BMC Medical Genomics

RESEARCH

Open Access

Major depletion of insulin sensitivity-associated taxa in the gut microbiome of persons living with HIV controlled by antiretroviral drugs

Eugeni Belda[1,2*†], Jacqueline Capeau[3†], Jean-Daniel Zucker[1,2], Emmanuelle Le Chatelier[4], Nicolas Pons[4], Florian Plaza Oñate[4], Benoit Quinquis[4], Rohia Alili[2], Soraya Fellahi[3,9], Christine Katlama[5], Karine Clément[2], Bruno Fève[3], Stéphane Jaureguiberry[6], Cécile Goujard[6], Olivier Lambotte[6,7], Joël Doré[4,8], Edi Prifti[1,2†] and Jean-Philippe Bastard[3,9†]

Abstract

Background Persons living with HIV (PWH) harbor an altered gut microbiome (higher abundance of *Prevotella* and lower abundance of *Bacillota* and *Ruminococcus* lineages) compared to non-infected individuals. Some of these alterations are linked to sexual preference and others to the HIV infection. The relationship between these lineages and metabolic alterations, often present in aging PWH, has been poorly investigated.

Methods In this study, we compared fecal metagenomes of 25 antiretroviral-treatment (ART)-controlled PWH to three independent control groups of 25 non-infected matched individuals by means of univariate analyses and machine learning methods. Moreover, we used two external datasets to validate predictive models of PWH classification. Next, we searched for associations between clinical and biological metabolic parameters with taxonomic and functional microbiome profiles. Finally, we compare the gut microbiome in 7 PWH after a 17-week ART switch to raltegravir/maraviroc.

Results Three major enterotypes (Prevotella, Bacteroides and Ruminococcaceae) were present in all groups. The first Prevotella enterotype was enriched in PWH, with several of characteristic lineages associated with poor metabolic profiles (low HDL and adiponectin, high insulin resistance (HOMA-IR)). Conversely butyrate-producing lineages were markedly depleted in PWH independently of sexual preference and were associated with a better metabolic profile (higher HDL and adiponectin and lower HOMA-IR). Accordingly with the worst metabolic status of PWH, butyrate production and amino-acid degradation modules were associated with high HDL and adiponectin and low HOMA-IR. Random Forest models trained to classify PWH vs. control on taxonomic abundances displayed high generalization

†Eugeni Belda, Jacqueline Capeau, Edi Prifti and Jean-Philippe Bastard contributed equally to this work.

*Correspondence:
Eugeni Belda
eugeni.belda@ird.fr

Full list of author information is available at the end of the article

performance on two external holdout datasets (ROC AUC of 80–82%). Finally, no significant alterations in microbiome composition were observed after switching to raltegravir/maraviroc.

Conclusion High resolution metagenomic analyses revealed major differences in the gut microbiome of ART-controlled PWH when compared with three independent matched cohorts of controls. The observed marked insulin resistance could result both from enrichment in *Prevotella* lineages, and from the depletion in species producing butyrate and involved into amino-acid degradation, which depletion is linked with the HIV infection.

Keywords HIV, Human gut microbiome, Metabolic status, Machine learning

Background

Current suppressive antiretroviral therapy (ART) allows a sustained efficient long-term control of HIV infection in persons living with HIV (PWH). However, when aging, they often present a dysmetabolic profile with truncal fat redistribution associated with insulin resistance [1–4]. In particular, the level of insulin resistance, as evaluated by the HOMA-IR index, appears disproportionally high according to their body mass index (BMI) when compared with the general population [5–7]. This insulin-resistant profile can explain, at least partly, the increased risk of diabetes and cardio-vascular disease reported in aging ART-controlled PWH [8]. It likely results from multiple factors, including an altered gut microbiota, the residual HIV infection, the HIV reservoir size or some ART. Altered gut microbiota has been largely demonstrated in PWH, dysbiosis in terms of decreased levels of alpha-diversity and altered microbiome composition vs. non-infected individuals being present in ART-naïve subjects and persisting in ART-controlled PWH [9, 10]. This altered gut microbiota is characterized by decreased levels of microbiome diversity and loss of microbial species associated to the production of Short Chain Fatty Acids (SCFA) from dietary fibers in parallel with enrichment of pro-inflammatory lineages like *Enterobacteriaceae*, *Desulfovibrionaceae*, and Fusobacteria, and has been shown correlated with markers of inflammation and disease progression ([11] and references therein). These alterations are further exacerbated in PWH developing further comorbidities in terms of cardiovascular events, kidney diseases or cancer [12], or in PWH with severe immunocompromised status [13].

In PWH with immunocompromised status or ART-naive common co-infections are frequent with opportunistic pathogens not only from microbial origin but also from fungal, viral or protozoan origin. Hepatitis C virus co-infection in PWH is estimated to affect 2.3 million individuals worldwide, particularly in intravenous drug users or in men having sex with men (MSM), with an immunopathology more deleterious than each infection separately [14]. Also, fungi are major opportunistic pathogens in PWH with advanced infection or with low CD4 count, mainly *Pneumocystis jirovecii* (pneumocystosis), *Cryptococcus neoformans* (cryptococcosis), *Histoplasma capsulatum* (histoplasmosis), and *Talaromyces (Penicillium) marneffei* (talaromycosis) [15, 16]. Altered fungal microbiome (mycobiome) composition in the oral and respiratory tract ecosystems has been also observed in PWH characterized by increased colonization of *Candida* and other fungal lineages in the oral cavity that dramatically decreases after ART therapy [17, 18], and pathogens like *Pneumocystis*, *Cryptococcus*, and *Aspergillus* in the lung [19, 20]. At the gut level, a recent study on 24 PWH and 12 healthy controls from Spain has shown that PWH exhibited a higher mycobiome richness with an enrichment of *Debaryomyces hansenii*, *Candida albicans*, and *Candida parapsilosis*, with a positive correlation between *Candida* species and the levels of pro-inflammatory cytokines, interleukin 22, and CD8+T cell counts [21].

At prokaryotic level, increased *Prevotella* and decreased *Bacteroides* at the genus level, associated with increased inflammation and immune activation [9, 22] are constant findings in PWH as compared with non-HIV-infected controls. This profile has been clearly associated with men having sex with men (MSM) sexual preference [23, 24]. In addition, a decreased level of *Faecalibacterium* species and/or of other butyrate-producing bacteria has been observed in PWH as well as a global *Ruminococcaceae* depletion correlating with the presence of noncommunicable diseases, and this, independently of sexual preferences [24]. Moreover, two other studies in PWH reported decreased *Faecalibacterium* presence in individuals with metabolic syndrome [25, 26]. However, the direct correlation with insulin resistance and the metabolic profile was not precisely studied. Moreover, in PWH, the relative abundance of *Faecalibacterium* species and/or butyrate producing bacteria was reported not modified in individuals with diabetes compared with nondiabetics in one study [27], whereas it tended to be decreased in diabetics in another one, which focused only on women. Similarly, a recent prospective study on the Women's Interagency HIV Study cohort (WIHS) have shown this decrease in butyrate-producing bacterial lineages (*Ruminococcus* genus) in diabetic women with and without HIV infection [28]. However, evaluation in women allowed to consider mainly the effect of diabetes, while decreasing the interference of microbiota

modifications linked to anal intercourse in MSM [29]. In this context, these butyrate-producing lineages like *Ruminococcaceae* and *Lachnospiraceae* were included in an HIV-related Microbiota Index derived from metagenomic signatures altered in PWH vs. controls in MSM and non-MSM groups based on 16 S profiling on the Copenhagen Comorbidity in HIV Infection cohort (COCOMO), an index that was positively associated with an excess risk of metabolic syndrome and that is stronger in individuals with severe immunocompromised status [13].

In the present study, using shotgun metagenomics, we estimated the relative abundance of the microbial species as well as their functional potential in stools, and compared their distribution between PWH and matched HIV-negative controls by age, gender and BMI from three independent cohorts. We confirmed previous findings with increased *Prevotella* lineages, possibly linked to the MSM status, and revealed novel association with insulin resistance. We confirmed the reduced level of *Faecalibacterium* and other butyrate-producing bacteria in PWH, both in the MSM and not-MSM groups, and revealed a novel association with an insulin-resistant profile associating increased HOMA-IR to reduced HDL and adiponectin. We hypothesize that the marked insulin resistance observed in PWH could result from the sexual preference associated with increased *Prevotella* lineages and from the deep defect in gut butyrate-producing bacteria associated with HIV infection. Finally, by using different machine learning approaches, we built classification models discriminating PWH from controls based on several of these signatures and which generalize well on two external cohorts with 80% ROC AUC.

Methods

Subjects HIV-infected

We evaluated at first 11 subjects issued from the ROCnRAL ANRS-157 study, a phase II, single-arm, multicenter clinical trial designed to evaluate the capacity of the Raltegravir (RAL) plus Maraviroc (MVC) combination to maintain HIV-viremia below 50 copies/ml in PWH with controlled viral load under ART and with trunk fat redistribution [30] (ClinicalTrials.gov registration number: NCT01420523, first trial registration: 18/08/2011). Stool collection was performed as described below. Patients had a second evaluation after a median of 17 weeks post RAL/MVC initiation (mean duration = 124 days; minimum = 84 days; maximum = 168 days) and a second stool sample was recovered for 7 of them (64%) for metagenomics analysis. We also collected stools from 14 additional PWH, ART-controlled, issued from the outpatients of the infectious diseases and the internal medicine and immunology departments of Bicêtre Hospital. Clinical and biological parameters were analyzed on fasted blood samples. All patients were ART-controlled with a viral load < 50 copies/ml. 60% were receiving a protease inhibitor (PI), 16% an integrase strand transfer inhibitors (INSTI) 80% nucleosidic reverse transcriptase inhibitors (NRTIs) and 44% a non-nucleosidic reverse transcriptase inhibitor (NNRTI). In terms of co-infection with Hepatitis C (HCV), 22/25 PWH were HCV negative or cured, 1/25 was HCV positive and 2 were unknown for HCV status. In terms of co-infection with Hepatitis B (HBV), 16/25 PWH were HBV negative, 4/25 were HBV positive and 5/25 were unknown for HBV status. In terms of ethnicity, 22/25 PWH were Caucasian, 2/25 were Black and 1/25 was unknown for ethnicity status.

Overall, 25 PWH were included in this study for shotgun metagenomic profiling, of which 7 were profiled 17 weeks post RAL/MVC (n = 32 fecal samples for shotgun metagenomic profiling).

Regarding the ethical aspects, the trial was approved by ANRS and CPP Ile de France VI. All patients gave their written approbation to the protocol.

Selection of controls from MetaHit, MicroObes and MetaCardis studies

We defined 3 different control groups of non-HIV individuals from 3 public external cohorts: n = 177 healthy Danish individuals from the MetaHit study [31, 32], n = 95 healthy French individuals from MetaCardis study [33] and n = 35 French individuals of the MicroObese study, corresponding to baseline samples of individuals with moderate obesity before a nutritional intervention [34], for which quantitative metagenomic profiles were available using the Integrated human Gut microbiome Catalog (IGC) [35].

For each external cohort, 25 individuals matched by age, gender and BMI with the 25 PWH cases of this study were selected using two different approaches. In the first approach, a matrix of Euclidean distances derived from age, gender and BMI between the 25 PWH and each of the reference control cohorts were computed with base R *dist* function. For each PWH individual, the control sample with the smallest Euclidean distance was selected iteratively, excluding at each iteration the selected control sample. In the second approach, we carried out 100 iterations of the propensity score method from the *Matching* R package [36]. The final selection of each control cohort (n = 25) between those obtained by the Euclidean approach and the Propensity score approach was achieved by evaluating differences between PWH and selected controls (Wilcoxon rank-sum test for age and BMI; Chi-square test for gender) and selecting the one that minimizes the overall differences defined as the maximum average p-value from the 3 tests. This workflow allowed us to select control groups from MetaCardis and MetaHit cohorts that were fully paired on the three

covariates (non-significant differences vs. the 25 PWH at significance level of 0.05, Table 1), whereas the in the case of MicroObese cohort, the 25 selected controls could be matched on age but still discordant on gender (most women) and BMI (higher levels in selected MicroObese controls; Table 1).

Stool collection, DNA extraction and metagenomic sequencing

PWH collected fecal samples in two 20 mL tubes within 24 h before each visit. Samples were either stored immediately at −80 °C or briefly conserved in home freezers, in anaerobic conditions, before transport to the laboratory where they were immediately frozen at −80 °C following guidelines. The total fecal DNA from 32 samples ($n=25$ baseline samples and 7 follow-up samples, 20 weeks after RAL/MVC initiation) was extracted, sequenced and analyzed. DNA extraction used quenching solutions to protect DNA from degradation by DNases and a bead-beating step that ensures the lysis of particularly robust cells. A barcoded fragment library was prepared for each sample and DNA sequencing data were generated using SOLiD 5500xl.

Metagenomic analyses

Primary analysis of shotgun metagenomic data from the $n=32$ PWH samples, from read quality control to mapping against the Integrated Gene Catalog (IGC) of the human gut microbiome [35] and quantification of gene abundances was performed using METEOR [37]. First, read quality control was performed by excluding low quality reads having an average quality score < 20. In addition, reads aligned to the human genome (GRCh38.p13) or to the SOLiD adapter sequences were discarded. Second, host or food associated reads were filtered out by aligning reads against the *Homo sapiens* (GRCh37.p10), *Bos taurus* (Btau_4.6.1) and *Arabidopsis thaliana* reference genomes with bowtie [38]. Third, an average of 103.6 million high-quality reads per sample (sd. 31.8 M) were mapped against the IGC 9.9 million gene reference catalogue [35] using bowtie with a maximum of three mismatches. An average of 47.48 million high-quality reads per sample (sd. 18.33 M) were mapped to the IGC catalog (45.02% of high-quality reads per sample on average, sd. 7.13%). Full details of the number of total reads, high-quality reads and mapped reads on the IGC catalog for the 32 PWH samples are available in Supplemental Table 1. A raw gene abundance table of PWH samples was generated by means of a two-step procedure previously described that handles multi-mapped reads [39]. First, uniquely mapping reads (reads that maps to a single gene in the IGC catalogue) were attributed to their corresponding genes. Second, shared reads (reads that mapped with the same alignment score to multiple genes) were attributed according to the ratio of their unique mapping counts.

The raw gene abundance table of PWH samples ($n=25$ PWH and $n=7$ follow-up) was integrated with the equivalent raw gene abundance tables of the $n=75$ matched controls from the MetaHit, MicroObes and MetaCardis studies into a unified raw gene count table of $n=107$ samples. To decrease technical bias due to different

Table 1 Clinical and biological characteristics of the subjects

	PWH vs. Microobes			PWH vs. Metahit			PWH vs. Metacardis		
	PWH *N*=25	Microobes *N*=25	*p.* overall	PWH *N*=25	Metahit *N*=25	*p.* overall	PWH *N*=25	Metacardis *N*=25	*p.* overall
BMI (kg/m^2)	25.0 (4.2)	32.7 (3.7)	8.7e-07	25.0 (4.2)	26.0 (4.4)	0.64	25.0 (4.2)	23.2 (1.58)	0.071
Waist circumference (cm)	97.9 (6.2)	105 (9.2)	0.024	97.9 (6.2)	. (.)	.	97.9 (6.2)	83.1 (7.17)	1.1e-05
Hip circumference (cm)	98.7 (3.6)	114 (9.3)	5.3e-06	98.7 (3.6)	. (.)	.	98.7 (3.6)	96.3 (5.2)	0.24
CRPus (mg/L)	5.9 (10.4)	5.3 (6.9)	0.23	5.9 (10.4)	. (.)	.	5.9 (10.4)	1.5 (1.7)	0.082
INS (µU/ml)	19.2 (20.1)	8.11 (3.7)	0.0031	19.2 (20.1)	. (.)	.	19.2 (20.1)	5.23 (3.35)	2.2e-06
LDL-cholesterol (mmol/L)	2.83 (0.92)	3.30 (0.75)	0.066	2.83 (0.92)	. (.)	.	2.83 (0.92)	2.82 (0.70)	0.77
HDL-cholesterol (mmol/L)	1.22 (0.32)	1.41 (0.36)	0.037	1.22 (0.32)	. (.)	.	1.22 (0.32)	1.66 (0.43)	0.00036
SBP (mm Hg)	138 (22.6)	119 (13.0)	0.026	138 (22.6)	. (.)	.	138 (22.6)	123 (15.8)	0.074
Glycemia (mmol/L)	5.79 (1.02)	5.31 (0.46)	0.14	5.79 (1.02)	. (.)	.	5.79 (1.02)	4.92 (0.52)	0.00031
IL-6 (pg/mL)	3.73 (5.72)	2.17 (1.72)	0.99	3.73 (5.72)	. (.)	.	3.73 (5.72)	1.67 (1.05)	0.62
sCD14 (ng/mL)	1595 (226)	1382 (390)	0.009	1595 (226)	. (.)	.	1595 (226)	1470 (305)	0.039
Adiponectin (µg/mL)	3.30 (2.04)	14.1 (6.93)	1.3e-07	3.30 (2.04)	. (.)	.	3.30 (2.04)	6.28 (4.18)	0.0027
HOMA-IR	5.12 (5.03)	1.93 (0.91)	0.00081	5.12 (5.03)	. (.)	.	5.12 (5.03)	1.16 (0.73)	1.7e-06
Gender:			0.00023			0.7			1
Female	5 (20.0%)	19 (76.0%)		5 (20.0%)	3 (12.0%)		5 (20.0%)	5 (20.0%)	
Male	20 (80.0%)	6 (24.0%)		20 (80.0%)	22 (88.0%)		20 (80.0%)	20 (80.0%)	
Age (years)	53.1 (7.0)	47.9 (10.3)	0.14	53.1 (7.0)	52.7 (8.2)	0.83	53.1 (7.0)	53.2 (12.5)	0.55

Results are given as mean (SD). Pairwise differences in clinical variables between PWH and each of the control groups were evaluated by Wicoxon rank-sum test (continuous variables) and Chi-square tests (categorical variables). BMI: body mass index; INS: fasting insulinemia; SBP: systolic blood pressure; sCD14: soluble CD14

sequencing depth, rarefaction (or downsizing) was performed by randomly selecting 11 million reads were for each sample using a draw without replacement. Finally, rarefied gene abundances were normalized according to the FPKM (fragments per kilobase of transcript per million mapped reads) strategy (normalization by the gene size and the number of total mapped reads reported in frequency) to give the gene abundance profile. Microbial gene richness was measured by counting the number of genes that were present for a given sample after rarefaction (i.e. with at least one read aligned on them; Supplemental Table 2).

Computation of microbial species abundance was performed using the metagenomic species (MGS) approach [40] implemented in the *momr* R package [41]. Abundance for each MGS was computed as the average abundance of the 50 most correlated genes as proposed in the original study [40]. Only MGS with more than 500 genes were considered for downstream statistical analyses. MGS taxonomic annotation was performed by aligning their genes against the GTDB r214 representative genomes [42] with blastn [43]. A species-level assignment was given if > 50% of the genes matched the same reference genome at a threshold of 95% of identity and 90% of gene length coverage. The remaining MGS were assigned to a given taxonomical level from genus to superkingdom level, if more than 50% of their genes had the same level of assignment. MGS abundance table and taxonomic annotations are available in Supplemental Table 3.

Functional profiles based on KEGG ortholog groups were derived from functional annotations of the IGC gene catalog obtained by the MOCAT2 framework [44]. Gut Metabolic Modules (GMMs) were computed from KO abundance matrix using the GOmixer R package [45] and KEGG functional modules were quantified from KO abundance matrix as the harmonic average of module KO abundances using the FAM R package (https://git.ummisco.fr/pipelines/fam.git). GMM and KEGG module abundance tables are available in Supplemental Tables 4 and 5 respectively.

Enterotyping of the $n = 107$ samples was performed following the Dirichlet Multinomial Mixture (DMM) method of Holmes et-al. [46] using MGS abundance matrix collapsed at the genus level. The DMM approach groups samples if their taxon abundances can be modeled by the same Dirichlet-Multinomial (DM) distribution. Classification at K = 3 groups (3 DM distributions) showed best model fit performance based on the Laplace metric (Supplemental Fig. 1A). Clustering in three groups (k = 3) was driven by the main bacterial lineages (*Prevotella*, *Bacteroides*, *Ruminococcaceae*) that defined the original enterotypes [47] based on non-parametric Kruskal-Wallis test and environmental fitting on the PCoA ordination (Fig. 1B), including also a significantly high microbial gene richness in *Ruminococcaceae* enterotype (Supplemental Fig. 1B). Enterotype assignments are available in Supplemental Table 2.

The estimation of the explanatory power of clinical covariates on microbiome profiles derived from genus-level MGS abundance data was performed using distance-based redundancy analyses (dbRDA) as implemented in the R package *vegan* v2.6-4 [48]. Microbiome inter-individual variation was visualized by principal coordinates analysis using Bray–Curtis dissimilarity on the genus-level MGS abundance data. Environmental fitting of clinical covariates with significant impact on microbiome composition based on dbRDA analyses over PCoA ordination from Bray-Curtis inter-sample dissimilarity matrix was computed with the *vegan::envfit* function.

External validation datasets

Two external datasets were used to validate the classification models. The cohort from the Bai *et-al* study [49] (PRJNA692830 ENA study accession) consisted of 24 Swedish individuals including 12 PWH under antiretroviral therapy and 12 controls. The cohort from the Lu et-al study [50] (PRJNA391226 ENA study accession) consisted of 71 Chinese individuals (61 HIV and 10 healthy controls).

Raw reads from each study were downloaded from the ENA repository and processed with NGLess v1.5.0 [51] for quality trimming (minimum read quality = 25; minimum read length = 40), host contaminant removal vs. reference human genome (min. identity = 90%, min. match size = 45), alignment filtered reads over the ICG 9.9-million gene catalog [35] (min. identity = 95%, min. match size = 45), and generation of raw gene abundance tables with the *dist1* metric (first the gene abundances are computed from unique mapped reads and then are corrected by the multiple mapped reads weighted by the coverage of unique mapped reads). The raw gene abundance tables were processed as described above for rarefaction (11 M reads x sample), normalization and MGS abundance calculation with the *momr* R package [41].

Classification analyses

The Predomics R package v1.01 [52] was used to build binary classification models of PWH (baseline) vs. each of the three control groups based on MGS abundance profiles. Models were trained using three different algorithms including Random Forest as state-of-the-art (SOTA) method, and the Binary and Ternary native Predomics models that were learned with the Terbeam heuristic. The Binary and Ternary models describe simple ecological relationships in microbial ecosystems. Details of the different heuristics and BTR models can be found in [52]. Each sample group (baseline PWH and 3 control

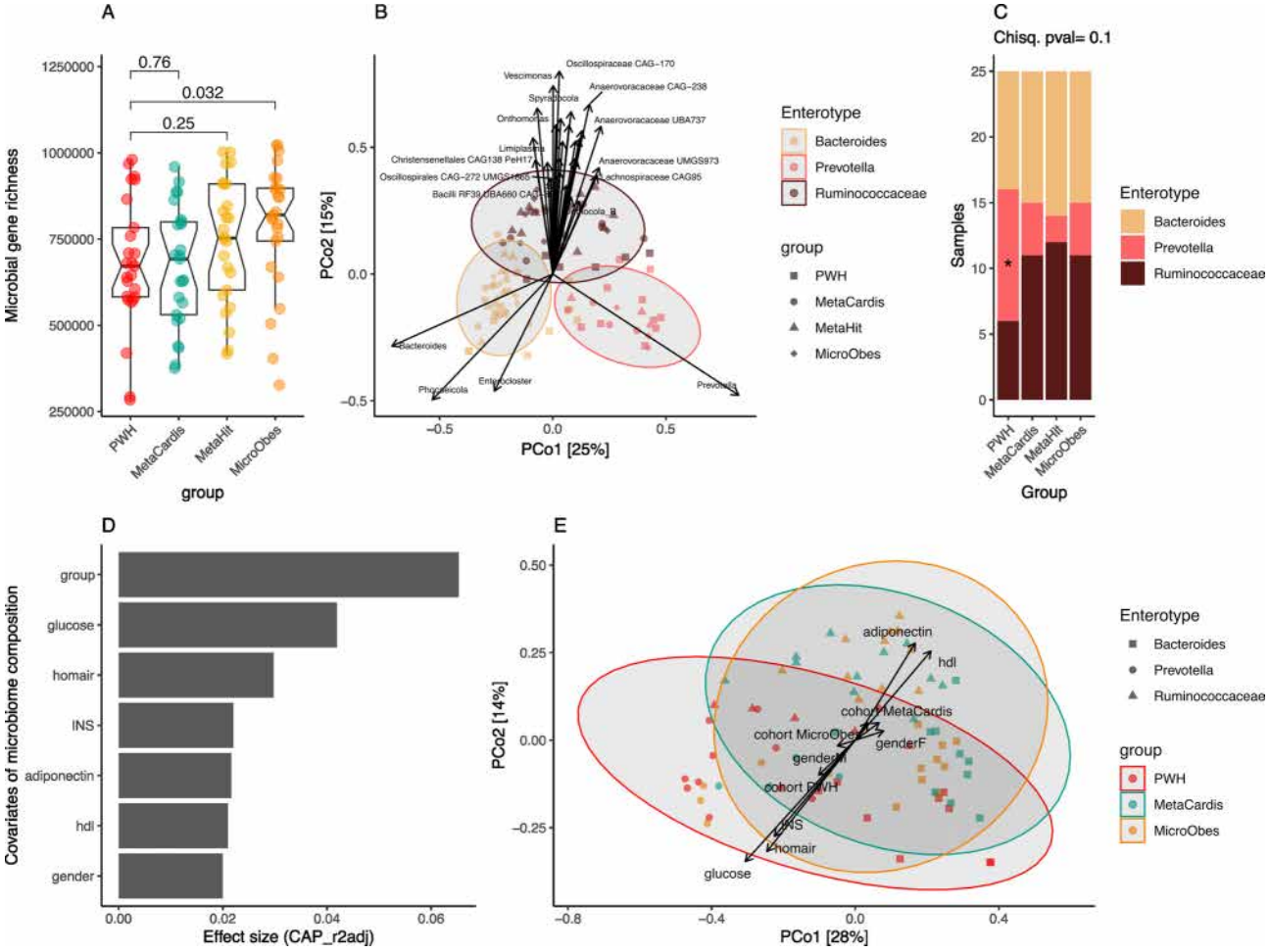

Fig. 1 General overview of microbiome composition of ROCnRAL cohort. (**A**) Gene richness distribution across study groups ($n=25$ individuals per group; p-values from pairwise Wilcoxon rank-sum test are shown in brackets). (**B**) Principal coordinates analysis (PCoA) analyses of samples in panel **A** with arrows representing the effect sizes of the 30 genus features with highest significant differences across enterotypes (FDR < 0.05; Kruskal-Wallis test) product of environmental fitting over ordination plot. (**C**) Enterotype compositions in each study group ($n=25$ individuals per group). *P*-value of Chi-squared test is shown in the top of the panel; *=FDR < 0.05, post-hoc analysis for Pearson's Chi-squared Test for Count Data (**D**) Clinical variables explaining the microbiome compositional variation across baseline individuals of PWH, MetaCardis and MicroObes groups ($n=25$ individuals each; distance-based redundancy analysis, dbRDA; genus-level Bray-Curtis dissimilarity; FDR < 0.05). (**E**) PCoA of inter-individual differences (genus-level Bray-Curtis beta-diversity) across samples in panel D with samples colored by study group and shaped by enterotypes. Arrows represent effect sizes of the significant variables identified by dbRDA analyses in panel **D** product of environmental fitting over ordination plot

groups) was divided into a training set of $n=20$ samples and a holdout set of $n=5$ samples. Models were trained on the training data and evaluated for accuracy and AUC on a 10 times 10-fold cross-validation schema, and the results of the best algorithm were further explored in the case of Binary and Ternary models to study the family of best models (FBM), described as models whose accuracy is within a given window of the best model's accuracy. This window is defined by computing a significance threshold, assuming that accuracy follows a binomial distribution ($p<0.05$). No hyperparameter optimization was performed. Features included in the FBM were further explored in terms of prevalence across models and groups as well as their importance, described as the mean decrease accuracy (MDA) of the model after feature removal. The best model in each prediction task was evaluated on two different holdout datasets consisting in the 5 samples of each group that were excluded from the training process and the two external datasets described above.

Statistical analyses

Wilcoxon rank-sum tests were used to compare microbial gene richness between PWH baseline group and each of the 3 control groups, and chi-square test followed by post-hoc analyses for interpreting chi-square contingency-table test results [53] was used to evaluate differences in enterotype composition between PWH and the 3 control groups. Wilcoxon rank-sum tests followed by multiple test correction using the Benjamini-Hochberg

method were used to identify taxonomic (MGS) and functional modules (GMM, KEGG modules) with significant differences between PWH and each control group. Features with FDR < 0.05 were considered as significantly different between PWH and each of the control groups. Similar non-parametric analyses were performed stratifying the PWH into MSM and no-MSM individuals to evaluate the impact of MSM status on the results observed with the entire PWH group as this variable (MSM status) was not available for the control groups (no possible adjustment). Wilcoxon signed-rank tests followed by FDR correction by Benjamini-Hochberg method were used to identify taxonomic (MGS) and functional modules (GMM, KEGG) with significant differences in PWH before and 20 weeks after RAL/MVC initiation, and features with FDR < 0.05 were considered as significantly different between both time points. Differences in microbial gene richness and enterotype composition between MSM and non-MSM groups were evaluated with Wilcoxon rank-sum tests and Fisher test respectively. Spearman correlations followed by FDR correction by Benjamini-Hochberg method were used to evaluate the association of MGS and functional modules derived from the univariate analyses with clinical covariates available for both PWH and their matched controls of the MetaCardis and MicroObese cohorts, and to evaluate the similarity in effect sizes of feature change (Cliff's delta) between non-MSM and MSM individuals vs. the three control groups for both MGS and GMM abundance tables (Supplemental Fig. 4). Differences in clinical, anthropometric and demographic variables in study groups are reported in mean ± standard deviation. All analyses were performed in R v4.2.2.

Results

Cohort description

The main subjects' characteristics are given on Table 1. Using quantitative metagenomics, we investigated the gut microbiome of 25 PWH, ART-controlled, aged 53.1±7.0 years, with an average BMI of 25±4.2 kg/m^2 and an increased level of insulin resistance (HOMA-IR 5.12±5.03) enrolled in the ROCnRAL study [30] and from the Internal Medicine and Immunology department of Bicêtre Hospital (Kremlin-Bicêtre, France). They were 20 men and 5 women. Regarding the HIV mode of transmission, their status was: 13 men having sex with men (MSM), 9 heterosexuals (4 men and 5 women), 1 blood transfusion, and 2 unknowns.

They were compared with three groups of HIV-negative individuals from three different studies (25/177 healthy Danish individuals from MetaHit study [31], 25/95 healthy French individuals from MetaCardis study [33] and 25/35 individuals of the MicroObes study corresponding to baseline timepoint [34]) paired on age, gender and BMI: Quantitative metagenomic profiles for all the samples were derived using the Integrated human Gut microbiome Catalog (IGC) [35] (see Methods).

We observed that PWH presented significantly higher levels of insulin and of HOMA-IR than control individuals from MetaCardis and MicroObes cohorts for which clinical data were available (Table 1). HOMA-IR values were also higher (p=0.05) in the MSM group (7.1±1.7) compared with the non-MSM group (2.7±0.4), while the average BMI was not significantly different (respectively 25.7±1.2 and 23.6±1.6 kg/m^2).

Among the 7 PWH from the ROCnRAL study who had been tested at baseline, there was no significant variation in BMI or HOMA-IR index after 17±4 weeks following the switching to dual therapy RAL/MVC.

Microbial gene richness and composition of PWH vs. control groups and after ART switch

Significant differences in microbial gene richness were observed across study groups (p-value=0.047; Kruskal-Wallis test, Fig. 1A). Pairwise comparison of PWH with different control groups showed a significant depletion of microbial gene richness only in comparison with the matched controls from the MicroObes cohort (p-value=0.032; Wilcoxon rank-sum test; Fig. 1A), suggesting no major dysbiosis in terms of microbiome diversity in PWH in comparison with healthy controls (MetaHit, MetaCardis groups).

Stratification of the individuals into different microbiome composition groups (i.e. enterotypes) based on the DMM approach [46] showed the presence of 3 enterotypes driven by *Bacteroides*, *Prevotella* and different lineages of the *Ruminococcaceae*, *Oscillospiraceae and Christensenellaceae* families that reproduce well known community types of the human gut microbiome [47, 54] (Fig. 1B, Supplemental Fig. 1B). Chi-square tests showed non-significant differences in enterotype composition between study groups (p-value=0.14), but post-hoc testing on the adjusted standardized residuals of the Chi-square test [53] showed that Prevotella enterotype was significantly enriched in the PWH group (FDR=0.046; Fig. 1C).

Distance-based redundancy analysis (dbRDA) of study groups and 19 clinical variables available in PWH and control individuals from MetaCardis and MicroObes cohorts, for which clinical data were available, showed seven variables with significant impact on microbiome composition (FDR < 0.05, Fig. 1D). Study group was the variable with the highest impact (adjR^2=0.065; FDR=0.01) followed by glucose levels (adjR^2=0.041; FDR=0.01), HOMA-IR (adjR^2=0.029; FDR=0.01), insulin levels (adjR^2=0.022; FDR=0.03), adiponectin levels (adjR^2=0.021; FDR=0.032), HDL (adjR^2=0.021; FDR=0.034) and gender (adjR^2=0.017; FDR=0.033),

this last variable possibly associated to the difficulties encountered in the matching of controls from the MicroObese cohort (see Methods; Table 1). Environmental fitting of these variables on the compositional landscape of microbiome variation defined by the PCoA of these 75 individuals (PWH, MetaCardis and MicroObes controls) showed that metabolic health variables (glucose levels, HOMA-IR, insulin) were associated with regions of the compositional space dominated by PWH individuals (Fig. 1E).

Longitudinal analyses on the 7 PWH switched to RAL/MVC therapy showed that the treatment didn't significatively impact the microbial gene richness (Supplemental Fig. 2A; p-value=0.375 Wilcoxon signed-rank test) nor the global microbiome composition of (p-value=1; PERMANOVA test). This was reflected by the proximity of baseline and 17-weeks microbiome samples of each individual in ordination analyses (Supplemental Fig. 2B) as well as by the stable enterotype assignment at both time points for 4 of the 7 PWH with follow-up (Supplemental Fig. 2C). Univariate analyses on taxonomic and functional profiles didn't show significant alterations after FDR correction (Supplemental Tables 6–8).

Overall, these results suggest that HIV infection (study group) and metabolic health status (glucose levels) play a significant explanatory role on microbiome compositional variation, with PWH showing strong enrichment in the *Prevotella* enterotype, which is in line with previous reports [9, 10].

Finally, we compared our data in MSM subjects ($n=13$) and non-MSM subjects (heterosexual men, women or infected through blood transfusion, $n=10$), excluding 2 subjects for which no information was available. We did not observe any significant difference in microbiome composition (p-value=0.107; PERMANOVA test; Supplemental Fig. 3A) as well as in microbial gene richness (p-value=0.41; Wilcoxon rank-sum test, Supplemental Fig. 3B) by MSM status in PWH. However, in terms of enterotype composition, we observed a tendency towards an enrichment in Prevotella enterotype in MSM subjects and Bacteroides enterotype in no-MSM subjects, although overall differences in enterotype composition by MSM status were not statistically significant, possibly due to the low number of participants (p-value=0.5; Fisher's exact test; Supplemental Fig. 3C).

Taxonomic and functional differences between PWH individuals and control groups

We identified 121 metagenomic species (MGS) whose abundance was significantly different between PWH and at least one of the control groups (FDR<0.05, Wilcoxon rank-sum tests). Of these, 27 MGS showing significant variations in abundance (FDR<0.05; Wilcoxon rank-sum test) in the same direction in PWH vs. all three control groups were identified as core MGS (Fig. 2A, Supplemental Table 9). Among these robust associations, PWH showed significant increase in 12 MGS belonging to the *Prevotella* genus and a MGS belonging to *Succinivibrionaceae* family of gamma-proteobacteria (*CAG00726:Succinivibrio sp000431835*) together with a significant decrease in several MGS assigned to the *Bacillota_A* phylum, including some butyrate producing lineages like *Ruthenibacterium*, *Gemmiger* and *Faecalibacterium* genus, as well as *Actinobacteriota* lineages like *CAG0059: Bifidobacterium longum* and CAG01272: *Adlercreutzia hattorii* (Fig. 2A).

Out of these 27 core MGS, 16 (i.e., 60%) were found to be significantly different in individual comparisons of MSM and no-MSM PWH individuals vs. at least one of the control groups (FDR<0.1 in both comparisons; Fig. 2A). Among them, *CAG00726: Succinivibrio sp000431835* (Fig. 2B) and *CAG00323: Prevotella sp900543975* (Fig. 2C) were significantly higher in both groups of PWH individuals vs. all three control groups whereas *CAG01272: Adlercreutzia hattorii* (Fig. 2D) and *CAG00755: Faecalibacterium sp900539945* (Fig. 2E) showed the opposite pattern (significantly depleted in both groups of PWH individuals vs. the three control groups). Comparison of the effect sizes of MGS changes between PWH and control groups (Cliff's Delta) in MSM and no-MSM individuals by means of Spearman correlations showed however a strong and significant positive correlation (Supplemental Fig. 4A), which suggests that the sign of the variations in MGS abundances between PWH and control groups are also robust to MSM status despite the loss of statistical power due to reduced group sizes.

At the functional level we observed a consistent depletion of the functional potential of the gut microbiome of PWH individuals when compared with the three different control groups. 77 Gut Metabolic Modules (GMM) showed significant differences in abundances between PWH and at least one control group (FDR<0.05; Wilcoxon rank-sum test, Supplemental Table 10), of which 29 core modules displayed robust and significant variations when compared against all three control groups, all of them being depleted in PWH (Fig. 3A). Among them we observed an overall depletion of the potential for amino acid degradation and for butyrate production in the gut microbiome of PWH individuals (Fig. 3A). Similar analyses stratified by MSM status of PWH individuals showed that the differences in 17 of the 29 core GMM were reproduced in individual groups of MSM and no-MSM PWH vs. the three control groups (FDR<0.1 in both comparisons, left panel in Fig. 3A). Butyrate production modules were retained in this subset of 17 core GMM with robust association to MSM preferences (Fig. 3B, C), which suggests that depletion of butyrate

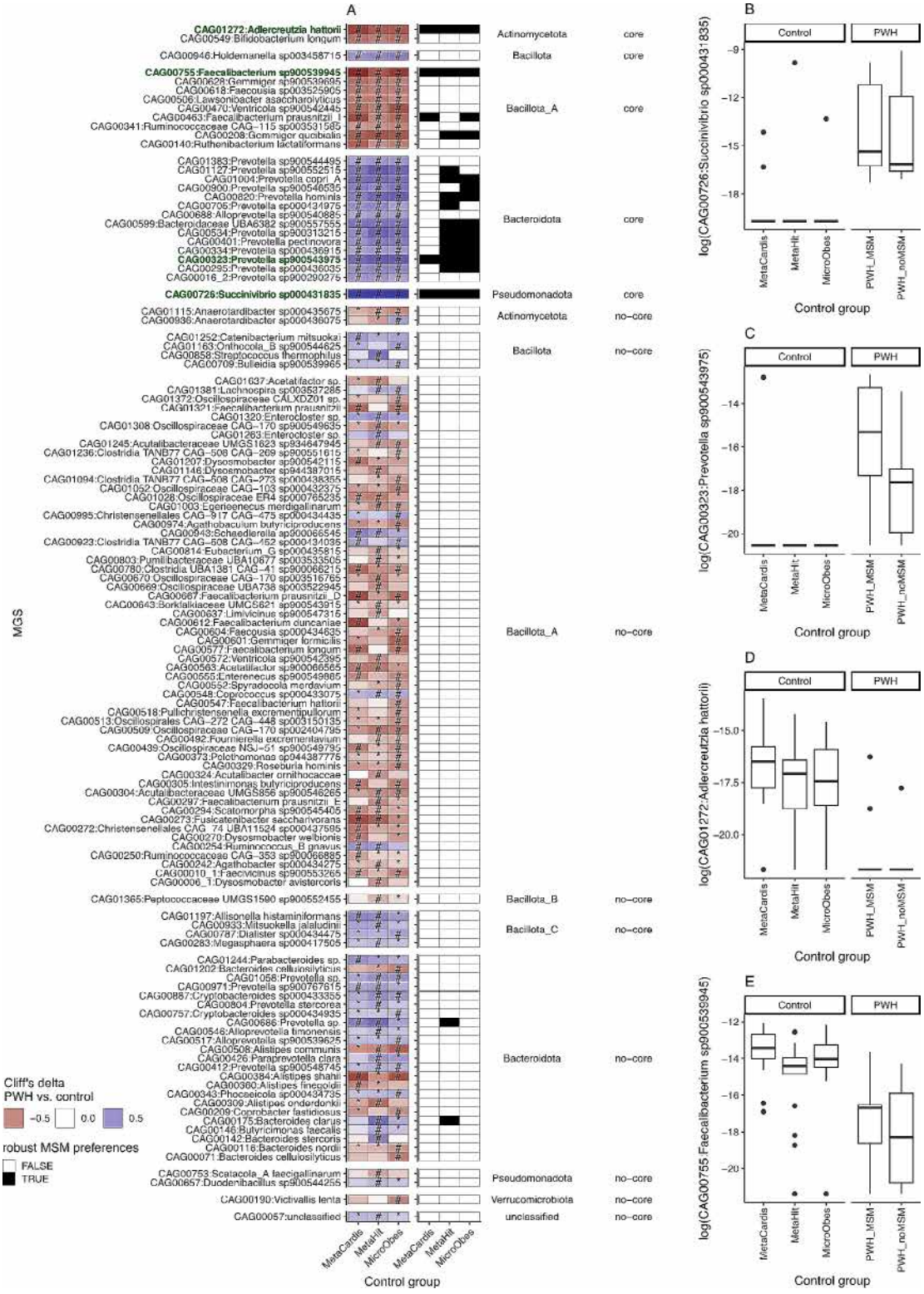

Fig. 2 Differentially abundant metagenomic species (MGS) across study groups. (**A**) Heatmap of Cliff's Delta effect sizes (left panel) of 121 MGS with significant differences in abundance between HIV individuals ($n = 25$) and MicroObes ($n = 25$), MetaHit ($n = 25$) or MetaCardis ($n = 25$) individuals (#=FDR < 0.05; *=p-value < 0.05; Wilcoxon rank-sum test). Positive values correspond to MGS significantly increased in the PWH group, whereas negative values correspond to MGS significantly depleted in the PWH group. Right panel represents the robustness of the associations to MSM status (FDR < 0.1; Wilcoxon rank-sum test in MSM vs. control and no-MSM vs. control comparisons; $n = 13$ MSM, $n = 10$ no-MSM). (**B**-**E**) Boxplots of log-transformed MGS abundances across study groups for the 4 MGS highlighted in green in panel **A**

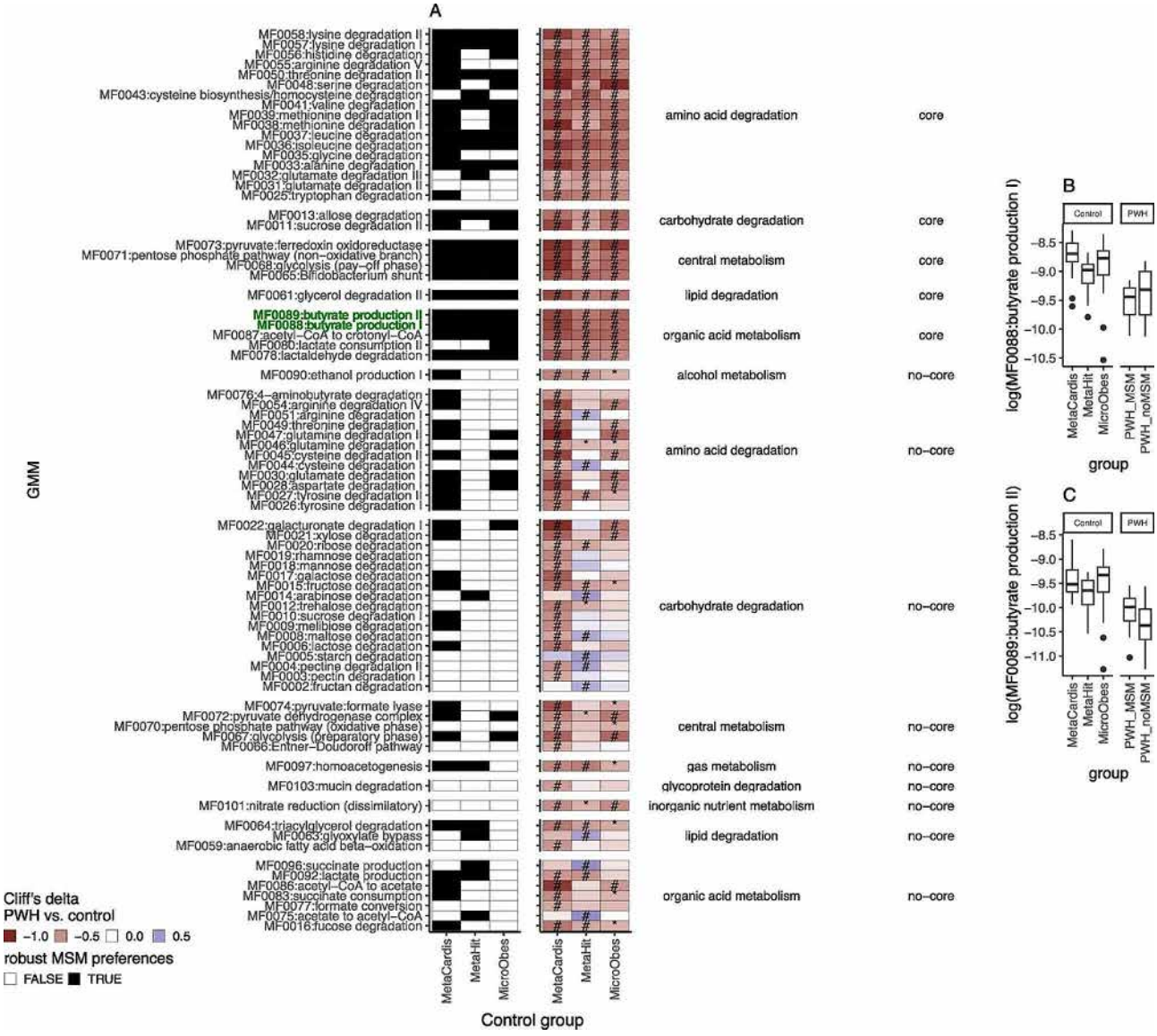

Fig. 3 Gut Microbiome Modules (GMMs) with significant differences across study groups. (**A**) Heatmap of Cliff's Delta effect sizes (right panel) of 77 GMMs with significant differences in abundance between PWH individuals ($n = 25$) and MetaCardis ($n = 25$), MetaHit ($n = 25$) and MicroObes ($n = 25$) individuals (#=FDR < 0.05; *=*p*-value < 0.05; Wilcoxon rank-sum test). Positive values correspond to GMMs significantly increased in PWH, whereas negative values correspond to GMMs significantly depleted in PWH. Left panel represents the robustness of the associations to MSM status (FDR < 0.1; Wilcoxon rank-sum test in MSM vs. control and no-MSM vs. control comparisons; $n = 13$ MSM, $n = 10$ no-MSM) (**B**-**C**) Boxplots of GMMs abundances across study groups for Butyrate production modules (significant decrease in PWH individuals vs. all three control groups and robust to MSM status; highlighted in green in panel **A**)

production potential is a strong signature of the microbiome of PWH and this independently of MSM preferences. Similar comparison of the effect sizes of GMM changes between PWH and control groups (Cliff's Delta) in MSM and no-MSM individuals by means of Spearman correlations showed less strong positive correlation than with the MGS data, although still significant, with all GMM showing differences in MSM and no-MSM groups vs. each control group being consistently depleted in the groups of PWH individuals (Supplemental Fig. 4B).

We observed similar results when a second functional space, corresponding to KEGG functional modules, was analyzed (Supplemental Fig. 5). 41 KEGG modules showed significant differences in abundance levels between PWH and all three control groups (FDR < 0.05; Wilcoxon rank-sum tests, Supplemental Table 11), of which 39 were consistently depleted and just 2 modules were consistently enriched in PWH individuals (Supplemental Fig. 5A). Among these, we observed an enrichment of a module involved in the biosynthesis of the ADP-L-glycero-beta-D-manno-heptose

(ADP-L-beta-D-heptose) precursor of the inner core lipopolysaccharide (LPS). This increase was observed only in the MSM group (Supplemental Fig. 5B).

Association between metagenomic features and clinical variables

Next, we looked for associations between core metagenomic features, significantly enriched or depleted in PWH individuals vs. all three control groups, and 13 clinical variables that were available in PWH individuals and controls from the MetaCardis and MicroObes cohorts. At the taxonomic level, MGS enriched in PWH were positively associated with insulin levels (INS) and insulin resistance (HOMA-IR) and negatively associated with adiponectin and HDL levels in stratified analyses combining PWH and MetaCardis controls and PWH and MicroObes controls (Fig. 4A, Supplemental Table 12). The opposite associations were observed for butyrate producing lineages of the Bacillota_A phylum, significantly depleted in PWH individuals (Fig. 4A). In contrast, associations with corpulence variables were less consistent, depending on the control group, which can be explained by the high corpulence profile of individuals of the MicroObes cohort, despite the efforts in matching for BMI in the selection of non-PWH samples (Fig. 4A).

At the functional level, the 29 modules significantly depleted in PWH vs. all three control groups, including butyrate production potential, were all consistently negatively associated with metabolic health variables (insulin levels, insulin resistance, glucose levels) and positively associated with adiponectin and HDL levels in all stratified analyses combining PWH with MetaCardis and MicroObes controls (Fig. 4B, Supplemental Table 12). Overall, these results show that the depletion in butyrate producing lineages and butyrate production potential in parallel with enrichment of *Prevotella* and gammaproteobacterial lineage of the Pseudomonadata phylum (CAG00726:*Succinivibrio sp000431835*) in PWH is consistently associated with a worse metabolic profile.

Predictive power of the gut microbiome for classification of PWH individuals vs. controls

We next explored the predictive power of the gut microbiome composition in predicting the PWH status in binary classification tasks vs. each of the three different control groups using a suite of machine learning (ML) algorithms implemented in the Predomics R package [52]. We tested random forest models as well as two ecosystem-inspired (binary and ternary) models initially introduced in a previous article. These are interpretable simplifications of linear models that generalize well [52] and trained with different heuristic algorithms. Specifically, in binary models we looked for combinations of MGS whose cumulative abundance allows classifying individuals as PWH or controls if the score is above or below a given threshold that is learned with the model, whereas in ternary models we looked for a difference between the cumulative abundance of two groups of features (MGS) as a signature to classify individuals as PWH or control (difference above or below a threshold that is learned by the model).

Random forest models showed the best overall performance during training on cross-validation (avg AUC +- std error=0.98 +- 0.003, 1.00+-0.00 and 0.99+-0.001 in PWH vs. MetaCardis, MetaHit and MicroObes controls respectively) followed by bininter (avg AUC +- std error=0.88 +- 0.011, 0.94+-0.009 and 0.94+-0.01 in PWH vs. MetaCardis, MetaHit and MicroObes controls respectively) and terinter models (avg AUC +- std error=0.88 +- 0.012, 0.91+-0.011 and 0.86+-0.01 in PWH vs. MetaCardis, MetaHit and MicroObes controls respectively; Supplemental Fig. 6A). However, when focused the analyses on the Family of Best binary and ternary Models (FBM) learned by Predomics (subset of models whose accuracy is statistically non-different from the best model) their performances overall increased to accuracy levels similar to those learned with random forest (close to 1) but with a number of features ranging from 2 to 10 (Supplemental Fig. 6B).

Predomics generated 718 best binary models (FBM), which selected a total of 39 different MGS in the classification tasks of PWH vs. each of the 3 control groups (Fig. 5A). Among these features we observed several of the core MGS retrieved in univariate analyses with similar variations in PWH vs. each of the three control groups and robust to MSM status, including *CAG00323: Prevotella sp90054375* and *CAG00755: Faecalibacterium sp900539945* retained in 62.23% of the FBM of PWH vs. MetaCardis controls, or *CAG00726: Succinivibrio sp000431835*, which is retained in all FMB in PWH vs. MicroObes controls, showing the overall highest mean feature importance (Fig. 5A). Most of the binary models in FBM of PWH vs. MetaHit and MicroObes controls included exclusively MGS enriched in the PWH group, several of them like *CAG00726: Succinivibrio sp000431835* or *CAG00323: Prevotella sp90054375* and the other 2 core MGS retained in at least one of the FBM in all three binary prediction tasks being highly specific of PWH even in terms of prevalence (rightmost panel in Fig. 5A; MGS with prevalence near zero in the three control groups vs. 100% prevalence in PWH). *CAG00175: Bacteroides clarus* seems to be specifically retained in the FBM trained to classify PWH vs. MetaHit controls, explained by the specific absence of this MGS in the MetaHit controls (0% prevalence vs. 45% prevalence in the MetaCardis and MicroObes controls and 80% prevalence in PWH; rightmost panel in Fig. 5A). In contrast, FBM of the binary method identified when classifying

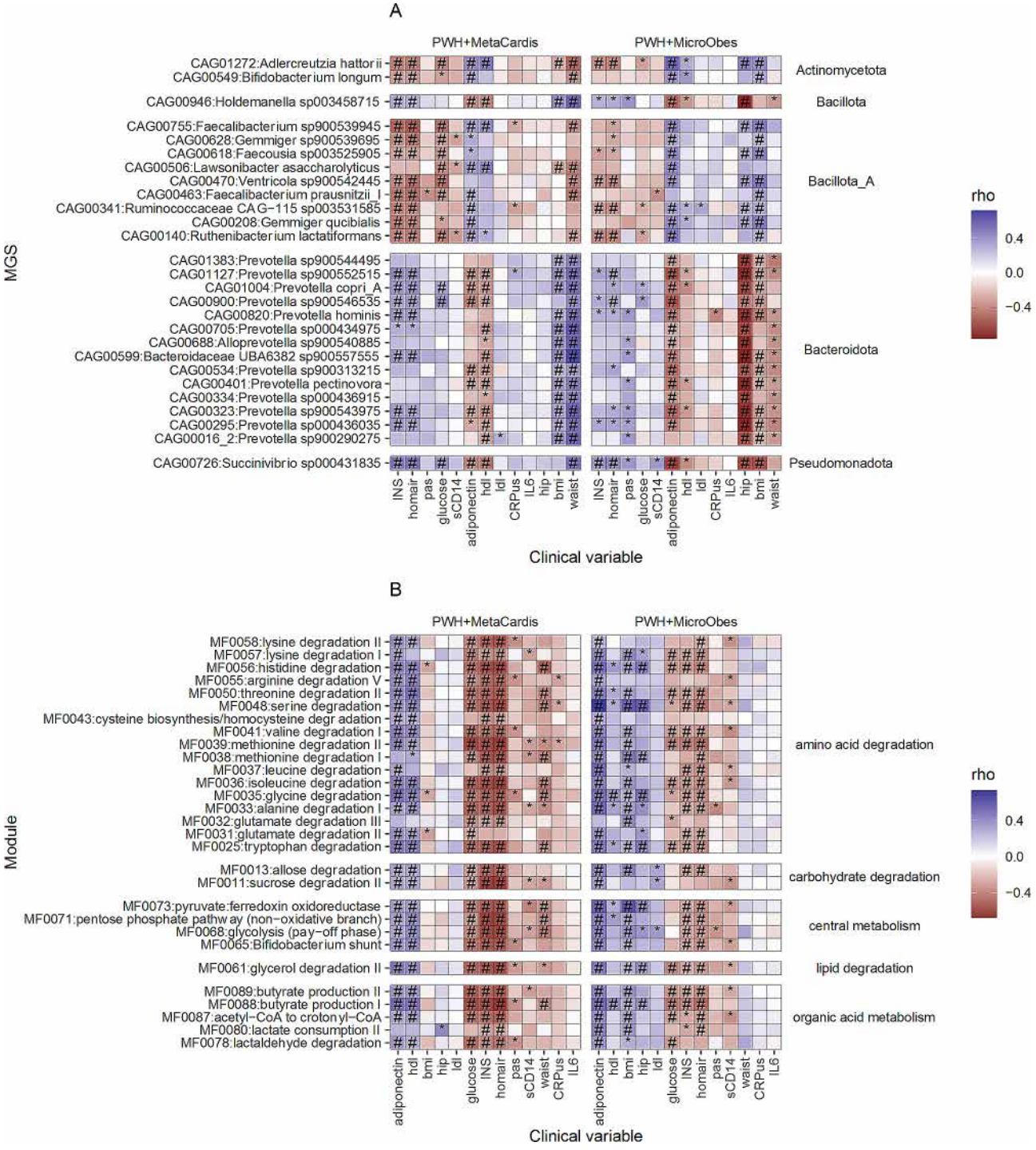

Fig. 4 Associations between metagenomic features and clinical variables in ROCnRAL sub-cohorts. (**A**) Heatmap of spearman correlations between 27 core MGS (y-axis) from Fig. 2A (significant differences in abundance between PWH ($n = 25$) and all three control groups) and clinical variables (y-axis) in PWH and MetaCardis individuals (left panel) and PWH and MicroObes individuals (right panel). #=FDR < 0.05, *=p-value < 0.05, Spearman correlation tests. (**B**) Same as panel A for 29 core Gut Metabolic Modules (GMM) in Fig. 3A

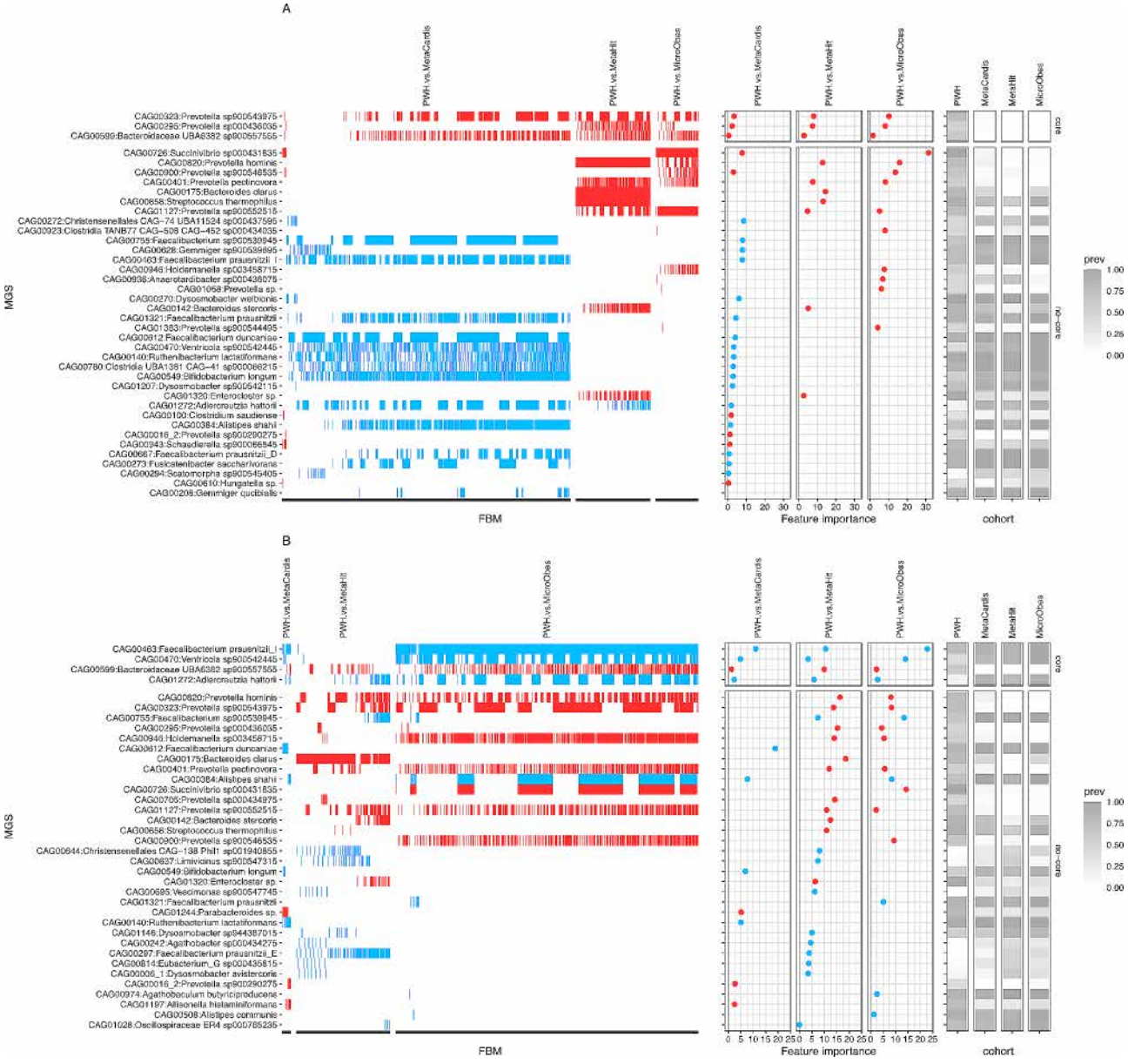

Fig. 5 Composition of Predomics Family of Best Models (FBM) in PWH classification vs. control groups. (**A**) Summary plot of the feature presence/absence (left panel), mean feature importance (center panel) and feature prevalence across study groups (right panel) for the 39 MGS (y-axis) retained in 718 bininter models included in the FBM in each prediction task (PWH vs. different control groups; $n = 511$ models in PWH vs. MetaCardis, 132 models in PWH vs. Metahit; 75 models in PWH vs. MicroObes). (**B**) Same as panel B for the 37 MGS retained in 669 terinter models included in the FBM in each prediction task (PWH vs. different control groups; $n = 14$ models in PWH vs. MetaCardis, 155 models in PWH vs. Metahit; 500 models in PWH vs. MicroObes). Blue and red MGS in model presence/absence and feature importance panels represent MGS with high mean abundances in control or PWH groups respectively. Core facet corresponds to MGS retained at least once in the FBM of each prediction task

PWH from MetaCardis controls included mainly MGS enriched in the MetaCardis controls, including multiple butyrate-producing lineages from the *Faecalibacterium*, *Gemmiger* and *Ruthenibacterium* genus (leftmost panel of Fig. 5A).

When focusing on the Predomics ternary models, 37 MGS were observed in the 669 FMB trained to classify PWH vs. each of the three control groups. These models combine groups of MGS enriched in PWH with groups of MGS enriched in the different control groups (red and blue MGS in the leftmost panel of Fig. 5B respectively). Among these MGS we continued to observe several of the core MGS described above in binary models, as well as notably *CAG01272: Adlercreutzia hattorii*, which is retained here in 21.42% 18.70% and 49.6% of the FBM of PWH vs. MetaCardis, MetaHit and MicroObes controls respectively (Fig. 5B). Also, we observed an interesting co-presence of *CAG00726: Succinivibrio sp000431835*

(enriched in PWH group) and *CAG00384: Alistipes shahii* (enriched in the control group) in the FBM issued from the PWH vs. MicroObes binary classification (co-present in 51.4% of FBM and co-absent in 48.2% of FBM), indicating potentially specific trade-offs between the abundances of these two MGS that changes between PWH and MicroObes controls (Fig. 5B).

Finally, we evaluated the performance of the best prediction models on different sets of unseen samples during the training process (testing sets of cross validation, see methods). The best binary models were extremely simple, with 3, 4 and 2 MGS to classify respectively PWH vs. MetaCardis, MetaHit and MicroObes controls, and retained core MGS retrieved in univariate analyses significantly enriched in PWH vs. the three control groups and robust to MSM status like *CAG00726 Succinivibrio sp000431835* and *CAG00323: Prevotella sp900543975*, being also the ones with the highest feature importance (Fig. 6A, left panel). Best models including *CAG00726: Succinivibrio sp000431835* performed at AUC of 100% when tested with 5 PWH and 5 control samples from each group that were excluded from the training process (internal holdouts), but generalizes poorly (AUC = 53–65) when applied to two external holdout datasets (study PRJNA692830: 13 PWH ART-controlled and 12 healthy controls; study PRJNA391226: 61 PWH and 10 healthy controls) (Fig. 6A, right panel). The best ternary models are equally sparse (3 MGS each), similar in composition to the binary ones in the inclusion of MGS enriched in the PWH group (*CAG00726: Succinivibrio*

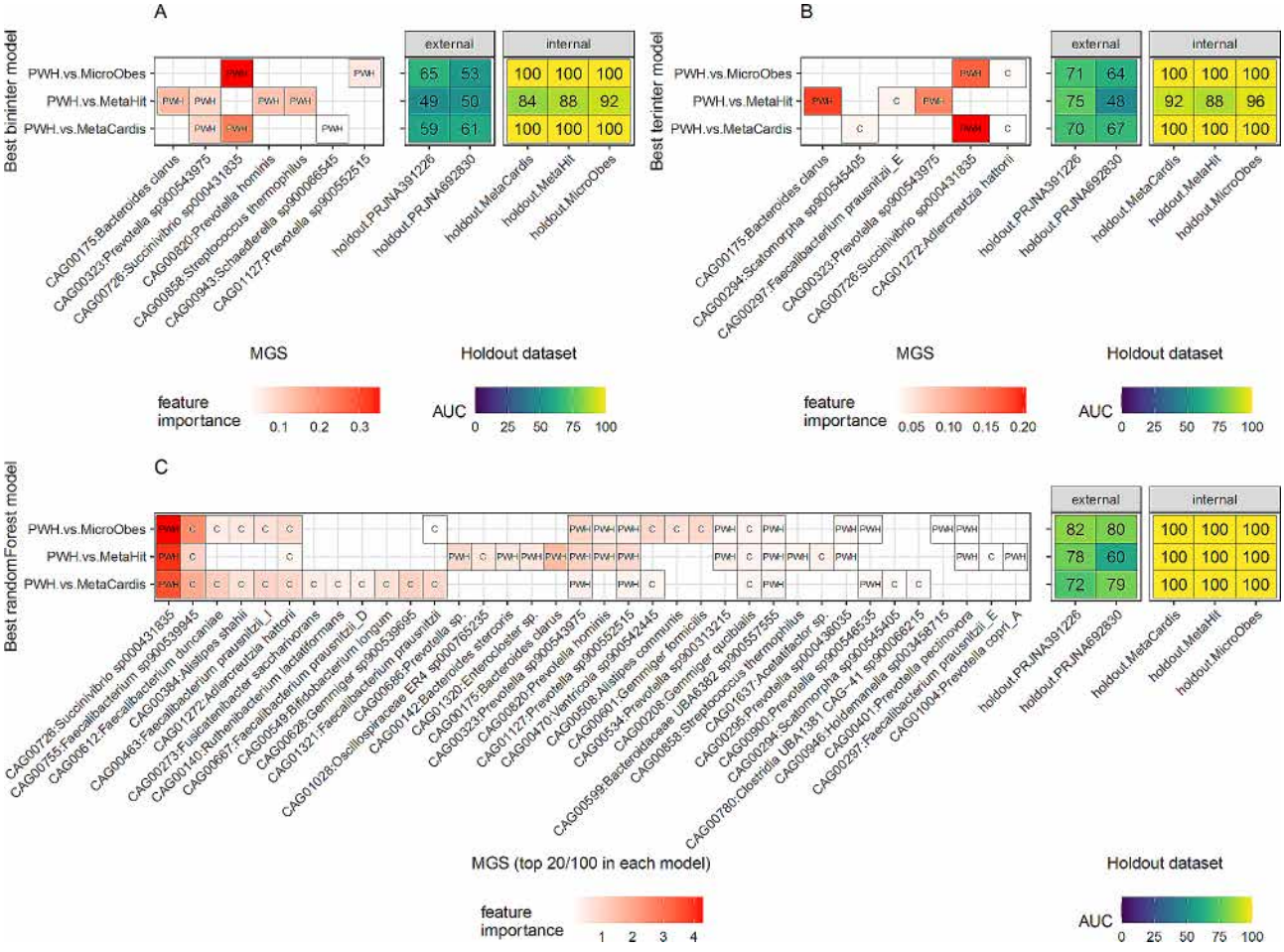

Fig. 6 Performance of best prediction models (binary and ternary Predomics models and random forest models) of PWH vs. control groups on hold-out datasets. (**A**) Best Predomics bininter models for the classification of PWH vs. the 3 control groups. Left panel represents the MGS (x-axis) included in each model (y-axis) with colors representing the feature importance (mean decrease accuracy) and labels representing the sign of the association between groups (PWH = increased in PWH group; C = increased in control group). Right panel represents the AUC of each best model vs. different hold-out datasets (x-axis). Internal holdouts correspond to 5 PWH and 5 control samples from each of the three control groups unseen during the training process. External holdouts correspond to PWH and control samples coming from two external quantitative metagenomics studies (PRJNA692830: 13 PWH ART-controlled and 12 healthy controls; PRJNA391226: 61 PWH and 10 healthy controls) for which MGS abundances in the IGC reference space has been generated (see methods for details). (**B**) and (**C**) represent the same results as (**A**) for the best Predomics terinter models and random forest models respectively. In panel **C**, the left panel is limited to the top 20 MGS with the highest GINI score in each model

sp000431835 and *CAG00323: Prevotella sp900543975)* but including also MGS enriched in all three control groups and robust to MSM status like *CAG01272: Adlercreutzia hattorii* (Fig. 6B, left panel). In terms of performance, models including *CAG00726: Succinivibrio sp000431835* and *CAG01272: Adlercreutzia hattorii* performed at AUC of 100% with internal holdout data, showing better generalization (AUC=64–71) than binary models on the two external holdout datasets (Fig. 6B, right panel).

Random forest models showed the best performance on the external holdout datasets, reaching AUC levels of 82.04% and 80.06% on the PRJNA391226 and PRJNA692830 studies respectively for the best model trained on PWH vs. MicroObes controls (Fig. 6C, right panel). However, these models are much more complex than those learned by predomics (n=100 MGS; number of trees=500; avg ± sd nodes=6.2±1.57, 5.78±1.52, 5.52±1.49 for random forest models of PWH vs. MetaCardis, MetaHit and MicroObes classification), but among the top 20 MGS with the highest contribution in terms of mean decrease Gini Score are commonly retained core MGS retrieved in univariate analyses with similar variations in PWH vs. the three control groups, robust to MSM status, and included in the sparse Predomics models like *CAG00726: Succinivibrio sp000431835*, *CAG00755:Faecalibacterium sp900539945*, *CAG00323: Prevotella sp900543975* and *CAG01272: Adlercreutzia hattorii.*

Overall, these results showed that the microbiome of PWH of this study harbors highly specific metagenomic signatures (*CAG00726: Succinivibrio sp000431835*, *CAG00755:Faecalibacterium sp900539945*, *CAG00323:Prevotella sp900543975* and *CAG01272: Adlercreutzia hattorii)* robustly identified in univariate analyses vs. the three control groups and robust to MSM status that are extremely powerful at predictive level but that generalizes poorly on external datasets by itself in the context of simple predictive models (Predomics models). However, the generalization on these external datasets improves markedly with more complex models (random forest), where these key MGS harbor the highest weight in terms of feature importance (Fig. 6C, left panel).

Discussion

In this study we show that the gut microbiome of 25 aging ART-controlled PWH compared to 3 independent groups of paired non-HIV infected controls, presents major alterations linked with insulin resistance and a dysmetabolic profile. First, a core set of metagenomic species, some of which belonging to the *Prevotella* genus, were robustly enriched in PWH vs. each of the 3 matched control groups. These markers are also related with worsened insulin resistance profiles. This signature does not seem to be explained by MSM sexual preferences alone as tested based on stratified analyses by this variable. Second, depleted butyrate-producing lineages of the microbiome could also be responsible for worsened metabolic profile and insulin resistance in PWH individuals, consistent across MSM sexual preferences. This defect is probably related to the HIV infection, even though all patients were controlled with ART modification. And third, machine learning methods uncovered robust predictive signatures, consistent with univariate analyses, which generalized relatively well in two additional external cohorts. This indicates the specific impact of the HIV infection in the microbiome ecosystem.

An initial limitation of this study was that only PWH were recruited, so to characterize the specific microbial signatures in the microbiome of PWH we capitalized on the availability of quantitative metagenomic profiles from three studies (MetaHit, MicroObese, MetaCardis). These external cohorts included individuals from different countries (French in MicroObese and MetaCardis; Danish in MetaHit), with variations in corpulence profiles (non-obese in MetaCardis and MetaHit; moderately obese and obese in MicroObes), and with metagenomic data generated with different sequencing technologies (SOLiD in MicroObese, Illumina in MetaHit, IonTorrent in MetaCardis). These as well as other technical and clinical/lifestyle covariates (stool collection methods, library preparation protocol, nutritional profiles) could have a major impact on the metagenomic profiles and derived signatures associated to study groups in quantitative metagenomic studies [55–57].

The reference databases may also play a major role in quantitative metagenomic profiling [58, 59], although important effort is put to harmonize metagenomic data from different studies into common reference space [60]. Moreover, different library preparation methods and sequencing technologies may have an important effect on the final abundance profiles. All these different aspects would be impossible to de-confound in this study if a single control group was included. Indeed, the integration of our PWH data with the metagenomic data in three control groups on the same reference space allowed to identify a core set of taxonomic and functional signatures that characterizes the microbiome of PWH with robust variations across all three control groups. We cannot exclude that other potential cofounder factors not accounted for in our study like diet, lifestyle, ethnicity or and medication adherence could impact the gut microbiome composition and metabolic parameters in PWH. These factors could explain the decrease in performance of predictive models (Predomics and Random Forest models) fitted on classification PWH from controls with our data when applied on two independent external studies, particularly

of simple models (Predomics) that penalizes larger models and as consequence contains a limited number of species (2–4 MGS). However, we observed that complex models derived from state-of-the-art machine learning approaches (Random Forest) allowed to correctly classify the samples of PWH vs. healthy controls in these independent external studies (80–82% AUC), showing better generalization of the microbiome signatures identified in our PWH.

Among these signatures we observed a strong enrichment of *Prevotella* lineages in PWH that is reflected at both MGS and enterotype levels and that constitutes a signature that has already been identified as enriched in the microbiome of PWH in other studies [22] but that has been also shown as potentially cofounded by the sexual preferences [23]. Here we lack information for this variable in the control groups, but stratified analyses on PWH by sexual preference shows a non-significant tendency towards higher enrichment of Prevotella enterotype in PWH of the MSM group (Supplemental Fig. 2) that would be in line with previous reports. However, we also observe that the increase of several core MGS of the *Prevotella* genus in the PWH was robust to sexual preferences (Fig. 2B, Supplemental Fig. 4).

In addition to these well-known signatures, our study identified other MGS that were highly specific of the microbiome of PWH and robust to MSM status like notably *CAG00726: Succinivibrio sp000431835*, an MGS that was predominantly absent from the three control groups and that showed the highest importance in the classification models of PWH vs. different control groups. In this context, a longitudinal study of PWH with 16 S profiling showed drastic increase in the *Succinivibrio* genus under conventional ART therapy [61], and a 16 S-based cross-sectional study showed also that *Succinivibrio* genus abundance was specifically associated to treatment- naïve PWH specifically infected by HIV-C subtype62. On the other side of the balance, the depletion of butyrate-producing firmicutes lineages of the Bacillota_A phyla, notably from the *Faecalibacterium* genus and evolutionary close lineages like *Gemmiger and Ruthenibacterium* [63], that were described in previous studies characterizing the microbiome of PWH [22, 25] were also reproduced in our cohort. This includes MGS like *CAG00755: Faecalibacterium SP900539945*, *CAG00463: Faecalibacterium prausnitzii_I or CAG00208: Gemmiger qucibialis* included in the core MGS consistently depleted in PWH in comparison to the three control groups and robust to MSM preferences (Fig. 2). But we also observed the near complete absence of *CAG01272: Adlercreutzia hattorii* in the microbiome of PWH individuals (MSM and no-MSM). This lineage was recently characterized as a close relative (93% ANI values) of *Adlercreutzia equolifaciens* [64], a bacterium with anti-inflammatory properties both in vitro and in vivo in a humanized mouse model of NAFLD whose abundance decrease with the severity of hepatic disease [65]. In the context of PWH, 16 S metabarcoding studies have previously shown a depletion of *Adlercreutzia* genus in women with HIV infection with or at high risk of HIV having developed carotid artery atherosclerosis [66] as well as a consistent depletion in women with HIV with and without T2D in comparison with non-HIV and non-T2D controls [29]. However, a recent study also shown that *Adlercrutzia equolifaciens* abundance was significantly reduced in the context of T2D in women with and without HIVS from the Women's Interagency HIV Study (WIHS), being included as signature of gut dysbiosis associated to T2D independent of HIV infection [28]. Here, thanks to the availability of shotgun data and precise taxonomic annotation based on the Genome Taxonomy DataBase42, we report to our knowledge the first evidence of the near complete absence *Adlercreutzia hattorii* in the microbiome of PWH, which could be associated to the inflammatory profile and altered metabolic status of PWH.

In the present study our focus has been the comprehensive profiling of the microbial gut ecosystem of PWH by using a comprehensive non-redundant gene catalog of human gut microbes (IGC) [35], which has the drawback of not covering other taxonomic groups relevant in the context of PWH etiology. Opportunistic infections (OIs) by a variety of bacteria, viruses, fungi and protozoan are common in individuals with highly immunocompromised status due to advanced disease or naïve to ART [67], and co-infection with Hepatitis C and B viruses (HBV, HCV) are also common among PWH ([14] and references therein). In addition to prokaryotic communities, fungal communities has been shown altered in PWH not only on the gut but on other human ecological niches like notably the oral cavity or the respiratory tract ([68] and references therein). In this context, the IGC catalog only includes 4 non-prokaryotic metagenomic species belonging to Blastocystis genus, a common intestinal protist lineage in human feces that has been punctually studied in relation with HIV infection ([69–73]), and for which we found no significant differences between PWH and none of the control groups (FDR > 0.05, Wilcoxon-rank sum tests; Supplemental Table 9). Future studies with more taxonomically broad reference catalogs as well as direct metagenomic assembly on large cohorts of PWH individuals will be needed to fully evaluate the contribution of non-prokaryotic communities of the human gut to PWH clinical profile.

Our study strongly suggests that insulin resistance can be partly related to the HIV status and partly to sexual preference in MSM, a highly represented population in Caucasian PWH in Western countries, but this latter point has not been precisely addressed. First, a high

level of insulin resistance has been commonly reported in PWH as compared to the general population. In the general population, the average HOMA-IR was 1.1 in individuals with normal BMI and low body fat and of 1.6 in those with high body fat [6]. The HOMA-IR value was markedly higher in the Modena cohort in Italy: among 2000 ART-controlled PWH, 68% men, aged in median 45 years old and with a median BMI of 23 kg/m^2, median HOMA-IR was 2.4 (IQR 1.4-4) and half patients were considered as insulin-resistant (HOMA-IR>2) [5]. In the TANGO trial including more than 700 ART-controlled PWH, 92% men, 79% Caucasian, with a BMI of 26, median HOMA-IR was 2.6 with about 70% being insulin-resistant [74]. Second, the role played by the sexual preference in the level of insulin resistance in PWH has not been directly explored. Interestingly, in MSM from the MACS cohort, HOMA-IR was high in non-HIV infected MSM aged 44.9 years at 2.9 for a BMI of 26.3 kg/m^2, and was significantly increased in MSM PWH aged 40.6 years at 3.3 for a significantly lower BMI at 25.4 kg/m^2 [7], strongly suggesting that sexual preference in MSM could enhance insulin resistance. Accordingly, a high level of HOMA-IR was not found in the corresponding WIHS cohort of women infected or not by HIV, with an HOMA-IR of 2.0 for a BMI of 29.4 kg/m^2 in non-infected women and a HOMA-IR of 1.9 for a BMI of 26.2 kg/m^2 for women PWH [75]. Thus, elevated HOMA-IR in MSM PWH could result both from sexual preference and from HIV infection.

Here, in 25 PWH with a mean BMI of 25 kg/m^2 we observed a very high level of insulin-resistance (mean HOMA-IR 5.1) that is not clearly explained by BMI or metabolic disorders. Moreover, HOMA-IR was higher in the MSM vs. the non-MSM group. We propose that altered microbiota, with increased *Prevotella*, which could be influenced by sexual preference (we observed a non-significant potential enrichment of Prevotella enterotype in PWH-MSM group; Supplemental Fig. 3), and decreased species producing butyrate, which could be linked to HIV infection, could both play a role in the insulin-resistant profile observed in PWH.

The relationship between gut dysbiosis and insulin resistance has been poorly addressed in PWH. A decreased level of butyrate-producing bacteria has been reported in several studies [10, 12] being independent on sexual preference [12, 24] with no reference to insulin resistance. Two studies evaluated gut microbiota in PWH with or without metabolic syndrome, in the absence of non-infected controls. They reported that *Faecalibacterium* was reduced in those with a metabolic syndrome [25] and was related to the liver fat index [26]. In patients with diabetes, one study evaluating women with HIV and diabetes reported a marginal decrease in *Faecalibacterium* as compared to non-diabetics ($p=0.07$) [29]. However, in another study evaluating PWH and non-infected controls, with or without diabetes, the overall capacity for butyrate metabolism as predicted by a PICRUSt analysis did not differ between the groups [27]. Conversely, such bacteria have been negatively linked to prediabetes and also with diabetes in non-infected individuals (for a review see [76]) in most studies but a positive relationship between *Faecalibacterium* and HOMA-IR has been recently reported in non-diabetic subjects with overweight or obesity [77]. In the present study, we present the clear link between their major reduction in PWH as compared to the three control groups and higher insulin resistance. This is translated at the functional level into an overall depletion of butyrate production potential in the microbiome of PWH, independent of sexual preferences. In this context, a recent study has shown that this depletion in the butyrate production potential of the microbiome of PWH was not reflected into systemic nor fecal levels of butyrate determined by metabolomics experiments, but that a depletion of propionate derived from microbiome lactate consumption was better reflected at both metagenomic and metabolomic level as signature of ART-treated PWH, with levels that decreases in PWH preceding morbidity and mortality [12]. Interestingly, among the functional modules significantly depleted in PWH vs. the three control groups we observed a signature of lactate consumption, robust to MSM preferences only in comparisons vs. the MicroObes control cohort (MF0080: Lactate consumption II; Fig. 3). We also report an association between several modules linked to amino-acid degradation and insulin sensitivity, although such an association has been previously reported in non-diabetic overweight or obese subjects [77].

Regarding microbial diversity, we observed a tendency to a decreased microbial gene richness in PWH vs. the MetaCardis and MetaHit group, but a significant decrease was only observed vs. the MicroObes group. We cannot exclude that potential biases associated with different sequencing technologies could influence these results, even if data was properly downsized to same equal sequencing depth and normalized. In previous studies with PWH, discordant results have been reported regarding microbial diversity, being either decreased or not [9]. This discrepancy could be also related to sexual preference, MSM individuals presenting higher diversity compared with non-MSM individuals [23, 24]. Our stratified analyses of microbial gene richness between MSM and non-MSM individuals revealed no significant difference between both groups. The results on diversity appear to be counterbalanced between sexual preference and HIV infection.

We observed that the module accounting for LPS production was higher in PWH as compared to controls.

However, when considering the MSM and non-MSM groups, LPS production was higher in the MSM group but not in the non-MSM group as compared to the 3 control groups. This suggests that higher LPS production, in favor of a pro-inflammatory profile, could be linked to the *Prevotella* enrichment in the MSM group, in good accordance with previous studies reporting the pro-inflammatory profile of *Prevotella* [10]. As well, we do not observe a strong association between tested inflammatory markers (IL-6, CRP) and MGS abundances in our group of PWH.

We were also able to analyze the evolution of gut microbiome according to ART. In fact, we analyzed 7 ART-controlled PWH first when talking ART including a protease inhibitor and second after a mean duration of 17 weeks when taking a dual therapy including an integrase strand transfer inhibitor (RAL) and a CCR5 inhibitor (MVC). While this shift was associated with a major modification in the profile of genes expressed in subcutaneous adipose tissue [78], with a profile indicating higher insulin resistance and decreased T lymphocytes activation, we did not observe major differences in the gut microbiome profile. In particular, the level of butyrate-producing bacteria was not modified, in accordance with the unmodified HOMA-IR index. While several studies compared gut dysbiosis before and after ART initiation and indicated the persistence of dysbiosis [9, 10], only a few studies evaluated gut microbiome modifications according to the nature of ART. Lower microbiota richness as compared to HIV-negative controls was observed in PWH receiving protease inhibitors while the richness was similar in those receiving integrase strand transfer inhibitors [79], but the sexual preference was not taken into consideration. The effect of maraviroc on gut microbiota has not been previously evaluated in PWH, so we cannot exclude that the 17 weeks period is not enough to observe a significant alteration of the gut microbiome after ART switch. Importantly, we analyzed the same subjects twice. Nevertheless, the low number of PWH in our study precludes the identification of small differences related to ART. More comprehensive longitudinal datasets would be needed to fully evaluate the impact of RAL/MVC therapy on gut microbiome composition.

Conclusion

We identified major modifications in gut microbiome in PWH when compared to three independent groups of paired non-infected individuals robust to sexual preferences and that remains unaltered 17 weeks after switching to RAL/MVC therapy. Among these signatures we confirm previously described alterations in the microbiome of PWH like strong enrichment of Prevotella lineages in parallel with a decrease of butyrate-producing lineages, together with novel and highly specific signatures of the microbiome of PWH like the absence of *Adlercrutzia equolifaciens/hattorii* or a strong enrichment of a metagenomic specie of *Succinivibrio* genus which show a strong predictive power in random forest models of classification that generalizes to AUC levels of 80–82% on two external cohorts of PWH/Control individuals. This is translated to an overall depletion of the potential for butyrate production and amino-acid degradation of the microbiome that robustly correlates with an altered metabolic status, suggesting that besides BMI, marked insulin resistance in PWH could result both from enrichment in Prevotella lineages, and from the depletion in species producing butyrate and involved into amino-acid degradation. PWH currently encounter a higher prevalence of cardiometabolic complications as they age than paired non-infected controls. Our study suggests that dysbiosis may play a role in this status. Reversing dysbiosis may be an important goal to reduce cardiometabolic burden. Future studies with large sample sizes and more extensive phenotyping of PWH and control individuals in terms of lifestyle, ethnicity, or diet would be needed to confirm this hypothesis towards a better understanding of the role of gut microbiome in the interplay of HIV infection with metabolic health of individuals.

Abbreviations

PWH	Persons living with HIV
ART	Antiretroviral-treatment
BMI	Body mass index
MSM	Men having sex with men
RAL	Raltegravir
MVC	Maraviroc
IGC	Integrated human Gut microbiome Catalog
MGS	Metagenomic species
GMMs	Gut Metabolic Modules
DMM	Dirichlet Multinomial Mixture
dbRDA	Distance-based redundancy analyses
FBM	Family of best models
MDA	Mean decrease accuracy

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-024-01978-5.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

Not applicable.

Author contributions

JC, JPB and EP designed the study. JC and JPB collected clinical information, drafted and reviewed the paper. EP and EB performed bioinformatic and statistical analyses, interpretation, made the graphics, drafted and reviewed the paper. JD extracted stool material from PWH from the ROCnRAL study and BQ performed sequencing. NP, FPO and ELC performed the bioinformatic analyses and reviewed the paper. CK drafted the protocol and was the PI for the ROCnRAL study. CK, CG, SJ and OL recruited PWH individuals. RA and KC contributed to cohort recruitment and reviewed the paper. SF and JPB

performed the metabolic and inflammatory blood samples analyses. JDZ and BF contributed to the manuscript redaction. All authors read and approved the final manuscript.

Funding
The study was funded in part by ANRS (ANRS 157 study trial).

Data availability
Raw sequence data of samples from PWH newly sequenced in this study are available on the French national data portal (https://entrepot.recherche.data.gouv.fr) under the DOI https://doi.org/10.57745/UOVSPM.. Microbial gene richness, enterotype assignment, study group (PWH, MicroObes, MetaCardis, MetaHit) and sampling point (for PWH) are available in Supplemental Table 2. Source MGS, GMM and KEGG module abundance tables are available in Suppelemental Tables 3, 4 and 5. Clinical data is available upon request.

Declarations

Ethics approval and consent to participate
ROCnRAL ANRS-157 study trial was approved by ANRS and CPP Ile de France VI (ANRS 157 study trial). All patients gave their informed consent to the protocol.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Human ethics and consent to participate
The Institutional Review Board of Pitié-Salpêtrière Hospital approved the study protocol (ClinicalTrials.gov: NCT01420523, first trial registration: 18/08/2011). The reference protocol (EudraCT 2011-002483-24 protocol 3.0 16/01/2012) and the amendment for the microbiote study were approved 29/02/2012 by AFSSAPS (Agence Française de Securité Sanitaire des produits de Santé) and DEMEB (Département de l'évaluation des essais cliniques et des médicaments à statut particulier). All patients provided written informed consent.

Author details
[1]IRD, Sorbonne Université, Unité de Modélisation Mathématique et Informatique des Systèmes Complexes, UMMISCO, Bondy F-93143, France
[2]Sorbonne Université, INSERM, Nutrition et Obesities, Systemic Approaches, NutriOmique, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France
[3]INSERM UMR_S938, Centre de Recherche Saint-Antoine, Institut Hospitalo-Universitaire de Cardio-Métabolisme et Nutrition (ICAN), Sorbonne Université, Paris F-75012, France
[4]Université Paris-Saclay, INRAE, MGP, Jouy-en-Josas 78350, France
[5]INSERM, Institut Pierre Louis d'Épidémiologie et de Santé Publique, AP-HP, Pitié Salpétrière Hospital, Department of Infectious Diseases, Sorbonne Université, Paris F-75013, France
[6]AP-HP, Hôpital Bicêtre, Service de Médecine Interne et Immunologie Clinique et Service de Maladies Infectieuses et Transmissibles, Kremlin-Bicêtre, France
[7]Université Paris Saclay, Inserm, CEA, UMR1184, Le Kremlin Bicêtre, France
[8]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, Jouy-en-Josas 78350, France
[9]Assistance Publique-Hôpitaux de Paris, Hôpitaux Universitaires Henri Mondor, Département de biochimie-pharmacologie, FHU-SENEC, INSERM U955 and Université Paris Est (UPEC), UMR U955, Faculté de Santé, Créteil F-93010 cedex, France

Received: 10 April 2024 / Accepted: 1 August 2024
Published online: 13 August 2024

References

1. Schouten J, Wit FW, Stolte IG, Kootstra NA, van der Valk M, Geerlings SE, et al. Cross-sectional comparison of the prevalence of age-associated comorbidities and their risk factors between HIV-infected and uninfected individuals: the AGEhIV cohort study. Clin Infect Dis. 2014;59:1787–97.
2. Pedro MN, Rocha GZ, Guadagnini D, Santos A, Magro DO, Assalin HB et al. Insulin resistance in HIV-Patients: causes and consequences. Front Endocrinol. 2018;9.
3. Lagathu C, Béréziat V, Gorwood J, Fellahi S, Bastard J-P, Vigouroux C, et al. Metabolic complications affecting adipose tissue, lipid and glucose metabolism associated with HIV antiretroviral treatment. Expert Opin Drug Saf. 2019;18:829–40.
4. Bastard J-P, Couffignal C, Fellahi S, Bard J-M, Mentre F, Salmon D, et al. Diabetes and dyslipidaemia are associated with oxidative stress independently of inflammation in long-term antiretroviral-treated HIV-infected patients. Diabetes Metab. 2019;45:573–81.
5. Milic J, Renzetti S, Ferrari D, Barbieri S, Menozzi M, Carli F, et al. Relationship between weight gain and insulin resistance in people living with HIV switching to integrase strand transfer inhibitors-based regimens. AIDS. 2022;36:1643–53.
6. Martinez KE, Tucker LA, Bailey BW, LeCheminant JD. Expanded normal weight obesity and insulin resistance in US adults of the National Health and Nutrition Examination Survey. J Diabetes Res. 2017;2017:9502643.
7. Longenberger A, Lim JY, Brown TT, Abraham A, Palella FJ, Effros RB, et al. Low physical function as a risk factor for incident diabetes mellitus and insulin resistance. Future Virol. 2011;6:439–49.
8. Shah ASV, Stelzle D, Lee KK, Beck EJ, Alam S, Clifford S, et al. Global Burden of Atherosclerotic Cardiovascular Disease in people living with HIV: systematic review and Meta-analysis. Circulation. 2018;138:1100–12.
9. Gootenberg DB, Paer JM, Luevano J-M, Kwon DS. HIV-associated changes in the enteric microbial community: potential role in loss of homeostasis and development of systemic inflammation. Curr Opin Infect Dis. 2017;30:31–43.
10. Dillon SM, Frank DN, Wilson CC. The gut microbiome and HIV-1 pathogenesis: a two-way street. AIDS. 2016;30:2737–51.
11. Vujkovic-Cvijin I, Somsouk M. HIV and the gut microbiota: composition, consequences, and avenues for amelioration. Curr HIV/AIDS Rep. 2019;16:204–13.
12. Sereti I, Verburgh ML, Gifford J, Lo A, Boyd A, Verheij E et al. Impaired gut microbiota-mediated short-chain fatty acid production precedes morbidity and mortality in people with HIV. Cell Rep. 2023;0.
13. Gelpi M, Vestad B, Hansen SH, Holm K, Drivsholm N, Goetz A, et al. Impact of human immunodeficiency virus–related gut microbiota alterations on metabolic comorbid conditions. Clin Infect Dis. 2020;71:e359–67.
14. Gobran ST, Ancuta P, Shoukry NH. A tale of two viruses: immunological insights into HCV/HIV coinfection. Front Immunol. 2021;12:726419.
15. Limper AH, Adenis A, Le T, Harrison TS. Fungal infections in HIV/AIDS. Lancet Infect Dis. 2017;17:e334–43.
16. Li S, Yang X, Moog C, Wu H, Su B, Zhang T. Neglected mycobiome in HIV infection: alterations, common fungal diseases and antifungal immunity. Front Immunol. 2022;13:1015775.
17. Mukherjee PK, Chandra J, Retuerto M, Sikaroodi M, Brown RE, Jurevic R, et al. Oral mycobiome analysis of HIV-Infected patients: identification of Pichia as an antagonist of opportunistic Fungi. PLoS Pathog. 2014;10:e1003996.
18. Chang S, Guo H, Li J, Ji Y, Jiang H, Ruan L, et al. Comparative analysis of salivary mycobiome diversity in human immunodeficiency virus-infected patients. Front Cell Infect Microbiol. 2021;11:781246.
19. Cui L, Lucht L, Tipton L, Rogers MB, Fitch A, Kessinger C, et al. Topographic diversity of the respiratory tract mycobiome and alteration in HIV and lung disease. Am J Respir Crit Care Med. 2015;191:932–42.
20. Bittinger K, Charlson ES, Loy E, Shirley DJ, Haas AR, Laughlin A, et al. Improved characterization of medically relevant fungi in the human respiratory tract using next-generation sequencing. Genome Biol. 2014;15:487.
21. Gosalbes MJ, Jimenéz-Hernandéz N, Moreno E, Artacho A, Pons X, Ruíz-Pérez S et al. Interactions among the mycobiome, bacteriome, inflammation, and diet in people living with HIV. Gut Microbes. 14:2089002.
22. Vázquez-Castellanos JF, Serrano-Villar S, Latorre A, Artacho A, Ferrús ML, Madrid N, et al. Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. Mucosal Immunol. 2015;8:760–72.
23. Noguera-Julian M, Rocafort M, Guillén Y, Rivera J, Casadellà M, Nowak P, et al. Gut microbiota linked to sexual preference and HIV infection. EBioMedicine. 2016;5:135–46.
24. Vujkovic-Cvijin I, Sortino O, Verheij E, Sklar J, Wit FW, Kootstra NA, et al. HIV-associated gut dysbiosis is independent of sexual practice and correlates with noncommunicable diseases. Nat Commun. 2020;11:2448.
25. Villanueva-Millán MJ, Pérez-Matute P, Recio-Fernández E, Lezana Rosales J-M, Oteo J-A. Characterization of gut microbiota composition in HIV-infected patients with metabolic syndrome. J Physiol Biochem. 2019;75:299–309.

26. Amador-Lara F, Andrade-Villanueva JF, Vega-Magaña N, Peña-Rodríguez M, Alvarez-Zavala M, Sanchez-Reyes K, et al. Gut microbiota from Mexican patients with metabolic syndrome and HIV infection: an inflammatory profile. J Appl Microbiol. 2022;132:3839–52.
27. Hoel H, Hove-Skovsgaard M, Hov JR, Gaardbo JC, Holm K, Kummen M, et al. Impact of HIV and Type 2 diabetes on Gut Microbiota Diversity, Tryptophan Catabolism and endothelial dysfunction. Sci Rep. 2018;8:6725.
28. Luo K, Peters BA, Moon J-Y, Xue X, Wang Z, Usyk M, et al. Metabolic and inflammatory perturbation of diabetes associated gut dysbiosis in people living with and without HIV infection. Genome Med. 2024;16:59.
29. Moon J-Y, Zolnik CP, Wang Z, Qiu Y, Usyk M, Wang T, et al. Gut microbiota and plasma metabolites associated with diabetes in women with, or at high risk for, HIV infection. EBioMedicine. 2018;37:392–400.
30. Katlama C, Assoumou L, Valantin M-A, Soulié C, Martinez E, Béniguel L, et al. Dual therapy combining raltegravir with etravirine maintains a high level of viral suppression over 96 weeks in long-term experienced HIV-infected individuals over 45 years on a PI-based regimen: results from the phase II ANRS 163 ETRAL study—authors' response. J Antimicrob Chemother. 2020;75:3699–700.
31. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. Nature. 2013;500:541–6.
32. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BAH, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. Nature. 2016;535:376–81.
33. Vieira-Silva S, Falony G, Belda E, Nielsen T, Aron-Wisnewsky J, Chakaroun R, et al. Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. Nature. 2020;581:310–5.
34. Cotillard A, Kennedy SP, Kong LC, Prifti E, Pons N, Le Chatelier E, et al. Dietary intervention impact on gut microbial gene richness. Nature. 2013;500:585–8.
35. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol. 2014;32:834–41.
36. Sekhon JS. Multivariate and propensity score matching Software with Automated Balance optimization: the matching package for R. J Stat Softw. 2011;42:1–52.
37. Pons N. METEOR, a platform for quantitative metagenomic profiling of complex ecosystems. Journees Ouvertes en Biologie, Informatique et Mathematiques.
38. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.
39. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, et al. Alterations of the human gut microbiome in liver cirrhosis. Nature. 2014;513:59–64.
40. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol. 2014;32:822–8.
41. Prifti E, Pons N, LeChatelier E, Almeida M, Leonard P, Batto J-M, et al. MetaOMineR: a fine-tuned pipeline for whole metagenomic data analyses. 4 ed. China: International Human Microbiome Congress, Hangzhou; 2013.
42. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022;50:D785–94.
43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
44. Kultima JR, Coelho LP, Forslund K, Huerta-Cepas J, Li SS, Driessen M, et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. Bioinformatics. 2016;32:2520–3.
45. Darzi Y, Falony G, Vieira-Silva S, Raes J. Towards biome-specific analysis of meta-omics data. ISME J. 2016;10:1025–8.
46. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS ONE. 2012;7:e30126.
47. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473:174–80.
48. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR et al. vegan: Community Ecology Package. 2022.
49. Bai X, Narayanan A, Nowak P, Ray S, Neogi U, Sönnerborg A. Whole-genome metagenomic analysis of the gut microbiome in HIV-1-Infected individuals on antiretroviral therapy. Front Microbiol. 2021;12.
50. Lu W, Feng Y, Jing F, Han Y, Lyu N, Liu F et al. Association between Gut Microbiota and CD4 recovery in HIV-1 infected patients. Front Microbiol. 2018;9.
51. Coelho LP, Alves R, Monteiro P, Huerta-Cepas J, Freitas AT, Bork P. NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language. Microbiome. 2019;7:84.
52. Prifti E, Chevaleyre Y, Hanczar B, Belda E, Danchin A, Clément K et al. Interpretable and accurate prediction models for metagenomics data. Gigascience. 2020;9.
53. Beasley TM, Schumacker RE. Multiple regression Approach to analyzing contingency tables: Post Hoc and Planned Comparison procedures. J Experimental Educ. 1995;64:79–93.
54. Costea PI, Hildebrand F, Arumugam M, Bäckhed F, Blaser MJ, Bushman FD, et al. Enterotypes in the landscape of gut microbial community composition. Nat Microbiol. 2018;3:8–16.
55. Voigt AY, Costea PI, Kultima JR, Li SS, Zeller G, Sunagawa S, et al. Temporal and technical variability of human gut metagenomes. Genome Biol. 2015;16:73.
56. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol. 2017;35:1069–76.
57. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. Nature. 2020;587:448–54.
58. Méric G, Wick RR, Watts SC, Holt KE, Inouye M. Correcting index databases improves metagenomic studies. 2019;:712166.
59. Alili R, Belda E, Le P, Wirth T, Zucker J-D, Prifti E, et al. Exploring semi-quantitative metagenomic studies using Oxford Nanopore sequencing: a computational and experimental protocol. Genes (Basel). 2021;12:1496.
60. Pasolli E, Schiffer L, Manghi P, Renson A, Obenchain V, Truong DT, et al. Accessible, curated metagenomic data through ExperimentHub. Nat Methods. 2017;14:1023–4.
61. Imahashi M, Ode H, Kobayashi A, Nemoto M, Matsuda M, Hashiba C, et al. Impact of long-term antiretroviral therapy on gut and oral microbiotas in HIV-1-infected patients. Sci Rep. 2021;11:960.
62. do Nascimento WM, Machiavelli A, Ferreira LGE, Cruz Silveira L, de Azevedo SSD, Bello G, et al. Gut Microbiome profiles and Associated Metabolic pathways in HIV-Infected Treatment-Naïve patients. Cells. 2021;10:385.
63. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning Age, Geography, and Lifestyle. Cell. 2019;176:649–e66220.
64. Sakamoto M, Ikeyama N, Yuki M, Murakami T, Mori H, Iino T, et al. Adlercreutzia hattorii sp. nov., an equol non-producing bacterium isolated from human faeces. Int J Syst Evol MicroBiol. 2021;71:005121.
65. Oñate FP, Chamignon C, Burz SD, Lapaque N, Monnoye M, Philippe C, et al. Adlercreutzia equolifaciens is an anti-inflammatory commensal bacterium with decreased abundance in gut microbiota of patients with metabolic liver disease. Int J Mol Sci. 2023;24:12232.
66. Wang Z, Peters BA, Usyk M, Xing J, Hanna DB, Wang T et al. Gut microbiota, plasma metabolomic profiles, and carotid artery atherosclerosis in HIV infection. Arteriosclerosis, thrombosis, and Vascular Biology. 2022;42:1081–93.
67. Justiz Vaillant AA, Naik R. In: StatPearls, editor. HIV-1–Associated opportunistic infections. Treasure Island (FL): StatPearls Publishing; 2024.
68. Hager CL, Ghannoum MA. The mycobiome in HIV. Curr Opin HIV AIDS. 2018;13:69–72.
69. Albrecht H, Stellbrink HJ, Koperski K, Greten H. Blastocystis Hominis in human immunodeficiency virus-related diarrhea. Scand J Gastroenterol. 1995;30:909–14.
70. Di Cristanziano V, D´Alfonso R, Berrilli F, Sarfo FS, Santoro M, Fabeni L, et al. Lower prevalence of Blastocystis sp. infections in HIV positive compared to HIV negative adults in Ghana. PLoS ONE. 2019;14:e0221968.
71. Fontanelli Sulekova L, Gabrielli S, Furzi F, Milardi GL, Biliotti E, De Angelis M, et al. Molecular characterization of Blastocystis subtypes in HIV-positive patients and evaluation of risk factors for colonization. BMC Infect Dis. 2019;19:876.
72. Xu N, Jiang Z, Liu H, Jiang Y, Wang Z, Zhou D, et al. Prevalence and genetic characteristics of Blastocystis Hominis and Cystoisospora Belli in HIV/AIDS patients in Guangxi Zhuang Autonomous Region, China. Sci Rep. 2021;11:15904.
73. Zhang S-X, Kang F-Y, Chen J-X, Tian L-G, Geng L-L. Risk factors for Blastocystis infection in HIV/AIDS patients with highly active antiretroviral therapy in Southwest China. Infect Dis Poverty. 2019;8:89.
74. van Wyk J, Ajana F, Bisshop F, De Wit S, Osiyemi O, Portilla Sogorb J, et al. Efficacy and safety of switching to Dolutegravir/Lamivudine fixed-dose 2-Drug regimen vs continuing a Tenofovir Alafenamide-based 3- or 4-Drug regimen

for maintenance of Virologic Suppression in adults living with human immunodeficiency virus type 1: phase 3, Randomized, Noninferiority TANGO Study. Clin Infect Dis. 2020;71:1920–9.
75. Hessol NA, Ameli N, Cohen MH, Urwin S, Weber KM, Tien PC. The association between diet and physical activity on insulin resistance in the women's interagency HIV Study. J Acquir Immune Defic Syndr. 2013;62:74–80.
76. Portincasa P, Bonfrate L, Vacca M, De Angelis M, Farella I, Lanza E, et al. Gut microbiota and short chain fatty acids: implications in glucose homeostasis. Int J Mol Sci. 2022;23:1105.
77. Atzeni A, Bastiaanssen TFS, Cryan JF, Tinahones FJ, Vioque J, Corella D et al. Taxonomic and functional fecal microbiota signatures Associated with insulin resistance in non-diabetic subjects with Overweight/Obesity within the Frame of the PREDIMED-Plus study. Front Endocrinol. 2022;13.
78. Bastard J-P, Pelloux V, Alili R, Fellahi S, Aron-Wisnewsky J, Capel E, et al. Altered subcutaneous adipose tissue parameters after switching ART-controlled HIV + patients to raltegravir/maraviroc. AIDS. 2021;35:1625–30.
79. Villanueva-Millán MJ, Pérez-Matute P, Recio-Fernández E, Lezana Rosales JM, Oteo JA. Differential effects of antiretrovirals on microbial translocation and gut microbiota composition of HIV-infected patients. J Int AIDS Soc. 2017;20:21526.

Publisher's Note