ARTICLE   OPEN

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection

Yolanda Guillén[1,2], Marc Noguera-Julian[1,2,3], Javier Rivera[1,3], Maria Casadellà[1,2], Alexander S. Zevin[4], Muntsa Rocafort[1,2], Mariona Parera[1], Cristina Rodríguez[1], Marçal Arumí[1], Jorge Carrillo[1,2], Beatriz Mothe[1,3,5], Carla Estany[5], Josep Coll[1,5], Isabel Bravo[5], Cristina Herrero[5], Jorge Saz[6], Guillem Sirera[5], Ariadna Torrella[7], Jordi Navarro[2,7], Manuel Crespo[8], Eugènia Negredo[2,3,5], Christian Brander[1,2,3,9], Julià Blanco[1,2,3], Maria Luz Calle[3], Nichole R. Klatt[4], Bonaventura Clotet[1,2,3,5] and Roger Paredes[1,2,3,5]

Human immunodeficiency virus (HIV)-1 infection causes severe gut and systemic immune damage, but its effects on the gut microbiome remain unclear. Previous shotgun metagenomic studies in HIV-negative subjects linked low-microbial gene counts (LGC) to gut dysbiosis in diseases featuring intestinal inflammation. Using a similar approach in 156 subjects with different HIV-1 phenotypes, we found a strong, independent, dose–effect association between nadir CD4+ T-cell counts and LGC. As in other diseases involving intestinal inflammation, the gut microbiomes of subjects with LGC were enriched in gram-negative Bacteroides, acetogenic bacteria and Proteobacteria, which are able to metabolize reactive oxygen and nitrogen species; and were depleted in oxygen-sensitive methanogenic archaea and sulfate-reducing bacteria. Interestingly, subjects with LGC also showed increased butyrate levels in direct fecal measurements, consistent with enrichment in *Roseburia intestinalis* despite reductions in other butyrate producers. The microbiomes of subjects with LGC were also enriched in bacterial virulence factors, as well as in genes associated with beta-lactam, lincosamide, tetracycline, and macrolide resistance. Thus, low nadir CD4+ T-cell counts, rather than HIV-1 serostatus per se, predict the presence of gut dysbiosis in HIV-1 infected subjects. Such dysbiosis does not display obvious HIV-specific features; instead, it shares many similarities with other diseases featuring gut inflammation.

INTRODUCTION

In settings with adequate access to antiretroviral therapy (ART), the main clinical problems of people living with human immunodeficiency virus (HIV)-1 (PLWH) derive either from late HIV-1 diagnosis[1,2] or from premature aging.[3,4] Up to one-third of all HIV-1-related deaths in Western countries can be attributed to late HIV-1 diagnosis.[5] Low nadir CD4+ T-cell counts are a major risk factor for developing AIDS- and non-AIDS-defining illnesses,[6–8] bacterial infections,[9] ART failure,[5] and suboptimal immune reconstitution with ART.[10,11] Premature aging involves the precocious development of type 2 diabetes, dyslipidemia, cardiovascular diseases, osteoporosis, and frailty syndrome.[4] In PLWH, such illnesses have been linked to chronic inflammation, immune activation, and endotoxemia.[4,12–14] In non-HIV-infected subjects, they have been related to shifts in the gut microbiome associated with reduced microbial gene richness.[15–20]

The gut microbiome is believed to play a central role in HIV-1 immunopathogenesis as well as in HIV-associated chronic complications. HIV-1 infection damages the intestinal mucosal barrier and gut-associated lymphoid tissues[14,21] enabling translocation of bacterial products to regional lymph nodes and systemic circulation.[22,23] Infection also induces inflammation of the mesenteric fat and promotes extensive deregulation of local immune responses.[24] In all, this contributes to chronic inflammation and immune activation, which are only partially restored by ART.[25] Immune activation hinders immune reconstitution,[26,27] diminishes the efficacy of HIV-1 preventive and curative strategies,[21] fosters immune senescence[28–30], and at the mucosal level, is a major risk factor for HIV-1 transmission.[21]

However, the effects of HIV-1 infection on the gut microbiome are not well understood and, to date, a consistent pattern of gut dysbiosis has not been identified. Unlike the vaginal microbiome, where compositional shifts from homeostasis are often recognizable as dysbiotic community types,[31,32] dysbiosis in the gut is more difficult to detect and even to define. As a matter of fact, the debate on what characterizes a healthy gut microbiome is still ongoing[33], and over 20% of the interperson microbiome variability is associated with factors related to diet, drugs, and anthropometric measurements,[34] which may confound associations between gut microbiome shifts and diseases.

[1]irsiCaixa AIDS Research Institute, Ctra de Canyet s/n, Badalona 08916 Catalonia, Spain; [2]Universitat Autònoma de Barcelona, Bellaterra 08193 Catalonia, Spain; [3]Universitat de Vic-Universitat Central de Catalunya, C. Sagrada Família 7, Vic 08500 Catalonia, Spain; [4]University of Washington, 3018 Western Avenue, Seattle, WA 98121, USA; [5]Infectious Diseases Service & Lluita contra la SIDA Foundation, Hospital Universitari Germans Trias i Pujol, Ctra de Canyet s/n, Badalona 08916 Catalonia, Spain; [6]BCN Checkpoint, Carrer del Comte Borrell, 164, Barcelona 08015 Catalonia, Spain; [7]Infectious Diseases Unit, Hospital Universitari Vall d'Hebrón, Passeig de la Vall d'Hebrón, 119-129, Barcelona 08035 Catalonia, Spain; [8]Infectious Diseases Unit, Internal Medicine Department, Complexo Hospitalario Universitario, Vigo. IIS Galicia Sur, Estrada de Clara Campoamor, 341, Vigo 36312 Pontevedra, Spain and [9]Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, Barcelona 08010 Catalonia, Spain
Correspondence: Roger Paredes (rparedes@irsicaixa.es)
These authors contributed equally: Yolanda Guillén, Marc Noguera-Julian.

Received: 17 May 2018 Revised: 3 August 2018 Accepted: 8 August 2018
Published online: 31 August 2018

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

233

Table 1. Study cohort characteristics

		All subjects (n = 156)	Low-gene counts (n = 103)	High-gene counts (n = 53)	p Value
Age, years, median (IQR)		43 (35; 51)	46 (36; 53)	38 (35; 46)	0.024
Gender, n (%)	Men	124 (79.5%)	77 (74.8%)	47 (88.7%)	0.058
	Women	31 (19.9%)	26 (25.2%)	5 (9.4%)	0.020
	Transgender women	1 (0.6%)	0	1 (1.9%)	0.340
Ethnicity, n (%)	Asiatic	1 (0.6%)	0	1 (1.9%)	0.340
	Caucasian	124 (79.5%)	88 (85.4%)	36 (67.9%)	0.013
	Hispanic-latin	28 (17.9%)	12 (11.7%)	16 (30.2%)	0.007
	Other	3 (1.9%)	3 (2.9%)	0	0.551
HIV-1 risk group, n (%)	MSM	100 (64.1%)	54 (52.4%)	46 (86.8%)	<0.001
	Non-MSM	56 (35.9%)	49 (47.6%)	7 (13.2%)	
HIV-1 status, n (%)	HIV-1-negative	27 (17.3%)	11 (10.7%)	16 (30.2%)	0.004
	HIV-1-positive	129 (82.7%)	92 (89.3%)	37 (69.8%)	
HIV-1 phenotype, n (%)	Late presenter	11 (7.1%)	9 (8.7%)	2 (3.8%)	0.335
	Discordant	18 (11.5%)	15 (14.6%)	3 (5.7%)	0.118
	Concordant	53 (34.0%)	42 (40.8%)	11 (20.8%)	0.277
	Early-treated	13 (8.3%)	8 (7.8%)	5 (9.4%)	0.277
	ART-naive	15 (9.6%)	8 (7.8%)	7 (13.2%)	0.783
	Viremic controller	11 (7.1%)	5 (4.9%)	6 (11.3%)	0.033
	Elite controller	8 (5.1%)	5 (4.9%)	3 (5.7%)	0.421
	HIV-1-negative	27 (17.3%)	11 (10.7%)	16 (30.2%)	0.004
BMI, kg/m^2, median (IQR)		23.8 (22.0; 26.1)	23.7 (21.8; 25.5)	24.4 (22.3; 26.3)	0.215
History of allergy, n (%)	Yes	30 (19.2%)	21 (20.4%)	9 (17.0%)	0.673
	No	122 (78.2%)	78 (75.7%)	44 (83.2%)	0.413
	NA	4 (2.6%)	4 (3.9%)	0	0.300
Antibiotic intake, previous 3 months, n (%)	Yes	2 (1.3%)	2 (1.9%)	0	0.549
	No	154 (98.7%)	101 (98.1%)	53 (100.0%)	
Antibiotic intake, previous 6 months, n (%)	Yes	35 (22.4%)	26 (25.2%)	9 (17.0%)	0.312
	No	121 (77.6%)	77 (74.8%)	44 (83.0%)	
Fecal consistency, n (%)	Hard	56 (35.9%)	33 (32.0%)	23 (43.4%)	0.217
	Soft	91 (58.3%)	63 (61.2%)	28 (52.8%)	0.392
	Liquid	5 (3.2%)	4 (3.9%)	1 (1.9%)	0.662
	NA	4 (2.6%)	3 (2.9%)	1 (1.9%)	1.000
Abdominal transit alterations, n (%)	Yes	23 (14.7%)	14 (13.6%)	9 (17.0%)	0.636
	No	127 (81.4%)	85 (82.5%)	42 (79.2%)	0.666
	NA	6 (3.8%)	4 (3.9%)	2 (3.8%)	1.000
CD4+ T-cell counts, median (IQR)	Current, cells/mm^3	700.0 (462.0; 860.0)	635.5 (331.0; 851.2)	750.0 (558.0; 930.0)	0.126
	Current, %	33.0 (24.2; 40.2)	32.8 (23.6; 40.2)	33.8 (25.6; 40.3)	0.474
	Nadir, cells/ mm^3	337.0 (140.0; 528.5)	280.0 (113.0; 491.5)	443.0 (339.0; 601.0)	0.002
CD8+ T-cell counts, median (IQR)	Current, cells/ mm^3	776.5 (576.0; 1012.0)	791.5 (558.8; 991.2)	748.5 (604.2; 1158.0)	0.294
	Current, %	39.4 (33.0; 48.92)	38.2 (31.6; 49.7)	41.6 (34.9; 47.1)	0.607
CD4+/CD8+ ratio		0.9 (0.5; 1.2)	0.9 (0.5; 1.3)	0.8 (0.5;1.2)	0.943
HIV-1 RNA, copies/mL, median (IQR)		<40.0 (40.0; 711.5)	<40.0 (40.0; 79.5)	<40.0 (40.0; 1339.0)	0.074

MSM men-who-have-sex-with-men; *ART* antiretroviral therapy; *BMI* body mass index; *SD* standard deviation; *IQR* 25–75% interquartile range.

Initial HIV gut microbiome studies using 16S rRNA gene sequencing suggested a shift from *Bacteroides* to *Prevotella* dominance in HIV-1 infection.[22,35–40] More recently, our group was able to demonstrate that such changes were confounded by sexual preference—men-who-have-sex-with-men (MSM) were enriched in *Prevotella*—and, after controlling for HIV-1 risk group, there was no consistent microbial dysbiosis pattern discernible by 16S rRNA gene sequencing.[41] One implication of our findings is that we should no longer expect "community-type structures" or "enterotype"-like microbiome clustering by HIV-infected vs. uninfected individuals. Considering the large biomass, ecological complexity and functional redundancy of the gut microbiome, it is conceivable that altered microbiome states might be subtler, and possibly imply shifts in discrete groups of microbial species or changes in functional features which might not necessarily be evident in genus-level ordination analyses. Moreover, gut dysbiosis might not necessarily be present in all HIV-1-infected subjects, but instead, be more frequently found in subjects with
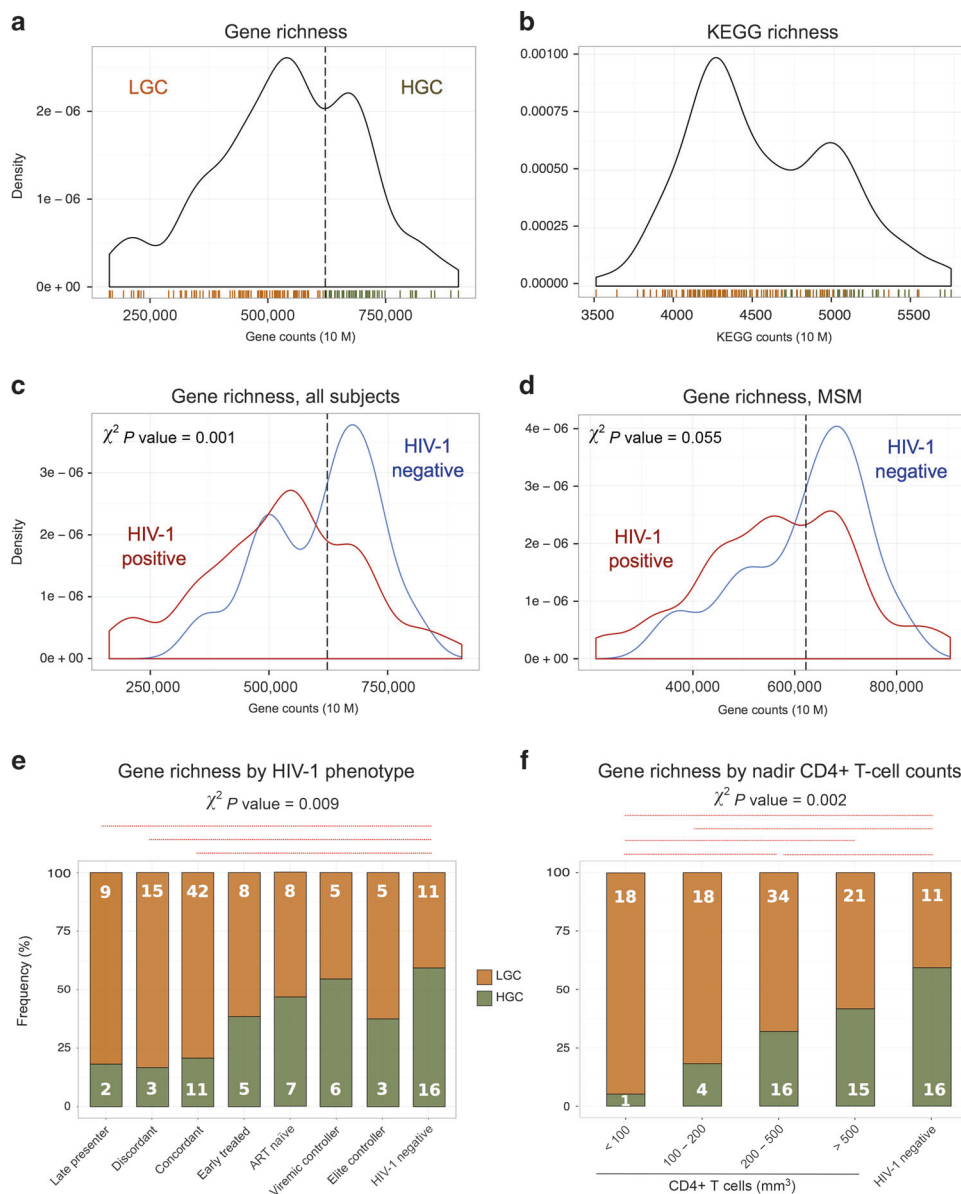
SPRINGER NATURE   SOCIETY FOR MUCOSAL IMMUNOLOGY

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

234

Fig. 1 Low-microbial gene richness is linked to HIV-1-mediated immune suppression. **a** Probability density function showing a bimodal distribution of the study population according to gut microbial gene richness, i.e., the number of different microbial genes observed in each individual at 10 million sequence downsampling size. The local minimum between the two modes (621,808 genes in this analysis) was used as a threshold to define people with high (HGCs) and low-gene counts (LGCs). **b** The bimodal distribution became even more aparent when KEGG enzyme richness, i.e., number of different enzymes found in the KEGG database was used. **c** HIV-1-infected subjects ($n = 156$) were more likely to be classified as LGCs; in comparison, most HGCs were HIV-1 negative. **d** The association between LGC and HIV-1 infection remained when only men-who-have-sex-with-men (MSM) were analyzed ($n = 100$). **e** HIV-1 phenotypes associated with immune deterioration were enriched in LGCs. **f** Subjects with lower nadir CD4+ counts were gradually more likely to be classified as LGCs. Late presenter, subjects with CD4+ T-cell counts <200 cells/mm³ at diagnosis and no ART exposure; discordant, with HIV-1 RNA <50 copies/mL and CD4+ T-cell counts <300 cells/mm³; concordant, with HIV-1 RNA levels <50 copies/mL and achieving CD4+ T-cell counts >500 cells/mm³; early treated, HIV-1-infected subjects who initiated ART during the first 6 months after the infection, achieving HIV-1 RNA levels <50 copies/mL during at least 3 months and with no HIV-1 RNA blips after achieving HIV-1 RNA <50 copies/mL; ART naïve, with HIV-1 RNA >10,000 copies/mL, nadir CD4+ T-cell counts >500 cells/mm³ and no ART exposure; viremic controller, with HIV-1 RNA between 50 and 2000 copies/mL during at least 2 years in the absence of ART; elite controller, subjects with HIV-1 RNA <50 copies/mL during at least 2 years in the absence of ART

higher degree of immune deficiency, longer duration of HIV-1 infection or late presentation, as previously suggested by using markers of immune activation.[41–43] Thus, it may be oversimplistic rendering HIV-1 vs. non-HIV-1 comparisons.

A highly consistent feature of the gut microbiomes of people with different gut inflammatory diseases in studies using shotgun metagenomics is a reduction in microbial gene richness,[15–20] which allows classifying individuals as having high-gene count

(HGC) or low-gene counts (LGC). Such studies have shown that subjects with LGC consistently display changes in their gut microbiomes that most authors would accept as "dysbiosis", including increases in Proteobacteria and other microbial species able to thrive in the presence of gut inflammation. In the present study, we used fecal whole metagenome shotgun sequencing and direct short-chain fatty acid (SCFA) measurements to characterize the nature and possible mechanisms driving gut microbiome

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

235

Table 2. Factors associated with low-microbial gene counts[a]

		Univariate			Multivariate		
		OR	95% CI	p Value	OR	95% CI	p Value
Age	Per each additional year	1.03	[1–1.07]	0.074	–	–	–
Gender	Female	1					
	Male	0.32	[0.10–0.82]	0.028	–	–	–
	Transgender woman	–	–	–			
Ethnic Group	Caucasian	1			1		
	Hispanic-Latin	0.27	[0.11–0.64]	0.003	0.26	[0.10–0.67]	0.006
	Asiatic and others[b]	–	–	–	–	–	–
HIV-1 risk group	Non-MSM	1			1		
	MSM	0.17	[0.06–0.39]	<0.001	0.20	[0.07–0.51]	0.002
HIV-1 status	Negative	1					
	Positive	3.68	[1.57–8.89]	0.003	–	–	–
Nadir CD4+ T-cell count, cells/mm^3	HIV-1 negative	1			1		
	>500	2.04	[0.75–5.74]	0.169	2.13	[0.73–6.45]	0.173
	200–500	3.09	[1.19–8.37]	0.006	2.92	[1.03–8.62]	0.047
	100–200	6.55	[1.86–27.71]	0.023	5.55	[1.40–26.15]	0.020
	<100	26.18	[4.40–506.6]	0.003	14.00	[2.02–288.71]	0.023

[a]Full dataset analysis, $n = 156$ subjects
[b]Analysis does not apply because all subjects are included in the same response group
MSM men-who-have-sex-with-men

shifts in HIV-1 infection, identify clinical factors associated with them, and explore their potential consequences on microbial pathogenesis and human health.

RESULTS

Study subjects
The study included 156 individuals: 129 (82.7%) HIV-1-infected with different immune phenotypes and 27 (17.3%) not infected with HIV-1 (Table 1). Participants were mostly MSM (64%) of Caucasian ethnicity (79%), with median 43 years of age, and median body mass index (BMI) of 23.8 kg/m^2. HIV-1-infected subjects were recruited from two tertiary HIV-1 clinics in Barcelona, Spain. Most HIV-1-negative controls were enrolled from a prospective cohort of HIV-negative MSM who attend quarterly medical and counseling visits at a community-based center in Barcelona.[44,45]

Microbial gene richness
Microbial gene richness values obtained by whole fecal metagenome shotgun sequencing followed a bimodal distribution across the study cohort (Fig. 1a). Using 10 million (M)-sequence downsampling, the local intermodal minimum was 621,808 genes. This threshold was used to classify participants as having HGC ($n = 53$) or LGC ($n = 103$), as in ref. 18. Subjects in the HGC group also had higher enzyme and metabolic pathway richness than LGCs (Fig. 1b, Supplementary Table 1). Gene richness measurements were not biased by sequencing yield (Supplementary Fig. 1) or gene catalog used (Supplementary Fig. 2).

Factors associated with microbial gene richness
HIV-1-infected subjects were more likely to be LGCs than HIV-1-negative controls, regardless of whether all subjects (71.3% vs. 40.7%, $p = 0.001$, Fig. 1c) or only MSM (59.7% vs. 34.8%, $p = 0.055$, Fig. 1d) were considered.[41] Subjects with LGCs were more frequently observed in individuals with more advanced HIV-1 disease stages ($p = 0.009$, Fig. 1e) and in gradually lower nadir CD4+ T-cell count ranks ($p = 0.002$, Fig. 1f). In addition, LGC subjects

were more likely to be older ($p = 0.024$), Caucasian ($p = 0.013$), non-MSM ($p < 0.001$) and women ($p = 0.020$) than HGC individuals (Table 1). Average microbial gene richness was also significantly lower in HIV-1-infected subjects, particularly in those with more advanced disease, immune discordant response, and lower nadir CD4+ T-cell counts (Supplementary Figs. 3 and 4, $p < 0.05$ in all cases). Neither current CD4+ T-cell counts nor CD4% were associated with microbial richness after stratifying for nadir CD4 + T-cell counts (Supplementary Fig. 5). Similarly, we did not observe an effect of ART or previous antibiotic exposure on gene richness after stratifying for nadir CD4+ T-cell counts (Supplementary Figs. 6 and 7).

Dose-dependent association between LGC and nadir CD4+ T-cell counts
In multivariate regression (Table 2) there was a dose–effect relationship between nadir CD4+ T-cell counts and microbial gene richness. Subjects with nadir CD4+ T-cell counts between 200 and 500 cells/mm^3 (OR = 2.9 [95 CI: 1.1–8.6], $p = 0.047$), 100 and 200 cells/mm^3 (OR = 5.5 [95 CI: 1.4–26.1], $p = 0.020$) and <100 cells/mm^3 (OR = 14.0 [95 CI: 2.02–288.7], $p = 0.023$) were more likely to be LGC than HIV-negative controls. Additional factors independently associated with LGC were HIV-1 risk group (OR for MSM vs. non-MSM = 0.2 [95 CI: 0.1–0.5], $p = 0.002$) or ethnicity (OR for Hispanic vs. Caucasian = 0.3 [95 CI: 0.1–0.7], $p = 0.006$). As expected, time between the study baseline and dates of HIV-1 diagnosis and nadir CD4+ T-cell counts were also longer in subjects with lower nadir CD4+ T-cell counts (Supplementary Fig. 8).

Changes in microbial species composition
Fecal microbiomes clustered by gene richness category rather than by HIV-1 serostatus (Fig. 2a, Supplementary Fig. 9). Microbiomes from HGC subjects were more similar among themselves than those from LGC individuals, which displayed many different configurations.

The microbial species composition of HGC and LGC individuals was remarkably different (Fig. 2b, c, Supplementary Figs.
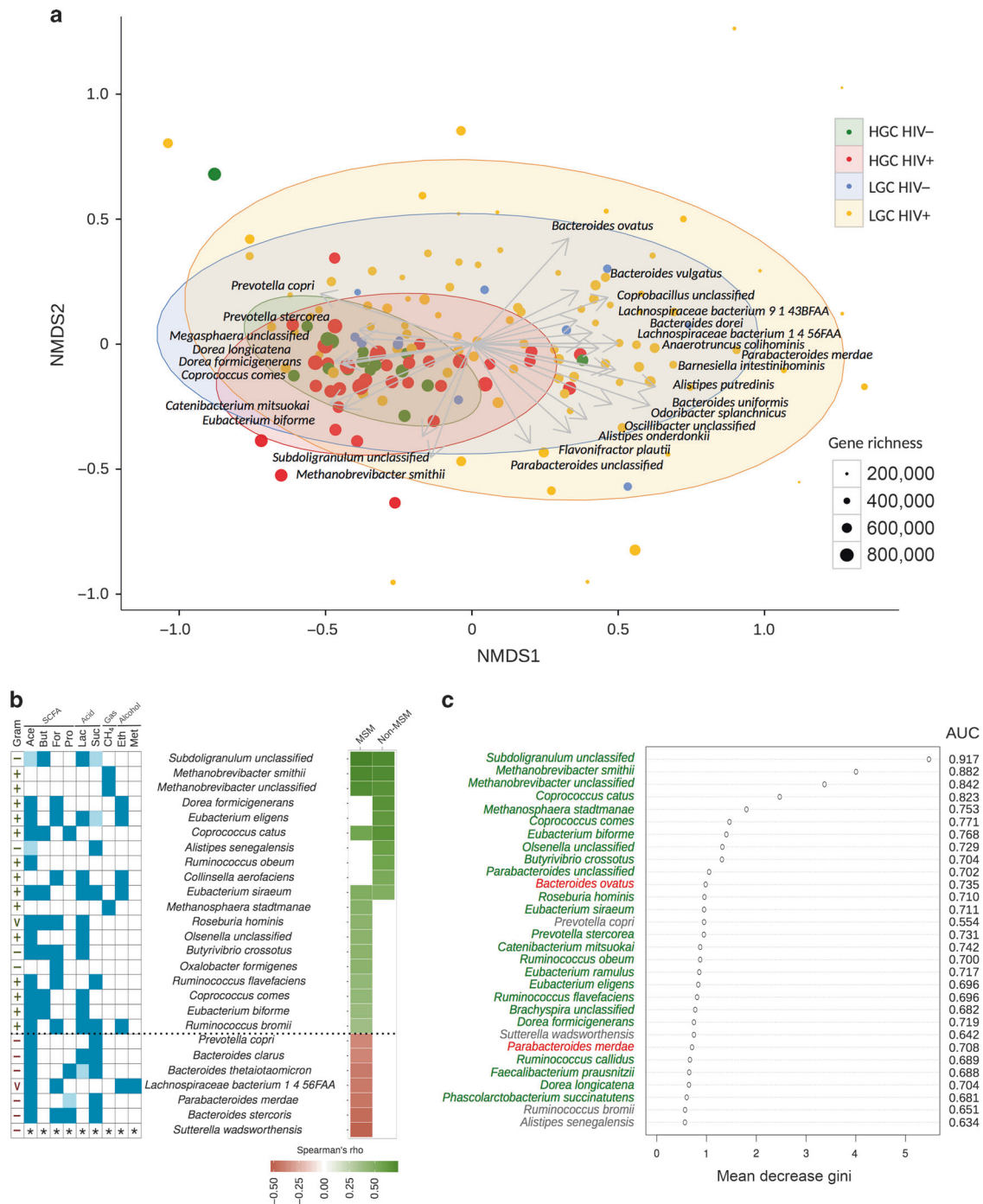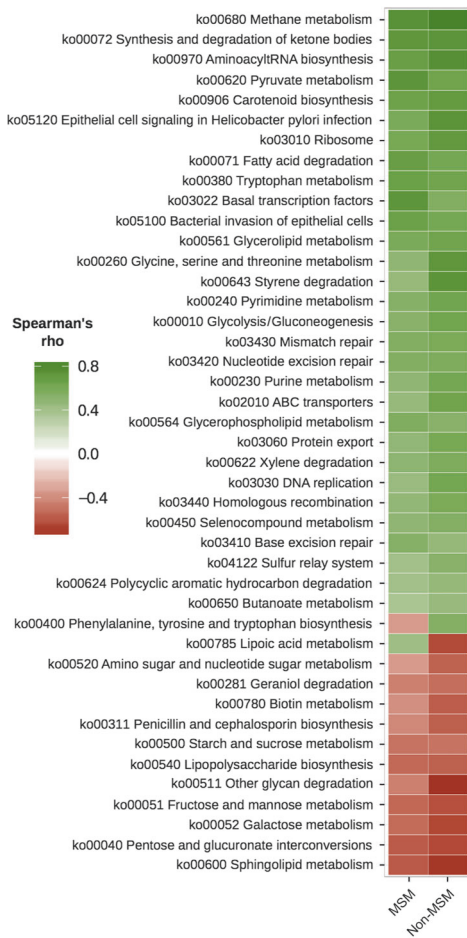
Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

236

Fig. 2 Microbial species associated with gut microbial gene richness. **a** Nonmetric multidimensional scaling plot based on Bray–Curtis microbiome distances showing more different microbiome configurations in subjects with LGCs than in HGCs. Dot sizes are proportional to microbiome gene richness. Ellipses represent 95% confidence intervals. Microbial species associated to an environmental vector of more than 0.35 NMDS coordinate length are shown. **b** Microbial species positively (green) and negatively (red) correlated with gene richness (Benjamini–Hochberg corrected Spearman Rank correlation p value < 0.05) in MSM ($n = 100$) and non-MSM ($n = 56$) subjects. Most microbial species positively correlated with richness were GRAM-positive, whereas most species negatively correlated with richness were GRAM-negative (V is for "variable" GRAM staining). Blue squares on the left handside summarize some of the main microbial metabolism products of each species, including the short-chain fatty acids (SCFAs) acetate (Ace), butyrate (But), formate (For), and propionate (Pro); the acids lactate (Lac) and succinate (Suc); the gas methane (CH4), and the alcohols ethanol (Eth) and methanol (Met); * is for unknown. **c** Random forest analysis showing the contribution of each microbial species to overall gene richness. In green, red, and gray, respectively, microbial species enriched in HGCs, LGCs and not significantly enriched in either category. AUC area under the curve
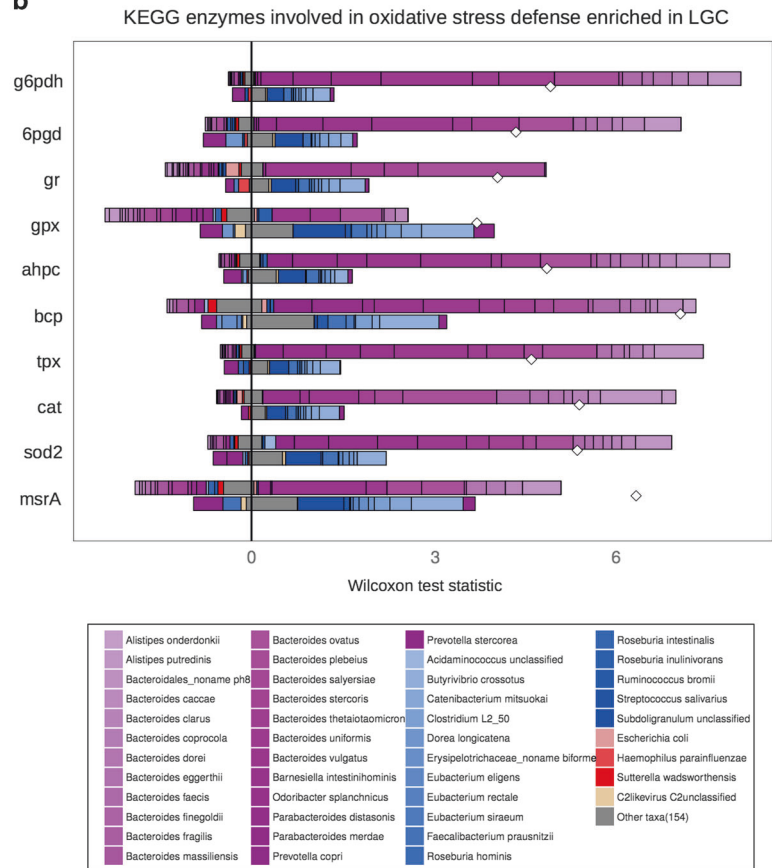
10–14). Gut microbiomes of HGC subjects were enriched in methanogenic archaea (*Methanobrevibacter* spp. *and Methanosphaera stadtmanae*), sulfate-reducing bacteria (*Desulfovibrio piger and Desulfovibrio desulfuricans*) (Supplementary Fig. 14),

and butyrate-producing species from the Ruminococcaceae (*Subdoligranulum* spp.) and Lachnospiraceae (*Butyrivibrio crossotus, Coprococcus comes, Conus catus*) families (Fig. 2). In addition, subjects with HGCs showed enrichment in *Eubacterium*
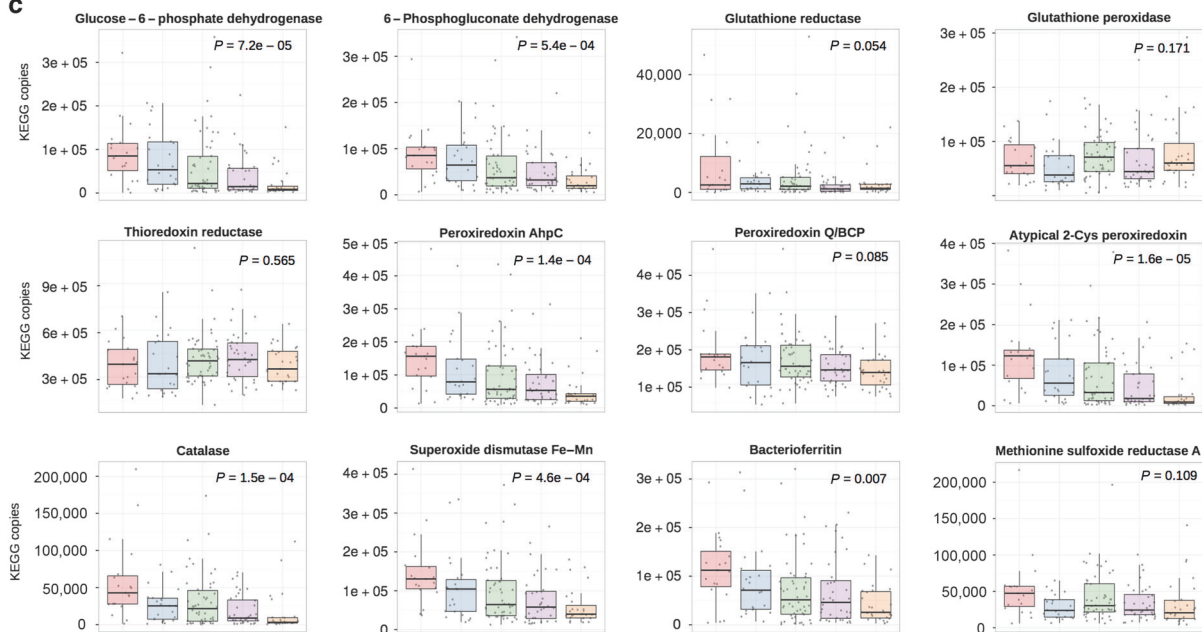
Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

237

a (heatmap of KEGG pathways with Spearman's rho scale)

b KEGG enzymes involved in oxidative stress defense enriched in LGC

c

Glucose – 6 – phosphate dehydrogenase *P* = 7.2e – 05
6 – Phosphogluconate dehydrogenase *P* = 5.4e – 04
Glutathione reductase *P* = 0.054
Glutathione peroxidase *P* = 0.171

Thioredoxin reductase *P* = 0.565
Peroxiredoxin AhpC *P* = 1.4e – 04
Peroxiredoxin Q/BCP *P* = 0.085
Atypical 2-Cys peroxiredoxin *P* = 1.6e – 05

Catalase *P* = 1.5e – 04
Superoxide dismutase Fe–Mn *P* = 4.6e – 04
Bacterioferritin *P* = 0.007
Methionine sulfoxide reductase A *P* = 0.109

siraeum, *Eubacterium eligens* and species from Dorea genera (*Dorea formicigenerans* and *Dorea longicatena*), as well as in cellulolytic bacteria from the Ruminococcus genus (*Ruminococcus flavefaciens*).

Conversely, subjects in the LGC group had microbiomes enriched in Bacteroides (*Bacteroides thetaiotaomicron*, *Bacteroides clarus*, *Bacteroides stercoris*), Parabacteroides (*Parabacteroides merdae*) and acetogenic bacterial species (*Blautia*

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

238

Fig. 3 Functional unbalance and microbial adaptation to oxidative stress. **a** Microbial functions significantly correlated with gene richness (Benjamini–Hochberg-corrected Spearman Rank correlation p value < 0.05) in both MSM and non-MSM subjects. **b** Differential enrichment in microbial enzymes involved in metabolism of reactive oxygen species (ROS) in LGCs (upper horizontal bars) vs. HGCs (lower horizontal bars), and contribution of microbial species to such enrichment. The plot shows significant differences in ROS gene enrichment between LGC and HGC (diamond, Wilcoxon test statistic) using a false discovery rate p value threshold of 0.05. Microbes are colored by phyla, i.e., Bacteroidetes in purple; Firmicutes in blue; Proteobacteria in red, Others in gray. **c** Abundance of microbial enzymes involved in ROS metabolism by nadir CD4+ T-cell count categories. G6PHD, glucose-6-phosphate dehydrogenase; 6pgd, 6-phosphogluconate dehydrogenase; gr, glutatione reductase; gpx, glutathione peroxidase; ahpc, peroxiredoxin AhpC; bcp, peroxiredoxin BCP; tpx, atypical 2-Cys peroxiredoxin; cat, catalase; sod2, superoxide dismutase Fe–Mn; msra, methionine sulfoxide reductase A

hydrogenotrophica, Blautia producta, Marvinbryantia formatexigens, and *Eubacterium limosum*) (Supplementary Fig. 14); as well as in Proteobacteria (Supplementary Fig. 13), including the Betaproteobacteria *Sutterella wadsworthensis*.

Interestingly, some of the best-known butyrate-producing microbial species[46,47] were present at significant amounts across the whole gene richness continuum. Whereas *Faecalibacterium prausnitzii* was more abundant with higher gene richness, *Roseburia intestinalis* and *Roseburia inulinivorans* were more prevalent with lower gene counts, and *Eubacterium rectale* was homogeneously distributed across gene richness.

Gene richness and inflammation
Individuals with lower gene richness had higher levels of systemic inflammation (IL-6, $p = 0.020$) (Supplementary Fig. 15). HIV-1 infection and particularly low nadir CD4 T-cell counts were also significantly associated with higher levels of systemic inflammation (IL-6 and IP-10), and enterocyte damage (intestinal fatty acid-binding protein (IFABP)) (Supplementary Fig. 16).

Diet, richness, and microbial composition
Microbial gene richness directly correlated with the dietary intake of monounsaturated fats, carotenoids, iron, fiber and Vitamin A (Supplementary fig. 17a). Dirichlet-multinomial regression identified a positive association between iron intake and levels of *Prevotella copri, E. eligens, Ruminococcus bromii* and *Ruminococcus callidus*, and a negative association with *Bacteroides ovatus*. Ethanol consumption was negatively associated with *Faecalibacterium prausnitzii* (Supplementary Fig. 17b). There was a positive correlation between fiber consumption and the overall abundance of methanogenic archaea, including species from both *Methanobrevibacter* and *Methanosphaera* genera (Supplementary Fig. 18). However, in a sensitivity analysis only iron intake remained associated with LGCs in a multivariate regression model (Supplementary Table 2).

Shifts in microbial functional pathways
Consistent with the stark differences in microbial species composition, key microbial functional pathways were also strongly correlated with gene richness, both in MSM and non-MSM (Fig. 3a). Functional differences between HGC and LGC (Supplementary material 2) indicated that microbiomes of both groups were able to use different energy resources and might provide different metabolic end products to their hosts.

Microbiomes from HGC subjects were enriched in genes involved in methanogenesis; biosynthesis of indole derivatives and carotenoids; and metabolism of pyruvate, fatty acids, glycerolipids, and glycerophospholipids, the latter needed for the biosynthesis of lipid components of the bacterial cell wall. They were also enriched in pathways related to microbial growth, replication and protein production, i.e., DNA replication, aminoacyl-tRNA biosynthesis, ribosome, purine, and pyrimidine metabolism, and mismatch, base and nucleotide excision repair.

Whereas HGCs were enriched in genes coding for enzymes implicated in glucose, starch, and cellulose utilization, LGC microbiomes were enriched in genes involved in the metabolism of fructose, and a wide range of disaccharides (sucrose, maltose, cellobiose, and lactose) and polysaccharides (pectin, galactan, and amylopectin). LGCs were also characterized by an enrichment of genes involved in biosynthesis of lipopolysaccharide (LPS), the main component of the outer membrane of Gram-negative bacteria. Tryptophan-degrading enzymes such as tryptophanase and aromatic-L-amino-acid decarboxylase, which catalyze the degradation of tryptophan to indole and tryptamine, respectively, were increased in LGCs.

### Contribution of microbial species to functional shifts
In an integrative taxonomic and functional comparative analysis to estimate taxon-level contributions to functional shifts (FishTaco),[48] LGCs were enriched in genes involved in reactive oxygen and nitrogen species (ROS/RNS) metabolism, encoded by different Bacteroides and Proteobacteria species (Fig. 3b, Supplementary Fig. 20). Moreover, ROS/RNS enzymes were significantly more abundant as microbial gene richness (Supplementary Figs. 19 and 21) and nadir CD4+ T-cell counts (Fig. 3c, Supplementary Fig. 22) decreased, in a dose–effect relationship.

Bacterial antioxidant enzymes enriched in LGCs included redox enzymes such as thioredoxin reductase and peroxiredoxins from the thioredoxin system; glutathione peroxidase and reductase; and glucose 6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase, the latter involved in the oxidative phase of the pentose phosphate pathway. Catalase and superoxide dismutase are ROS-scavenging enzymes that directly eliminate non-radical and radical ROS, respectively. RNS metabolism included enzymes involved in dissimilatory nitrate reduction, nitrite detoxification, regulation of nitrate/nitrite responses and nitrogen availability, molybdate transport and dimethyl sulfoxide (DMSO), trimethylamine N-oxide (TMAO), and tetrathionate respiration. RNS and their derivatives, as well as S- and N-oxides and tetrathionate, are used by Proteobacteria and other bacteria enriched in LGC as alternative electron acceptors in anaerobic respiration.[49]

LPS biosynthesis in LGC subjects was also mainly encoded by Bacteroidetes and Proteobacteria (Supplementary Fig. 23). A comprehensive list of all functional shifts identified by FishTaco is available (Supplementary material 3).

Differences in fecal SCFA content
We then measured the abundance of acetic, butyric, propionic, valeric, isobutyric, and isovaleric acids in stool samples. Overall, the molar ratio of acetic:butyric:propionic acids was approximately 3:1:1, consistent with previous studies.[50] Concentrations of butyric and propionic acids were significantly higher in LGCs, whereas isovaleric acid increased in HGCs (Fig. 4). The acetic/butyric ratio increased in HGC individuals, which could indicate that the contribution of acetate to butyrate formation is higher in LGCs. No significant SCFA variations were found between nadir CD4+ T-cell categories. Butyric acid levels were significantly correlated with abundance of *Roseburia intestinalis* and *Roseburia inulinivorans*, which are known to be able to generate large amounts of butyrate from acetate consumption.[47,51]

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

239

a

b

c

d
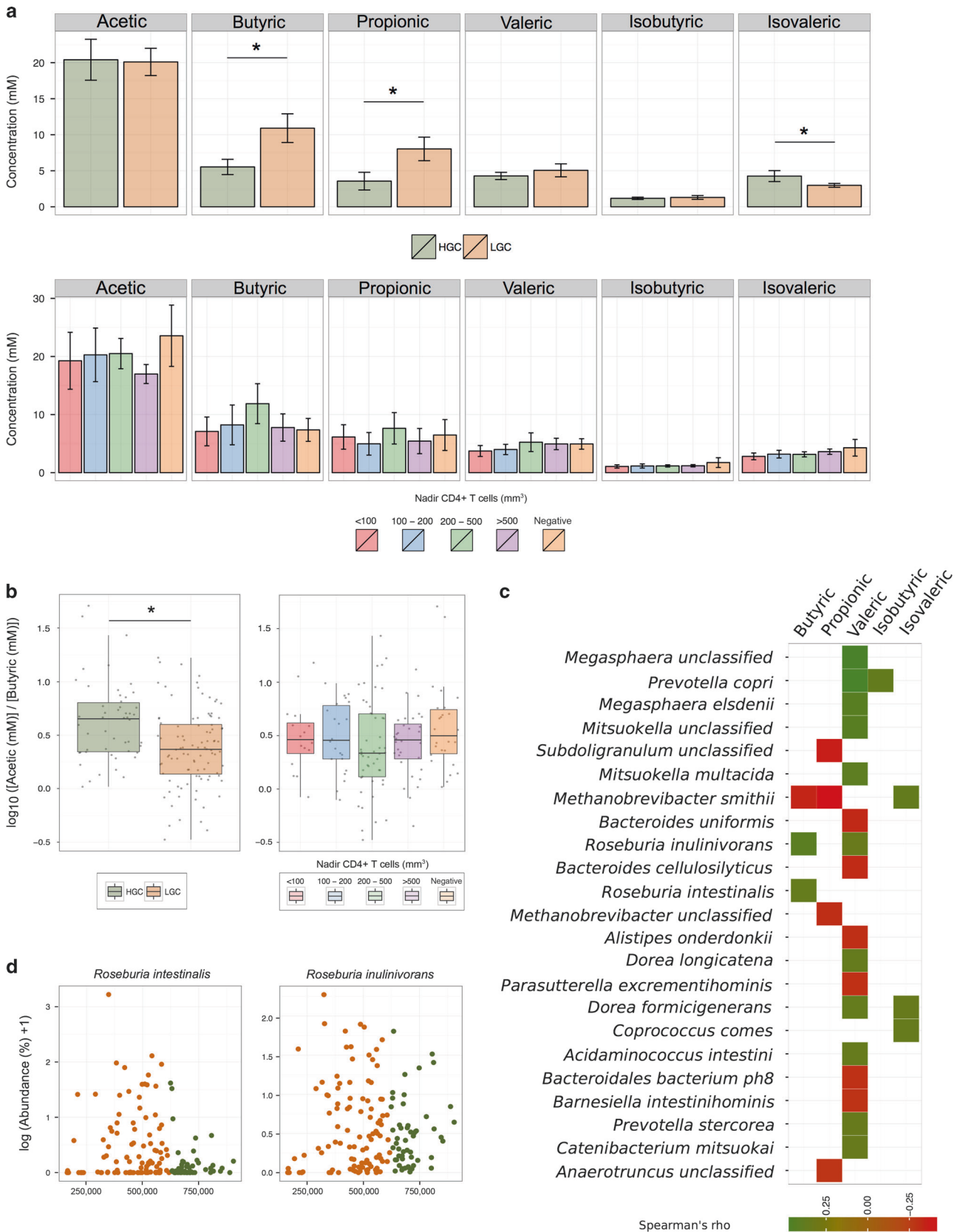
Bacterial virulence factors and antimicrobial resistance

Microbial shifts associated with immune deficiency also implied increases in bacterial virulence factors as well as changes in the

gut microbial resistome (Supplementary Figs. 24 and 25). By FishTaco,[48] subjects with LGCs were enriched in intracellular toxins (tccC), and toxins able to damage cell membranes (hlyIII, slo, rtxA1,

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

240

Fig. 4 Fecal short-chain fatty acid (SCFA) levels. **a** Fecal butyric and propionic acid levels were significantly higher in subjects with low-gene counts (LGC), whereas isovaleric acid levels were higher in subjects with high gene counts (HGC); no significant differences in SCFAs were observed between different nadir CD4+ T-cell count categories. **b** The acetyc/butyric ratio was significantly higher in HGCs than in LGCs, but was not significantly different by nadir CD4+ T-cell counts. **c** Butyrate levels were positively correlated with abundance of *Roseburia intestinalis* and *Roseburia inulinivorans*, two well-known butyrate producers. Only statistically significant correlations are shown ($p < 0.05$). **d** Abundance of *R. intestinalis* and *R. inulinivorans* across microbial gene richness values (HGC in green, LGC in orange). *($p < 0.05$)
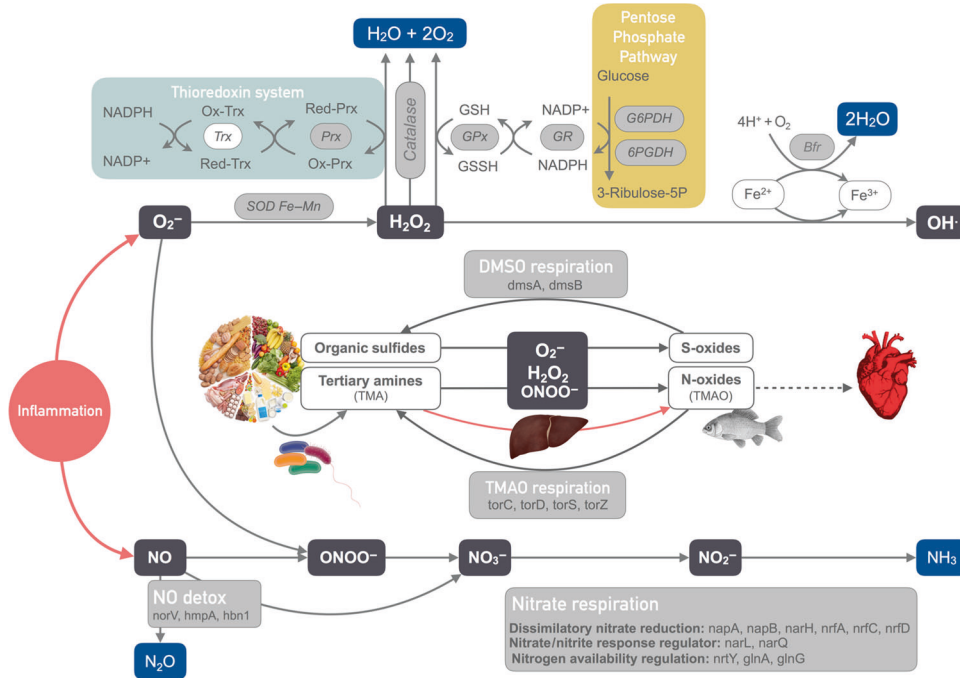
Fig. 5 Summary of enzymes and metabolic processes related with ROS/RNS clearance and enriched in subjects with LGCs (gray boxes). Bacteria enriched in LGC reduce ROS and RNS and generate less toxic metabolites that are eventually excreted from the body. Moreover, different Proteobacteria utilize S-oxides and N-oxides as electron acceptors for anaerobic DMSO and TMAO respiration. S-oxides derive from dietary organic sulfides, which are oxidized by ROS. The N-oxide trimethylamine N-oxide (TMAO) is generated by oxidation of diet-derived tertiary amines like trimethylamine (TMA), either by hepatic flavin-containing monooxigenases or by the direct effect of ROS. TMAO concentrations, however, are also high in fish, from which it can be directly absorbed. Plasma TMAO levels have been epidemiologically linked (dashed line) to severe cardiovascular disease and metabolic disorders, which are also prevalent in HIV-1-infected subjects. G6PDH glucose-6-phosphate dehydrogenase; 6PGDH 6-phosphogluconate dehydrogenase; GR glutathione reductase; GPx glutathione peroxidase; Trx thioredoxin reductase; Prx peroxiredoxin; SOD superoxide dismutase, Fe–Mn family; Bfr bacterioferritin; NO nitric oxide; TMA trimethylamine; TMAO trimethylamine oxide; DMSO dimethyl sulfoxide

tlyC, and plcC) and the extracellular matrix (cloSI, hya, and NEU1). Similarly, they were also enriched in a number of enzymes involved in antimicrobial resistance (Supplementary Fig. 24), including single antimicrobial resistance genes like Fsr, involved in biofilm formation,[52] and genes encoding for multidrug efflux pumps.[53] The latter comprises ATP-binding cassette (ABC) transporters (ABC-2 type permease), and genes from the four families of secondary multidrug transporters, i.e., major facilitator superfamily (YebQ, Blt, Mef, and TetA), small multidrug resistance (SMR: TC.SMR, SugE), multidrug and toxic compound extrusion (MATE: TC.MATE), and resistance-nodulation-cell division (EmrAB-TolC; AcrA-TolC, MdtA, and MacA). Gut dysbiosis was also associated with increases in genes encoding for other bacterial survival mechanisms like the multiple peptide resistance factor (MprF), which confers resistance to antimicrobial peptides,[54] and metal ion efflux systems like the CusCBA copper efflux complex.[55]

In a targeted exploration of virulence and resistance factors included, respectively, in the virulence factor database (VFDB) and the comprehensive antimicrobial resistance database (CARD) curated databases, subjects with nadir CD4+ T-cells <100 cells/mm³ had higher counts of important bacterial virulence genes, such as catalase, ClpC ATPase,[56] Type IV pili,[57] type III secretion system,[58] HSI-I,[59] and Flagella[60] (Supplementary Fig. 26). Similarly,

higher antimicrobial resistance gene copies were observed in subjects with LGCs, as well as in individuals with progressively lower nadir CD4+ T-cell counts (Supplementary Fig. 25). While subjects with HGCs were enriched in genes involved in self-antibiotic resistance, which control the overgrowth of other bacterial species[61] individuals with LGCs were enriched in genes associated with beta-lactam, lincosamide, tetracycline, and macrolide resistance.

DISCUSSION
Using shotgun metagenomics and direct measurements of fecal SCFA levels, we found evidence of a strong, dose-dependent association between reductions in gut microbial gene richness,[18,19] and nadir CD4+ T-cell counts, which reflected complex compositional and functional changes occurring in the gut microbiome alongside HIV-1-associated immune deficiency. The observed reductions in oxygen-sensitive, Gram-positive syntrophic gut microbes like methanogenic archaea, together with increases in ROS/RNS-resistant generalists, like Bacteroides, Parabacteroides, and Proteobacteria, suggests adaptation of the gut microbial ecosystem to oxidative stress as a putative, not necessarily exclusive, driver of such changes.

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

241

HIV-1 infection is a well-established source of oxidative stress. It triggers gut inflammation by damaging the intestinal epithelial lining and submucosal and lymphoid tissues, allowing enteric adenovirus expansion,[35,62] NF-kβ activation,[63] and release of cytokines and chemokines,[49] arachidonic acid[64] and neopterin,[65] which induce migration and accumulation of neutrophils and macrophages in gut tissues. Antimicrobial ROS and RNS are produced by the NADPH-oxidase in neutrophils and inducible NO-synthases in macrophages. Production of free radicals is further enhanced by the effect of HIV-1 proteins,[66–68] microbial products like LPS,[69,70] and HIV-1-associated mitochondrial dysfunction.[71,72] Increased residency times and reduced apoptosis of macrophages in the context of gut dysbiosis extend the production of ROS/RNS over time.[73] The persistent accumulation of free radicals ultimately overwhelms antioxidant defenses in host cells,[74,75] and hampers the growth of microbes deficient in ROS/RNS-resistance mechanisms.[49,76,77]

An oxidative environment thus favors the growth of commensals with antioxidant capabilities and higher metabolic versatility.[46,78,79] Bacteroides spp. or Proteobacteria, for example, are ROS/RNS-resistant bacteria able to use a wide range of final electron acceptors during anaerobic respiration, including nitrate, TMAO, DMSO and tetrathionate.[80] In our study, genes encoding for enzymes implicated in the utilization of the latter compounds were enriched in LGC individuals, and were mainly encoded by Bacteroides and Proteobacteria (Fig. 5).

Of note, alongside with pro-inflammatory Proteobacteria like *Sutterella wadsworthensis*,[79] subjects with LGCs also had increased butyrate and propionate levels in direct fecal SCFA measurements. Bacteroides spp. are propionate producers,[81] which may explain the increased fecal propionate levels. However, butyrate results might seem more counterintuitive given the anti-inflammatory properties of butyrate.[82] Total fecal SCFAs measurements must be interpreted with caution because they neither measure direct SCFA production nor colonocyte uptake.[83] Overall, our findings support the notion that shifts in hydrogenotrophic microbes in the context of exogenous gut inflammation may modify SCFA metabolism in vivo.

Hydrogenotrophic microbes, i.e., methanogenic archaea, sulfate-reducing bacteria and acetogenic bacteria, compete for gut microbial hydrogen. Methanogenic archaea are promoted by fiber consumption[84] possibly because it leads to increased hydrogen production by primary fiber degraders. In our study, fiber consumption was indeed associated with abundance of methanogenic archaea. Methanogenic archaea, however, are also highly intolerant to oxygen[85] being outcompeted by alternative hydrogenotrophic microbes like acetogenic bacteria in oxidative environments.

In this study, subjects with LGCs had lower amounts of methanogens and sulfate-reducing bacteria. As seen elsewhere,[86] methanogenic archaea were, indeed, negatively correlated with fecal butyrate concentrations. In contrast, acetogenic bacteria were more abundant in LGC subjects. Acetogens compete with methanogens for hydrogen and are more tolerant to oxidative stress and oxygen fluxes.[85,87–89] They produce large quantities of acetate through acetogenesis,[89] which is not only a source of energy for gut epithelial cells, but also contributes to butyrate formation by butyrate-producing bacteria like Roseburia spp.[86,90,91] Therefore, we suggest that, in the absence of methanogens, acetogenic bacteria increase their contribution to interspecific H_2 transfer. This increases butyrate production by acetate-consuming butyrogenic bacteria like Roseburia spp., as evidenced by lower acetate:butyrate ratios in LGCs, also observed in previous studies.[92]

The strong association between intestinal microbial shifts and nadir CD4+ counts found in this study may be of importance in HIV medicine. Nadir CD4+ T-cell counts decrease with longer times since HIV-1 diagnosis and thus reflect the extent of HIV-1-induced immune damage, being a major surrogate marker of excess mortality, systemic inflammation, and clinical complications in chronically HIV-1-infected subjects.[11,93] Increases in genes implicated in LPS biosynthesis, bacterial virulence factors, and changes in the gut bacterial resistome, including depletion in genes involved in self-antibiotic resistance, and enrichment in genes associated with beta-lactam, lincosamide, tetracycline, and macrolide resistance in subjects with LGCs, could be related to the increased susceptibility to bacterial infections seen in subjects with severe immune suppression.[11,93]

Importantly, the compositional and functional shifts associated with gene richness in our study were highly consistent across HIV-1 transmission groups, ruling out confounding by HIV-1 risk group,[41] and have been described in other diseases characterized by gut inflammation,[18,19] such as obesity and inflammatory bowel disease. This suggests that gut microbiome shifts observed in our study are not necessarily unique to HIV-1 infection, but instead, share important characteristics with those seen in other diseases featuring gut inflammation. Moreover, the existence of gut microbiome shifts in some but not all HIV-1-infected subjects, particularly in those with higher immune damage in the past, suggests that gut microbiome testing has the potential to become a useful clinical tool to identify which individuals are more likely to develop gut microbiome-related diseases, including those associated with inflammaging and accelerated aging, impaired immune reconstitution, and severe bacterial infections. Prospective cohort studies linking gut microbiome data with AIDS and non-AIDS-related diseases are thus essential to validate and extend our findings. Such studies are critical to evaluate potentially unmeasured confounders, limit information recall biases, and include comprehensive information on other drugs that might modify the microbiome and might have not been evaluated herein. In addition, the use of colonic biopsies might provide further insights into the factors that condition gut microbial shifts during HIV-1 infection. In conclusion, low nadir CD4+ T-cell counts, rather than HIV-1 serostatus per se, predict the presence of gut dysbiosis in HIV-1 infected subjects. None of the compositional and functional shifts associated with such dysbiosis seem to be HIV-specific; instead, they are similar to those found in other diseases featuring gut inflammation, suggesting adaptation to oxidative stress as an important, not necessary exclusive, driver of gut microbial shifts in HIV-1 infection.

METHODS
Study design
This was a cross-sectional study in HIV-1-infected subjects with different virologic and immunologic phenotypes and HIV-negative controls. Comprehensive details of the cohort design and characteristics have been published elsewhere.[41] Study participants were recruited in Barcelona, Catalonia, Spain, between January and December 2014. HIV-1 infected subjects were recruited from the HIV Clinics of the tertiary care hospitals Germans Trias i Pujol and Vall d'Hebrón. HIV-1-negative controls were mainly recruited from a prospective cohort of HIV-negative MSM at risk of becoming infected by HIV-1,[44] who attend quarterly medical and counseling visits including HIV-1 testing (Alere Determine HIV-1/2 Ag/Ab Combo, Orlando, FL) at, a community-based center in Barcelona.[45] Additional controls were HIV-negative partners from HIV-1-infected subjects attending the HIV clinics.

The study included participants between 18 and 60 years and BMI within 18.5 and 30. Exclusion criteria were: (a) any gross dietary deviation from a usual diet, or any specific regular diet, i.e., vegetarian, low-carbohydrates, etc.; (b) antibiotic use during the previous 3 months (with the exception of late presenters, who could receive antibiotics to treat opportunistic infections); (c) pregnancy or willingness to become pregnant; (d) current drug

consumption or alcohol abuse; (e) any chronic digestive disease such as peptic ulcer, Crohn's disease, ulcerative colitis, or coeliac disease; (f) any surgical resection of the intestines except for appendectomy; (g) any autoimmune disease; and (h) any symptomatic chronic liver disease or presence of hepatic insufficiency defined as a Child-Pugh C score.

HIV-1-infected subjects were classified into the following seven mutually excluding categories based on their virological and immunological phenotype: (a) *elite controllers*: HIV-1 RNA < 50 copies/mL during at least 2 years in the absence of ART; (b) *viremic controllers*: HIV-1 RNA between 50 and 2000 copies/mL during at least 2 years in the absence of ART; (c) *early-treated*: ART initiation during the first 6 months after HIV-1 infection, HIV-1 RNA levels <50 copies/mL during at least the 3 last months and with no HIV-1 RNA blips after achieving HIV-1 RNA <50 copies/mL; (d) *ART-naïve*: HIV-1 RNA >10,000 copies/mL, nadir CD4+ T-cell counts >500 cells/mm^3 and no ART exposure; (e) *immune concordant*: HIV-1 RNA levels <50 copies/mL, and CD4+ T-cell counts >500 cells/mm^3 during at least 2 years; (f) *immune discordant*: HIV-1 RNA <50 copies/mL and CD4+ T-cell <300 cells/mm^3 during at least 2 years under ART; and (g) *late presenters*: CD4+ T-cell counts <200 cells/mm^3 at HIV-1 diagnosis and no ART exposure.

Data collection
Clinical and laboratory data were collected in a centralized database specifically designed for this study (OpenClinica, 2015 OpenClinica, LLC) using a standardized questionnaire.[41]

Dietary evaluation
Study participants received a thorough dietary and nutritional assessment by a specialized dietitian/nutritionist using two standardized and validated questionnaires:

a. A prospective dietary nutrient survey aimed at recording, as precisely as possible, any food, supplement or liquid intake during 3–5 consecutive days, including at least one weekend day.
b. A second questionnaire was a recall of food portions taken per week, on average, over the last year. To a great extent, this questionnaire is equivalent to the one used by the Health Professionals Follow-up Study of the Harvard School of Public Health (http://www.hpfstudy.org),[94] but has been adapted to include local dietary products and exclude items that are almost not consumed at all in our environment (e.g., peanut butter, etc).

After completion of both questionnaires, the study subject reviewed them with the study dietitian. Individual nutritional data was standardized by energy intake by fitting a linear model and taking residuals as new nutritional values (mean = 0 and SD = 1).[95] This ensured that amounts of nutrients were comparable between subjects with different total energy intake. A Dirichlet-multinomial distribution was assumed for species counts to account for data overdispersion, as suggested previously.[41,96]

Fecal DNA extraction
Study participants collected fecal samples at home in sterile fecal collection tubes the same day or the day before their medical appointment, before the proctology exam, and following pre-specified standard operating procedures. If required, samples were stored at 4 °C overnight until DNA extraction. Total DNA contained in aliquots of ~200 mg of each fecal sample was extracted using the PowerSoil DNA Extraction Kit (MO BIO Laboratories, Carlsbad, CA, USA) and then cryopreserved at −80 °C until sequencing.

Microbial DNA sequencing
Whole fecal DNA was fragmented with the Nextera-XT® Illumina kit. One library of approximately 300-basepair-clone insert size was constructed for each sample. Total fecal DNA was sequenced in an Illumina Hi-Seq® platform at the *Institut de Medicina Predictiva del Cancer* (IMPPC), Badalona, Spain.

Sequence quality assessment
Sequence quality was assessed using the FastQC software.[97] Total sequences were filtered by length and quality using Trimmomatic[98] ensuring a minimum base quality of Q30 for both leading and trailing bases, a minimum length of 75 basepairs and a minimum sequence quality average of Q20 for 30 basepair sliding windows across sequences. Human contamination was removed by mapping filtered sequences against the human genome using the *bwa*[99] software and discarding reads that uniquely aligned to the human reference with mapping quality higher than Q20. After filtering, a mean (standard deviation) of 31.0 (9.3) million sequences were obtained for each individual.

Taxonomical classification
The Metaphlan v2 software[100] was used to infer the microbial composition of each stool sample.

Microbial gene richness assessment
Filtered sequences were mapped against the integrated gene catalog (IGC)[101] using the *bwa* software.[99] Unique alignments with a minimum quality of Q20 were selected for subsequent analyses. An average of 24.8 million sequences (~80 % of total filtered sequences) mapped against the IGC reference catalog. To be able to compare the microbial gene content across samples we used a downsampling size of 10 million aligned sequences (Supplementary Fig. 2).

Gene richness was measured as the total number of different genes present in the sample regardless of their abundance and length, as described previously.[18] A minimum of one filtered mapped sequence was set to consider the presence of a gene. The copy number of each gene was estimated by dividing the total reads mapping to a gene divided by the gene's length. Gene relative abundance was measured as its copy number divided by the sum of the total gene copies in the sample. As in previous studies,[18] gene richness followed a bimodal distribution in a probability density function. The local minimum of the density function for gene richness distribution between the two modes, was considered the threshold value to classify individuals in two groups according to their microbial gene richness content: HGC and LGC.

To rule out possible biases due to the use of an external reference catalog not necessarily enriched in microbiome sequences from PLWH, we also constructed our own microbial gene catalog by de novo assembly and annotation of metagenomic sequences using the MOCAT2 pipeline.[102] This catalog was used to confirm the gene richness categorizations obtained using the IGC catalog.[101]

Microbial functional profile
Genes found in metagenomes downsampled at 10 M sequences were associated to one or multiple KEGG categories according to the IGC reference database.[101] To measure KEGG abundances, all genes associated to the same KEGG category were collapsed and their copy numbers were summed. If a gene was associated to more than one KEGG category, all categories were considered separately. The relative abundance of KEGG categories was measured as in the relative gene abundance approach detailed above. To analyze the metabolic pathways represented in all samples, we fed the HUMAnN pipeline[103] with the KEGG copy number table. The HUMAnN output included the relative abundances of metabolic modules and pathways for each subject. KEGG functions that were not associated to prokaryotic or archaeal metabolism according to KEGG catalog information[104] were manually filtered out, to ensure that only microbial functions were analyzed.

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

243

Contribution of microbial species to functional shifts
Bidirectional imbalances in microbial enzyme abundance between HGC and LGC groups were identified using the FishTaco software,[48] which also estimates the contribution of specific microbial species to such imbalances. The genomic content of the member species was inferred *de novo* by coupling the initial taxonomic and functional abundance profiles. Taxonomic profiles were generated considering only microbial species present in at least 5% of all patient samples. A false discovery rate threshold *q* < 0.05 was used to correct for multiple comparisons. Functional unbalances of interest were plotted using the FishTacoPlot R package.

Microbial virulence factors and antimicrobial resistance
To further characterize the presence of bacterial virulence factors and antimicrobial resistance genes in subjects with and without dysbiosis, we mapped 10 M randomly selected, quality-filtered, sequence reads from each sample against the VFDB[105] and the CARD,[106] respectively. Gene copy number was measured as described above.

Soluble markers of enterocyte integrity, bacterial translocation, and systemic inflammation
Plasma samples were collected and cryopreserved at −80 °C until quantification. Six different soluble plasma markers were measured in all study participants. Plasma levels of markers of enterocyte damage (IFABP), microbial translocation (soluble CD14 (scCD14) and LPS binding protein (LBP)) and systemic inflammation (interleukin-6 (IL-6), C-reactive protein (CRP), and interferon-gamma-inducible protein-10 (IP-10)) were analyzed using commercial ELISA kits in accordance with manufacturer's instructions [IL-6: high sensitivity ELISA with signal amplification, BMS213HS, eBioscience (San Diego, CA); scCD14: 0014-27 (ng/mL), Diaclone (Besançon, France); IFABP: DuoSet ELISA development system (pg/mL), R&D systems (Minneapolis, MN); CRP: KHA0031 (mg/mL), Invitrogen (Carlsbad, CA); IP-10: 1P10-09 (pg/mL), Diaclone (Besançon, France); LBP: KA0448 (ng/mL), Abnova (Taipei, Taiwan)]. If plasma levels were above the upper detection limit, the concentration considered was the maximum quantifiable value, i.e., the highest concentration included in the standard curve.

Fecal short-chain fatty acid levels
Acetic, butyric, propionic, valeric, isobutyric, and isovaleric acids concentrations were measured in 146 stool samples. Approximately ~100 mg of stool were suspended in sterile acidified water (pH 2.5) at a final concentration of 100 mg stool/mL and vigorously vortexed at room temperature for 5 min until the stool was completely homogenized. The homogenate was then centrifuged at 13 K RPM for 3 min and transferred 500 μL of the supernatant into a fresh microcentrifuge tube. Then 30 μL of 8 mM 2-methylvaleric acid (Sigma Aldrich) was added in ethyl acetate as an internal standard and 500 μL of ethyl acetate to the supernatant and vigorously vortexed this mixture for 10 min. These extractions were centrifuged at 13 K RPM for 3 min and transferred 300 μL of the upper organic phase to a glass autosampler vial. The SCFA were then analyzed on an HP6890 GC (Agilent) using a Stabilwax column (30 m x 0.32 mmID x 0.25 μm df, Restek) and 1 μL splitless injections. The oven temperature was as follows: 100 °C for 2 min then increased by 25 °C/min to 200 °C for 1 min followed by 250 °C for 1 min. The detector was an HP5973 MSD (Agilent). We prepared standard solutions of acetic, propionic, isobutyric, butyric, isovaleric, and valeric acids in acidified water and treated them as described above.

Statistical analyses of the microbiome composition and function
Microbiome samples were clustered according to their species composition using a nonmetric multidimensional scaling (NMDS) approach based on ecological distance matrices calculated by Bray–Curtis dissimilarities, as implemented R (Vegan, metaMDS and ggplot2 packages). NMDS ellipses were drawn based on a confidence interval (CI) of 0.95, and only species associated to an environmental vector with a length higher than 0.35 were represented.

The association between baseline variables and the HGC and LGC categories was evaluated using the two-tailed Fisher's exact test or the Wilcoxon rank sum test, as needed. To determine factors independently associated with LGC, we first fitted a univariate logistic regression model including each baseline variable showing a significantly different distribution between HGC and LGC. Variables significantly associated with LGC in univariate regression were then included in a multivariate logistic regression analysis. If two variables were redundant or collinear (e.g., MSM vs. non-MSM and gender) only one of them was included in the multivariate regression analysis. Odds ratios and 95% CIs were used to evaluate the weight of each variable relative to its reference.

Correlations between gene richness and other continuous variables, including microbial species, metabolic functions, and specific microbial enzymes were evaluated with the Spearman rank correlation test. Differences between HGC and LGC groups in abundance of microbial species, KEGG counts, antimicrobial resistance and bacterial virulence genes were evaluated using a Wilcoxon rank sum test. Differences in microbial species, functions and enzymes by nadir CD4+ T-cell count categories were evaluated using a Kruskal–Wallis test with post hoc comparison tests as required. In all previous evaluations the Benjamini–Hochberg correction was applied to avoid type I error inflation due to multiple comparisons. A random forest analysis was used to determine the relative contribution of each microbial species associated with richness to the classification of subjects as HGC or LGC. Area-under-the-curve values were used to compare the classification accuracy of the different predictors in both machine-learning approaches.

As in previous publications from our group,[41] nutrient consumption was standardized by energy intake by fitting a linear model and taking residuals as new nutritional values. Dirichlet-multinomial regression[107] was implemented to identify the strongest relationships between nutritional standardized nutrient intake and species abundance and variance, penalizing weak associations using a penalized likelihood approach.[41] In addition, nutrients independently correlated with gene richness were identified using Spearman correlation or Wilcoxon Mann–Whitney nonparametric tests, as needed.

Ethics and community involvement
The study was reviewed and approved by the Institutional Review Boards of the Hospital Universitari Germans Trias i Pujol (reference PI-13-046) and the Hospital Vall d'Hebrón (reference PR(AG)109/2014). All participants provided written informed consent in accordance with the World Medical Association Declaration of Helsinki. The study concept, design, patient information and results were discussed with the HIVACAT Community Advisory Committee (CAC), who provided input on these aspects as well as on the presentation and dissemination of study results.

Sequence and data availability
Raw Illumina MiSeq sequences and study metadata were deposited in the National Center for Biotechnology Information - NCBI repository (Bioproject accession number: PRJNA307231, SRA accession number: SRP068240).

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

244

AUTHOR CONTRIBUTIONS

ADDITIONAL INFORMATION

REFERENCES
1. Antinori, A. et al. Late presentation of HIV infection: a consensus definition. *HIV Med.* **12**, 61–64 (2011).
2. Mocroft, A. et al. Risk factors and outcomes for late presentation for HIV-positive persons in Europe: results from the Collaboration of Observational HIV Epidemiological Research Europe Study (COHERE). *PLoS Med.* **10**, e1001510 (2013).
3. Paiardini & Müller-Trutwin, M. HIV-associated chronic immune activation. *Immunol. Rev.* **254**, 78–101 (2013).
4. Deeks, S. G. HIV infection, inflammation, and aging. *Annu. Rev. Med.* **62**, 141–155 (2011).
5. Battegay, M., Fluckiger, U., Hirschel, B. & Furrer, H. Late presentation of HIV-infected individuals. *Antivir. Ther.* **12**, 841–851 (2007).
6. Ingle, S. M. et al. Impact of risk factors for specific causes of death in the first and subsequent years of antiretroviral therapy among HIV-infected patients. *Clin. Infect. Dis.* **59**, 287–297 (2014).
7. The INSIGHT START Study Group. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *N. Engl. J. Med.* **373**, 795–807 (2015).
8. May, M. T. et al. Mortality according to CD4 count at start of combination antiretroviral therapy among HIV-infected patients followed for up to 15 years after start of treatment: Collaborative Cohort Study. *Clin. Infect. Dis.* **62**, 1571–1577 (2016).
9. Connor, J. O. et al. Effect of immediate initiation of antiretroviral therapy on risk of severe bacterial infections in HIV-positive people with CD4 cell counts of more than 500 cells per μL: secondary outcome results from a randomised controlled trial. *Lancet HIV* **3018**, 4–11 (2017).
10. Robbins, G. K. et al. Incomplete reconstitution of T-cell subsets on combination antiretroviral therapy in the AIDS Clinical Trials Group protocol 384. *Clin. Infect. Dis.* **48**, 350–361 (2009).
11. Negredo, E. et al. Nadir CD4 T-cell count as predictor and high CD4 T-cell intrinsic apoptosis as final mechanism of poor CD4 T-cell recovery in virologically suppressed HIV-infected patients: clinical implications. *Clin. Infect. Dis.* **50**, 1300–1308 (2010).
12. Brenchley, J. M. et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nat. Med.* **12**, 1365–1371 (2006).
13. Douek, D. C. Disrupting T-cell homeostasis: how HIV-1 infection causes disease. *AIDS Rev.* **5**, 172–177 (2003).
14. Sandler, N. G. & Douek, D. C. Microbial translocation in HIV infection: causes, consequences and treatment opportunities. *Nat. Rev. Microbiol.* **10**, 655–666 (2012).
15. Claesson, M. J. et al. Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184 (2012).
16. Tang, W. H. W. et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368**, 1575–1584 (2013).
17. Koeth, R. A. et al. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 533–534 (2013).
18. Chatelier, E. Le et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
19. Cotillard, A. et al. Dietary intervention impact on gut microbial gene richness. *Nature* **500**, 585–588 (2013).
20. Shoaie, S. et al. Quantifying diet-induced metabolic changes of the human gut microbiome. *Cell Metab.* **22**, 320–331 (2015).
21. Klatt, N. R., Chomont, N., Douek, D. C. & Deeks, S. G. Immune activation and HIV persistence: implications for curative approaches to HIV Infection. *Immunol. Rev.* **254**, 326–342 (2013).
22. Lozupone, C. A. et al. HIV-induced alteration in gut microbiota: driving factors, consequences, and effects of antiretroviral therapy. *Gut Microbes* **5**, 562–570 (2014).
23. Brenchley, J. M. et al. CD4+ T-cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J. Exp. Med.* **200**, 749–759 (2004).
24. Morais, D. et al. Microbiota-dependent sequelae of acute infection compromise tissue-specific immunity article microbiota-dependent sequelae of acute infection compromise tissue-specific immunity. *Cell* **163**, 354–366 (2015).
25. Althoff, K. N., Smit, M., Reiss, P. & Justice, A. C. HIV and ageing: improving quantity and quality of life. *Curr. Opin. HIV AIDS* (2016). https://doi.org/10.1097/COH.0000000000000305
26. Corbeau, P. & Reynes, J. Review article Immune reconstitution under antiretroviral therapy: the new challenge in HIV-1 infection. *Blood* **117**, 5582–5591 (2017).
27. Massanella, M., Negredo, E., Clotet, B. & Blanco, J. Immunodiscordant responses to HAART—mechanisms and consequences. *Expert Rev. Clin. Immunol.* **9**, 1135–1149 (2013).
28. Kilbourne, A. M. et al. General medical and psychiatric comorbidity among HIV-infected veterans in the post-HAART era. *J. Clin. Epidemiol.* **54**(Suppl 1), S22–S28 (2001).
29. Sudano, I. et al. Cardiovascular disease in HIV infection. *Am. Heart J.* **151**, 1147–1155 (2006).
30. Guaraldi, G. et al. Premature age-related comorbidities among HIV-infected persons compared with the general population. *Clin. Infect. Dis.* **53**, 1120–1126 (2011).
31. Klatt, N. R. et al. Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *Science* **356**, 938–945 (2017).
32. Gosmann, C. et al. Lactobacillus-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young south african women report lactobacillus-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition. *Immunity* **46**, 29–37 (2017).
33. Costea, P. I. et al. Enterotypes in the landscape of gut microbial community composition. *Nat. Microbiol.* **3**, 8–16 (2018).
34. Rothschild, D. et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
35. 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).
36. Williams, B., Landay, A. & Presti, R. M. Microbiome alterations in HIV infection a review. *Cell Microbiol.* **18**, 645–651 (2016).
37. Williams, B. et al. A summary of the first HIV microbiome workshop 2015. *AIDS Res. Hum. Retroviruses* **32**, 935-941 (2016).
38. 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).
39. Vázquez-Castellanos, J. F. et al. Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. *Mucosal Immunol.* **8**, 760–772 (2015).
40. 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).
41. Noguera-Julian, M. et al. Gut microbiota linked to sexual preference and HIV infection. *EBioMedicine* **5**, 135–146 (2016).
42. Dillon, S. M., Frank, D. N. & Wilson, C. C. The gut microbiome and HIV-1 pathogenesis: a two-way street. *AIDS* **30**, 2737–2751 (2016).

Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection
Y Guillén et al.

245

43. Vujkovic-Cvijin, I. et al. Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. *Sci. Transl. Med.* **5**, 193ra91 (2014).

44. Coll, J. et al. Early detection of HIV infection and of asymptomatic sexually transmitted infections among men who have sex with men. *Clin Microbiol Infect.* **24**, 540-545 (2018).

45. Meulbroek, M. et al. BCN Checkpoint, a community-based centre for men-who-have-sex-with-men in Barcelona, Catalonia, Spain, shows high efficiency in HIV detection and linkage to care. *HIV Med.* **14**, 25–28 (2013).

46. Barcenilla, A. et al. Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl. Environ. Microbiol.* **66**, 1654–1661 (2000).

47. Duncan, S. H., Barcenilla, A., Stewart, C. S., Pryde, S. E. & Flint, H. J. Acetate utilization and butyryl coenzyme A (CoA): acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Appl. Environ. Microbiol.* **68**, 5186–5190 (2002).

48. Manor, O. & Borenstein, E. Systematic characterization and analysis of the taxonomic drivers of functional shifts in the human resource systematic characterization and analysis of the taxonomic drivers of functional shifts in the human microbiome. *Cell Host Microbe.* **21**, 1–14 (2017).

49. Winter, S. E., Lopez, C. A. & Bäumler, A. J. The dynamics of gut-associated microbial communities during inflammation. *EMBO Rep.* **14**, 319–327 (2013).

50. Canfora, E. E., Jocken, J. W. & Blaak, E. E. Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat. Rev. Endocrinol.* **11**, 577–591 (2015).

51. Bergey, D. H., Buchanan, R. E. & Gibbons, N. E. *Bergey's Manual of Systematic Bacteriology.* (Springer-Verlag, New York, NY, 1974).

52. Ahmed, M., Lyass, L., Markham, P. N., Taylor, S. S. & Neyfakh, A. A. Two highly similar multidrug transporters of *Bacillus subtilis* whose expression is differentially regulated. *J. Bacteriol.* **177**, 3904–3910 (1995).

53. Fernández, L. & Hancock, R. E. W. Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. *Clin. Microbiol. Rev.* **25**, 661–681 (2012).

54. Andra, J., Goldmann, T., Ernst, C. M., Peschel, A. & Gutsmann, T. Multiple peptide resistance factor (MprF)-mediated resistance of *Staphylococcus aureus* against antimicrobial peptides coincides with a modulated peptide interaction with artificial membranes comprising lysyl-phosphatidylglycerol. *J. Biol. Chem.* **286**, 18692–18700 (2011).

55. Rensing, C. & Grass, G. *Escherichia coli* mechanisms of copper homeostasis in a changing environment. *FEMS Microbiol. Rev.* **27**, 197–213 (2003).

56. Neckers, L. & Tatu, U. Review molecular chaperones in pathogen virulence: emerging new targets for therapy. *Cell Host Microbe* **4**, 519–527 (2008).

57. Craig, L., Pique, M. E. & Tainer, J. A. Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* **2**, 363–378 (2004).

58. Coburn, B., Sekirov, I. & Finlay, B. B. Type III secretion systems and disease. *Clin. Microbiol. Rev.* **20**, 535–549 (2007).

59. Mougous, J. D. et al. A virulence locus of pseudomonas. *Science* 1526–1530 (2013). https://doi.org/10.1126/science.1128393

60. Rossez, Y., Wolfson, E. B., Holmes, A., Gally, D. L. & Holden, N. J. Bacterial Flagella: twist and stick, or dodge across the kingdoms. *PLoS Pathog.* **11**, 1–15 (2015).

61. Hopwood, D. A. How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them? *Mol. Microbiol.* **63**, 937–940 (2007).

62. Handley, S. A. et al. Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell* **151**, 253–266 (2012).

63. Fiume, G. et al. Human immunodeficiency virus-1 Tat activates NF-??B via physical interaction with IκB-α and p65. *Nucleic Acids Res.* **40**, 3548–3562 (2012).

64. Serrano-Villar, S. et al. Gut bacteria metabolism impacts immune recovery in HIV-infected Individuals. *EBioMedicine* **8**, 203–216 (2016).

65. Baier-Bitterlich, G., Fuchs, D. & Wachter, H. Chronic immune stimulation, oxidative stress, and apoptosis in HIV infection. *Biochem. Pharmacol.* **53**, 755–763 (1997).

66. Miller, R. A. & Britigan, B. E. Role of oxidants in microbial pathophysiology. *Clin. Microbiol. Rev.* **10**, 1–18 (1997).

67. Vilhardt, F. et al. The HIV-1 Nef protein and phagocyte NADPH-oxidase activation. *J. Biol. Chem.* **277**, 42136–42143 (2002).

68. Mangino, G. et al. HIV-1 myristoylated nef treatment of murine microglial cells activates inducible nitric oxide synthase, NO2 production and neurotoxic activity. *PLoS ONE* **10**, 1–21 (2015).

69. Cassatella, M. A. et al. Molecular basis of interferon-g and lipopolysaccharide enhancement of phagocyte respiratory burst capability. *J. Biol. Chem.* **265**, 20241–20246 (1990).

70. Jacob, B. A. et al. HIV-1-induced pulmonary oxidative and nitrosative stress: exacerbated response to endotoxin administration in HIV-1 transgenic mouse model. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **291**, L811–L819 (2006).

71. Arunagiri, C., Macreadie, I., Hewish, D. & Azad, A. A. C- terminal domain of HIV-1 accessory protein Vpr is involved in penetration, mitochondrial dysfunction and apoptosis of human CD4+ lymphocytes. *Apoptosis* **2**, 69–76 (1997).

72. Ivanov, A. et al. Oxidative stress during HIV infection: mechanisms and consequences. *Oxid. Med. Cell Longev.* **2016**, 1–18 (2016).

73. Kanther, M. et al. Commensal microbiota stimulate systemic neutrophil migration through induction of serum amyloid A. *Cell Microbiol.* **16**, 1053–1067 (2015).

74. Flores, S. C. et al. Tat protein of human immunodeficiency virus type 1 represses expression of manganese superoxide dismutase in HeLa cells. *Proc. Natl Acad. Sci. USA* **90**, 7632–7636 (1993).

75. Sandstrom, P. A., Roberts, B., Folks, T. M. & Buttke, T. M. HIV gene expression enhances T-cell susceptibility to hydrogen peroxide- induced apoptosis. *AIDS Res. Hum. Retrovir.* **9**, 1107–1113 (1993).

76. Bäumler, A. J. & Sperandio, V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**, 85–93 (2016).

77. Lemaitre, B. & Hoffmann, J. The Host Defense of *Drosophila melanogaster*. *Annu. Rev. Immunol.* **25**, 697–743 (2007).

78. Smith, C. J., Rocha, E. R. & Paster, B. J. The medically important bacteroides spp. in health and disease. *Prokaryotes* 381–427 (2006). https://doi.org/10.1007/0-387-30746-x

79. Mukhopadhya, I. et al. A comprehensive evaluation of colonic mucosal isolates of sutterella wadsworthensis from inflammatory bowel disease. *PLoS ONE* **6**, 1–10 (2011).

80. Winter, S. E. & Bäumler, A. J. A breathtaking feat: to compete with the gut microbiota, salmonella drives its host to provide a respiratory electron acceptor. *Gut Microbes* **2** (2011).

81. Reichardt, N. et al. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J.* **8**, 1323–1335 (2014).

82. Canani, R. B. et al. Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World J. Gastroenterol.* **17**, 1519–1528 (2011).

83. Cummings, J. & Macfarlane, G. The control and consequences of bacterial fermentation in the human colon. *J. Appl. Bacteriol.* **70**, 443–459 (1991).

84. Fernandes, J., Wolever, T. M. S. & Rao, A. V. Interrelationships between age, total dietary fiber intake and breath methane in humans. *Nutr. Res.* **20**, 929–940 (2000).

85. Jarrell, K. F. Extreme oxygen sensitivity in methanogenic archaebacteria. *Bioscience* **35**, 298–302 (1985).

86. Abell, G. C. J., Conlon, M. A. & Mcorist, A. L. Methanogenic archaea in adult human faecal samples are inversely related to butyrate concentration. *Microb. Ecol. Health Dis.* **18**, 154–160 (2006).

87. Karnholz, A., Küsel, K., Gößner, A., Schramm, A. & Drake, H. L. Tolerance and metabolic response of acetogenic bacteria toward oxygen. *Appl. Environ. Microbiol.* **68**, 1005–1009 (2002).

88. Dolla, A., Fournier, M. & Dermoun, Z. Oxygen defense in sulfate-reducing bacteria. *J. Biotechnol.* **126**, 87–100 (2006).

89. Drake, H. L. et al. Acetogenic bacteria: what are the in situ consequences of their diverse metabolic versatilities? *Biofactors* **6**, 13–24 (1997).

90. Duncan, S. H. et al. Contribution of acetate to butyrate formation by human faecal bacteria. *Br. J. Nutr.* **91**, 915 (2004).

91. Chassard, C. & Bernalier-Donadille, A. H2 and acetate transfers during xylan fermentation between a butyrate-producing xylanolytic species and hydrogenotrophic microorganisms from the human gut. *FEMS Microbiol. Lett.* **254**, 116–122 (2006).

92. Serrano-Villar, S. et al. The effects of prebiotics on microbial dysbiosis, butyrate production and immunity in HIV-infected subjects. *Mucosal Immunol.* **10,** 1279-1293 (2017).

93. Bray, S. et al. Predictive value of CD4 cell count nadir on long-term mortality in HIV-positive patients in Uganda. *HIV AIDS* **4**, 135–140 (2012).

94. Wu, G. D. et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).

95. Willett, W., Howe, G. & Kushi, L. Adjustment for total energy intake in epidemiologic studies. *Am. J. Clin. Nutr.* **4 Suppl**, S1220–S1231 (1997).

96. la Rosa, P. S. et al. Hypothesis testing and power calculations for taxonomic-based human microbiome data. *PLoS ONE* **7**, 1–13 (2012).

97. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010) http://www.bioinformatics.babraham.ac.uk/projects/fastqc

98. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

99. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

100. Truong, D. T. et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903 (2015).

101. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).

102. Kultima, J. R. et al. Genome analysis MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* **32**, 2520–2523 (2016).

103. Abubucker, S. et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* **8**, e1002358 (2012).

104. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

105. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* **44**, 694–697 (2016).

106. Jia, B. et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, 566–573 (2017).

107. Chen, J. & Li, H. Variable selection for sparse Dirichlet-multinomial regression with an applicaton to microbiome data analysis. *Ann. Appl. Stat.* **7** (2013).