

RESEARCH ARTICLE

Open Access



# The bronchial epithelial cell bacterial microbiome and host response in patients infected with human immunodeficiency virus

Marc A. Sze<sup>1\*</sup>, Stella Xu<sup>1</sup>, Janice M. Leung<sup>1</sup>, Emily A. Vucic<sup>2</sup>, Tawimas Shaipanich<sup>3</sup>, Aida Moghadam<sup>4</sup>, Marianne Harris<sup>4,5,6</sup>, Silvia Guillemi<sup>4,5,6</sup>, Sunita Sinha<sup>7</sup>, Corey Nislow<sup>7</sup>, Darra Murphy<sup>8</sup>, Cameron Hague<sup>8</sup>, Jonathon Leipsic<sup>8</sup>, Stephen Lam<sup>2</sup>, Wan Lam<sup>2</sup>, Julio S. Montaner<sup>6,9</sup>, Don D. Sin<sup>1,3</sup> and S. F. Paul Man<sup>1,3</sup>

## Abstract

**Background:** Chronic Obstructive Pulmonary Disease (COPD) is an important comorbidity in patients living with human immunodeficiency virus (HIV). Previous bacterial microbiome studies have shown increased abundance of specific bacterium, like *Tropheryma whippelii*, and no overall community differences. However, the host response to the lung microbiome is unknown in patients infected with HIV.

**Methods:** Two bronchial brush samples were obtained from 21 HIV-infected patients. One brush was used for bacterial microbiome analysis using the Illumina MiSeq<sup>TM</sup> platform, while the other was used to evaluate gene expression patterns of the host using the Affymetrix Human Gene ST 2.0 array. Weighted gene co-expression network analysis was used to determine the relationship between the bacterial microbiome and host gene expression response.

**Results:** The Shannon Diversity was inversely related to only one gene expression module ( $p = 0.02$ ); whereas evenness correlated with five different modules ( $p \leq 0.05$ ). After FDR correction only the Firmicutes phylum was significantly correlated with any modules (FDR < 0.05). These modules were enriched for cilia, transcription regulation, and immune response. Specific operational taxonomic units (OTUs), such as OTU4 (Pasteurellaceae), were able to distinguish HIV patients with and without COPD and severe emphysema.

**Conclusion:** These data support the hypothesis that the bacterial microbiome in HIV lungs is associated with specific host immune responses. Whether or not these responses are also seen in non-HIV infected individuals needs to be addressed in future studies.

**Keywords:** Bronchial brushing, Bacterial microbiome, HIV, Lungs, Gene expression

## Background

The increased susceptibility of patients infected with human immunodeficiency virus (HIV) to lung diseases, in particular chronic obstructive pulmonary disease (COPD), has now been recognized in numerous epidemiological studies [1–3]. Because cigarette smoke exposure only partially explains this elevated risk [1], the

pathogenesis of comorbid lung disease in HIV is largely a mystery. Investigation of the HIV lung bacterial microbiome using bronchoalveolar lavage fluid has suggested a greater abundance of *Tropheryma whippelii* in the HIV lung [4] and no significant impact of anti-retroviral therapy on the bacterial community composition in HIV-infected individuals [5]. However, the impact of the lung microbiota on the pathogenesis of chronic lung diseases such as COPD in HIV is unclear. Moreover, there is little information whether the lung microbiome is associated with significant host responses in the lungs.

\* Correspondence: marcsze@med.umich.edu

<sup>1</sup>Centre for Heart Lung Innovation, St. Paul's Hospital & Department of Medicine, University of British Columbia, Rm 166 – 1081 Burrard St., Vancouver, BC V6Z 1Y6, Canada

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



Although characterization of bacterial community composition in a disease state is an important first step in uncovering the possible clinical relevance of the lung microbiome [4, 6], the next logical step is to discover whether or not changes in the lung microbiome induce a host response that may be important in disease pathogenesis. We have recently shown, using lung tissue samples from non-HIV infected individuals with COPD, that shifts in the lung microbiome are associated with important changes in inflammatory response in these lungs [7]. One important limitation of that study was that the microbiome was characterized in a block of lung tissue and as such cell-specificity could not be ascertained. Moreover, this study did not include any patients with HIV infection. Here, we extend these observations by investigating the interactions between the host gene expression response and the bacterial microbiome in bronchial epithelial cells of small airways collected from the same site in patients infected with HIV. The specific aims of this study were to describe the bacterial community composition of the HIV bronchial epithelium and to determine whether the bacterial microbiome of the HIV bronchial epithelium is associated with specific gene expression signatures of the host that may reveal the underlying pathogenesis of chronic airways disease in HIV-infected individuals.

## Methods

### Patient population

All subjects provided written informed consent for the collection of cytologic brushings for research purposes under the UBC Providence Health Care ethics protocol H14-03267. Subjects were recruited from patients undergoing bronchoscopy for pulmonary nodules, masses, or pneumonia (all conditions were diagnosed radiographically by computed tomography (CT) imaging at St. Paul's Hospital, Vancouver, BC). Entry criteria into the study included documented HIV-1 infection and  $\geq 19$  years of age. All subjects performed spirometry according to the American Thoracic Society/European Respiratory Society guidelines [8] within three months, except for five subjects who underwent bronchoscopy for acute infection. COPD was defined by post-bronchodilator forced expiratory volume in one second (FEV1)/forced vital capacity (FVC) ratio of less than 70 %.

Patients underwent thoracic CT imaging using a 64 detector CT scanner (Discovery HD 750 or a VCT, GE Healthcare, Milwaukee, WI). A central imaging core laboratory (SPH CT Corelab), blinded to spirometry and clinical data, interpreted the CT images for emphysema based on a modified method of Kazerooni, et al. [9]. Emphysema severity was qualitatively scored according to an established algorithm (see Additional file 1). CT scans were also qualitatively scored for respiratory bronchiolitis

(none, trivial, mild, moderate, and severe) and bronchiectasis (presence or absence). Details on bronchoscopy and specimen collection can be found in the Additional file 1. Bronchial epithelial cells were obtained from sites away from the acute infection, masses or nodules.

### Bacterial microbiome analysis

DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Toronto, Ontario) from both patient samples and background negative environmental controls. Total 16S load was quantified using a droplet digital polymerase chain reaction (ddPCR) assay [10]. These background controls were used to assess whether the bacterial community of the HIV samples were impacted by the instruments and reagents used during the extraction and PCR process. To assess the 16S load within the samples the average 16S load from the negative controls were subtracted from each HIV 16S sample. Touchdown PCR [11] of the 16S rRNA gene V4 region was used to generate a DNA template for sequencing. Cycle conditions for the touchdown PCR can be found in the Additional file 1. Sequencing was performed on an Illumina MiSeq<sup>TM</sup> (Illumina, Redwood City, CA, USA) with  $2 \times 250$  paired end-read chemistry. The protocol established by Kozich, et al. was used for the sequencing and subsequent data cleanup within the program mothur (V1.34.4) [12]. After processing, sequence cleanup, and chimera removal, a total of 3,559,398 reads remained. Data analysis was performed in R (V3.2.0) and R studio (V0.99.441) employing the vegan (V2.3-0) package [13, 14]. In order to adequately perform alpha and beta diversity analysis subsampling to the lowest total reads (3164) was performed [15]. Along with a 97 % similarity threshold, a total of 451 different Operational Taxonomic Units (OTUs) were identified. Sequence data has been deposited in the NCBI sequence read archive under the accession number SRP068430. The corresponding metadata can be found at <http://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP068430&go=go>. Alternatively, one can use the following link <http://www.ncbi.nlm.nih.gov/Traces/study/> and search for SRP068430 to bring up the relevant information needed. The relevant samples used for this study are BIDC1-4, BIDC7-10, BIDC14-26 for the HIV samples and Neg1-4 for the background negative control samples.

### Microarray analysis

RNA was extracted using the Qiagen RNeasy Plus Universal Kit (Qiagen, Toronto, Ontario). 1  $\mu$ g of RNA was processed and hybridized onto the Affymetrix Human Gene ST 2.0 array (Affymetrix Inc, Santa Clara, USA) according to the manufacturer's protocol at the Hospital for Sick Children, Centre for Applied Genomics

(Toronto, Ontario). Raw CEL files were processed and RMA normalized in R (V3.2.0) and R studio (V0.99.441) using a standard protocol from the *oligo* package (V1.32.0) [16]. Gene symbols and names were obtained from the *hugene20sttranscriptcluster.db* from bioconductor [17].

### Data analysis and integration

For phyla level comparisons a *t*-test with Bonferroni correction was applied. A random forest algorithm with Boruta feature selection [18, 19] was used to identify any OTUs that could be discriminative of specific clinical traits in the patient population (e.g. smoking status, COPD, etc.). Traits were chosen for Boruta feature selection analysis based on whether or not their PERMANOVA value was less than or equal to 0.1. We determined differences in the bacterial community composition between groups by a Bonferroni corrected PERMANOVA [20] of  $\leq 0.0125$ . Robustly co-expressed sets of genes (i.e. modules) were identified in airway expression data using a weighted gene co-expression network analysis (WGCNA) [21, 22]. Modules eigengene vector values were then compared to alpha diversity measures (Shannon Diversity, OTU richness, and evenness) and phyla measures using the WGCNA (V1.46) R package [21, 22]. For this analysis no grouping was performed by smoking, CD4 cell count, or viral load status as these variables were not significantly associated with microbiome measures. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) [23] was used to identify the most relevant pathway clusters for each module that were significantly correlated to the specific bacterial microbiome measurements. The top 10 enrichment clusters were used as a guide to discovering pathways that were most strongly associated with each module. A false discovery rate (FDR) of less than 0.05 was considered significant. In addition to this module comparison versus microbiome metrics for our network used *p*-values of less than 0.05 as well. The OTU data was reported to the lowest taxonomic identification, either within the text or in the respective figure.

## Results

### Overview of the bacterial microbiome in the HIV cohort

An overview of the clinical characteristics of the study subjects showed that all individuals were between 40 and 75 years of age with a majority on highly active anti-retroviral therapy (HAART) at the time of assessment [Table 1]. The total 16S concentration in each subject following background negative control subtraction was  $0.42 \pm 1.39$  16S/ng of DNA (mean  $\pm$  standard deviation). The Shannon Diversity was  $2.13 \pm 0.54$ , OTU richness was  $37.52 \pm 11.83$ , and evenness was  $0.59 \pm 0.13$  for this patient population (mean  $\pm$  standard deviation). The

distribution of Shannon Diversity, OTU richness, and evenness can be found in the Additional file 1: Table S1. On a cursory overview, the phyla distribution seems to be quite different than the experimental background negative controls. However, there was only a significant difference in the relative abundance of the Actinobacteria phylum, following Bonferroni correction, between the HIV group and background negative controls ( $p = 0.003$ ) [Fig. 1a]. This would suggest that apart from the Actinobacteria phylum the other phyla distributions are similar to the background negative controls. A total of 23.8 % of HIV subjects contained OTUs that aligned to *Tropheryma*.

### Airway microbiome comparisons between those with and without COPD by spirometry

There was no difference in Shannon Diversity, evenness, and OTU richness between HIV patients with and without COPD ( $p > 0.05$ ). The diagnosis of COPD had no influence on the phyla observed ( $p > 0.05$ ) [Fig. 1b]. However, in COPD, there was a trend towards greater abundance of the Actinobacteria and Proteobacteria phyla. Using a Bray-Curtis dissimilarity matrix and Non-Metric Multidimensional Scaling (NMDS) with PERMANOVA, we found no significant difference in the bacterial community composition between those with and without COPD (PERMANOVA = 0.10) [Fig. 2a]. However, analysis of specific OTUs in relation to COPD status revealed 3 OTUs that were able to discriminate HIV patients with and without COPD [Fig. 2b]: OTU4 (Pasteurellaceae), OTU15 (Brachybacterium), and OTU38 (Yersinia). In COPD samples, there was a paucity of OTU4 and OTU15, and a slight enrichment of OTU38 [Fig. 2b]. Ribosomal database classifier [24, 25] revealed that OTU4 contained sequences of bacteria in the genus for *Haemophilus*

### Airway microbiome comparisons based on CT presence of emphysema or bronchiectasis

We did not detect a significant difference in the bacterial community composition based on severe emphysema status that was detected on CT scans (PERMANOVA = 0.06) [Fig. 2c]. However, there were two OTUs that discriminated samples from those with and without emphysema: OTU4 (*Pasteurellaceae-Haemophilus*) and OTU30 (*Pedobacter*). There was no difference in bacterial community composition in relation to emphysema distribution (whether centrilobular or paraseptal) or across respiratory bronchiolitis severity (both PERMANOVA  $> 0.10$ ) (data not shown).

In those with bronchiectasis on CT, the bacterial community composition was also not significantly different from those without any bronchiectasis (PERMANOVA = 0.04)

**Table 1** An overview of clinical traits of HIV infected patients sampled in this study

Age	Vital Status	Current VL	Current CD4	Bronchoscopy Indication	Smoking Status	Pack-Years	Current HAART	CT Emphysema	CT Bronchiectasis	FEV1 (L)	FEV1/FVC (%)
60-69	Alive	<40	400-499	Cancer	Current	30	Yes	Yes	Yes	1.53	33.04
50-59	Alive	1000-9999	100-199	Pneumonia	Current	39	No	Yes	No	1.11	90.00
70-79	Alive	<40	500-599	Cancer	Current	130	Yes	Yes	No	2.99	70.55
50-59	Alive	<40	600-699	Cancer	Current	12.5	Yes	Yes	No	3.35	56.63
50-59	Alive	40-1000	500-599	Cancer	Current	37.5	Yes	Yes	No	2.71	64.18
70-79	Alive	10000-99999	200-299	Cancer	Current	30	No	No	No	N/A	N/A
50-59	Alive	<40	900-999	Cancer	Current	15	Yes	No	No	3.33	70.95
50-59	Alive	10000-99999	200-299	Pneumonia	Current	19.5	No	Yes	No	N/A	N/A
60-69	Alive	<40	700-799	Cancer	Past	20	Yes	Yes	No	2.54	59.86
60-69	Alive	<40	800-899	Cancer	Past	3	Yes	Yes	No	2.87	71.44
60-69	Alive	40-999	200-299	Cancer	Past	45	Yes	No	No	2.41	76.80
60-69	Alive	<40	≥1000	Bronchiectasis	Past	12	Yes	No	Yes	3.06	70.57
60-69	Alive	<40	300-399	Pneumonia	Past	75	Yes	No	Yes	2.47	85
40-49	Alive	<40	100-199	Cancer	Current	30	Yes	Yes	No	2.41	51.56
40-49	Alive	10000-99999	100-199	Pneumonia	Current	115	Yes	Yes	Yes	2	78.28
50-59	Alive	<40	400-499	Cancer	Past	90	Yes	Yes	No	3.33	75.76
70-79	Deceased	<40	400-499	Cancer	Past	20	Yes	Yes	Yes	2.75	69.09
60-69	Deceased	<40	100-199	Cancer	Past	4	Yes	No	Yes	2.45	72.86
60-69	Deceased	<40	100-199	Cancer	None	0	Yes	Yes	No	N/A	N/A
50-59	Deceased	<40	100-199	Pneumonia	Current	N/A	Yes	Yes	No	N/A	N/A
40-49	Deceased	≥100000	<100	Cancer	None	0	No	Yes	No	N/A	N/A

[Fig. 2e]. Two OTUs were important for this discrimination: OTU1 (*Prevotella*), and OTU38 (*Yersinia*).

### The impact of acute lung infection on airway microbiome and host responses

We found no significant difference in the bacterial community composition between those with and without pneumonia (PERMANOVA = 0.30) probably because the bronchoscopic samples were obtained from the lung contralateral to the site of active infection. We also found no significant differences in bacterial community composition between those with CD4 counts above or below 200 (PERMANOVA = 0.84), across smoking status (current, past, or never smokers; PERMANOVA = 0.37), or whether Tropheryma was detected or not (PERMANOVA = 0.16) [Additional file 1: Figure S1–S5].

### Significant pathways in WGCNA modules that correlated with the bacterial microbiome

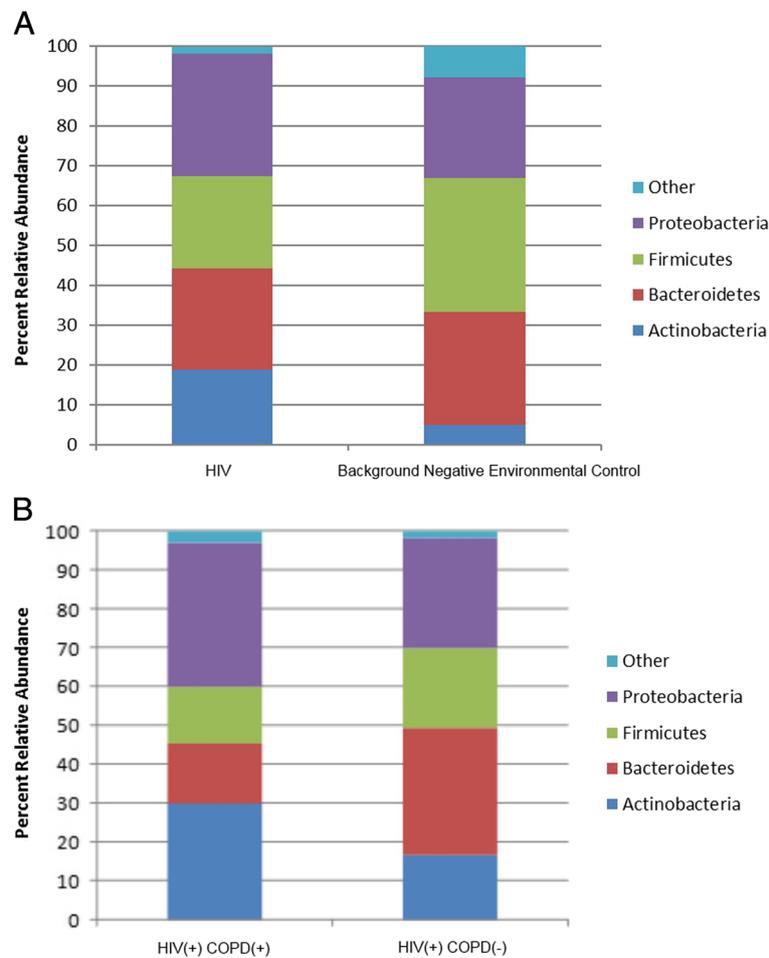
The power measurement of 6 was used to create the gene co-expression network and a single sample was excluded since it was an extreme outlier [21, 22] [Additional file 1: Figures S6 and S7]. This sample was considered an outlier since on the hierarchical clustering of the gene expression data it formed its own unique

branch on the tree versus all other samples [Additional file 1: Figure S7]. DAVID was used to assess the most relevant pathways involved in the WGCNA modules that correlated with the bacterial microbiome. In total there were 14/23 (60.8 %) gene expression modules that correlated with at least one measure of the bacterial microbiome [Table 2]. Most interesting were the immune pathways identified by the Tan, Red, Pink, and Green Yellow modules and the cilia pathways represented by the Green module [Table 2].

### WGCNA of alpha diversity and phyla with gene expression

Our analysis revealed a number of gene expression modules that correlated with the Firmicutes phylum. It was the only group that had significant correlations with any modules after FDR correction (two negative and two positive correlations). The negatively correlated modules were Green (FDR = 0.037,  $p = 4 \times 10^{-4}$ ), Midnight Blue (FDR = 0.037,  $p = 8 \times 10^{-4}$ ). The positively correlated modules were, Brown (FDR = 0.037,  $p = 8 \times 10^{-4}$ ), and Blue (FDR = 0.037,  $p = 5 \times 10^{-4}$ ) [Fig. 3 and Additional file 1: Figure S9].

When looking at those that had a  $p$  value under 0.05 but not an FDR under 0.05 there were additional



**Fig. 1** Breakdown of major phyla. **a** Comparison between HIV patient samples ( $n = 21$ ) and background negative controls ( $n = 4$ ). There was a significant difference in the Actinobacteria phylum between HIV and background negative controls ( $p = 0.003$ ). There were also slightly more Proteobacteria in the HIV group than in the background negative controls ( $p > 0.05$ ). **b** Comparison between HIV patients with ( $n = 6$ ) and without COPD ( $n = 10$ ). No difference between the different phyla was observed ( $p > 0.05$ )

correlations that have been summarized in Fig. 3. Briefly, the Tan module may be negatively correlated with both Shannon Diversity and evenness (FDR = 0.283,  $p = 0.02$  and FDR = 0.368,  $p = 0.03$  respectively) [Fig. 3a]. Evenness may also be correlated with the Midnight Blue module (FDR = 0.184,  $p = 0.009$ ) [Fig. 3a]. The Proteobacteria phylum may also be positively correlated with the Magenta module (FDR = 0.283,  $p = 0.02$ ) and the Turquoise module (FDR = 0.368,  $p = 0.03$ ) [Fig. 3b].

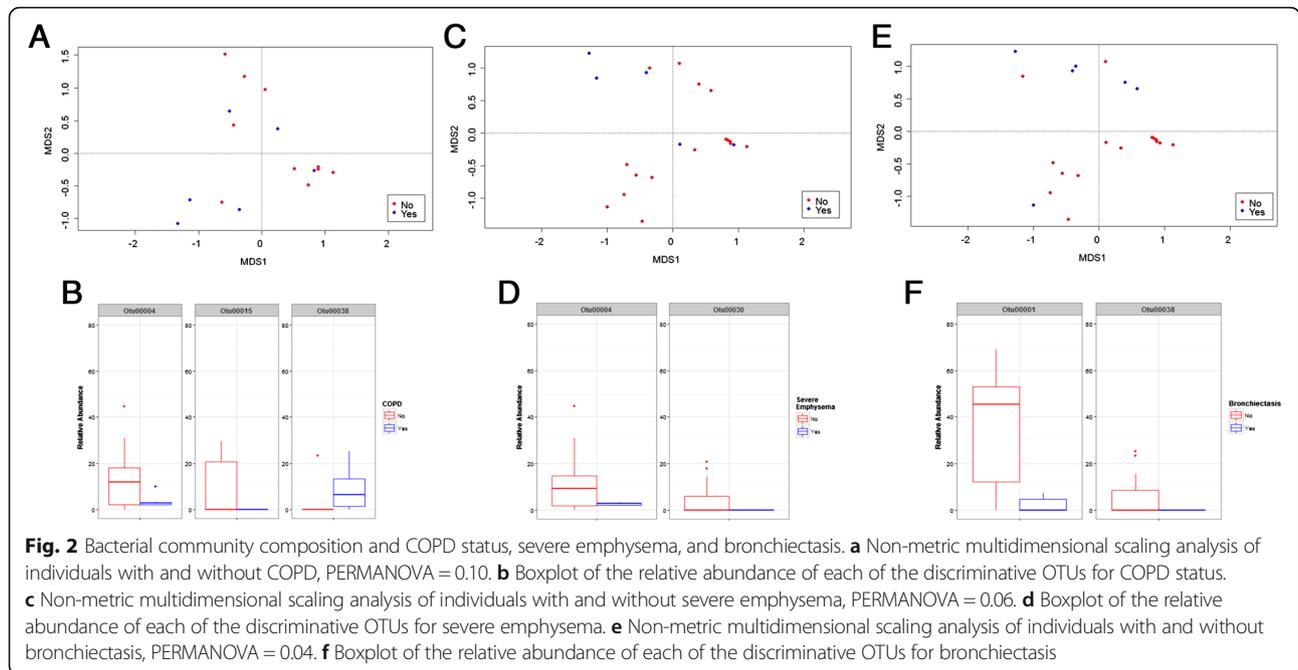
#### WGCNA of the important OTUs and gene expression

No modules and OTUs found to be predictive by random forest for COPD, severe emphysema, or bronchiectasis were found above an FDR of 0.05. are described in the Additional file 1: Figure S10. However, some correlations between modules and these OTUs occurred with a  $p$ -value under 0.05. This included OTU4 [Fig. 3c] which was negatively correlated with the Grey

(FDR = 0.941,  $p = 0.04$ ), Green (FDR = 0.941,  $p = 0.04$ ), and Green Yellow (FDR =  $p = 0.03$ ) modules. It was also positively correlated with the Blue ( $p = 0.004$ ) and Dark Green ( $p = 0.04$ ) modules.

#### Discussion

The interplay between the microbiome and host gene expression is increasingly recognized as a key element of health and disease. Our study extensively examined the relationship between the bacterial microbiome and host gene expression from bronchial epithelial cells taken from the same small airways of patients infected with HIV. We found that the small airway microbiome of HIV-infected patients demonstrated only modest differences in the global bacterial community composition compared with background negative controls [Fig. 1]. However, we did not find any significant differences in the global airway bacterial composition between those



with and without COPD, between those with elevated or reduced CD4 counts, between those with bronchiectasis, or between those with and without emphysema on CT scans [Fig. 2]. However, when we investigated individual OTUs using an uncorrected PERMANOVA threshold of 0.10 or below, we discovered OTU signatures that were distinct for those with COPD (measured by spirometry), severe emphysema (detected on CT), and bronchiectasis. Spirometry-based COPD was associated with OTU4, OTU15, and OTU38, severe emphysema was associated with OTU4 and OTU30, while bronchiectasis was

associated with OTU1 and OTU38. More importantly, we found that measures of the airway microbiome including alpha diversity measures, phyla, and OTUs, were significantly related to distinct host response in the same airway as captured by gene expression modules. Many of these modules involved immune and inflammatory responses, cell signaling, and cilia pathways suggesting immunomodulatory role of the airway microbiota in the host's ability to process and remove irritants and aeropathogens. Additional work will be needed to validate this hypothesis.

**Table 2** An overview of significant gene expression module pathways

Gene Expression Module	Number of Genes	Pathway Identified	FDR
Tan	201	Lysosome, Immune Response, Plasma Membrane	<5.0 × 10 <sup>-4</sup>
Red	554	Immune Response, Defense Response, Inflammatory Response, Response to Wounding	<1.0 × 10 <sup>-12</sup>
Midnight Blue	82	Magnesium Ion Binding	<0.05
Green	791	Cilia	<2 × 10 <sup>-4</sup>
Turquoise	6050	Intracellular Organelle, Membrane-Enclosed Lumen	<2 × 10 <sup>-7</sup>
Dark Green	43	None Identified	N/A
Black	452	Cell to Cell Signaling, Cell Membrane	<1.0 × 10 <sup>-5</sup>
Magenta	365	Oxidation/Reduction, Microsomes	<2.0 × 10 <sup>-2</sup>
Pink	427	Immune Response, Immunoglobulin, Antigen Presentation	<1.0 × 10 <sup>-3</sup>
Brown	1274	Glycoprotein, Plasma Membrane, Immune Response	<1.0 × 10 <sup>-6</sup>
Blue	5675	Nucleus, Transcription Regulation, Nuclear Lumen	<2 × 10 <sup>-4</sup>
Grey	5861	Olfactory Transduction	<4.9 × 10 <sup>-42</sup>
Green Yellow	245	Immunoglobulin, Antigen Processing and Presentation	<1 × 10 <sup>-4</sup>
Light Green	63	None Identified	N/A

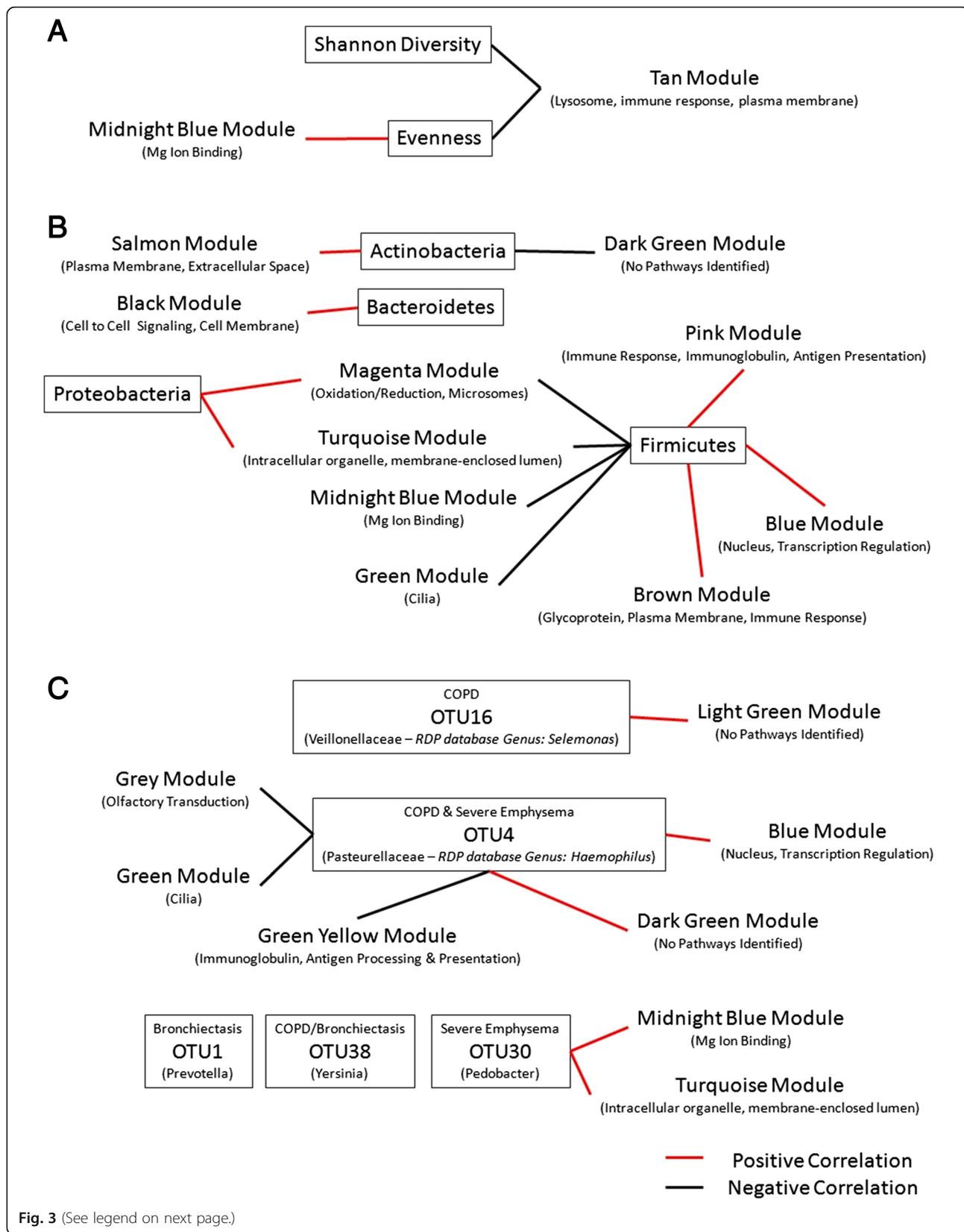


Fig. 3 (See legend on next page.)

(See figure on previous page.)

**Fig. 3** Network of the module correlations with bacterial microbiome measures. **a** Gene expression modules and alpha diversity measures. **b** Gene expression modules and bacterial phyla. **c** Gene expression modules and important OTUs for COPD, severe emphysema, and bronchiectasis. In brackets under each module is a brief description of pathways identified by DAVID for genes in the module. Red represents significant positive correlations while black represents significant negative correlations

Our findings may be consistent with previous studies on the lung microbiome in HIV, which found *Tropheryma whippelii* as a discriminative bacterium in bronchoalveolar lavage fluid (BALF), occurring in 13.4 % of HIV subjects versus only 1.3 % of HIV-uninfected subjects [4]. In our study, which used bronchial brushes rather than BALF, we demonstrated the presence of *Tropheryma* in 23.4 % of the samples. Although these data are in line with the previous literature, additional molecular studies such as a qPCR assay targeting a gene specific for the species would be needed to confirm that the *Tropheryma* we identified was indeed *T.whippelii*. We extend the previous findings by characterizing the host gene expression response to the bacterial microbiome. For instance, we found that Shannon Diversity and evenness were negatively correlated with genes involved with lysosome formation and immune response. This finding is consistent with the evolving concept that reduction in bacterial diversity is associated with an elevated risk of clinical infection and increased inflammatory response by the host [26, 27]. It should be noted that certain organisms independent of their numbers, are more likely to elicit an inflammatory response compared with others that are less “pathogenic”. For instance, although in our study we found that bacteria in the Actinobacteria phylum were significantly more abundant in HIV lungs than in the background negative environmental controls, these bacteria were not significantly associated with gene expression modules. In contrast, bacteria in the Firmicutes phyla (though less abundant compared with Actinobacteria) were significantly associated with several different gene expression modules. Firmicutes were negatively related to pathways governing cilium and positively associated with gene expression modules associated with immune response and transcription regulation. When we explored all correlations that had a *p*-value under 0.05 [Fig. 3] the Proteobacteria phylum was positively associated with gene expression pathways related to oxidation/reduction and intracellular orgnaelles, whereas the Firmicutes phylum was negatively associated with these pathways. This data is consistent with a previous study which reported a natural antagonism between the Firmicutes and Proteobacteria phyla in the oropharynx [28]. These data are also consistent with the evolving concept that the lung microbiome is propagated by upper airway seeding [29, 30]. We speculate that the host immune response is regulated in the HIV lung by the seeding of certain organisms from the upper airways into the lower airway tract. We posit that the predominance of Firmicutes phylum

leads to a heightened inflammatory state. Additional studies into the host interactions with the bacterial microbiome within the lung will need to be completed to confirm this hypothesis.

Most intriguingly, we found that OTU4 (*Pasteurellaceae-Haemophilus*) was predictive of both COPD (by spirometry) and severe emphysema (by CT). Previous studies suggest that *Haemophilus influenzae* is an important pathogen in COPD [31] and a recent study using whole lung tissue has shown that this organism is found in control subjects but not in patients with GOLD 4 severity [7]. Consistent with this observation, in our study *Haemophilus spp* was found in airways of patients without COPD by spirometry and without significant emphysema on CT scan [Figs. 2 and 3], although there were no correlations under and  $FDR < 0.05$  when exploring OTU correlations with gene expression modules. Intriguingly, those correlations that were under a *p* value of 0.05 showed that OTU4 negatively correlated with both pathways involved with cilia and antigen processing and presentation [Fig. 3]. This raises the tantalizing hypothesis that up-regulation of immune genes which activate the adaptive immune processes may enable processing and removal of *Haemophilus spp* in the airways. However, this result would need to be validated in a study with more power to asses this relationship. Up-regulation of genes involved in cilia may have a similar effect. We speculate that COPD airways have altered immune and/or cilia function that may prevent effective clearance of *Haemophilus spp*. Additional work will be needed to validate this hypothesis.

There are several limitations to this study. First, the findings pertain exclusively to HIV-infected patients. Thus it is possible that these OTUs may not help to distinguish COPD, severe emphysema, or bronchiectasis in HIV-uninfected patients. However, a recent study suggests that the bacterial microbiome between HIV-uninfected patients and HIV patients on successful antiretroviral therapy may be similar [5]. Second, no oral wash was performed prior to the bronchoscopy. This could have led to minor contaminations of the bronchial brush samples. However, it was reassuring that the findings of the present study were consistent with others that used mouth rinsing procedures [6, 29, 32]. Thirdly, all patients enrolled in this study had a clinical indication for bronchoscopy. While great care was taken to sample epithelial cells from uninfected regions of the lung, and far away from nodules or masses, our results could be confounded by these

underlying conditions. However, we did not find the bacterial species identified by routine clinical culture in those patients with a diagnosis of pneumonia in the analysis of 16S, which would support that sampling was indeed from the unaffected portions of the lung. This would explain the fact that there was no difference in the bacterial community composition between those with and without pneumonia. Another limitation is that we were unable to validate the microarray expression results with RT-PCR due to the large size of many of the modules. Thus it is possible that some of the genes within the modules are not accurate. However, a module is based on more than one gene and in order for a module to be wrong the majority of gene expression values within it would have to be incorrect. Finally, it is possible that some of the gene expression could be accounted for by infiltrating immune cells that were taken along with the epithelial cells during sampling. We cannot conclusively rule this possibility out but samples in this study were obtained away from locations with signs of clear inflammation. Future studies in which the bronchial epithelial cell microbiome is assessed in asymptomatic HIV-infected individuals would help to clarify the relationships between the microbiome and host response, and in certain pulmonary phenotypes.

Overall, this study provides a preliminary investigation into the host gene expression interaction with the bacterial microbiome in the small airways of HIV infected individuals. It supports the hypothesis that diversity and evenness of the community are important in modulating inflammatory responses of the host. This study also shows how bacteria in some phyla and OTUs may be important in disease pathogenesis by modifying either the host response and/or ecological niche areas. Our work supports the possibility that specific interactions between the bacterial microbiome and host cells within the airways of the lung occur and may be associated with distinct disease phenotypes; these findings would require additional studies for validation.

## Conclusions

In summary this study demonstrates that the bacterial microbiome and host gene expression may interact with one another in individuals with HIV infection. It identifies pathways, such as the mucociliary transport system, as important in the interaction between host and bacterial microbiome. However, more research into this specific area needs to be accomplished to confirm these results and observations.

## Additional file

**Additional file 1: Table S1.** Alpha diversity measures used in the analysis. **Figure S1.** NMDS of HIV patients based on pneumonia status. **Figure S2.** NMDS of HIV patients based on CD4 counts below 200.

**Figure S3.** NMDS of HIV patients based on smoking status. **Figure S4.** NMDS of HIV patients based on whether Emphysema was present on CT scan. **Figure S5.** NMDS of HIV patients based on whether Tropheryma was present. **Figure S6.** Power threshold calculation for WGCNA. **Figure S7.** Outlier detection for WGCNA. A single sample was removed (BIDC25) since it clustered quite differently than the rest of the data set. **Figure S8.** Heatmap showing that BIDC25 is also very different than other samples versus traits of interest. **Figure S9.** Module trait relationship with alpha diversity measures and phyla. Green squares represent negative correlations while red squares represent positive correlations. The two numbers displayed in each square is the R-value correlation and P-value (number in brackets) respectively. **Figure S10.** Module trait relationship with the important OTUs identified. Green squares represent negative correlations while red squares represent positive correlations. The two numbers displayed in each square is the R-value correlations and P-value (number in brackets respectively). (DOC 430 kb)

## Abbreviations

16S: 16 Svedberg; BC: British Columbia; BIDC: Internal naming system for patient samples; COPD: Chronic Obstructive Pulmonary Disease; CT: Computed Tomography; DAVID: Database for Annotation Visualization and Integrated Discovery; ddPCR: Droplet digital Polymerase Chain Reaction; FDR: False discovery rate; FEV1: Forced expiratory volume in 1 second; FVC: Forced vital capacity; HAART: Highly active antiretroviral therapy; HIV: Human immunodeficiency virus; NCBI: National Center for Biotechnology Information; Neg: Negative; NMDS: Non-metric multidimensional scaling; OTU: Operational taxonomic unit; PCR: Polymerase chain reaction; PERMANOVA: Permutation analysis of variance; RMA: Robust multi-array average; rRNA: Ribosomal ribonucleic acid; SPH: Saint Paul's Hospital; SRP: Sequence read project; UBC: University of British Columbia; WGCNA: Weighted gene co-expression network analysis

## Acknowledgements

None.

## Funding

Funding was provided by the Canadian Institutes of Health Research grant number 342422 and the BC Lung Association.

## Availability of data and materials

Data for this study have been provided at <http://www.ncbi.nlm.nih.gov/Traces/study/> (search for SRP068430).

## Authors' contributions

Study conception and design: MAS, JML, EAV, SL, WL, DDS, SFP. Data acquisition: SX, JML, EAV, TS, AM, MH, SG, SS, CN, DM, CH, JL, SL, WL, JSM, DDS, SFP. Data analysis: MAS, JML, EAV. Manuscript writing: MAS, JML. Manuscript editing: MAS, SX, JML, EAC, MH, SG, DM, CH, JL, SL, WL, JSM, DDS, SFP. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Consent for publication

Consent to publish was not obtained, therefore key baseline demographic information was summarized to prevent identification of subjects.

## Ethics approval and consent to participate

All patients provided written informed consent to participate. The study was approved by the University of British Columbia Providence Health Care ethics board (Protocol H14-03267).

## Author details

<sup>1</sup>Centre for Heart Lung Innovation, St. Paul's Hospital & Department of Medicine, University of British Columbia, Rm 166 – 1081 Burrard St., Vancouver, BC V6Z 1Y6, Canada. <sup>2</sup>Department of Integrative Oncology, BC Cancer Research Centre, Vancouver, BC, Canada. <sup>3</sup>Division of Respiratory Medicine, St. Paul's Hospital, University of British Columbia, Vancouver, BC, Canada. <sup>4</sup>AIDS Research Program, St. Paul's Hospital, Vancouver, BC, Canada. <sup>5</sup>Department of Family Medicine, Faculty of Medicine, University of British

Columbia, Vancouver, BC, Canada. <sup>6</sup>Division of HIV/AIDS, Department of Medicine, University of British Columbia, Vancouver, BC, Canada. <sup>7</sup>Faculty of Pharmaceutical Sciences, Pharmaceutical Sciences Building, University of British Columbia, Vancouver, BC, Canada. <sup>8</sup>Department of Radiology and Diagnostic Imaging, St. Paul's Hospital, Vancouver, BC, Canada. <sup>9</sup>British Columbia Centre for Excellence in HIV/AIDS, St. Paul's Hospital, Vancouver, BC, Canada.

Received: 20 April 2016 Accepted: 27 October 2016

Published online: 09 November 2016

## References

- Crothers K, Butt AA, Gibert CL, Rodriguez-Barradas MC, Crystal S, Justice AC, et al. Increased COPD among HIV-positive compared to HIV-negative veterans. *Chest*. 2006;130(5):1326–33.
- Crothers K, Huang L, Goulet JL, Goetz MB, Brown ST, Rodriguez-Barradas MC, et al. HIV infection and risk for incident pulmonary diseases in the combination antiretroviral therapy era. *Am J Respir Crit Care Med*. 2011;183(3):388–95.
- Diaz PT, King MA, Pacht ER, Wewers MD, Gadek JE, Nagaraja HN, et al. Increased susceptibility to pulmonary emphysema among HIV-seropositive smokers. *Ann Intern Med*. 2000;132(5):369–72.
- Lozupone C, Cota-Gomez A, Palmer BE, Linderman DJ, Charlson ES, Sodergren E, et al. Widespread colonization of the lung by *Tropheryma whipplei* in HIV infection. *Am J Respir Crit Care Med*. 2013;187(10):1110–7.
- Beck JM, Schloss PD, Venkataraman A, Twigg III H, Jablonski KA, Bushman FD, et al. Multi-center Comparison of Lung and Oral Microbiomes of HIV-infected and HIV-uninfected Individuals. *Am J Respir Crit Care Med*. 2015;192(11):1335–44.
- Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med*. 2013;187(10):1067–75.
- Sze MA, Dimitriu PA, Suzuki M, McDonough JE, Campbell JD, Brothers JF, et al. The Host Response to the Lung Microbiome in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2015;192(4):438–45.
- Miller MR, Crapo R, Hankinson J, Brusasco V, Burgos F, Casaburi R, et al. General considerations for lung function testing. *Eur Respir J*. 2005;26(1):153–61.
- Kazerouni EA, Martinez FJ, Flint A, Jamadar DA, Gross BH, Spizarny DL, et al. Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring. *AJR Am J Roentgenol*. 1997;169(4):977–83.
- Sze MA, Abbasi M, Hogg JC, Sin DD. A comparison between droplet digital and quantitative PCR in the analysis of bacterial 16S load in lung tissue samples from control and COPD GOLD 2. *PLoS One*. 2014;9(10):e110351.
- Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protoc*. 2008;3(9):1452–6.
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. 2013;79(17):5112–20.
- Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 2003;14(6):927–30.
- Oksanen J, Blanchet GF, Kindt R, Legendre P, Minchin PR, O'Hara, RB, et al. Community Ecology Package [Internet]. 2015. Available from: <http://cran.r-project.org>, <https://github.com/vegandevs/vegan>.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537–41.
- Carvalho BS. Preprocessing Affymetrix Exon ST and Gene ST Arrays. Available from: <http://www.bioconductor.org/packages//2.7/bioc/vignettes/oligo/inst/doc/V5ExonGene.pdf>.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80.
- Breiman L. Random forests. *Machine Learning*. 2001;45(1):5–32.
- Kursa MB, Rudnicki WR. Feature selection with the boruta package [Internet]. *Journal*; 2010 [cited 2013 Feb 9]. Available from: <http://www.jstatsoft.org/v36/i11/paper/>. Accessed 19th June 2015.
- McArdle BH, Anderson MJ. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*. 2001;82(Journal Article):290–7.
- Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusk AJ, Horvath S. Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome*. 2007;18(6–7):463–72.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
- Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 2003;4(5):3.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73(16):5261–7.
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014;42(Database issue):D633–42.
- Horwitz D, McCue T, Mapes AC, Ajami NJ, Petrosino JF, Ramig RF, et al. Decreased microbiota diversity associated with urinary tract infection in a trial of bacterial interference. *J Infect*. 2015;71(3):358–67.
- Pérez-Cobas AE, Artacho A, Ott SJ, Moya A, Gosalbes MJ, Latorre A. Structural and functional changes in the gut microbiota associated to *Clostridium difficile* infection. *Front Microbiol*. 2014;5:335.
- Lemon KP, Klepac-Ceraj V, Schiffer HK, Brodie EL, Lynch SV, Kolter R. Comparative analyses of the bacterial microbiota of the human nostril and oropharynx. *MBio*. 2010;1(3). pii: e00129-10.
- Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *MBio*. 2015;6(2):e00037.
- Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Beck JM, Huffnagle GB, et al. Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung Biogeography. *Ann Am Thorac Soc*. 2015;12(6):821–30.
- Sethi S, Evans N, Grant BJ, Murphy TF. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. *N Engl J Med*. 2002;347(7):465–71.
- Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, et al. The lung tissue microbiome in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;185(10):1073–80.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

