch S, **Vázquez-Castellanos JF**, Sainz T, Vera M, Moreno S, Estrada V, Gosalbes MJ, Latorre A, Margolles A, Seifert J, Barbas C, Moya A, Ferrer M.

Sci Rep. 2016 May 18;6:26192. doi: 10.1038/srep26192.

Abstract

Imbalances in gut bacteria have been associated with multiple diseases. However, whether there are disease-specific changes in gut microbial metabolism remains unknown. Here, we demonstrate that human immunodeficiency virus (HIV) infection (n = 33) changes, at quantifiable levels, the metabolism of gut bacteria. These changes are different than those observed in patients with the auto-immune disease systemic lupus erythaematosus (n = 18), and *Clostridium difficile*-associated diarrhoea (n = 6). Using healthy controls as a baseline (n = 16), we demonstrate that a trend in the nature and directionality of the metabolic changes exists according to the type of the disease. The impact on the gut microbial activity, and thus the metabolite composition and metabolic flux of gut microbes, is therefore disease-dependent. Our data further provide experimental evidence that HIV infection drastically changed the microbial community, and the species responsible for the metabolism of 4 amino acids, in contrast to patients with the other two diseases and healthy controls. The identification in this present work of specific metabolic deficits in HIV-infected patients may define nutritional supplements to improve the health of these patients.

Gut Bacteria Metabolism Impacts Immune Recovery in HIV-infected Individuals

Serrano-Villar S, Rojo D, Martínez-Martínez M, Deusch S, **Vázquez-Castellanos JF**, Bargiela R, Sainz T, Vera M, Moreno S, Estrada V, Gosalbes MJ, Latorre A, Seifert J, Barbas C, Moya A, Ferrer M.

EBioMedicine. 2016 Jun;8:203-16. doi: 10.1016/j.ebiom.2016.04.033. Epub 2016 Apr 27.

Abstract

While changes in gut microbial populations have been described in human immuno-deficiency virus (HIV)-infected patients undergoing antiretroviral therapy (ART), the mechanisms underlying the contributions of gut bacteria and their molecular agents (metabolites and proteins) to immune recovery remain unexplored. To study this, we examined the active fraction of the gut microbiome, through examining protein synthesis and accumulation of metabolites inside gut bacteria and in the bloodstream, in 8 healthy controls and 29 HIV-infected individuals (6 being longitudinally studied). We found that HIV infection is associated to dramatic changes in the active set of gut bacteria simultaneously altering the metabolic outcomes. Effects were accentuated among immunological ART responders, regardless diet, subject characteristics, clinical variables other than immune recovery, the duration and type of ART and sexual preferences. The effect was found at quantitative levels of several molecular agents and active bacteria which were herein identified and whose abundance correlated with HIV immune pathogenesis markers. Although, we cannot rule out the possibility that some changes are partially a random consequence of the disease status, our data suggest that most likely reduced inflammation and immune recovery is a joint solution orchestrated by both the active fraction of the gut microbiota and the host.

12.5. Other publications performed during the thesis

Active and Secretory IgA-Coated Bacterial Fractions Elucidate Dysbiosis in Clostridium difficile Infection.

Džunková M, Moya A, **Vázquez-Castellanos JF**, Artacho A, Chen X, Kelly C, D'Auria G.

mSphere. 2016 May 25;1(3). pii: e00101-16. doi: 10.1128/mSphere.00101-16. eCollection 2016 May-Jun.

Abstract

The onset of Clostridium difficile infection (CDI) has been associated with treatment with wide-spectrum antibiotics. Antibiotic treatment alters the activity of gut commensals and may result in modified patterns of immune responses to pathogens. To study these mechanisms during CDI, we separated bacteria with high cellular RNA content (the active bacteria) and their inactive counterparts by fluorescence-activated cell sorting (FACS) of the fecal bacterial suspension. The gut dysbiosis due to the antibiotic treatment may result in modification of immune recognition of intestinal bacteria. The immune recognition patterns were assessed by FACS of bacterial fractions either coated or not with intestinal secretory immunoglobulin A (SIgA). We described the taxonomic distributions of these four bacterial fractions (active versus inactive and SIgA coated versus non-SIgA coated) by massive 16S rRNA gene amplicon sequencing and quantified the proportion of C. difficile toxin genes in the samples. The overall gut microbiome composition was more robustly influenced by antibiotics than by the C. difficile toxins. Bayesian networks revealed that the *C. difficile* cluster was preferentially SIgA coated during CDI. In contrast, in the CDI-negative group Fusobacterium was the characteristic genus of the SIgA-opsonized fraction. Lactobacillales and *Clostridium cluster IV* were mostly inactive in CDI-positive patients. In conclusion, although the proportion of C. difficile in the gut is very low, it is able to initiate infection during the gut dysbiosis caused by environmental stress (antibiotic treatment) as a consequence of decreased activity of the protective bacteria.

IMPORTANCE

C. difficile is a major enteric pathogen with worldwide distribution. Its expansion is associated with broad-spectrum antibiotics which disturb the normal gut microbiome. In this study, the DNA sequencing of highly active bacteria and bacteria opsonized by intestinal secretory immunoglobulin A (SIgA) separated from the whole bacterial community by FACS elucidated how the gut dysbiosis promotes C. difficile infection (CDI). Bacterial groups with inhibitory effects on C. difficile growth, such as Lactobacillales, were mostly inactive in the CDI patients. C. difficile was typical for the bacterial fraction opsonized by SIgA in patients with CDI, while Fusobacterium was characteristic for the SIgA-opsonized fraction of the controls. The study demonstrates that sequencing of specific bacterial fractions provides additional information about dysbiotic processes in the gut. The detected patterns have been confirmed with the whole patient cohort independently of the taxonomic differences detected in the nonfractionated microbiomes.

Carriage of Enterobacteria Producing Extended-Spectrum β-Lactamases and Composition of the Gut Microbiota in an Amerindian Community

Gosalbes MJ, **Vázquez-Castellanos JF**, Angebault C, Woerther PL, Ruppé E, Ferrús ML, Latorre A, Andremont A, Moya A.

Antimicrob Agents Chemother. 2015 Nov 9;60(1):507-14. doi: 10.1128/AAC.01528-15.

Abstract

Epidemiological and individual risk factors for colonization by enterobacteria producing extended-spectrum beta-lactamases (E-ESBL) have been studied extensively, but whether such colonization is associated with significant changes in the composition of the rest of the microbiota is still unknown. To address this issue, we assessed in an isolated Amerindian Guianese community whether intestinal carriage of E-ESBL was associated with specificities in gut microbiota using metagenomic and metatranscriptomic approaches. While the richness of taxa of the active microbiota of carriers was similar to that of noncarriers, the taxa were less homogeneous. In addition, species of four genera, *Desulfovibrio, Oscillospira, Parabacteroides*, and *Coprococcus*, were significantly more abundant in the active microbiota of noncarriers than in the active microbiota of carriers, whereas such was the case only for species of *Desulfovibrio* and *Oscillospira* in the total microbiota. Differential genera in noncarrier microbiota could either be associated with resistance to colonization or be the consequence of the colonization by E-ESBL.

Fragmentation and Coverage Variation in Viral Metagenome Assemblies, and Their Effect in Diversity Calculations

García-López R*, **Vázquez-Castellanos JF***, Moya A.
* These authors contributed equally to this work.

Front Bioeng Biotechnol. 2015 Sep 17;3:141. doi: 10.3389/fbioe.2015.00141. ECollection 2015.

Abstract

Metagenomic libraries consist of DNA fragments from diverse species, with varying genome size and abundance. High-throughput sequencing platforms produce large volumes of reads from these libraries, which may be assembled into contigs, ideally resembling the original larger genomic sequences. The uneven species distribution, along with the stochasticity in sample processing and sequencing bias, impacts the success of accurate sequence assembly. Several assemblers enable the processing of viral metagenomic data de novo, generally using overlap layout consensus or de Bruijn graph approaches for contig assembly. The success of viral genomic reconstruction in these datasets is limited by the degree of fragmentation of each genome in the sample, which is dependent on the sequencing effort and the genome length. Depending on ecological, biological, or procedural biases, some fragments have a higher prevalence, or coverage, in the assembly. However, assemblers must face challenges, such as the formation of chimerical structures and intra-species variability. Diversity calculation relies on the classification of the sequences that comprise a metagenomic dataset. Whenever the corresponding genomic and taxonomic information is available, contigs matching the same species can be classified accordingly and the coverage of its genome can be calculated for that species. This may be used to compare populations by estimating abundance and assessing species distribution from this data. Nevertheless, the coverage does not take into account the degree of fragmentation, or else genome completeness, and is not necessarily representative of actual species distribution in the samples. Furthermore, undetermined sequences are abundant in viral metagenomic datasets, resulting in several independent contigs that cannot be assigned by homology or genomic information. These may only be classified as different operational taxonomic units (OTUs), sometimes remaining inadvisably unrelated. Thus, calculations using contigs as different OTUs ultimately overestimate diversity when compared to diversity calculated from species coverage. In order to compare the effect of coverage and fragmentation, we generated three sets of simulated Illumina paired-end reads with different sequencing depths. We compared different assemblies performed with RayMeta, CLC Assembly Cell, MEGAHIT, SPAdes, Meta-IDBA, SOAPdenovo, Velvet, Metavelvet, and MIRA with the best attainable assemblies for each dataset (formed by arranging data using known genome coordinates) by calculating different assembly statistics. A new fragmentation score was included to estimate the degree of genome fragmentation of each taxon and adjust the coverage accordingly. The abundance in the metagenome was compared by bootstrapping the assembly data and hierarchically clustering them with the best possible assembly. Additionally, richness and diversity indexes were calculated for all the resulting assemblies and were assessed under two distributions: contigs as independent OTUs and sequences classified by species. Finally, we search for the strongest correlations between

the diversity indexes and the different assembly statistics. Although fragmentation was dependent of genome coverage, it was not as heavily influenced by the assembler. The sequencing depth was the predominant attractor that influenced the success of the assemblies. The coverage increased notoriously in larger datasets, whereas fragmentation values remained lower and unsaturated. While still far from obtaining the ideal assemblies, the RayMeta, SPAdes, and the CLC assemblers managed to build the most accurate contigs with larger datasets while Meta-IDBA showed a good performance with the medium-sized dataset, even after the adjusted coverage was calculated. Their resulting assemblies showed the highest coverage scores and the lowest fragmentation values. Alpha diversity calculated from contigs as OTUs resulted in significantly higher values for all assemblies when compared with actual species distribution, showing an overestimation due to the increased predicted abundance. Conversely, using PHACCS resulted in lower values for all assemblers. Different association methods (random-forest, generalized linear models, and the Spearman correlation index) support the number of contigs, the coverage, and fragmentation as the assembly parameters that most affect the estimation of the alpha diversity. Coverage calculations may provide an insight into relative completeness of a genome but they overlook missing fragments or overly separated sequences in a genome. The assembly of a highly fragmented genomes with high coverage may still lead to the clustering of different OTUs that are actually different fragments of a genome. Thus, it proves useful to penalize coverage with a fragmentation score. Using contigs for calculating alpha diversity result in overestimation but it is usually the only approach available. Still, it is enough for sample comparison. The best approach may be determined by choosing the assembler that better fits the sequencing depth and adjusting the parameters for longer accurate contigs whenever possible whereas diversity may be calculated considering taxonomical and genomic information if available.

Effect of daily intake of pomegranate juice on fecal microbiota and feces metabolites from healthy volunteers

Mosele JI, Gosalbes MJ, Macià A, Rubió L, **Vázquez-Castellanos JF**, Jiménez Hernández N, Moya A, Latorre A, Motilva MJ.

Mol Nutr Food Res. 2015 Oct;59(10):1942-53. doi: 10.1002/mnfr.201500227. Epub 2015 Aug 26.

Abstract

SCOPE:

The purpose of the study was to evaluate the effect, regarding the metabolic and microbial profile of feces, of diet supplementation of healthy adults with pomegranate juice (PJ).

METHODS AND RESULTS:

Twelve healthy adults were recruited to the study, which consisted of the intake of 200 mL/day of PJ during 4 weeks. Feces were collected before and after the supplementation with PJ. Metabolites (phenolic catabolites, short-chain fatty acids, and fecal steroids) and microbial profile were analyzed at baseline and at 4 weeks. Fecal phenolic metabolites, 3-phenylpropionic acid, catechol, hydroxytyrosol, and urolithin A, showed a significant increase in their concentration after supplementation with PJ. Among fecal steroids, parallel to the significant increase of cholesterol concentration, a significant decrease of coprostanol was observed. Although no significant changes in the microbiota profile were observed, different relationships between initial microbiota and the metabolites produced were found. Catechol showed positive and negative correlation with *Oscillospora* and *Paraprevotella* genera, respectively, and 3-phenylpropionic acid was positively correlated with Odoribacter genus.

CONCLUSION:

Inclusion of PJ in the diet did not significantly alter the gut microbiota composition in healthy adults, but the individual bacterial composition could contribute to the generation of potential health-promoting phenolic metabolites.

Molecular epidemiology studies on the immigrant population in Spain

González-Candelas F, Alma Bracho M, Comas I, d'Auria G, D Unková M, García R, Gosalbes MJ, Isaac S, Latorre A, López-Labrador FX, Patiño Galindo JÁ, Palero F, Pérez-Brocal V, Pérez-Cobas AE, Sánchez-Busó L, Silva FJ, **Vázquez-Castellanos JF**, Moya A.

Rev Esp Salud Publica. 2014 Nov-Dec;88(6):819-28. doi: 10.4321/S1135-57272014000600013.

Abstract

BACKGROUND:

Molecular epidemiology is a new scientific discipline which allows to integrate information on the genetic variation of infectious pathogens with their diffusion in a population and its subgroups including, for instance, resistance mutations to antibiotics and antiretrovirals. We present the results of an analysis of scientific publications that analyze the health status of the immigrant population in Spain from a molecular epidemiology perspective.

METHODS:

We reviewed original articles published in 1998-2014 with the keywords "molecular epidemiology", "molecular typing", "sequencing", "immigrant", and "Spain".

RESULTS:

From a total of 267 articles identified initially, only 50 passed through the established filters. Most of them (36) analyzed infections by *Mycobacterium tuberculosis* (3) and HIV (3), followed at a large distance by Staphylococcus aureus and hepatitis B virus. The main goal of these works was the typing of the pathogen and to determine the frequency of resistance mutations.

CONCLUSION:

Is difficult to generalize the conclusions from the analyzed articles because most of them have a purely descriptive and quite restricted scope, considering the type and size of the samples studied. Several studies are focused on the most likely origin for the strains or variants of the pathogen but others also reveal transmissions from the local to the immigrant populations.

Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach

Pérez-Brocal V, García-López R, **Vázquez-Castellanos JF**, Nos P, Beltrán B, Latorre A, Moya A.

Clin Transl Gastroenterol. 2013 Jun 13;4:e36. doi: 10.1038/ctg.2013.9.

Abstract

OBJECTIVES:

This study aimed to analyze and compare the diversity and structure of the viral and microbial communities in fecal samples from a control group of healthy volunteers and from patients affected by Crohn's disease (CD).

METHODS:

Healthy adult controls (n=8) and patients affected by ileocolic CD (n=11) were examined for the viral and microbial communities in their feces and, in one additional case, in the intestinal tissue. Using two different approaches, we compared the viral and microbial communities in several ways: by group (patients vs. controls), entity (viruses vs. bacteria), read assembly (unassembled vs. assembled reads), and methodology (our approach vs. an existing pipeline). Differences in the viral and microbial composition, and abundance between the two groups were analyzed to identify taxa that are under- or over-represented.

RESULTS:

A lower diversity but more variability between the CD samples in both virome and microbiome was found, with a clear distinction between groups based on the microbiome. Only ≈5% of the differential viral biomarkers are more represented in the CD group (Synechococcus phage S CBS1 and Retroviridae family viruses), compared with 95% in the control group. Unrelated patterns of bacteria and bacteriophages were observed.

CONCLUSIONS:

Our use of an extensive database is critical to retrieve more viral hits than in previous approaches. Unrelated patterns of bacteria and bacteriophages may be due to uneven representation of certain viruses in databases, among other factors. Further characterization of Retroviridae viruses in the CD group could be of interest, given their links with immunodeficiency and the immune responses. To conclude, some methodological considerations underlying the analysis of the viral community composition and abundance are discussed.

Comparison of different assembly and annotation tools on analysis of simulated viral metagenomic communities in the gut

Vázquez-Castellanos JF*, García-López R*, Pérez-Brocal V, Pignatelli M, Moya A.
* These authors contributed equally to this work.

BMC Genomics. 2014 Jan 18;15:37. doi: 10.1186/1471-2164-15-37.

Abstract

BACKGROUND:

The main limitations in the analysis of viral metagenomes are perhaps the high genetic variability and the lack of information in extant databases. To address these issues, several bioinformatic tools have been specifically designed or adapted for metagenomics by improving read assembly and creating more sensitive methods for homology detection. This study compares the performance of different available assemblers and taxonomic annotation software using simulated viral-metagenomic data.

RESULTS:

We simulated two 454 viral metagenomes using genomes from NCBI's RefSeq database based on the list of actual viruses found in previously published metagenomes. Three different assembly strategies, spanning six assemblers, were tested for performance: overlap-layout-consensus algorithms Newbler, Celera and Minimo; de Bruijn graphs algorithms Velvet and MetaVelvet; and read probabilistic model Genovo. The performance of the assemblies was measured by the length of resulting contigs (using N50), the percentage of reads assembled and the overall accuracy when comparing against corresponding reference genomes. Additionally, the number of chimeras per contig and the lowest common ancestor were estimated in order to assess the effect of assembling on taxonomic and functional annotation. The functional classification of the reads was evaluated by counting the reads that correctly matched the functional data previously reported for the original genomes and calculating the number of over-represented functional categories in chimeric contigs. The sensitivity and specificity of tBLASTx, PhymmBL and the k-mer frequencies were measured by accurate predictions when comparing simulated reads against the NCBI Virus genomes RefSeq database.

CONCLUSIONS:

Assembling improves functional annotation by increasing accurate assignations and decreasing ambiguous hits between viruses and bacteria. However, the success is limited by the chimeric contigs occurring at all taxonomic levels. The assembler and its parameters should be selected based on the focus of each study. Minimo's non-chimeric contigs and Genovo's long contigs excelled in taxonomy assignation and functional annotation, respectively.tBLASTx stood out as the best approach for taxonomic annotation for virus identification. PhymmBL proved useful in datasets in which no related sequences are present as it uses genomic features that may help identify distant taxa. The k-frequencies underperformed in all viral datasets.

ABBREVIATIONS

ADMA	Asymmetric Dimethylarginine. It is an endothelial dysfunction and cardiovascular disease marker
AIDS	Acquired immunodeficiency syndrome
AMPs	Antimicrobial peptides
ART	Antiretroviral therapy
BCFA	Branched-chain fatty acids
BH	Benjamini-Hochberg
BIC	Bayesian Information Criterion
BN	Bayesian networks
BPI	Bactericidal permeability-increasing protein. Marker for bacterial translocation
CARD	Comprehensive Antibiotic Resistance Database
CCR5	CC-chemokine receptor 5
CD	Crohn's disease
CD14	Cluster of differentiation 14
CD25+	CD4 or CD8 percentage of cells expressing markers of activation CD25+
CD38+	CD4 or CD8 percentage of cells expressing markers of activation CD38+
CD38+/HLA-DR+	CD4 or CD8 percentage of cells expressing markers of activation HLADR+
CD4	CD4+ T-cell counts (cells/uL)
CD57+	CD4 or CD8 percentage of cells expressing markers of senescence CD57+
CD8	CD8+ T-cell counts (cells/uL)
CDI	Clostridium difficile infection
CID	Collision-induced dissociation
CPD	Conditional probability distribution
CRP	C-reactive protein
CXCR4	CXC-chemokine receptor 4
DCs	Dendritic cells
D-dimers	Fibrin degradation product that serves as a thrombosis marker
DGA	Probabilistic directed acyclic graphical
DNAseq	Metagenomic data sequences
FB	Filamentous bacteria
FMD	Flow mediated dilation

FMT	Fecal microbial transplantation
FOS	Fructo-oligosaccharides
FOXP3	Forkhead box P3
GALT	Gut-associated lymphoid tissue
GIT	Gastrointestinal tract
GLM	Generalized Linear Models
GOS	Galacto-oligosaccharides
H2S	Hydrogen-sulphide
HC	Hill-climbing score-based learning algorithm
HDAC	Deacetylases
HIV	Human Immunodeficiency Virus
HIV-	HIV-uninfected
HLADR+	CD4 or CD8 percentage of cells expressing markers of activation HLADR+
hs-CRP	High-sensitivity C-reactive protein. Marker for systemic inflammation
IBD	Inflammatory bowel disease dysbiosis
IDO1	Indoleamine 2,3-dioxygenase 1
IEC	Intestinal epithelial cells
IFN-γ	Interferon-gamma
IgA	Immunoglobulin A
IL-12	Interleukin-12
IL-17	Interleukin-17
IL-22	Interleukin-22
IL6	Interleukin-6
INR	Immunological ART non-responders
IQR	Interquartile range
IR	Immunological ART responders
JPD	Joint probability distribution
KO	KEGG orthologous
lasso	Least absolute shrinkage and selection operator
LC-ESI-QTOF-MS	Liquid chromatography–Electro Spray Ionization- Quadrupole Time of Flight- Mass Spectrometry
LC-MS	Liquid chromatography-mass spectrometry
LDA	Linear discriminative analysis
LEfSe	Linear discriminative analysis effect size
LPS	Lipopolysaccharide
MAMPs	Microbe-associated molecular patterns
MB	Markov blanket
MID	Multiplex Identifier
MSM	Men who have sex with men
Nadir CD4+ T-cells counts	The lowest point to which the CD4 count has dropped
NF-κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells

NK	Natural killer cells
NMDS	Non-metric multidimensional scaling analysis
nts	Nucleotides
ORF	Open reading frame
OTUs	Operational taxonomic units
PAM	Partitioning Around Medoids algorithm
PCA	Principal component analysis
PCoA	Principal coordinates analysis
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
qPCR	Quantitative PCR
RDP	Ribosomal Database Project-II
RNAseq	Metatranscriptomic sequencing data
sCD14	Soluble Cluster of differentiation 14. Marker for bacterial translocation
SCFA	Short-chain fatty acids
SIV	Simian immunodeficiency virus
sj/β-TREC ratio	The ratio of two different T-cell receptor excision circles (TRECs). The signal-joint TRECs (sj-TREC) and the β-TREC . It is a direct estimator of the thymic function
sj-TREC	Signal-joint T-cell receptor excision circles
SSU	16S small subunit rRNA
T3SS	Type 3 secretion system
Th cells	T helper cells
Th1	T helpers 1 cells
Th17	T helper 17 cell
Th2	T helpers 2 cells
TLR2	Toll-like receptor 2
TLR4	Toll-like receptor 4
TMA	Trimethylamine
TMAO	Trimethylamine N-oxide
TNF	Tumor necrosis factor
Treg	Regulatory T cells
UC	Ulcerative colitis
VL	Viral load
VU	Virus untreated
β-TREC	Product of the ß chain TCR rearrangement at the most immature thymocyte subset and the sj-TREC