

# SCIENTIFIC REPORTS

OPEN

## A Lachnospiraceae-dominated bacterial signature in the fecal microbiota of HIV-infected individuals from Colombia, South America

Homero San-Juan-Vergara<sup>1</sup>, Eduardo Zurek<sup>2</sup>, Nadim J. Ajami<sup>3</sup>, Christian Mogollon<sup>4</sup>, Mario Peña<sup>1</sup>, Ivan Portnoy<sup>2</sup>, Jorge I. Vélez<sup>2</sup>, Christian Cadena-Cruz<sup>1</sup>, Yirys Diaz-Olmos<sup>1</sup>, Leidy Hurtado-Gómez<sup>1</sup>, Silvana Sanchez-Sit<sup>5</sup>, Danitza Hernández<sup>4</sup>, Irina Urruchurtu<sup>4</sup>, Pierina Di-Ruggiero<sup>4</sup>, Ella Guardo-García<sup>6</sup>, Nury Torres<sup>6</sup>, Oscar Vidal-Orjuela<sup>1</sup>, Diego Viasus<sup>1</sup>, Joseph F. Petrosino<sup>3</sup> & Guillermo Cervantes-Acosta<sup>1</sup>

HIV infection has a tremendous impact on the immune system's proper functioning. The mucosa-associated lymphoid tissue (MALT) is significantly disarrayed during HIV infection. Compositional changes in the gut microbiota might contribute to the mucosal barrier disruption, and consequently to microbial translocation. We performed an observational, cross-sectional study aimed at evaluating changes in the fecal microbiota of HIV-infected individuals from Colombia. We analyzed the fecal microbiota of 37 individuals via 16S rRNA gene sequencing; 25 HIV-infected patients and 12 control (non-infected) individuals, which were similar in body mass index, age, gender balance and socioeconomic status. To the best of our knowledge, no such studies have been conducted in Latin American countries. Given its compositional nature, microbiota data were normalized and transformed using Aitchison's Centered Log-Ratio. Overall, a change in the network structure in HIV-infected patients was revealed by using the SPIEC-EASI MB tool. Genera such as *Blautia*, *Dorea*, *Yersinia*, *Escherichia-Shigella* complex, *Staphylococcus*, and *Bacteroides* were highly relevant in HIV-infected individuals. Differential abundance analysis by both sparse Partial Least Square-Discriminant Analysis and Random Forest identified a greater abundance of Lachnospiraceae-OTU69, *Blautia*, *Dorea*, *Roseburia*, and Erysipelotrichaceae in HIV-infected individuals. We show here, for the first time, a predominantly Lachnospiraceae-based signature in HIV-infected individuals.

The health complications arising from HIV infection associated to developing acquired immunodeficiency syndrome impose a heavy physical and psychological toll on people<sup>1</sup>. Approximately 78 million people have been diagnosed with HIV, of whom ~35 million people have died. In 2015 alone, there were 36.7 million people living with HIV<sup>1,2</sup>. Among the 4 HIV groups, M is the most widely distributed group and is comprised of nine subtypes. Subtype C is responsible for about one-half of all global infections, while subtype B, the prevalent virus in Colombia, is the most widespread<sup>3,4</sup>.

A mucosal barrier integrated by the intestinal epithelium and the underlying immune system favors tolerance towards bacteria present in the intestine lumen<sup>5-9</sup>. Commensal bacteria of the intestinal microbiota also contribute to the mucosal barrier by competing for space and resources with potentially pathobiontic bacteria<sup>10</sup>. HIV

<sup>1</sup>División Ciencias de la Salud, Fundación Universidad del Norte, Barranquilla, Colombia. <sup>2</sup>División de Ingenierías, Fundación Universidad del Norte, Barranquilla, Colombia. <sup>3</sup>Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, USA. <sup>4</sup>IPS Medicina Integral, Barranquilla, Colombia. <sup>5</sup>Maestría en Estadística Aplicada, Universidad del Norte, Barranquilla, Colombia. <sup>6</sup>IPS de la Costa, Barranquilla, Colombia. Correspondence and requests for materials should be addressed to H.S.-J.V. (email: [hsanjuan@uninorte.edu.co](mailto:hsanjuan@uninorte.edu.co)) or G.C.-A. (email: [guicerva@uninorte.edu.co](mailto:guicerva@uninorte.edu.co))

exhibits a predilection towards the gastrointestinal tract immune system because of the presence of highly susceptible cells<sup>11–13</sup>. In HIV-infected individuals, substantial decrease in the levels of Th17 lymphocytes, and albeit to a less extent, in the levels of Foxp3(+) T lymphocytes are found within Peyer's plaques and the intestine lamina<sup>1,14–16</sup>. Consequently, mucosal barrier is destroyed, and the permeability of the intestine is therefore increased. In turn, this results in a loss of tolerance and a chronic activation of CD4(+) T lymphocytes<sup>1,13,17,18</sup>.

Extrapolation from a population-level analysis on fecal microbiota composition suggested that 784 ± 40 genera of bacteria are capable of colonizing the human intestine<sup>19</sup>. Despite this enormous diversity, Falony *et al.* identified 17 bacterial genera that were common to 95% of individuals<sup>19</sup>. These genera, proposed to be the "core" of intestinal microbiota, include unclassified Lachnospiraceae, unclassified Ruminococcaceae, *Bacteroides*, *Blautia*, unclassified Erysipelotrichaceae, *Roseburia*, *Faecalibacterium*, unclassified Clostridiaceae, *Dorea*, *Alistipes*, unclassified Clostridiales, *Clostridium* XIVa, unclassified Veillonellaceae, unclassified Hyphomicrobiaceae, *Coprococcus*, *Clostridium* IV, and *Parabacteroides*<sup>19</sup>. Subsequently, this "core" was updated by removing *Alistipes*, *Clostridium* IV, and *Parabacteroides* after the incorporation of data obtained from 308 samples of Papua New Guinean, Peruvian, and Tanzanian origin<sup>19</sup>. Consequently, intestinal microbiota is influenced by the geographical location<sup>20,21</sup>. Environmental factors such as the type of diet or the presence of parasites, as well as to genetic factors, may contribute to such differences<sup>20,22</sup>.

Overall, the majority of studies found that HIV infection is associated with intestinal dysbiosis, characterized by increased numbers of Enterobacteriaceae and Enterococcaceae pathobionts, as well as by an enrichment of the genus *Prevotella*<sup>23–27</sup>. With the exception of three studies – two of which were conducted in China and the other based on a cohort in Uganda<sup>27–29</sup> – all studies were conducted on western countries where the diets are rich in fats but low in fiber and complex carbohydrates<sup>23–25</sup>. Interestingly, the results obtained in the Uganda study substantially differ from the consensus observed in HIV-infected individuals from developed countries. Particularly, this study did not find any differences between HIV-infected and non-infected individuals with respect to the genus *Prevotella*<sup>29</sup>.

Here, in an observational and cross-sectional study, we assayed the fecal microbiota composition of a group of 25 individuals infected with HIV and a control group of 12 non-infected individuals similar in age, socioeconomical status, diet, and gender distribution. All individuals lived in the Caribbean coastal region of Colombia. To the best of our knowledge, this is the first study of this type that has been conducted in a South American country. We detect a fecal signature that differentiates HIV-infected individuals from non-infected individuals. This signature is characterized by an increase in bacteria belonging to the genera Lachnospiraceae-OTU69, *Blautia*, *Dorea* and Erysipelotrichaceae-OTU109 in the fecal microbiota of HIV-infected individuals. These results are distinct from previously reported signatures, likely due to the differences that our study population might have regarding diet and genetic makeup.

## Results

### Demographic characteristics of the HIV-infected and control groups and sequencing approach.

We recruited 25 patients with chronic HIV infection living in Barranquilla and Cartagena – two closely located cities along the Colombian Caribbean Coast. These patients were being treated at two Healthcare Institutions which provide an integral care to HIV-infected patients (IPS Medicina Integral and IPS de la Costa). In addition, 12 non-infected individuals matched by gender distribution, body mass index (BMI), diet, and socioeconomic strata with the HIV-infected group were recruited from those who were blood donors of a blood bank (Table 1). Diet composition for both groups is shown as Supplementary Fig. S1.

La Rosa *et al.*<sup>30</sup> provided calculations for the power achieved in taxonomic-based human microbiome studies. La Rosa *et al.* found that studies with 5% significance level, 20,000 reads per subject and 25 subjects will have a power over 97%. In our study, we performed 30,000 reads per subject and included 37 individuals (25 HIV-infected and 12 controls). Kelly *et al.*<sup>31</sup> indicated that sample sizes above 10 individuals per group (n<sub>total</sub> = 20) likely afforded adequate statistical power (i.e., of at least 90%) for the primary outcome measure in microbiome studies. Thus, our sample size was enough to detect the differences to be reported.

We excluded those individuals who had a BMI ≤ 18.5 kg/m<sup>2</sup>, showed signs of active liver disease (determined by alanine aminotransferase levels), or took antibiotics during the last 30 days previous to sample collection. Three of the HIV-infected patients had consumed antibiotics 31 to 60 days prior to sample collection. More specifically, patient V.02 had undergone one cycle of trimethoprim/sulfamethoxazole, patient V.03 one cycle of dicloxacillin, and patient V.07 one cycle of ampicillin/azithromycin.

All recruited patients were recently diagnosed and they were in stage V according to Fiebig classification criteria. When recruited patients were clinically stable and had no acute illness. They were still in the assessment period for HAART assignment.

According to the Bristol index of stool consistency, the samples of 22 out of the 25 samples from HIV-infected patients were classified as normal (stool type 3 or 4), while samples V.15 and V.16 were classified as type 2 - slightly constipated, and only sample V.19 as type 6 - semi-liquid. All fecal samples obtained from the control group were classified as normal.

About 13.4 million pairs of reads were generated after sequencing the V4 region of 16S rRNA gene using the Illumina-MiSeq. 2 × 250 bp. After proper quality control and demultiplexing, the number of reads was reduced to 1,420,474, being equivalent to 264,027 unique sequences. Singletons and chimeras were also removed. This yielded 1,286,018 reads, or 726 operational taxonomic units (OTUs) after applying a similarity cut-off of 97% to define species. Approximately 29,856 reads were analyzed per sample (Supplementary Fig. S2).

There were no differences between HIV-infected and control groups in terms of observed species as well as in Chao and Shannon indexes (Supplementary Table S1). The complete list of the OTUs at the genus level is shown in Supplementary Table S2. The bacterial counts per each OTU at the genus level are shown in Supplementary Tables S3–S6.

	Non-Infected Control	HIV-Infected
Number of individuals	12	25
Age (years)	35.6 (18–54) Median: 39	29.4 (17–48) Median: 30
Male/Female ratio	7/5	15/10
Race	All Hispanics	All Hispanics
Body mass index (BMI kg/m <sup>2</sup> )	27.2 (23.5–33.9) Median: 26.0	24.1 (19.4–29.3) Median: 24.4
Socio-economic strata	Lower middle class: 33.3% Below poverty line: 66.7%	Lower middle class: 40% Below poverty line: 60%
Alcohol (low intake) (number of individuals)	2	1
Intermittent Smoking (number of individuals)	1	5
Viral load in blood plasma (copies/mL)	N/A	32,300.4 (0–240,000) Median: 58,698.8
CD4(+) T cell count (Cells/ $\mu$ L)	1,055.4 (501.2–2,619.3) Median: 841	546.5 (39.3–1,012.4) Median: 528.9
CD4/CD8 ratio	1.802 (0.76–2.60) Median: 1.80	0.606 (0.03–1.80) Median: 0.51
Percentage of Th-17 lymphocytes (%)	1.9 (0.4–4.4) Median: 1.55	2.5 (1.1–5.3) Median: 2.0
Percentage of activated CD4(+) T cells (%)	2.62 (1.0–3.9) Median: 2.6	9.6 (1.0–27.1) Median: 8.0
HIV-Low % activated CD4(+) T cells	NA	N = 4 (1 < 5%)
HIV-Medium % activated CD4(+) T cells	NA	N = 11 (5–10%)
HIV-High % activated CD4(+) T cells	NA	N = 10 (>10%)
Percentage of activated CD8(+) T cells (%)	8.8 (2.2–32.4) Median: 6.65	37.6 (4.0–82.4) Median: 35.3

**Table 1.** Descriptive data for subjects in the study. Data is presented as “mean(range) and median” for age, BMI, viral load, CD4(+) T cell count, CD4/CD8 ratio, percentage of Th17 lymphocytes, percentage of activated CD4(+) T cells, and percentage of activated CD8(+) T cells.

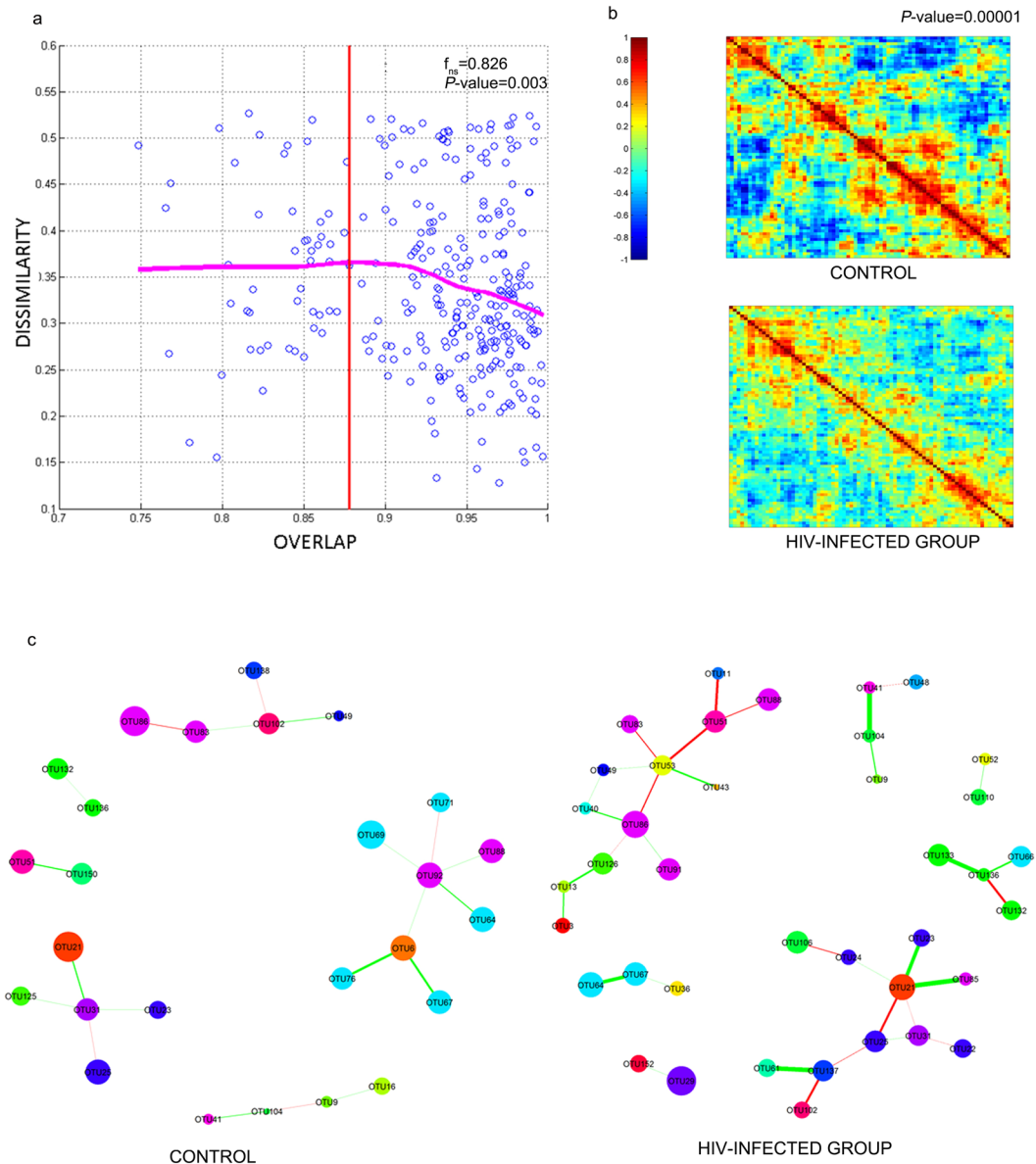
**The intestinal microbiota of HIV patients follows a universal pattern.** Bashan *et al.*<sup>32</sup> reported that as more genera are shared between people, each shared genus tend to have the same abundance in those individuals, reflecting a universal dynamics emerging among the various genera and species. This dynamic is lost in the presence of a strong destabilizing factor. This finding prompted us to investigate whether immune dysregulation due to HIV infection disrupted the dynamics of intestinal microbiota. Such immune dysregulation was evidenced by the inverse CD4/CD8 ratio and an increase in the percentage of activated CD4 T lymphocytes (Table 1). We built a Dissimilarity-Overlap Curve (DOC) for each pair of samples being compared, and calculated the fraction of genera that is shared versus how dissimilar the abundance is in those shared genera (“dissimilarity index”). DOC showed that 82.6% of sample pairs were located after the inflection point – the point at which the curve started taking a negative slope (Fig. 1a), meaning that the intestine of these individuals must be colonized by microorganisms that build relationships with each other, and thereby favor network building. The P-value of 0.003 was calculated as the fraction of bootstrap runs ( $n_{boot} = 1000$ ) resulting with non-negative slopes.

**Changes in the correlation patterns and network architecture of intestinal microbiota of HIV-infected patients.** Since DOC reflects an underlying network<sup>32</sup>, we compared the structure of correlation- and precision-matrices of the microbiota from HIV-infected individuals to the microbiota from the control group. Since the data had a compositional architecture, we first transformed it using Aitchison’s centered log-ratio (clr) procedure<sup>33–37</sup>. We then calculated the Pearson’s correlation matrices using heat maps (Fig. 1b). A Jennrich test<sup>38</sup> showed that the correlation structure for HIV-infected patients was significantly different from that of the control group (P-value < 0.00001). The microbiota of individuals from the control group presented more correlations compared to those of the HIV-infected group.

As correlation matrices included both direct and indirect relationships. We used the SPIEC-EASI-MB method<sup>39</sup> to construct a network assessing only direct relationships. A sparse network was built to highlight the more stable relationships (Fig. 1c). HIV-infected group displayed a far more enriched network architecture in terms of direct relations between OTUs. We also built networks whose nodes were fixed in their position to clearly show the differences between the network structures of the HIV-infected and control groups (Supplementary Fig. S3a). In HIV-infected group, seven clusters were found, two of which were composed by 14 and 11 OTUs, respectively. Meanwhile, six clusters were identified in the control group; the maximum number of OTUs observed for one cluster was eight (Supplementary Fig. S3b). In addition, both frequency-vs-edges and frequency-vs-degree plots confirm the higher correlated structure in terms of direct relations between nodes in the HIV network (Supplementary Fig. S3c,d).

When comparing both networks, we found differences in which genera took the role of major nodes and their relationships. For instance, the *Bifidobacterium* (OTU6), *Subdoligranulum* (OTU92) and *Enterobacter* (OTU132) genera that act as nodes in the control group disappeared in the HIV-infected group. In the control group, the genus *Bifidobacterium* establishes a mutualistic relationship with the genus *Dorea* (OTU67), while the genus *Subdoligranulum* establishes a weak relationship with the genus *Blautia* (OTU64). However, in HIV-infected patients, as the nodes corresponding to the *Bifidobacterium* and *Subdoligranulum* genera disappeared in HIV-infected group, a strong mutualistic relationship formed between two genera of the Lachnospiraceae family – *Blautia* and *Dorea*.

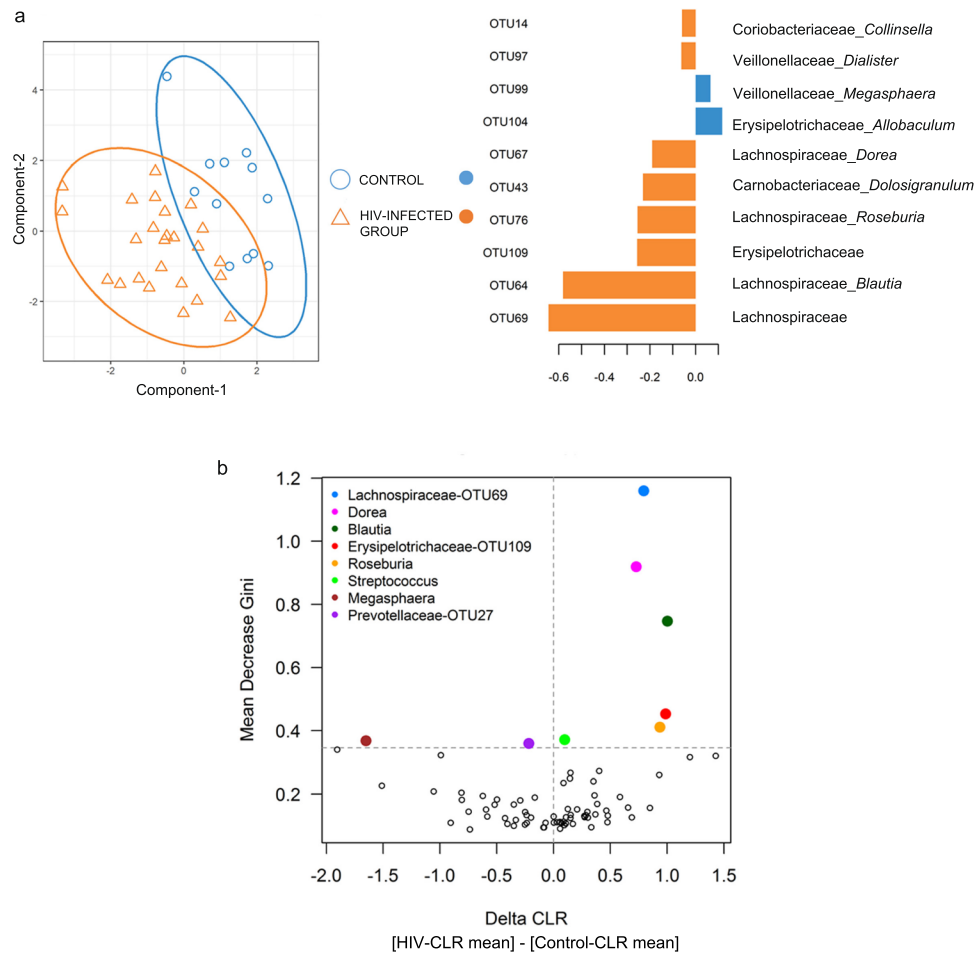
As the genus *Enterobacter* lost its major node property in HIV-infected patients, *Yersinia* (OTU136) established strong mutualistic relationships with the *Coprococcus* (OTU66) and the *Escherichia-Shigella* complex (OTU133). Both *Yersinia* and *Escherichia-Shigella* belong to the family Enterobacteriaceae. Although the



**Figure 1.** Network structure of the microbial communities found in HIV-infected and control individuals. **(a)** The fecal microbiota of HIV-infected individuals follows a pattern of universality as described by Bashan *et al.*<sup>32</sup>. Each point represents the comparisons between the samples of two individuals. The fraction of shared bacteria (*overlap*) is plotted against the dissimilarity index, which was evaluated based on the square root of the Jensen–Shannon divergence. The dissimilarity index calculates the distance between pairs of individuals in terms of abundance of shared species. The dissimilarity-overlap curve (DOC) is represented by a pink curve and was calculated using the robust LOWESS method. The fraction of points beyond the inflection point ( $O_c$ ) of the DOC curve – the point at which the curve shows a negative slope – was calculated by:  $f_{ns} = \frac{\text{number of pair samples with } O > O_c}{\text{total number of sample pairs}}$ . The  $P$ -values are finally calculated as the fraction of bootstrap runs resulting with non-negative slopes. **(b)** Heat map of Pearson Correlation of HIV-infected individuals and non-infected control group. We performed a Jennrich test to calculate the  $p$ -value associated to the correlation structures between HIV-infected patients control group. **(c)** Differences in the microbial networks of fecal samples obtained from HIV-infected and control individuals. Using the SPIEC-EASI MB method, networks were constructed from the table of clr-transformed OTUs.

genera *Staphylococcus* (OTU41) and *Allobaculum* (OTU104) are preserved in both groups, these genera became more collaborative in the HIV-infected group. Furthermore, the genus *Bacteroides* (OTU21) established associations with *Butyrivimonas* (OTU23) from the Bacteroidaceae family, and *Flavonifractor* (OTU 85) from the Ruminococcaceae family in HIV-infected group.

In summary, the microbiota network structure changed in patients infected with HIV, making the genera *Blautia* (OTU64), *Dorea* (OTU67), *Yersinia* (OTU136), *Escherichia-Shigella* (OTU133), *Staphylococcus* (OTU41), and *Bacteroides* (OTU21) relevant in the HIV network.



**Figure 2.** Identification of an OTU signature associated with HIV-infected individuals. **(a)** PCA analysis of the sPLS-DA model fed with clr-transformed OTUs (left), and the resulting contribution of OTUs of component 1 (right). **(b)** OTU importance according to the random forest method. The Mean Gini Decrease of each OTU is plotted versus the delta of the CLR averages for each OTU (i.e., for each OTU, the  $\Delta$ CLR was calculated subtracting the CLR mean of the HIV-infected group from the CLR mean of the control group).

### HIV-infected individuals have a distinct bacterial signature compared to control individuals.

Using the sparse-partial least squares-discriminant analysis (sPLS-DA) technique<sup>40,41</sup>, we found that the HIV-infected group separate from the control group (Fig. 2a). In order of decreasing influence, the former group consists of an undefined genus from the Lachnospiraceae family (OTU69), *Blautia* (OTU64), an undefined genus from the Erysipelotrichaceae family (OTU109), *Roseburia* (OTU76), *Dolosigranulum* (OTU43), *Dorea* (OTU67), *Dialister* (OTU97), and *Collinsella* (OTU14) (Fig. 2a). These genera belong to the families Lachnospiraceae (OTU69, *Blautia*, *Roseburia*, and *Dorea*), Erysipelotrichaceae (OTU109), Carnobacteriaceae (*Dolosigranulum*), Veillonellaceae (*Dialister*), and Coriobacteriaceae (*Collinsella*). Consequently, the order Clostridiales clustered most of the genera identified in the signature.

We used random forest (RF) classifiers<sup>42</sup>, a machine-learning technique – to evaluate whether the signature found by sPLS-DA could be replicated (Fig. 2b). The Mean Decrease Gini (MDG) criterion was used as a measure of relative importance for each OTU. We found that the results generated through RF nicely correspond with those of sPLS-DA. Five of the eight genera identified by sPLS-DA occupied the top positions having the highest MDG. In decreasing order these were Lachnospiraceae-OTU69 (MDG = 1.159), *Dorea* (MDG = 0.918), *Blautia* (MDG = 0.747), Erysipelotrichaceae-OTU109 (MDG = 0.453), and *Roseburia* (MDG = 0.411). The MDG values of *Dialister*, *Dolosigranulum* and *Collinsella* fell below the 90<sup>th</sup> percentile, with having MDG values of 0.32, 0.316 and 0.26 respectively. The  $\Delta$ -clr showed that the five concordant genera between the sPLS-DA and RF techniques were more abundant in the HIV-infected group. Interestingly, three of these genera (*Lachnospiraceae*-OTU69, *Blautia* and *Dorea*) belong to the Lachnospiraceae family.

**Specific bacteria are associated with the severity of HIV infection.** From the assessed immune parameters, both the percentage of activated CD4(+) T cells and CD4/CD8 ratio were the only factors able to discriminate between the HIV-infected and the control groups. Since CD4 loss is in great extent driven by immune activation<sup>43</sup>, we assessed CD4(+) T cell activation based on the coexpression of HLA-DR and CD38.

We represented the percent activation values as a continuous function and identified three distinct subgroups in the HIV-infected group using a mixture of Gaussian distributions (Fig. 3a). Subgroup “High” included 10 HIV-infected individuals with percentage of activated CD4(+) T cells levels greater than or equal to 10%; subgroup “Medium” consisted of 11 patients whose levels were  $\geq 5\%$  but  $< 10\%$ , and subgroup “Low” was comprised of four patients whose percentage values were comparable to those of the control group (i.e.,  $< 5\%$ ). The bacterial counts per each OTU at the genus level for the HIV subgroups and the control group are shown in Supplementary Tables 3–6 ANOVA-based analyses showed that the average percentage differed among subgroups ( $F_{3,33} = 38.86$ ,  $P < 0.00001$ ). This result was subsequently confirmed using a Tukey least significance difference test with 95% confidence intervals (CIs) was performed (Supplementary Table S7).

We found that HIV subgroups differed in indices of  $\alpha$ -diversity (Fig. 3b). In particular, the Chao, Shannon, and Simpson indices corresponded to the  $x$ -axis values of  $-2$ ,  $0.5$  and  $2$ , respectively. Based on Shannon and Simpson indexes, the subgroup-High showed a greater richness, homogeneity, and evenness in terms of bacterial OTUs compared with subgroup-Low ( $p = 0.0039$ ) (Supplementary Fig. S4). The highest Simpson indices and the lowest Shannon index score in subgroup-Low implied that a small share of the OTUs carried most of the bacterial counts ( $p = 0.0039$ ) (Supplementary Fig. S4). Nonetheless, this result should be taken with caution as this subgroup was comprised of only a small number of individuals.

We calculated  $\beta$ -diversity metrics by employing UNIFRAC weighted criteria and the principal coordinate analysis (PCoA) ordination method (Fig. 3c). The control group was distributed throughout the entire plot space, while the subgroups-High and -Low were located at the opposite sides on first principal coordinate (PERMANOVA  $p$ -value =  $0.002$ ). Although subgroup-Medium made a gradient between the two other subgroups, it was still significantly separated from subgroup-High (PERMANOVA  $p$ -value =  $0.045$ ) and from subgroup-Low (PERMANOVA  $p$ -value =  $0.035$ ).

We then probed the Firmicutes/Bacteroidetes (F/B) ratio across HIV-infected subgroups. We found an F/B higher than 1 in 90.0%, 54.4%, and none of the individuals from subgroups High, Medium, and Low, respectively. These suggest that the abundance of Firmicutes ( $P_{\text{ANOVA-FDR}} = 0.0268$ ), Clostridia ( $P_{\text{ANOVA-FDR}} = 0.0290$ ) and Clostridiales ( $P_{\text{ANOVA-FDR}} = 0.0456$ ) were significantly different among subgroups. Using Tukey’s test adjusted for False discovery rate (FDR), we found that subgroup-High had a significantly greater proportion of Firmicutes, Clostridia and Clostridiales than both the control group ( $P_{\text{Bonferroni}} = 0.008$ ) and the subgroup-Low ( $P_{\text{Bonferroni}} = 0.002$ ) (Fig. 3d). In contrast, for Bacteroidetes, we discriminated among subgroups only at the phylum level ( $P_{\text{ANOVA-FDR}} = 0.0378$ ). Particularly, subgroup-Low had significantly larger proportion of Bacteroidetes compared to the control ( $P_{\text{Bonferroni}} = 0.0428$ ) and subgroup-High ( $P_{\text{Bonferroni}} = 0.0025$ ) (Fig. 3d).

In our report, no difference was found in terms of Th17 and Treg between HIV-infected patients and control individuals. A clustergram in Supplementary Fig. S5 shows that both Th17 and Treg have no correlation with OTUs that separates control group and HIV-infected group. We employed the MetagenomeSeq tool<sup>44</sup> to assess whether the bacteria in the previously identified signature could discriminate between subgroups. This tool operates acceptably for datasets having a sparse and compositional nature<sup>45–47</sup> (Fig. 4). When comparing the HIV-infected group as a whole with the control group, four of the five genera from the signature were statistically significant: Lachnospiraceae-OTU-69 ( $P_{\text{adjusted}} = 0.00017$ ), *Blautia* ( $P_{\text{adjusted}} = 0.00023$ ), *Dorea* ( $P_{\text{adjusted}} = 0.003$ ) and Erysipelotrichaceae-OTU109 ( $P_{\text{adjusted}} = 0.047$ ) (Fig. 4a). Supplementary Table S8 shows the complete list of those OTUs with  $P_{\text{adjusted}} < 0.05$ . Comparison between subgroup-High and the control group revealed 32 genera with  $P_{\text{adjusted}} < 0.010$ . All genera belonging to this signature were statistically significant: Lachnospiraceae-OTU69 ( $P_{\text{adjusted}} = 2.95\text{E-}12$ ), *Blautia* ( $P_{\text{adjusted}} = 6.31\text{E-}10$ ), *Dorea* ( $P_{\text{adjusted}} = 4.27\text{E-}8$ ), *Roseburia* ( $P_{\text{adjusted}} = 2.83\text{E-}6$ ) and Erysipelotrichaceae-OTU109 ( $P_{\text{adjusted}} = 0.00033$ ) (Fig. 4b). Supplementary Table S9 shows the top-half of those OTUs with  $P_{\text{adjusted}} < 0.05$ .

## Discussion

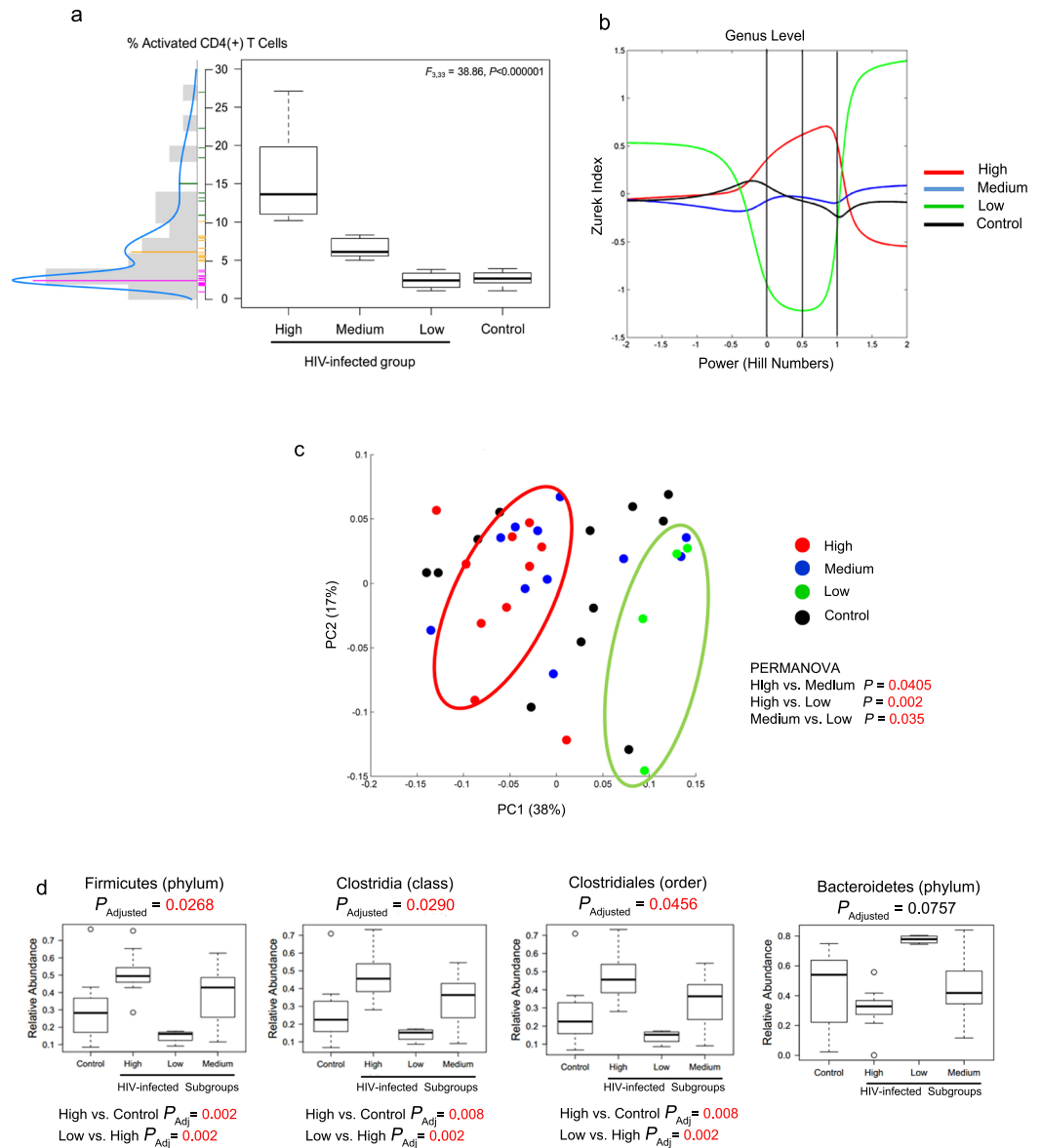
Here, we conducted an observational, cross-sectional study aimed to determine how the fecal bacteria of HIV-infected individuals differs from that of non-infected individuals in a Colombian setting. To the best of our knowledge, this is the first time that a study of this type has been conducted in a South American country. Our results indicate that the microbial fecal community of HIV-infected patients established associations among themselves which allowed us to probe for network building. Such relationships can respond in a dynamic way to environmental factors such as variations in diet and/or host health<sup>32,48,49</sup>.

Since sequencing data is of a compositional nature, we applied Aitchison’s centered log ratio-transformation (clr) to avoid spurious associations arising from such compositional structure and consequently be able to find true associations and correlations<sup>33</sup>. We found that the network structure of HIV-infected individuals highlights a dynamic in which the *Blautia*, *Dorea*, *Yersinia*, *Escherichia-Shigella*, *Staphylococcus* and *Bacteroides* genera all appear as nodes.

By applying the sPLS-DA and random forest statistical methods on a matrix of clr-transformed microbiota data<sup>33</sup>, we found a genus-level signature of OTUs discriminating between HIV- and control individuals. This signature consists of the genera from both the Lachnospiraceae family (Lachnospiraceae-OTU69, *Blautia*, *Dorea*, *Roseburia*) and the Erysipelotrichaceae family (Erysipelotrichaceae-OTU109).

In terms of  $\alpha$ -diversity, the individuals who had the lowest percentage of activated T cells had a heterogeneous distribution dominated by very few bacterial species. In contrast, the HIV-infected individuals belonging to subgroup-High had a homogeneous distribution comprised of a more diverse bacterial population.  $\alpha$ -diversity indexes reported by other studies have covered the entire spectrum of possibilities<sup>23,26–28,50,51</sup>, which might be due to the heterogeneity of the course of infection or to the immunological status of the individuals.

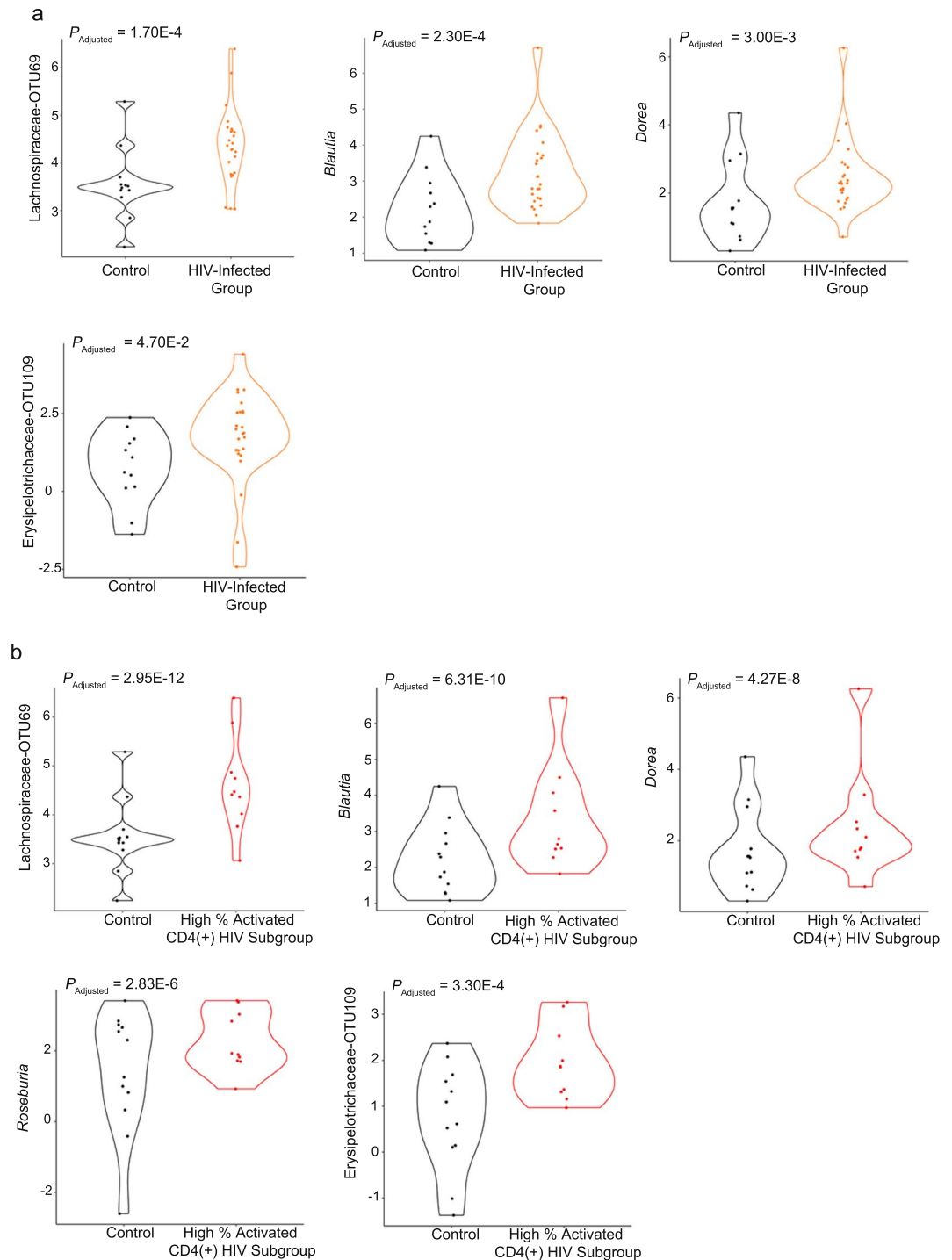
Overall, the bacterial signature we observed in HIV-infected patients can be summarized as an increase in the relative abundance of the genera Lachnospiraceae-OTU69, *Blautia*, *Dorea*, and Erysipelotrichaceae-OTU109. Genera within the Erysipelotrichaceae family have consistently been reported as part of the bacterial signature



**Figure 3.** Identification of OTU signatures associated with CD4(+) T cell activation status. **(a)** Subgroups of HIV-infected individuals according to their percentage of CD4(+) T cell activation. ANOVA-based analysis was performed to calculate statistical difference among groups. **(b)** Diversity indices, based on the intrinsic relationship of each one with respect to the transformation of Hill numbers, are plotted continuously on the x-axis. The Chao, Shannon and Simpson indices are shown when the value of the Hill numbers are -2, 0.5 and 2, respectively. The homogeneity with respect to the distribution of OTU abundance is an inverse function of the Simpson index. The richness of the OTUs is a direct function of the Shannon index. **(c)** PCA plot of  $\beta$ -diversity using the weighted UNIFRAC metric. PERMANOVA was used to estimate P-value. **(d)** Relative abundance of Firmicutes and Bacteroidetes in HIV-infected subgroups. Statistical difference among groups was estimated using ANOVA and  $P_{Adjusted}$  was calculated using false discovery rate and Bonferroni. Statistical difference between subgroups was estimated using Tukey least significance difference test with 95% confidence intervals (CIs) and P-value was adjusted by multiple comparisons.

associated with the microbiota dysbiosis in HIV-infected individuals<sup>23–25,27,52</sup>. Erysipelotrichaceae positively correlated with TNF- $\alpha$  and chronic intestinal inflammation in simian immunodeficiency virus-infected animals, suggesting its considerable potential to cause host perturbations<sup>53</sup>. Although most of the studies found Lachnospiraceae to be associated to non-infected individuals<sup>24,28,50,54,55</sup>, that was not our case. Although *Blautia-Dorea* pair has a beneficial effect for the intestinal mucosa<sup>56,57</sup>, their presence has been positively correlated with intestinal permeability in individuals with alcohol dependence or who suffered multiple sclerosis relapse episodes<sup>58–60</sup>, and is one of the colitogenic genera favoring Crohn's disease development<sup>61</sup>.

In contrast with other studies<sup>23,25–28,50,52</sup>, we did not find an increased abundance in the *Prevotella* genus or from genera belonging to the Enterobacteriaceae or Enterococcaceae families. However, enteropathogenic



**Figure 4.** MetagenomeSeq analysis. OTUs that were both identified as part of the signature by sPLS-DA/random forest and had a statistically significant  $P_{adj}$  value ( $P_{adj} < 0.05$ ) are shown. For better visualization, CLR values were used for plotting the abundance of each OTU. The  $P_{adj}$  values are shown in each respective panel. **(a)** Comparison of the control group with the HIV-infected group. **(b)** Comparison of the control group with the HIV-infected subgroup with high percentage of activated CD4(+) T cells.

bacteria may appear as the course of infection progresses as is shown by *Yersinia* and the *Escherichia-Shigella* appearing in HIV-infected group network structure. Overall, such dissimilarities might be due to differences in gut microbiota steady-state prior to infection-induced destabilization, disease status, the environmental factors, and environmental interaction with the unique genetic makeup of this hispanic population.

HIV-infected individuals showed a reduced capacity to produce IgA and IgG against new antigens derived from incoming Firmicutes and Proteobacteria<sup>58</sup>. This observation seemed to be compatible with our findings. Moreover, the levels of Firmicutes seemed to increase as individuals showed higher percentages of activated



CD4(+) T cells. As such, it would be interesting to investigate whether such expansion of Firmicutes communities could be explained by a lack of IgA and IgG updating in the face of a dynamic microbiota<sup>58</sup>.

Some genera of Lachnospiraceae family may produce SCFAs<sup>62</sup>. Under healthy conditions, SCFAs range from 70 to 140 mM in the proximal colon and from 20 to 70 mM in the distal colon<sup>57</sup>. No study has actually determined whether HIV-infected individuals microbiota has a diminished capacity to produce SCFAs. In our case, *Blautia* is capable of producing acetate from the end products of the glucose metabolism of *Dorea*<sup>63</sup>. Moreover, Erysipelotrichaceae may convert acetyl-CoA to butyryl-CoA via crotonyl-CoA. This pathway, in turn, leads to the production of butyrate<sup>64,65</sup>.

Intestinal microbiota trophic architecture would preserve its ability to produce SCFAs in HIV-infected group. For example, *Prevotella* has a high capacity to degrade fiber, whose products include succinate and acetyl-CoA<sup>66,67</sup>; while, bacteria of the genus *Phascolarctobacterium* can use succinate as a substrate to produce propionate in the intestine<sup>68</sup>. Interestingly, Ling *et al.* found an increase in both *Phascolarctobacterium* and *Prevotella* in the fecal samples of HIV-infected individuals from China<sup>27</sup>. Ling *et al.*<sup>27</sup> also reported an increase in *Faecalibacterium* and *Butyricicoccus*, which may produce butyric acid<sup>69,70</sup>.

SCFAs may not be so beneficial after all in HIV-infected individuals. Due to histone deacetylases inhibition, SCFAs can activate the transcription and production of viral particles from HIV-1 episomes<sup>71</sup>, reactivate the expression of the provirus integrated in the genome<sup>72,73</sup>, and improve the efficiency of post-entry viral events<sup>74</sup>. Furthermore, even the beneficial effect of inducing Foxp3(+) regulatory T cells<sup>75</sup> could be detrimental in the context of HIV infection, as this could decrease the response against infected cells, and thereby contribute to the persistence of the virus<sup>73-79</sup>.

Our study has the typical limitations of both an observational study with a cross-sectional design and case-control analyses. First of all, the cross-sectional design of this study only allows us to establish associations without being able to know the exact causal relationship between microbiota signature and immune deregulation. Likewise, the cross-sectional design does not allow us to define the gut microbiota evolutionary dynamics. Although longitudinal studies are desirable, ethical issues place a heavy restriction on conducting such design. The patient is recommended to initiate treatment as soon as a diagnosis is made. However, several studies have reported that the core ecosystem is stable in intervals of 1 year from the first sampling<sup>80-83</sup>. Taking the time it took for patients to be placed under treatment and the stability of the gut microbiota into consideration, a longitudinal study design conducted in a short time interval would not change our major findings. Secondly, fecal microbiota is only a proxy for the microbial community associated with the intestinal mucosa. As such, it is implicit to conduct further studies probing the intestinal mucosa adhered microbial community. Third, we did not directly assess microbial translocation. Fourth, this study did not analyze the changes associated with highly active antiretroviral therapy (HAART). Although it may be argued that a paired post-ART analysis would help in assessing HIV impact on gut microbiota clearly, the major problem is that some of the antiretroviral drugs have an impact on gut microbiota by themselves as each patient is usually subjected to a different treatment scheme based on HIV strain mutational profile that is infecting<sup>84-86</sup>. Since no individuals belonging to a control group is taking ART, it will be very difficult to assign microbiota modifications to viral load, immune changes, or the ART drug itself. Fifth, sample size might be a limitation in our study. Nonetheless, our findings are important to share with the scientific community as they highlight how HIV infection has a differential impact on the intestinal microbiota of a South American community, compared to that reported in studies conducted in developed Western societies.

Certain limitations and pitfalls of massively parallel 16S rRNA gene sequencing need to keep in mind. In contrast to 16S analyses, Whole Metagenome Sequencing (WMS) enables the analysis of the microbial phylogenetic composition and functional diversity<sup>87</sup>. WMS may detect microheterogeneity and genetic intraspecies variations, while bypassing the introduction of additional biases during the PCR amplification steps required in 16S rRNA gene analysis<sup>87,88</sup>. In addition, the high similarity of 16S rRNA amplicon sequences impairs the taxonomic resolution<sup>89</sup>.

Altogether, we have detected the presence of a bacterial signature that can differentiate HIV-infected individuals from non-infected individuals. We call into attention that a community previously overlooked shows a different profile in terms of the impact of HIV infection on gut microbiota. This signature is characterized by an increase in bacteria belonging to the Lachnospiraceae-OTU69, *Blautia*, *Dorea* and Erysipelotrichaceae-OTU109 genera. The Lachnospiraceae-based signature, in the light of previous reports, suggest the need for experimental studies probing intestinal microbiota mechanisms on the context of HIV infection.

## Methods

**Inclusion and exclusion criteria.** This study was approved by the ethics committee of the Universidad del Norte. We recruited 25 patients infected with HIV and 12 control individuals. All individuals provided informed consent to participate in the study. HIV-infected patients and control individuals were recruited using the framework and surveys associated to Human Microbiome Project<sup>90,91</sup>. We surveyed both HIV-infected patients and control individuals for demographic data covering socio-economic status, for systems reviews and life-styles (alcohol use and smoking), for drug and antibiotic use, and for diet composition in their meals. In addition, a research team-associated physician conducted a physical examination in each individual. The methods were carried out in accordance with the relevant guidelines and regulations. No patient was undergoing antiretroviral therapy, presented comorbidity with infectious agents such as the hepatitis B and C virus, was experiencing diarrhea at the time of fecal sampling, or had taken antibiotics within the previous 30 days.

We initially recruited all the HIV-infected patients and then control individuals were subsequently recruited based on the combined demographic information of the HIV-infected patients. Demographic information (i.e., age, male/female ratio, socio-economic status, body mass index, and diet composition) was used for matching purposes.

**Stool sample collection.** Each individual was asked for a stool sample. Individuals were given a sterile stool-collection recipient to deposit the sample without any contamination from urine. Stool samples while were transported from the collection site to the Universidad del Norte were kept at 4 °C, where they were aliquoted and stored at –150 °C.

**Blood sample collection.** For each individual (patients and controls), 30 mL of whole blood was taken, and 20 mL added to a tube containing EDTA and the remaining 10 mL in an empty tube without anticoagulant. Part of the blood that was mixed with EDTA was used to determine the activation state of the CD4 and CD8 cells, while the remainder was centrifuged at 2,500 rpm for 10 min to separate the plasma from the cells. The cell fraction was cultured to determine the percentage of Th17 versus T regulatory lymphocytes.

**Absolute CD4 and CD8 cell counts, CD4/CD8 ratio and percent activation of T cells.** Blood was incubated with anti-CD3, -CD4, -CD38, and -HLA-DR antibodies for 10 min using the kit Multitest CD4 FITC/CD38PE/CD3PerCP/anti-HLA-DR-APC (BD Biosciences). Anti-CD8-APC-Cy7 antibody was added to the test (BD Biosciences). BD FACS lysis solution was used to eliminate the red blood cells and to fix the lymphocytes. Absolute CD4 and CD8 cell quantification was carried out using Trucount tubes (BD Biosciences). Cells were processed through BD FACSCANTO II flow cytometer and the BD FACS DIVA software (BD Biosciences). CD4/CD8 ratio was calculated from the respective absolute counts. The percentage of CD3(+)/CD4(+) and CD3(+)/CD8(+) T cells simultaneously expressing HLA-DR(+) and CD38(+) was used to identify activated Th and Tc lymphocytes.

**Lymphocyte cell culture and percentage of Th17 versus regulatory T cells.** Peripheral blood mononuclear cells (PBMC) was recovered following Ficoll-Hypaque separation approach. Cells at a concentration of  $1 \times 10^6$ /mL were cultured in 100-mm dishes for 5 h in RPMI medium supplemented with 10% fetal bovine serum, 10 U/mL penicillin-streptomycin, L-glutamine, 20 mM HEPES. T cells were stimulated with PMA and Ionomycin. Golgi Stop was used to accumulate IL17 inside cells. Cells were pelleted, PFA-fixed, and saponin-permeabilized. Cells were labelled with anti-IL-17, anti-FoxP3, anti-CD3, and anti-CD4 antibodies (BD Biosciences Human Th17/Treg Phenotyping kit). Th17 and Treg (Foxp3+) cells percentages were determined using the BD FACSCANTO II flow cytometer (BD Bioscience) and the BD FACS Diva software (BD Biosciences).

**DNA extraction, 16S rRNA gene sequencing and bioinformatics analysis.** 16S rRNA gene compositional analysis provides a summary of the composition and structure of the bacterial component of the microbiome. Genomic bacterial DNA extraction methods were optimized to maximize the yield of bacterial DNA while keeping background amplification to a minimum. 16S rRNA gene sequencing methods were adapted from the methods developed for the NIH-Human Microbiome Project<sup>90,91</sup>. Briefly, bacterial genomic DNA was extracted using MO BIO PowerSoil DNA Isolation Kit (MO BIO Laboratories). The 16S rDNA V4 region was amplified by PCR and sequenced in the MiSeq platform (Illumina) using the  $2 \times 250$  bp paired-end protocol yielding pair-end reads that overlap almost completely. The primers used for amplification contain adapters for MiSeq sequencing and dual-index barcodes so that the PCR products may be pooled and sequenced directly<sup>92</sup>, targeting at least 10,000 reads per sample.

Our standard pipeline for processing and analyzing the 16S rRNA gene data incorporated phylogenetic and alignment-based approaches to maximize data resolution. The read pairs were demultiplexed based on the unique molecular barcodes, and reads were merged using USEARCH v7.0.1001<sup>93</sup> allowing zero mismatches and a minimum overlap of 50 bases. Merged reads were trimmed at first base with Q5. In addition, a quality filter was applied to the resulting merged reads and reads containing above 0.05 expected errors were discarded.

We used an in-house pipeline for 16S analysis as well as several other tools, including custom analytic packages developed at the CMMR to provide summary statistics and quality control measurements for each sequencing run, as well as multi-run reports and data-merging capabilities for validating built-in controls and characterizing microbial communities across large numbers of samples or sample groups.

16S rRNA gene sequences were assigned into Operational Taxonomic Units (OTUs) or phylotypes at a similarity cutoff value of 97% using the UPARSE algorithm. OTUs are then mapped to an optimized version (v.111) of the SILVA Database<sup>94,95</sup> containing only the 16S v4 region to determine taxonomies. Abundances were recovered by mapping the demultiplexed reads to the UPARSE OTUs. A custom script constructs an OTU table from the output files generated in the previous two steps, which is then used to calculate  $\alpha$ -diversity,  $\beta$ -diversity<sup>96</sup>, and provide taxonomic summaries that were leveraged for all subsequent analyses discussed below.

**Data analysis.** In order to construct DOC curves, the dissimilarity and overlapping measure must first be defined. Here, we used the measures proposed by Bashan *et al.*<sup>32</sup>. The measures were then plotted as points in a scatterplot. Subsequently, the DOC area was calculated from LOWESS method. The cutoff point,  $O_c$ , from which the slope of the curve is always negative is then calculated (i.e.,  $\frac{\partial D}{\partial O} < 0 | O > O_c$ ). The fraction  $f_{ns}$  is calculated as follows:

$$f_{ns} = \frac{\text{number of sample pairs with } O > O_c}{\text{total number of sample pairs}}$$

To calculate the p-values related to the DOC plot slope we carry out the procedure described in the paper published by Bashan *et al.*<sup>32</sup>. We retain only data points showing an overlap value greater than the median for both the control and disease groups. A bootstrap procedure repeating the aforementioned steps is performed and the slopes of the DOC plots are calculated at every bootstrap realization. The P-values are finally calculated as the fraction of bootstrap realizations resulting with non-negative slopes. The number of bootstrap runs was named n\_boot.

The compositional nature of the data restricts the data analysis to a simplex space. As such, we applied the Aitchison's centered log ratio transformation (CLR) to carry the data to a Euclidean space. In this way, each OTU is properly compared between the different samples<sup>34,35</sup>. The `logratio.transfo` function of the `mixOmics` package was used for CLR transformation<sup>97,98</sup>. A multiplicative Bayesian replacement strategy using the `cmultRepl` function of the `zCompositions` package allowed us to replace zeros<sup>99</sup>.

We performed a Jennrich test<sup>38</sup> to calculate the p-value associated to the correlation structures between HIV-infected patients control group.

SPiEC-EASI<sup>39</sup> was employed using the strategy of Meinshausen-Buhlmann for graph estimation of the network, which is a modified precision matrix built from  $\beta$  coefficients calculated from the average nearest neighbor. To reduce the number of false positive during the construction of networks, we eliminated from the analysis any OTU that was not present in at least 50% of the samples<sup>100</sup>. sPLS-DA and Random Forest were techniques used for feature selection. sPLS-DA used an approach that asks and identifies which features (OTUs) separate the HIV-infected group from control group based on a discriminant analysis of the partial least square metric<sup>40</sup>. Random Forest<sup>101</sup>, on the other hand, is a machine learning technique that also identifies features (OTUs) based on a construction of collection of decision trees with controlled variance<sup>102</sup>. Variables capable of differentiating between the groups were verified by Sparse PLS discriminant analysis (sPLS-DA)<sup>40</sup>. The most important bacteria were grouped in the first component. The Random Forest technique was also applied in order to determine the importance of bacteria in marking differences between the different patient groups. The performance of bacteria was measured using clr-normalized data and taking into account the mean decrease Gini<sup>78,79</sup>. To analyze the differences in microbial communities between the distinct patient groups, we used metagenomeSeq. For this, data was normalized by cumulative sum scaling using the `cumNorm` function, and analysis was performed with the zero-inflated log-normal fit `ZigModel` function<sup>44</sup>.

**Data availability.** The sequence data from this study are deposited in the Bioproject database (BioProject ID is PRJNA408085).

## References

- Cohen, M. S., Shaw, G. M., McMichael, A. J. & Haynes, B. F. Acute HIV-1 Infection. *N. Engl. J. Med.* **364**, 1943–1954 (2011).
- Mahy, M. *et al.* Producing HIV estimates: from global advocacy to country planning and impact measurement. *Glob. Health Action* **10**, 1291169 (2017).
- Hemelaar, J., Gouws, E., Ghys, P. D. & Osmanov, S., WHO-UNAIDS. Network for HIV Isolation and Characterisation. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS Lond. Engl.* **25**, 679–689 (2011).
- Villarreal, J.-L. *et al.* Characterization of HIV type 1 envelope sequence among viral isolates circulating in the northern region of Colombia, South America. *AIDS Res. Hum. Retroviruses* **28**, 1779–1783 (2012).
- Peterson, L. W. & Artis, D. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat. Rev. Immunol.* **14**, 141–153 (2014).
- Kato, L. M., Kawamoto, S., Maruya, M. & Fagarasan, S. Gut TFH and IgA: key players for regulation of bacterial communities and immune homeostasis. *Immunol. Cell Biol.* **92**, 49–56 (2014).
- Wells, J. M. *et al.* Homeostasis of the gut barrier and potential biomarkers. *Am. J. Physiol. Gastrointest. Liver Physiol.* **312**, G171–G193 (2017).
- Kumar, P. *et al.* Intestinal Interleukin-17 Receptor Signaling Mediates Reciprocal Control of the Gut Microbiota and Autoimmune Inflammation. *Immunity* **44**, 659–671 (2016).
- Bauché, D. & Marie, J. C. Transforming growth factor  $\beta$ : a master regulator of the gut microbiota and immune cell interactions. *Clin. Transl. Immunol.* **6**, e136 (2017).
- Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141 (2014).
- Perreau, M. *et al.* Follicular helper T cells serve as the major CD4 T cell compartment for HIV-1 infection, replication, and production. *J. Exp. Med.* **210**, 143–156 (2013).
- Sun, H. *et al.* Th1/17 Polarization of CD4 T Cells Supports HIV-1 Persistence during Antiretroviral Therapy. *J. Virol.* **89**, 11284–11293 (2015).
- Brenchley, J. M. & Douek, D. C. HIV infection and the gastrointestinal immune system. *Mucosal Immunol.* **1**, 23–30 (2008).
- Cassol, E. *et al.* Impaired CD4+ T-cell restoration in the small versus large intestine of HIV-1-positive South Africans receiving combination antiretroviral therapy. *J. Infect. Dis.* **208**, 1113–1122 (2013).
- Christensen-Quick, A. *et al.* Human Th17 Cells Lack HIV-Inhibitory RNases and Are Highly Permissive to Productive HIV Infection. *J. Virol.* **90**, 7833–7847 (2016).
- Kim, C. J. *et al.* Mucosal Th17 cell function is altered during HIV infection and is an independent predictor of systemic immune activation. *J. Immunol. Baltim. Md 1950* **191**, 2164–2173 (2013).
- Mudd, J. C. & Brenchley, J. M. Gut Mucosal Barrier Dysfunction, Microbial Dysbiosis, and Their Role in HIV-1 Disease Progression. *J. Infect. Dis.* **214**(Suppl 2), S58–66 (2016).
- Sankaran, S. *et al.* Rapid onset of intestinal epithelial barrier dysfunction in primary human immunodeficiency virus infection is driven by an imbalance between immune response and mucosal repair and regeneration. *J. Virol.* **82**, 538–545 (2008).
- Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
- Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).
- Shankar, V. *et al.* Differences in Gut Metabolites and Microbial Composition and Functions between Egyptian and U.S. Children Are Consistent with Their Diets. *mSystems* **2** (2017).
- Org, E. *et al.* Genetic and environmental control of host-gut microbiota interactions. *Genome Res.* **25**, 1558–1569 (2015).
- Lozupone, C. A. *et al.* Alterations in the Gut Microbiota Associated with HIV-1 Infection. *Cell Host Microbe* **14**, 329–339 (2013).
- Mutlu, E. A. *et al.* A Compositional Look at the Human Gastrointestinal Microbiome and Immune Activation Parameters in HIV Infected Subjects. *PLoS Pathog.* **10**, e1003829 (2014).
- Dinh, D. M. *et al.* Intestinal microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. *J. Infect. Dis.* **211**, 19–27 (2015).
- Dubourg, G. *et al.* Gut microbiota associated with HIV infection is significantly enriched in bacteria tolerant to oxygen. *BMJ Open Gastroenterol.* **3**, e000080 (2016).
- Ling, Z. *et al.* Alterations in the Fecal Microbiota of Patients with HIV-1 Infection: An Observational Study in A Chinese Population. *Sci. Rep.* **6**, 30673 (2016).
- Sun, Y. *et al.* Fecal bacterial microbiome diversity in chronic HIV-infected patients in China. *Emerg. Microbes Infect.* **5**, e31 (2016).

29. Monaco, C. L. *et al.* Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host Microbe* **19**, 311–322 (2016).
30. La Rosa, P. S. *et al.* Hypothesis testing and power calculations for taxonomic-based human microbiome data. *PLoS One* **7**, e25078 (2012).
31. Kelly, B. J. *et al.* Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinforma. Oxf. Engl.* **31**, 2461–2468 (2015).
32. Bashan, A. *et al.* Universality of human microbial dynamics. *Nature* **534**, 259–262 (2016).
33. Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V. & Egozcue, J. J. It's all relative: analyzing microbiome data as compositions. *Ann. Epidemiol.* **26**, 322–329 (2016).
34. Smith, P. F., Renner, R. M. & Haslett, S. J. Compositional data in neuroscience: If you've got it, log it! *J. Neurosci. Methods* **271**, 154–159 (2016).
35. Tsilimigras, M. C. B. & Fodor, A. A. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Ann. Epidemiol.* **26**, 330–335 (2016).
36. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**, 15 (2014).
37. Pesenson, M. Z., Suram, S. K. & Gregoire, J. M. Statistical analysis and interpolation of compositional data in materials science. *ACS Comb. Sci.* **17**, 130–136 (2015).
38. Jennrich, R. I. An Asymptotic  $\chi^2$  Test for the Equality of Two Correlation Matrices. *J. Am. Stat. Assoc.* **65**, 904–912 (1970).
39. Kurtz, Z. D. *et al.* Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **11**, e1004226 (2015).
40. Lê Cao, K.-A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**, 253 (2011).
41. Lee, S. C. *et al.* Helminth colonization is associated with increased diversity of the gut microbiota. *PLoS Negl. Trop. Dis.* **8**, e2880 (2014).
42. Menze, B. H. *et al.* A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* **10**, 213 (2009).
43. Joshi, A. *et al.* HIV-1 Env Glycoprotein Phenotype along with Immune Activation Determines CD4 T Cell Loss in HIV Patients. *J. Immunol. Baltim. Md 1950* **196**, 1768–1779 (2016).
44. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **10**, 1200–1202 (2013).
45. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
46. Lê Cao, K.-A. *et al.* MixMC: A Multivariate Statistical Framework to Gain Insight into Microbial Communities. *PLoS One* **11**, e0160169 (2016).
47. Thorsen, J. *et al.* Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. *Microbiome* **4**, 62 (2016).
48. Basson, A., Trotter, A., Rodriguez-Palacios, A. & Cominelli, F. Mucosal Interactions between Genetics, Diet, and Microbiome in Inflammatory Bowel Disease. *Front. Immunol.* **7**, 290 (2016).
49. Adair, K. L. & Douglas, A. E. Making a microbiome: the many determinants of host-associated microbial community composition. *Curr. Opin. Microbiol.* **35**, 23–29 (2017).
50. Nowak, P. *et al.* Gut microbiota diversity predicts immune status in HIV-1 infection. *AIDS Lond. Engl.* **29**, 2409–2418 (2015).
51. Noguera-Julian, M. *et al.* Gut Microbiota Linked to Sexual Preference and HIV Infection. *EBioMedicine* **5**, 135–146 (2016).
52. Vujkovic-Cvijin, I. *et al.* Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. *Sci. Transl. Med.* **5**, 193ra91 (2013).
53. Handley, S. A. *et al.* Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell* **151**, 253–266 (2012).
54. McHardy, I. H. *et al.* HIV Infection is associated with compositional and functional shifts in the rectal mucosal microbiota. *Microbiome* **1**, 26 (2013).
55. Vázquez-Castellanos, J. F. *et al.* Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. *Mucosal Immunol.* <https://doi.org/10.1038/mi.2014.107> (2014).
56. Jenq, R. R. *et al.* Intestinal Blautia Is Associated with Reduced Death from Graft-versus-Host Disease. *Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant.* **21**, 1373–1383 (2015).
57. Wong, J. M. W., de Souza, R., Kendall, C. W. C., Emam, A. & Jenkins, D. J. A. Colonic health: fermentation and short chain fatty acids. *J. Clin. Gastroenterol.* **40**, 235–243 (2006).
58. Leclercq, S. *et al.* Intestinal permeability, gut-bacterial dysbiosis, and behavioral markers of alcohol-dependence severity. *Proc. Natl. Acad. Sci. USA* **111**, E4485–4493 (2014).
59. Chen, J. *et al.* Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. *Sci. Rep.* **6**, 28484 (2016).
60. Buscarinu, M. C. *et al.* Altered intestinal permeability in patients with relapsing-remitting multiple sclerosis: A pilot study. *Mult. Scler. Houndmills Basingstoke Engl.* **23**, 442–446 (2017).
61. Palm, N. W. *et al.* Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000–1010 (2014).
62. Dillon, S. M. *et al.* Low abundance of colonic butyrate-producing bacteria in HIV infection is associated with microbial translocation and immune activation. *AIDS Lond. Engl.* **31**, 511–521 (2017).
63. Rajilić-Stojanović, M. & de Vos, W. M. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol. Rev.* **38**, 996–1047 (2014).
64. Vital, M., Howe, A. C. & Tiedje, J. M. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio* **5**, e00889 (2014).
65. Davids, M. *et al.* Functional Profiling of Unfamiliar Microbial Communities Using a Validated De Novo Assembly Metatranscriptome Pipeline. *PLoS One* **11**, e0146423 (2016).
66. Kovatcheva-Datchary, P. *et al.* Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of *Prevotella*. *Cell Metab.* **22**, 971–982 (2015).
67. De Vadder, F. *et al.* Microbiota-Produced Succinate Improves Glucose Homeostasis via Intestinal Gluconeogenesis. *Cell Metab.* **24**, 151–157 (2016).
68. Engels, C., Ruscheweyh, H.-J., Beerenwinkel, N., Lacroix, C. & Schwab, C. The Common Gut Microbe *Eubacterium hallii* also Contributes to Intestinal Propionate Formation. *Front. Microbiol.* **7**, 713 (2016).
69. Jeraldo, P. *et al.* Capturing One of the Human Gut Microbiome's Most Wanted: Reconstructing the Genome of a Novel Butyrate-Producing, Clostridial Scavenger from Metagenomic Sequence Data. *Front. Microbiol.* **7**, 783 (2016).
70. Antharam, V. C. *et al.* Intestinal dysbiosis and depletion of butyrogenic bacteria in *Clostridium difficile* infection and nosocomial diarrhea. *J. Clin. Microbiol.* **51**, 2884–2892 (2013).
71. Kantor, B., Ma, H., Webster-Cyriaque, J., Monahan, P. E. & Kafri, T. Epigenetic activation of unintegrated HIV-1 genomes by gut-associated short chain fatty acids and its implications for HIV infection. *Proc. Natl. Acad. Sci. USA* **106**, 18786–18791 (2009).
72. Ye, F. & Karn, J. Bacterial Short Chain Fatty Acids Push All The Buttons Needed To Reactivate Latent Viruses. *Stem Cell Epigenetics* **2**, (2015).
73. Das, B. *et al.* Short chain fatty acids potentially induce latent HIV-1 in T-cells by activating P-TEFb and multiple histone modifications. *Virology* **474**, 65–81 (2015).

74. Lucera, M. B. *et al.* The histone deacetylase inhibitor vorinostat (SAHA) increases the susceptibility of uninfected CD4+ T cells to HIV by increasing the kinetics and efficiency of postentry viral events. *J. Virol.* **88**, 10803–10812 (2014).
75. Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* **341**, 569–573 (2013).
76. Chevalier, M. F. & Weiss, L. The split personality of regulatory T cells in HIV infection. *Blood* **121**, 29–37 (2013).
77. Rueda, C. M., Vellilla, P. A., Chougnat, C. A. & Rugeles, M. T. Incomplete normalization of regulatory t-cell frequency in the gut mucosa of Colombian HIV-infected patients receiving long-term antiretroviral treatment. *PLoS One* **8**, e71062 (2013).
78. Elahi, S. *et al.* Protective HIV-specific CD8+ T cells evade Treg cell suppression. *Nat. Med.* **17**, 989–995 (2011).
79. He, T. *et al.* Cutting Edge: T Regulatory Cell Depletion Reactivates Latent Simian Immunodeficiency Virus (SIV) in Controller Macaques While Boosting SIV-Specific T Lymphocytes. *J. Immunol. Baltim. Md 1950* **197**, 4535–4539 (2016).
80. Rajlić-Stojanović, M., Heilig, H. G. H. J., Tims, S., Zoetendal, E. G. & de Vos, W. M. Long-term monitoring of the human intestinal microbiota composition. *Environ. Microbiol.* <https://doi.org/10.1111/1462-2920.12023> (2012).
81. Faith, J. J. *et al.* The long-term stability of the human gut microbiota. *Science* **341**, 1237439 (2013).
82. Martínez, I., Müller, C. E. & Walter, J. Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. *PLoS One* **8**, e69621 (2013).
83. Yin, Y. *et al.* Investigation into the stability and culturability of Chinese enterotypes. *Sci. Rep.* **7**, 7947 (2017).
84. Pinto-Cardoso, S., Klatt, N. R. & Reyes-Terán, G. Impact of antiretroviral drugs on the microbiome: unknown answers to important questions. *Curr. Opin. HIV AIDS* **13**, 53–60 (2018).
85. Pinto-Cardoso, S. *et al.* Fecal Bacterial Communities in treated HIV infected individuals on two antiretroviral regimens. *Sci. Rep.* **7**, 43741 (2017).
86. Villanueva-Millán, M. J., Pérez-Matute, P., Recio-Fernández, E., Lezana Rosales, J. M. & Oteo, J. A. Differential effects of antiretrovirals on microbial translocation and gut microbiota composition of HIV-infected patients. *J. Int. AIDS Soc.* **20**, 21526 (2017).
87. Hiergeist, A., Gläsner, J., Reischl, U. & Gessner, A. Analyses of Intestinal Microbiota: Culture versus Sequencing. *ILAR J.* **56**, 228–240 (2015).
88. Loong, S. K., Khor, C. S., Jafar, F. L. & AbuBakar, S. Utility of 16S rDNA Sequencing for Identification of Rare Pathogenic Bacteria. *J. Clin. Lab. Anal.* **30**, 1056–1060 (2016).
89. Shin, J. *et al.* Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Sci. Rep.* **6**, 29681 (2016).
90. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
91. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
92. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
93. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinforma. Oxf. Engl.* **26**, 2460–2461 (2010).
94. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
95. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–596 (2013).
96. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
97. Liqueur, B., Lê Cao, K.-A., Hocini, H. & Thiébaud, R. A novel approach for biomarker selection and the integration of repeated measures experiments from two assays. *BMC Bioinformatics* **13**, 325 (2012).
98. Rohart, F., Eslami, A., Matigian, N., Bougeard, S. & Lê Cao, K.-A. MINT: a multivariate integrative method to identify reproducible molecular signatures across independent experiments and platforms. *BMC Bioinformatics* **18**, 128 (2017).
99. Martín-Fernández, J.-A., Hron, K., Templ, M., Filzmoser, P. & Palarea-Albaladejo, J. Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat. Model. Int. J.* **15**, 134–158 (2015).
100. Weiss, S. *et al.* Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* **10**, 1669–1681 (2016).
101. Knights, D., Costello, E. K. & Knight, R. Supervised classification of human microbiota. *FEMS Microbiol. Rev.* **35**, 343–359 (2011).
102. Breiman, L. Machine Learning. *45*, 5–32 (2001).

## Acknowledgements

We are indebted to HIV-infected patients who contributed to the present research. In addition, we honor the non-infected blood donors who participated in our study. We also want to acknowledge the financial support from COLCIENCIAS grant number 121556934635 and Contract number 0770-2013. COLCIENCIAS by no means is responsible for what is reported in the present article.

## Author Contributions

Study design: H.S.V., N.A., C.M., G.C.A. Data analysis, statistics and modelling: E.Z., M.P., I.P., N.A., J.I.V., S.S.S. Writing of the manuscript: H.S.V., N.A., D.V., O.V.O. MiSeq sequencing and O.T.U. generation: N.A., J.F.P. Patient recruitment and information processing: C.M., D.H., I.U., P.D.R., E.G.G., N.T. Experimental work: C.C.C., Y.D.O., L.H.G.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-22629-7>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018