SUPPLEMENTARY MATERIALS

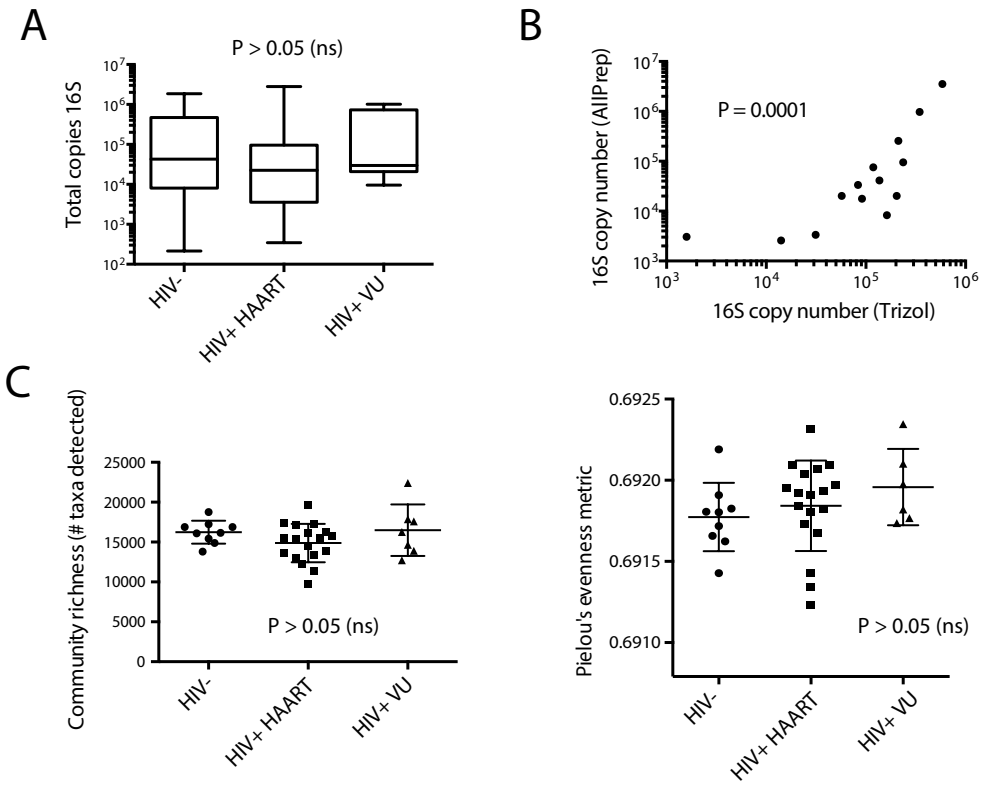


Figure S1: Gross bacterial community metrics do not differ significantly across subject groups or intra-subject samples. **(A)** Total bacterial abundances compared between subject groups as measured by TaqMan qPCR assay. The Kruskal-Wallis test was used to test differences in subject groups ($P > 0.05$). **(B)** Total bacterial load is concordant across different biopsies from the same individual and irrespective of DNA extraction method. Y-axis represents total bacterial load measures on biopsies with nucleic acids extracted using the AllPrep kit (Qiagen); X-axis represents separate biopsies from the same individual, taken at the same clinical visit and from the same relative anatomical location, with nucleic acids extracted using Trizol (Life Technologies). Correlation was tested by the Spearman method ($P = 0.0001$; Spearman rho correlation coefficient, $R_S = 0.86$). **(C)** Bacterial community evenness (Pielou's metric) and richness is not significantly different among subject groups. The Kruskal-Wallis test was used to detect significant differences across all subject groups ($P > 0.05$).

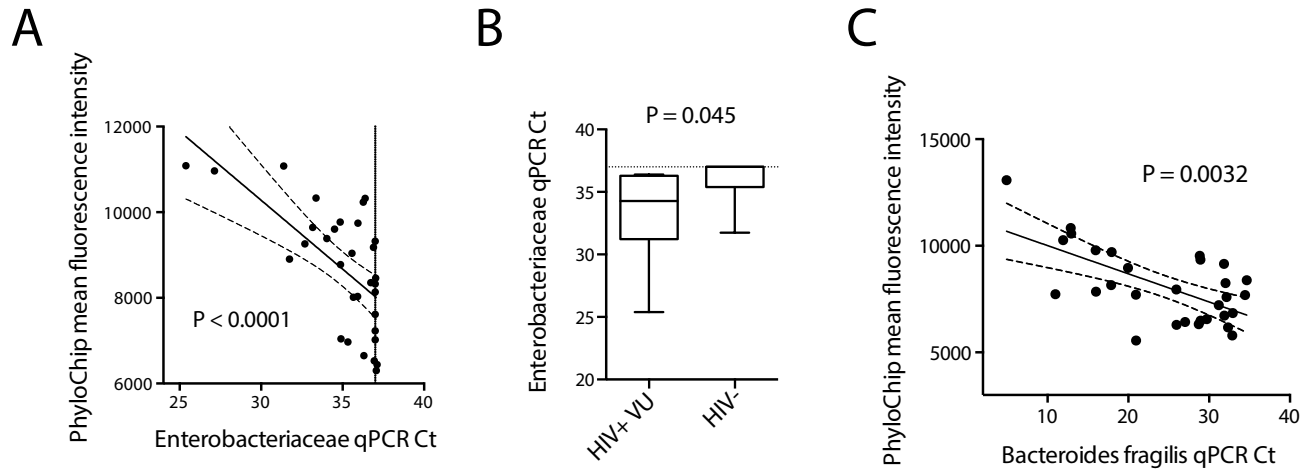
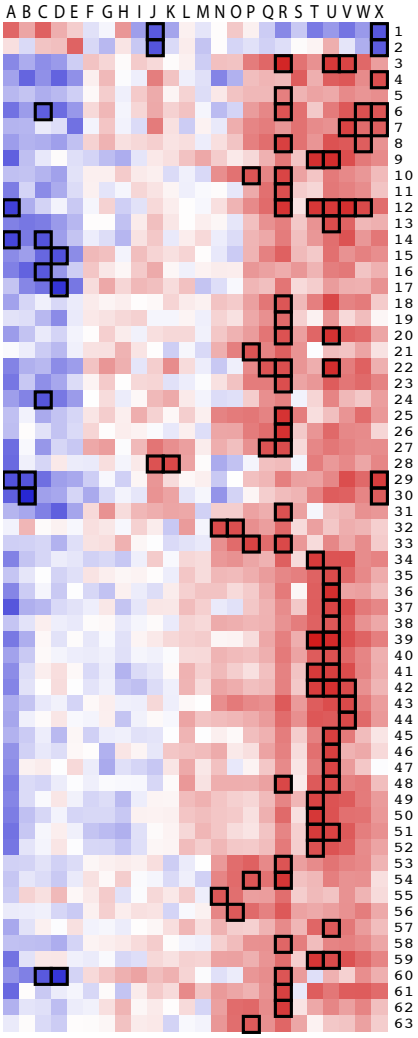


Figure S2: Comparisons of quantitative PCR (qPCR) measurements and PhyloChip relative abundance measurements [mean fluorescence intensity (MFI)] for selected taxa of interest. **(A)** PhyloChip MFI of Enterobacteriaceae members within the DMC were averaged and compared by Spearman correlation to qPCR measures of Enterobacteriaceae abundance using sample DNA extracted from mucosal biopsies, as determined by Ct (R_S = -0.66, p<0.0001), with linear regression line (p<0.0001). Fine dotted line represents qPCR limit of detection (Ct=37), as determined by Ct for which amplification in no-template controls was observed across replicates. Dotted lines represent 95% CI for linear regression coefficients. **(B)** qPCR measures of Enterobacteriaceae differ in VU subjects compared with HIV-uninfected risk-matched controls (P = 0.045). A Mann-Whitney non-parametric test was used to determine significance. Dotted line again represents limit of detection. **(C)** PhyloChip MFI of *Bacteroides fragilis* was compared to qPCR measurements for *B. fragilis* performed on pre-PhyloChip hybridization 16S amplicon samples. Spearman correlation was used to determine significance (R_S = -0.52, P = 0.0033), with linear regression line (P = 0.0002). Dotted lines represent 95% CI for linear regression coefficients.

IMUNOPATHOLOGIC DISEASE MARKERS

A	%IL-17+ in gut CD8 T cells
B	%IL-22+ in gut CD8 T cells
C	%IL-17+ in gut CD4 T cells
D	%IL-22+ in gut CD4 T cells
E	IFAB pg/ml plasma
F	%Treg in gut (CD25+FoxP3+)
G	Interferon-stimulated gene expression
H	sCD14 ng/mL plasma
I	HIV DNA copies/1e6 cells
J	HIV RNA copies/1e6 cells
K	Gut IDO1 relative mRNA
L	CD4 T cell density in gut
M	CD4 count
N	Zonulin ng/mL plasma
O	D-Dimer ng/ml plasma
P	IL-6 pg/ml plasma
Q	Stnf-RII pg/ml plasma
R	Kyn/Trp
S	%38+DR+ in blood memory CD4 T cells
T	%38+DR+ in gut CD8 T cells
U	%38+DR+ in gut CD4 T cells
V	%38+DR+ in blood memory CD8 T cells
W	IP-10 pg/ml plasma
X	Self-reported CD4 nadir



BACTERIAL TAXONOMIC CLASSIFICATIONS

1	c__Bacteroidia, f__Porphyromonadaceae, g__unclassified
2	c__Clostridia, f__Lachnospiraceae, g__Ruminococcus
3	c__Actinobacteria, f__Actinosynnemataceae, g__Actinokineospora
4	c__Actinobacteria, f__Bogoriellaceae, g__Georgenia
5	c__Actinobacteria, f__Corynebacteriaceae, g__Corynebacterium
6	c__Actinobacteria, f__Microbacteriaceae, g__unclassified
7	c__Actinobacteria, f__Micrococcaceae, g__Arthrobacter
8	c__Actinobacteria, f__Mycobacteriaceae, g__Mycobacterium
9	c__Alphaproteobacteria, f__Caulobacteraceae, g__unclassified
10	c__Alphaproteobacteria, f__Pelagibacteraceae, g__CandidatusPelagibacter
11	c__Alphaproteobacteria, f__Rhizobiaceae, g__Rhizobium
12	c__Alphaproteobacteria, f__Rhodobacteraceae, g__unclassified
13	c__Bacilli, f__Bacillaceae, g__Bacillus
14	c__Bacilli, f__Bacillaceae, g__Oceanobacillus
15	c__Bacilli, f__Bacillaceae, g__Virgibacillus
16	c__Bacilli, f__Planococcaceae, g__Sporosarcina
17	c__Betaproteobacteria, f__Alcaligenaceae, g__Castellaniella
18	c__Betaproteobacteria, f__Neisseriaceae, g__unclassified
19	c__Betaproteobacteria, f__Oxalobacteraceae, g__unclassified
20	c__Betaproteobacteria, f__Rhodocyclaceae, g__Dechloromonas
21	c__Betaproteobacteria, f__unclassifieg__sfA
22	c__Clostridia, f__ClostridialesFamilyXI, IncertaeSedis, g__Anaerococcus
23	c__Clostridia, f__Lachnospiraceae, g__unclassified
24	c__Clostridia, f__Peptococcaceae, g__Desulfitobacterium
25	c__Deltaproteobacteria, f__Desulfuromonadaceae, g__unclassified
26	c__Deltaproteobacteria, f__Nitrospinaceae, g__unclassified
27	c__Deltaproteobacteria, f__unclassifieg__sfA
28	c__Epsilonproteobacteria, f__Campylobacteraceae, g__Campylobacter
29	c__Erysipelotrichi, f__Erysipelotrichaceae, g__Allobaculum
30	c__Erysipelotrichi, f__Erysipelotrichaceae, g__unclassified
31	c__Gammaproteobacteria, f__Acidithiobacillaceae, g__Acidithiobacillus
32	c__Gammaproteobacteria, f__Alteromonadaceae, g__Cellvibrio
33	c__Gammaproteobacteria, f__Colwelliaceae, g__Colwellia
34	c__Gammaproteobacteria, f__Ectothiorhodospiraceae, g__unclassified
35	c__Gammaproteobacteria, f__Enterobacteriaceae, g__Averyella
36	c__Gammaproteobacteria, f__Enterobacteriaceae, g__Baumannia
37	c__Gammaproteobacteria, f__Enterobacteriaceae, g__Dickeya
38	c__Gammaproteobacteria, f__Enterobacteriaceae, g__Leclercia
39	c__Gammaproteobacteria, f__Enterobacteriaceae, g__Pantoea
40	c__Gammaproteobacteria, f__Enterobacteriaceae, g__Raoultella
41	c__Gammaproteobacteria, f__Enterobacteriaceae, g__Salmonella
42	c__Gammaproteobacteria, f__Enterobacteriaceae, g__Trabulsiella
43	c__Gammaproteobacteria, f__Halomonadaceae, g__Chromohalobacter
44	c__Gammaproteobacteria, f__Halomonadaceae, g__Cobetia
45	c__Gammaproteobacteria, f__Idiomarinaceae, g__unclassified
46	c__Gammaproteobacteria, f__J115, g__unclassified
47	c__Gammaproteobacteria, f__Moraxellaceae, g__Moraxella
48	c__Gammaproteobacteria, f__Oceanospirillaceae, g__unclassified
49	c__Gammaproteobacteria, f__Pasteurellaceae, g__Actinobacillus
50	c__Gammaproteobacteria, f__Pasteurellaceae, g__Aggregatibacter
51	c__Gammaproteobacteria, f__Pasteurellaceae, g__Haemophilus
52	c__Gammaproteobacteria, f__Pasteurellaceae, g__unclassified
53	c__Gammaproteobacteria, f__Piscirickettsiaceae, g__Thiomicrospira
54	c__Gammaproteobacteria, f__SUP05, g__unclassified
55	c__Gammaproteobacteria, f__Thiotrichaceae, g__Leucothrix
56	c__Gammaproteobacteria, f__unclassifieg__sfAI
57	c__Gammaproteobacteria, f__Xanthomonadaceae, g__Stenotrophomonas
58	c__Gammaproteobacteria, f__Xanthomonadaceae, g__unclassified
59	c__Gammaproteobacteria, f__Xanthomonadaceae, g__Xanthomonas
60	c__Oscillatoriophycideae, f__Phormidiaceae, g__Phormidium
61	c__Planctomycea, f__Planctomycetaceae, g__unclassified
62	c__Synechococcophycideae, f__Synechococcaceae, g__Prochlorococcus
63	c__TM7-3, f__unclassifieg__sfA

Figure S3: Genus and immunologic parameter identities for genera bearing one or more correlation with unadjusted $P < 0.005$, from analyses including all HIV-infected subjects studied. As was done in Figure 3A, Spearman correlation coefficients for taxa with representative species belonging to the same genus and displaying similar MFI trends were averaged, while the median P value of the same Spearman correlations was used to define genera with unadjusted $P < 0.005$ (boxed intersections).

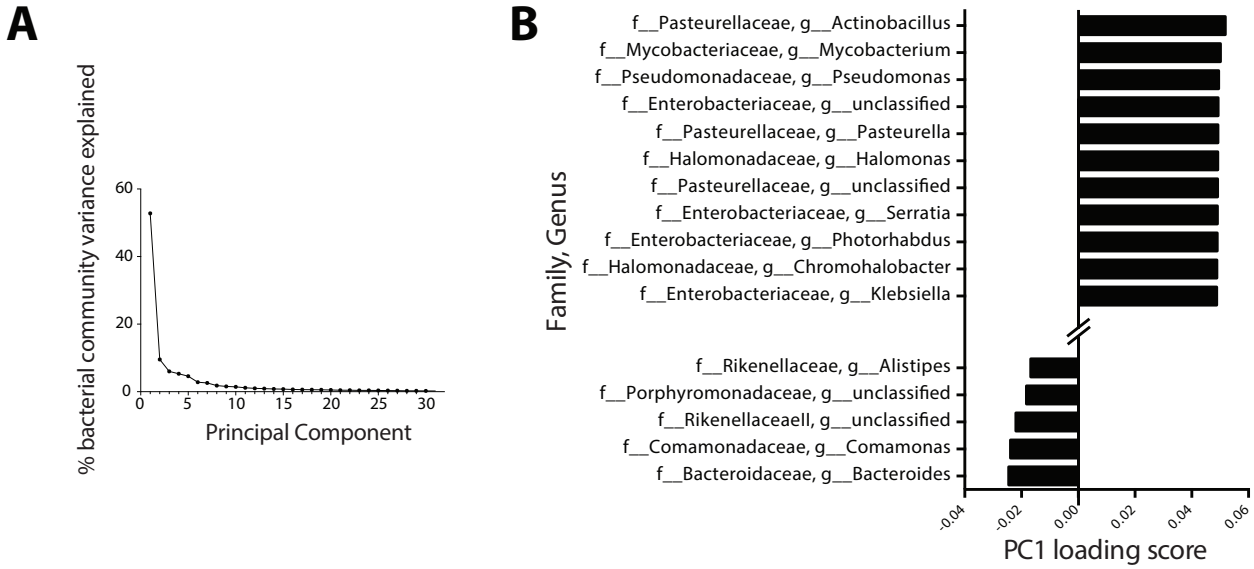


Figure S4: Principal component analysis characteristics. **(A)** Proportions of covariance in DMC taxon abundance data explained by first thirty principal components. A principal component analysis was performed as detailed in Methods, the first thirty of which represented 99.45% of DMC community covariance, while the first five principal components together represented 78.23% of covariance. **(B)** Loading scores for Principal Component 1 (PC1). Values of factor loading scores were ranked and maxima for the top genera loading most heavily toward the positive and negative extremes of PC1 are displayed.

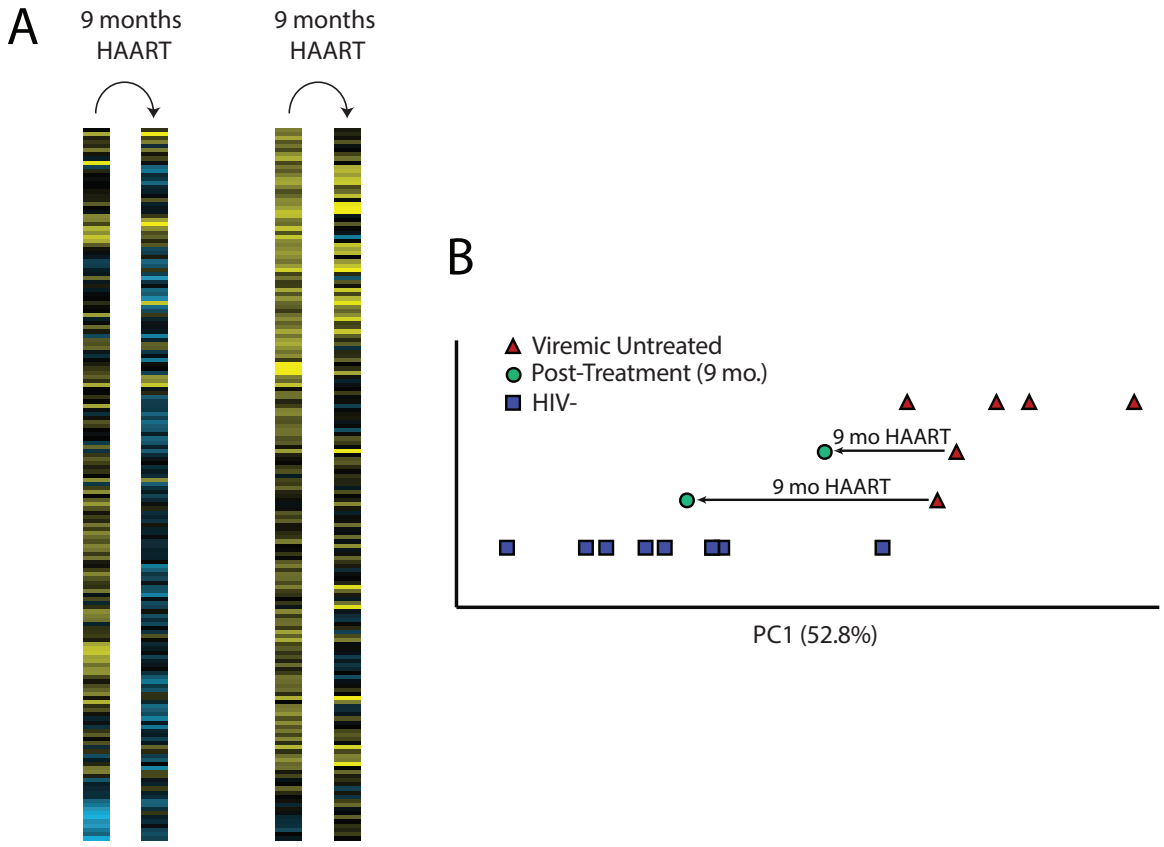


Figure S5: Highly active antiretroviral therapy causes a shift in DMC composition toward the healthy, uninfected state in two subjects. **(A)** Two viremic, untreated (VU) subjects returned for follow-up after 9 months of successful HAART (as determined by sustained undetectable plasma viral load). The abundance of genera within the DMC diminished significantly in both subjects upon treatment, as determined by a two-tailed, paired T-test (***$P < 5x10^{-10}$). **(B)** PC1 values are shown for VU (top row), HIV-uninfected (bottom row), and two subjects who were sampled when viremic and untreated and then again after nine months of HAART (middle two rows).

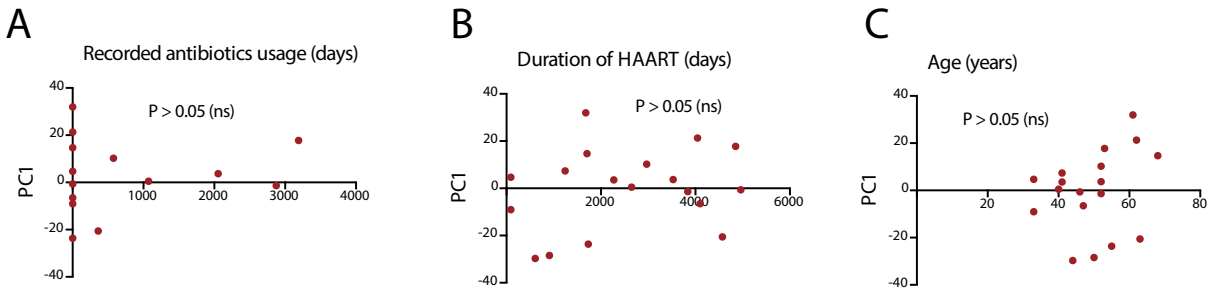


Figure S6: Comparison of adjustment variables (antibiotics usage, days on HAART, age) to PC1. The Spearman correlation method was used to test trends between antibiotics usage history (data available for 13 out of 18 subjects; P = 0.38), days on HAART (data available for all subjects; P = 0.47), and age (data available for all subjects; P = 0.21).

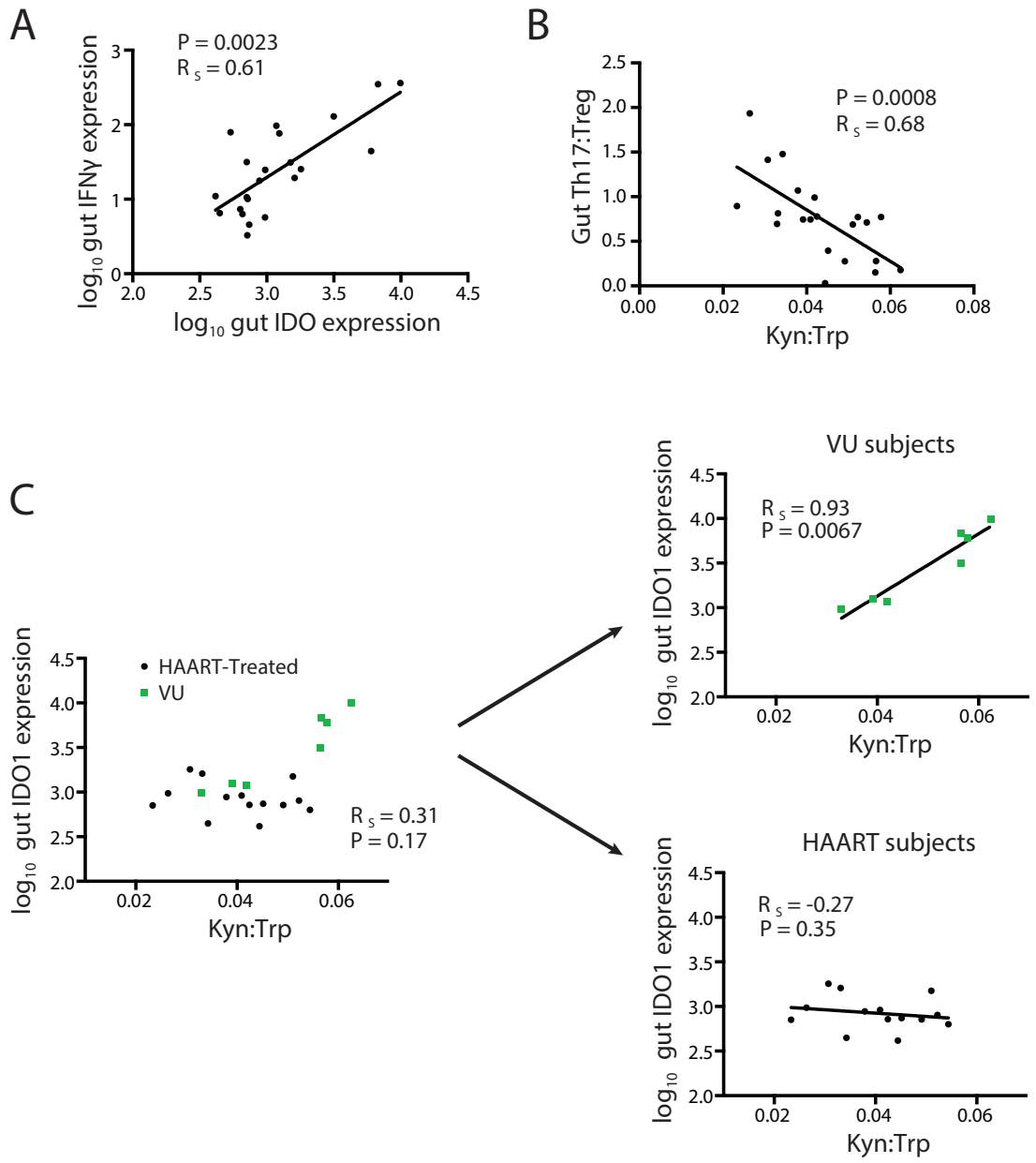


Figure S7: Relationships between IDO1 expression, IDO1 activity, T cell differentiation, and type II interferon. **(A)** Interferon γ expression and IDO1 expression was measured in all HIV-infected subjects studied by quantitative polymerase chain reaction from RNA extracted from whole rectosigmoid biopsy specimens, normalized to the housekeeping gene, HPRT, and compared using a Spearman correlation test. **(B)** Th17 and T_{reg} cells in the gut mucosa of all HIV-infected subjects were quantified using flow cytometry as described in Materials & Methods and compared. **(C)** Gut IDO1 expression

was compared to Kyn:Trp in all HIV-infected subjects, in VU subjects alone, and in HAART subjects alone.

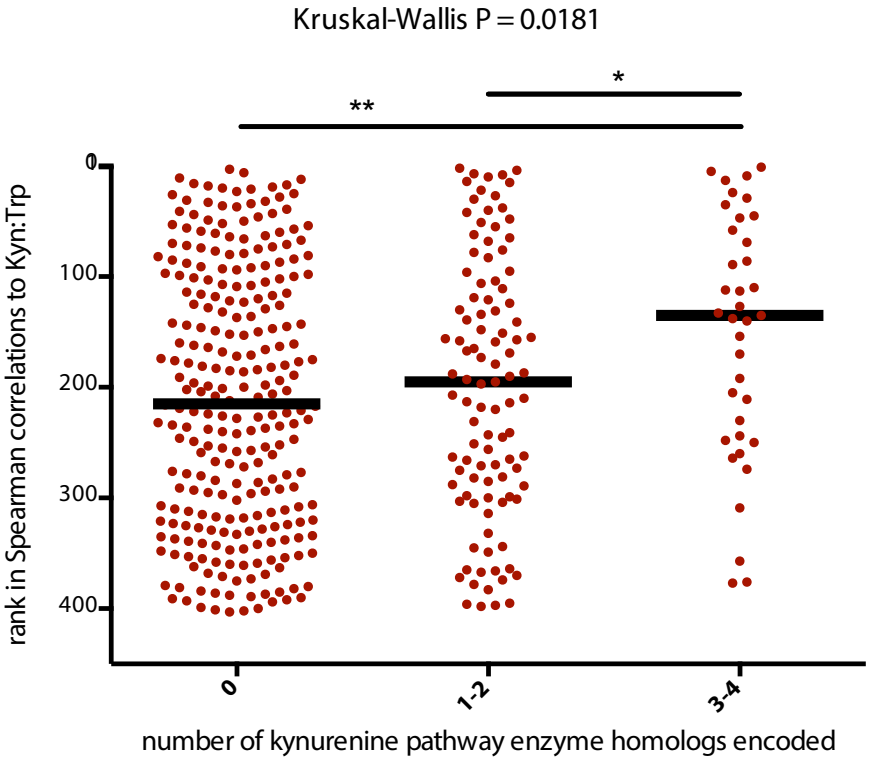


Figure S8: Bacterial taxa that predict plasma Kyn:Trp ratios in a multivariate machine learning analysis preferentially encode genetic homologs of tryptophan catabolism enzymes involved in the kynurenine pathway. The R package 'randomForest' was used to incorporate analysis of the effects of complex dependencies in microbial abundance data, and allowing for high-resolution multivariate assessment of the predictive capacity of microbial abundances toward plasma Kyn:Trp ratios. Results were similar to those using the univariate Spearman correlation method, in that genera that genetically encoded greater numbers of tryptophan-catabolizing enzymes better predicted Kyn:Trp ratios in plasma of HIV-infected subjects than did those with fewer or no genetic homologs to tryptophan-catabolizing enzymes.

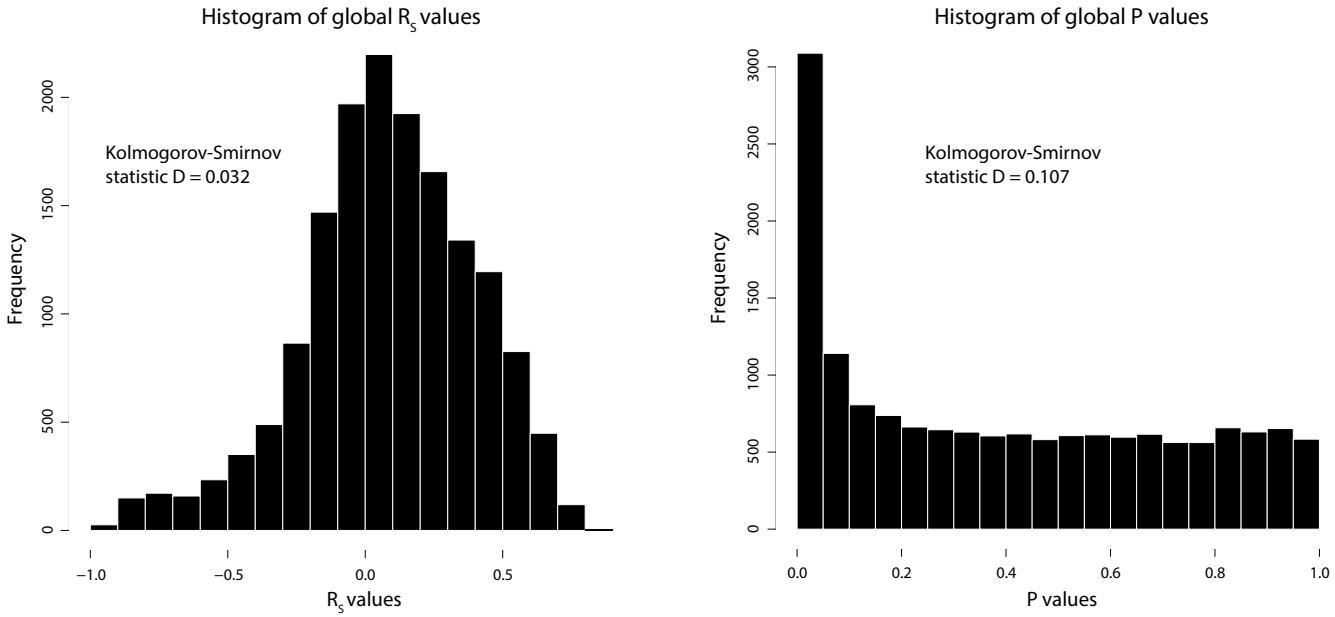


Figure S9: Data distribution of R_S and P values for all taxa as compared to all immunologic variables in Spearman correlation tests. Histograms depict frequency (y-axis) of each value for R_S and P values (x-axes), and show a normal distribution for R_S values and a non-normal distribution for P values.

Table S1: Patient cohort data**.** Paired blood and rectosigmoid biopsy specimens were collected from a cohort of male HIV-infected and uninfected subjects receiving care at San Francisco General Hospital with characteristics described below:

Median (Range)	**Age**	**Viral Load**	**Duration of HAART**	**CD4 counts, cells/ul blood**	**Years since first seropositive test**
Viremic Untreated (n=6)	33 (28-55)	11716 copies/ml (1077-83800)	0 years	356.3 (313-819)	2.20 (0.37-5.81)
HAART-Treated (n=16)*	52 (40-68)	<40 copies/ml	> 2 years	374.5 (251-1110)	21.22 (7.81-32.17)
Uninfected (n=9)	46 (36-59)	<40 copies/ml	0 years	803 (460-1487)	0
Long Term Non-Progressor (n=1)	34	820 copies/ml	0 years	505	21.12

Ethnicity (n)	**African American**	**Asian**	**Caucasian**	**Hispanic/ Latino**	**Multiracial**	**Native American**	**Pacific Islander**
Viremic Untreated (n=6)	1	1	3	0	1	0	0
HAART-Treated (n=16)*	3	0	10	2	0	0	1
Uninfected (n=9)	4	1	3	1	0	0	0
Long Term Non-Progressor (n=1)	0	0	1	0	0	0	0

***Note:** Two viremic untreated subjects returned after nine months of effective HAART (as determined by lack of detectable plasma viremia), at which point blood and rectosigmoid biopsy samples were obtained (thus totaling 18 HAART samples for analysis).

Table S2: DMC taxon phylogenetic classifications and statistical analyses. (Separately uploaded)

Table S3: Spearman correlations between individual taxa of the DMC and markers of disease progression, as calculated among all HIV-infected subjects. (Separately uploaded)

Table S4: Spearman correlations between PC1 and immunologic variables among all HIV-infected subjects.

Immune Variable	Spearman R_S value	Unadjusted P value	Benjamini-Hochberg adjusted P value
%CD38+DR+ amongst gut CD4+ T cells	0.738	0.00047	0.00746
Kyn:Trp ratio	0.684	0.00062	0.00746
%CD38+DR+ amongst blood memory CD8+ T cells (all CD8+ after exclusion of CD27+CD45RO-)	0.7	0.00121	0.00966
%CD38+DR+ amongst gut CD8+ T cells	0.649	0.00354	0.02121
IP-10 pg/ml	0.543	0.00907	0.04355
Soluble TNF-Receptor II pg/ml plasma	0.493	0.01975	0.07902
%IL-17+ in gut CD8+ T cells	-0.423	0.06344	0.21691
Self-reported CD4+ T cell nadir	0.366	0.0723	0.21691
IL-6 pg/ml plasma	0.379	0.08205	0.21881
%CD38+DR+ in blood memory CD4+ T cells (all CD4+ after exclusion of CD27+CD45RO-)	0.389	0.11054	0.2653
D-Dimer ng/ml plasma	0.223	0.31842	0.63683
Zonulin ng/ml plasma	0.223	0.31842	0.63683
CD4+ T cell density in gut	0.222	0.37623	0.65517
%IL-22+ amongst gut CD8+ T cells	-0.196	0.38218	0.65517
%IL-17+ amongst gut CD4+ T cells	-0.162	0.45994	0.69009
HIV RNA copies/1e6 cells	0.162	0.46006	0.69009
IFAB pg/ml plasma	-0.15	0.50629	0.69746
%IL-22+ amongst gut CD4+ T cells	-0.14	0.5231	0.69746
Cumulative measure of interferon-stimulated gene (ISG) expression*	0.104	0.65403	0.80936
sCD14 ng/ml plasma	-0.097	0.67447	0.80936
%Treg in gut (CD25+FoxP3+) of total CD4+ T cells	0.078	0.71192	0.81363
Gut IDO1 relative mRNA	0.069	0.74976	0.81792
HIV DNA copies/1e6 cells	0.057	0.80098	0.8358
CD4+ T cell count	0.031	0.88452	0.88452

*A cumulative measure of ISG expression was obtained by quantitative PCR of four selected ISGs (GBP1, IFI27, MX1, and OAS1). Cts were subtracted from the housekeeping gene HPRT, and a geometric mean was calculated for all four genes on a per subject basis, yielding the final measure used for comparison to PC1.

Table S5: Spearman correlations between PC1 and immunologic variables among HIV+ subjects on HAART.

Immune Variable	Spearman R_S value	Unadjusted P value	Benjamini-Hochberg adjusted P value
IL-6 pg/ml plasma	0.743	0.0015	0.036
Kyn:Trp ratio	0.727	0.0032	0.038
Soluble TNF-Receptor II pg/ml plasma	0.654	0.0082	0.066
D-Dimer ng/ml plasma	0.607	0.0164	0.095
Gut IDO1 relative mRNA	-0.559	0.0197	0.095
IP-10 pg/ml	0.561	0.0297	0.119
%CD38+DR+ amongst gut CD4+ T cells	0.556	0.0389	0.134
HIV RNA copies/1e6 cells	-0.491	0.0534	0.16
Zonulin ng/ml plasma	0.486	0.0664	0.177
%CD38+DR+ in blood memory CD8+ T cells (all CD8+ after exclusion of CD27+CD45RO-)	0.489	0.076	0.182
%IL-17+ amongst gut CD8+ T cells	-0.382	0.1598	0.349
%CD38+DR+ amongst gut CD8+ T cells	0.337	0.2392	0.478
sCD14 ng/ml plasma	0.213	0.4643	0.812
%CD38+DR+ amongst blood memory CD4+ T cells (all CD4+ after exclusion of CD27+CD45RO-)	0.209	0.4738	0.812
HIV DNA copies/1e6 cells	-0.186	0.5075	0.812
%Treg in gut (CD25+FoxP3+) of total CD4+ T cells	0.123	0.6273	0.889
IFAB pg/ml plasma	-0.136	0.6296	0.889
Self-reported CD4+ T cell nadir	0.088	0.7293	0.966
CD4 T+ cell density in gut	0.077	0.8122	0.966
Cumulative measure of interferon-stimulated gene (ISG) expression*	-0.053	0.8456	0.966
%IL-17+ amongst gut CD4+ T cells	0.041	0.8797	0.966
%IL-22+ amongst gut CD8+ T cells	-0.039	0.8894	0.966
%IL-22+ amongst gut CD4+ T cells	-0.021	0.9397	0.966
CD4+ T cell count	0.011	0.9665	0.966

*****A cumulative measure of ISG expression was obtained by quantitative PCR of four selected ISGs (GBP1, IFI27, MX1, and OAS1). Cts were subtracted from the housekeeping gene HPRT, and a geometric mean was calculated for all four genes on a per subject basis, yielding the final measure used for comparison to PC1.

Table S6: Comparisons of PC1 marker of dysbiosis to gut inflammatory gene expression in whole rectosigmoid biopsy specimens.

Variable	Spearman R_S value	Unadjusted P value
TNFa	0.35	0.27
CYBB	-0.28	0.29
IFNa2	-0.2	0.45
IFNb	0.15	0.57
IFNg	-0.15	0.59
Arginase 1	-0.14	0.62
NOS2	0.08	0.76

Table S7: Spearman correlation to Kyn:Trp genus ranks, and analysis of kynurenine pathway metabolic enzymes (those involved in metabolism of tryptophan to 3-hydroxyanthranilic acid) in bacteria. Spearman correlation tests were performed on all detected taxa as compared to Kyn:Trp ratios in plasma of HIV-infected subjects. P values were adjusted for multiple comparisons using the Benjamini-Hochberg technique (*80*). The UniProt Consortium database (*52*) (www.uniprot.org) was queried for all annotated bacterial enzymes within the pathway to catabolize tryptophan to 3-hydroxyanthranilic acid as described in Methods. (Separately uploaded)