

RESEARCH

Open Access



# Airway host-microbiome interactions in chronic obstructive pulmonary disease

Zhang Wang<sup>1,2</sup>, Barbara Maschera<sup>3</sup>, Simon Lea<sup>4</sup>, Umme Kolsum<sup>4</sup>, David Michalovich<sup>3</sup>, Stephanie Van Horn<sup>5</sup>, Christopher Traini<sup>5</sup>, James R. Brown<sup>1\*</sup> , Edith M. Hessel<sup>3†</sup> and Dave Singh<sup>4†</sup>

## Abstract

**Background:** Little is known about the interactions between the lung microbiome and host response in chronic obstructive pulmonary disease (COPD).

**Methods:** We performed a longitudinal 16S ribosomal RNA gene-based microbiome survey on 101 sputum samples from 16 healthy subjects and 43 COPD patients, along with characterization of host sputum transcriptome and proteome in COPD patients.

**Results:** Dysbiosis of sputum microbiome was observed with significantly increased relative abundance of *Moraxella* in COPD versus healthy subjects and during COPD exacerbations, and *Haemophilus* in COPD ex-smokers versus current smokers. Multivariate modeling on sputum microbiome, host transcriptome and proteome profiles revealed that significant associations between *Moraxella* and *Haemophilus*, host interferon and pro-inflammatory signaling pathways and neutrophilic inflammation predominated among airway host-microbiome interactions in COPD. While neutrophilia was positively correlated with *Haemophilus*, interferon signaling was more strongly linked to *Moraxella*. Moreover, while *Haemophilus* was significantly associated with host factors both in stable state and during exacerbations, *Moraxella*-associated host responses were primarily related to exacerbations.

**Conclusions:** Our study highlights a significant airway host-microbiome interplay associated with COPD inflammation and exacerbations. These findings indicate that *Haemophilus* and *Moraxella* influence different components of host immune response in COPD, and that novel therapeutic strategies should consider targeting these bacteria and their associated host pathways in COPD.

**Keywords:** Chronic obstructive pulmonary disease, COPD, Microbiome, Exacerbations, Clinical study, Transcriptome, Proteome, Healthy, Smokers, Next-generation sequencing technologies

## Background

Chronic obstructive pulmonary disease (COPD) is a heterogeneous lung disease in which recurrent bacterial infections are a major etiological factor [1–4]. The human microbiome in the respiratory tract differs between healthy subjects and COPD patients [5–7], shifts in composition during COPD exacerbations [8–12] and varies among exacerbation subtypes [9], all suggesting a close association between the lung microbiome and COPD pathophysiology with potential involvement of

host immunity and inflammatory responses. It is thought that disruption of microbiome, known as dysbiosis, could trigger a dysregulated host immune response that results in infection susceptibility, inflammation and negative effects on host biology [13].

A systematic understanding of airway host-microbiome interaction in relation to COPD pathogenesis could provide the mechanistic basis for modulation of host-microbe interactions as a potential novel therapeutic strategy for COPD. A previous study on COPD patients showed that the lung microbiome was significantly associated with sputum pro-inflammatory markers especially interleukin-8 (IL8/CXCL-8, 9). In particular, there is a significant correlation between sputum interleukin-8 (IL-8/CXCL-8) with both alpha and beta diversity of the airway

\* Correspondence: [James.R.Brown@gsk.com](mailto:James.R.Brown@gsk.com)

†Edith M. Hessel and Dave Singh are Co-senior authors

<sup>1</sup>Computational Biology, Human Genetics, Research and Development (R&D), GlaxoSmithKline (GSK), 1250 S. Collegeville Road, Collegeville, PA 19426-0989, USA

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



microbiome in COPD. In the correlation network, sputum IL-8/CXCL-8 showed the highest degree of microbiota connectivity with a significant negative correlation to 15 bacterial operational taxonomic units (OTUs), suggesting sputum IL-8/CXCL-8 could be an indicator of microbiome community structure and diversity.

Few studies have simultaneously characterized both lung microbiome and human multi-omics profiles in COPD, and in other respiratory diseases in general. Sze et al. measured the lung microbiome and host transcriptome in COPD and found Firmicutes and Proteobacteria were associated with different host gene expression profiles [14]. Molyneaux et al. profiled both lung microbiome and peripheral whole-blood transcriptome for idiopathic pulmonary fibrosis patients and identified two gene modules involved in host defense that are strongly associated with the microbiome profile [15]. However, a comprehensive understanding of the collective host response at both transcriptional and protein expression levels to the lung microbiome community is lacking. A systems biology approach integrating lung microbiome and host multi-omics datasets is necessary to better understand host-microbiome interactions in COPD.

Here we performed a 16S ribosomal RNA (rRNA) gene-based survey on sputum microbiome from 16 healthy subjects and 43 COPD patients. Host sputum cell counts, transcriptome and proteome were also characterized for COPD patients. To our knowledge, this is the first study that characterizes both lung microbiome and host transcriptome and proteome profiles in stable COPD and during exacerbations. We found significant interplay between lung microbiome composition and host response in COPD that is potentially important to current treatments and future therapeutic strategies.

## Methods

### Patient selection

The presented study was conducted in accordance with the Declaration of Helsinki [16] and Good Clinical Practice [17]. The human biological samples were sourced ethically and in accord with the terms of the informed consents under the University of Manchester and University Hospital of South Manchester IRB/EC approved protocol (Approval number: 10/H1003/108).

Healthy subjects and COPD patients were enrolled at the Medicines Evaluation Unit (Manchester University Foundation NHS Trust Hospital). Patients with asthma, or significant respiratory disease other than COPD, or the inability to produce sputum after sputum induction were excluded from the study. Patients were seen at stable at least 6 weeks after the use of any short term antibiotics. Patients contacted the research team if they experienced a change in symptoms consistent with an acute exacerbation. Daily diary cards were used. Patients

were assessed by a clinician and exacerbations defined as an increase in respiratory symptoms for two consecutive days. Smoking status, historical exacerbation frequency, GOLD status, inhaled corticosteroid (ICS) administration, Quality of Life (QoL) scores and lung function measurements (FEV<sub>1</sub>, FVC and FEV<sub>1</sub>/FVC ratio) were recorded for COPD patients (Table 1, Additional file 1: Table S1). Smoking status and lung function measurements were recorded for healthy subjects.

### Sputum collection

Sputum samples were collected at a single time-point from 16 healthy subjects and longitudinally from 43 COPD patients. Sputum sampling were performed prior to any systemic therapy including treatment with oral corticosteroids and/or antibiotics. Sputum samples were obtained by spontaneous expectoration or induced. For COPD patients, spontaneous expectoration was attempted first, if no sputum or too little sputum was produced, induction was then performed. For healthy subjects, only induction method was performed. Sputum samples from COPD patients were collected at stable (defined as no evidence of symptom-defined exacerbations in the preceding 4 weeks and the subsequent 2 weeks post-clinic visit), exacerbations (defined according to Anthonisen criteria [18] and/or healthcare utilization [19]), two and 6 week post-exacerbations and 6 months from first stable visit. All exacerbation sputum samples were collected prior to the institution of any exacerbation treatment. The missing samples are mostly due to patients unable to produce sufficient amount of sputum for downstream experiments (Additional file 1: Figure S1).

### Sputum processing

Sputum samples were processed to obtain cell pellets and supernatant, for immune cell counting, host transcriptome and proteome analysis, according to a previous method [20]. Briefly, sputum plugs were selected from saliva and put on ice (minimum weight 0.1 g). Eight times volume of phosphate-buffered saline (PBS) was added to the sputum. The mixture was incubated in a roller mixer for 15 min on ice, vortexed every 5 min and centrifuged at 790 g for 10 min. The supernatant was split into aliquots and stored at -80 °C for sputum proteome analysis. For cell pellets, a four-fold volume of 0.2% DTT was added and the mixture was incubated for 15 min in a roller mixer on ice, vortexed every 5 min, filtered using 48 µm nylon-mesh filter and centrifuged. Cell pellets were resuspended in 1 ml of PBS to perform haemocytometer cell counts, cyto-spin differential cell counts and stored at -80 °C for transcriptomic assays.

### Microbiome 16S rRNA gene sequencing

For quality control purposes, bacterial DNA extractions, sequencing and data analyses were performed in a single,

**Table 1** Major demographic and baseline clinical features of all subjects in this study

Demographic and baseline clinical features	Healthy Controls (N = 16)	COPD Patients (N = 43)
Age, years <sup>a</sup>	55.0 (8.8)	65.0 (4.8)
Gender, n (% Male) <sup>b</sup>	11 (68.8)	31 (72.1)
Current Smoker, n (%)	8 (50.0)	16 (37.2)
Number of cigarette packs per year	34.4 (11.2)	50.9 (28.2)
GOLD Stage: I/II/III/IV, n (%)	NA	2 (4.7)/19 (44.2)/16 (37.2)/6 (14.0)
Inhaled steroid use, n (%)	NA	33 (76.7)
LABA use, n (%)	NA	34 (79.1)
LAMA use, n (%)	NA	35 (81.4)
Number of exacerbations per year	NA	1.9 (1.4)
CAT score	NA	21.8 (8.3)
SGRQ total score	NA	49.0 (22.9)
mMRC score	NA	2.1 (1.2)
Pre FEV <sub>1</sub> (L)	3.0 (0.7)	1.1 (0.5)
Pre FVC (L)	4.1 (1.0)	2.8 (0.9)
Pre FEV <sub>1</sub> /FVC ratio	0.73 (0.0)	0.4 (0.1)
Post FEV <sub>1</sub> (L)	3.1 (0.8)	1.2 (0.6)
Post FVC (L)	4.0 (0.9)	3.3 (1.0)
Post FEV <sub>1</sub> /FVC ratio	0.76 (0.0)	0.4 (0.1)

<sup>a</sup> Continuous data present as mean (SD)

<sup>b</sup> Categorical data present as number (proportion)

GOLD Global Initiative for Chronic Obstructive Lung Disease, LABA long-acting beta-agonist, LAM long-acting muscarinic antagonist, CAT COPD Assessment Test, SGRQ St. George's Respiratory Questionnaire, mMRC modified Medical Research Council, FEV<sub>1</sub> forced expiratory volume in one second, FVC forced vital capacity

centralised lab at the GlaxoSmithKline (GSK) R&D facility in Collegeville, PA, USA. The detailed procedure of bacterial genomic DNA isolation, 16S library preparation, sequencing, reagent controls, and sequence data processing was provided in the supplementary material of our previous study [11]. Bacterial genomic DNA was extracted from healthy and COPD sputum samples using Qiagen DNA Mini kit. The variable 4 (V4) region of the 16S rRNA gene was PCR-amplified with the appropriate reagent controls [9, 11], and was sequenced using Illumina Miseq. The demultiplexed and quality-filtered sequencing reads were subject to open-reference operational taxonomic unit picking (97% identity cutoff) using QIIME 1.9 [21].

Seven OTUs were detected with > 10 sequencing reads in the negative reagent controls (Additional file 1: Table S2). Although negative reagent controls were performed for all DNA isolation, extraction and PCR amplification step, we performed further analyses to ensure that potential contamination risks were minimized. We compared our results against the 92 contaminant genera detected in sequenced negative 'blank' controls by Salter et al. [22]. We failed to detect 56 out of the 92 contaminant genera in our dataset (Additional file 1: Table S3). Of the remaining genera that were found in our data, none had an average relative abundance greater than 0.0004, or had a relative abundance greater than 0.1 in a particular

sample, except for *Streptococcus* which contains known lung pathogens.

#### Bacterial qPCR assays

All qPCR assays were performed using 384-well microbial DNA qPCR arrays (Qiagen, Germantown, MD) on a QuantStudio 12 K Flex Real-Time PCR System (Life Technologies, Carlsbad, California, USA). The 10 µl reaction mixture contained 5 µl of Microbial qPCR master mix with ROX and 5 µl of Microbial-free water (Qiagen, Germantown, MD). Each well was spotted with a mix of two PCR primers (10 µM each) and one 5'-hydrolysis probe (5 µM) with 10 ng of added sample DNA. The following cycling parameters were used: initial cycle of 95 °C for 10 min; 40 cycles of 95 °C for 15 s; and 60 °C for 2 min. All qPCR templates were run in duplicate and tested for amplification inhibition by use of a positive PCR Control (Qiagen, Germantown, Maryland, USA). For standard curve calculation, each plate run included a decimal serial dilution of double-stranded DNA oligos (Integrated DNA Technologies, Skokie, Illinois, USA) designed from the 16S rRNA gene of pan-bacteria, *Haemophilus influenzae*, *Moraxella catarrhalis*, *Streptococcus pneumoniae*, *Prevotella melaninogenica* and *Veillonella dispar*. The cycle threshold values and DNA copy numbers were calculated using the QuantStudio

12 K Flex software (Life Technologies, Carlsbad, California, USA).

### Host RNA microarray analysis

Host transcriptome was profiled for 38 COPD sputum samples (Additional file 1: Figure S1). Total RNA was extracted using Trizol reagent (Invitrogen) from sputum cell pellets and further purified with a RNeasy mini kit (Qiagen, Valencia, California, USA) according to the manufacturer's instructions. RNA quality was evaluated on the Agilent 2100 Bioanalyzer and quantitated by OD260. For samples passing RNA QC criteria (RIN > 5.5, A260/280 value 1.6–2.4, total RNA > 50 ng, presence of distinct 28S and 18S ribosomal RNA peaks), 50 ng RNA was used for NuGEN amplification and labeling of probes using the NuGEN Ovation RNA Amplification System (NuGEN Technologies). The amplified sscDNA was purified using the Agencourt RNAClean magnetic bead clean-up system. The sscDNA samples were quantified by spectrophotometry and profiled on Agilent 2100 Bioanalyser prior to array hybridization. The array hybridization was performed using Affymetrix GeneChip HG-U133 Plus 2.0 microarray (Affymetrix, Santa Clara, California, USA), which contains 54,675 probe-sets interrogating 50,155 human transcripts. The raw microarray data (CEL files) were corrected for background signal, quantile normalized and summarized using robust multiarray average (RMA) normalization to generate probe-set-level microarray data using Array Studio v10.0 (OmicSoft, Cary, North Carolina, USA). The probe-set-level microarray data were log<sub>2</sub> transformed and converted to gene-level (24,442 genes) by selecting the probe with greatest inter-quantile range for its corresponding gene as suggested previously [23]. The microbiome and microarray data are deposited at the National Centre for Biotechnology Information Sequence Read Archive (SRP136124) and Gene Expression Omnibus databases (GSE112165), respectively.

### Proteomic assays

Host proteome was characterized for 37 sputum samples using the SOMAscan<sup>®</sup> platform (Somalogic, Additional file 1: Figure S1). The SOMAscan<sup>®</sup> assay has been described in detail previously [24–26]. The assay quantitatively transforms the proteins present in a biological sample into a specific SOMAmer-based DNA signal. Briefly, each SOMAmer<sup>®</sup> reagent binds a target protein (in total 1310 proteins) and is quantified on a custom Agilent microarray hybridization chip. Normalization and calibration were performed according to SOMAscan<sup>®</sup> Data Standardization and File Specification Technical Note (SSM-020). The output of the SOMAscan<sup>®</sup> assay was reported in relative fluorescent units and was log<sub>2</sub> transformed for downstream analysis.

### Statistical analysis

Differentially represented bacterial taxa were identified using edgeR [27]. Differentially expressed genes in microarray were identified using limma in R bioconductor [28] and were enriched for pathways using MetaCore v5.0 (Thomson Reuters). Multivariate modeling was performed to associate microbiome, host transcriptome and proteome data. The power calculation was performed following the procedure of Morgan et al. [29]. Specifically, correlated variable pairs were simulated with standard normal distribution and a sample size of 40, the number of samples that have both transcriptome/proteome and microbiome data. The 80th percentile of raw *P*-values of the Spearman correlation test was calculated as a function of true covariance of the variables. The number of allowable tests for 80% power and 5% type I error rate was estimated by Bonferroni correction, which is 0.05 divided by the 80th percentile of raw *P*-values calculated as above.

To reduce dimensionality, a Principal Component Analysis (PCA) was performed on the gene-level microarray data. PCA was also performed on the Somalogic proteome profile of 1310 proteins. The transcriptome Principal Components (tPCs) and proteomic Principal Components (pPCs) with proportion of variance > 2% were selected for association testing. For the microbiome datasets, 9 bacterial genera with average relative abundance > 1% were selected for association testing. A variance-stabilizing arcsin square root transformation was applied to the microbiome proportion data. All continuous variables were scaled to unit variance. For each genus and Shannon diversity, a general linear mixed model (GLMM) was established associating the variable with tPCs and pPCs adjusting for timepoints and patient demographic factors including smoking status, GOLD status and exacerbation frequency using lme4 in R [30]. Subject ID was included as a random variable to adjust for multiple measures per subject. The model was optimized in terms of Akaike information criterion (AIC) through backward elimination of non-significant effects in a stepwise algorithm using the “step” function in the R lmerTest package [31]. The same GLMM was applied for associating each pPCs with all tPCs. For association within stable samples, as no repeated measures were involved, a general linear model (GLM) was established using glm in R [32]. The model was optimized in terms of AIC through backward elimination of non-significant effects in a stepwise algorithm using the “step” function in the R stats package [32].

To assess functional enrichment of each tPC, all 24,442 genes were ranked by their loadings in that tPC and a Gene Set Enrichment Analysis (GSEA) [33] was performed on the ranked gene list using concatenated MetaCore (GeneGo), KEGG, Reactome, BioCarta and Pathway Interaction Database (PID) pathways (a total of

2809 pathways) using GSEAP reranked 6.0.10 in GenePattern [34]. The enrichment scoring scheme was set to 'classic' as suggested in the program instructions. One thousand permutations were performed for each run. Gene sets larger than 500 genes or smaller than 15 genes were excluded from the analysis.

The false discovery rate (FDR) method was used to adjust *P*-values for multiple testing wherever applicable [35].

## Results

### Sputum microbiome between healthy and COPD and during exacerbations

Sputum microbiome was characterized for 101 sputum samples (Fig. 1a) from COPD patients (Fig. 1b) and healthy controls (Fig. 1c). A total of 16,386,538 reads were generated after demultiplexing and quality filtering. After rarefaction to 89,462 reads per sample, 2807 OTUs were identified among all samples. Similar to other lung microbiome studies [5, 6, 8–10, 36–39], the majority of OTUs belonged to Firmicutes (53.6%), Bacteroidetes (21.9%) and Proteobacteria (19.5%) at the phylum level, and *Veillonella* (37.7%), *Prevotella* (15.3%), *Haemophilus* (14.0%), *Streptococcus* (8.6%) and *Moraxella* (2.9%) at the genus level. Quantitative PCR showed significant correlations between the absolute quantities of all five species and their relative abundances in the microbiome data (Spearman's rho  $\geq 0.43$ , FDR  $P = 2.57E-3$ ).

Significantly increased relative abundance of *Haemophilus* was observed in healthy smokers versus non-smokers ( $\log_2FC = 3.36$ , FDR  $P = 0.041$ ), and in COPD ex-smokers versus current smokers ( $\log_2FC = 2.49$ , FDR  $P = 0.025$ , Fig. 2a). Comparison of the microbiome profiles between healthy subjects and stable COPD patients showed a significantly increased relative abundance of the genera *Moraxella*, *Streptococcus* and Actinobacteria ( $\log_2$ Fold Change ( $\log_2FC$ )  $\geq 1.32$ , FDR  $P = 0.026$ , Additional file 1: Table S4) and decreased alpha diversity (Shannon,  $P = 0.036$ ) in stable COPD patients (Fig. 2b). A significantly increased *Moraxella* was observed at stable state in GOLD III versus II patients and in inhaled corticosteroids (ICS) versus non-ICS exposed patients (Additional file 1: Figure S2).

During COPD exacerbations, increased *Moraxella* ( $\log_2FC = 3.14$ , FDR  $P = 0.019$ ) and decreased alpha diversity was observed compared to stable state (unpaired analysis, Fig. 2b, paired analysis see Additional file 1: Figure S3), along with significantly increased neutrophil and decreased macrophage percentage ( $FC \geq 1.2$ ,  $P \leq 0.05$ , Additional file 1: Figures S4–S5). A non-significant increase of total bacterial load was observed during exacerbations (Additional file 1: Figure S6). Conversely, the trend of increased *Moraxella* and decreased alpha diversity was reversed at post-exacerbation time points (Fig. 2b).

Sputum neutrophil counts were most significantly associated with microbiome compositions, with positive correlations with *Haemophilus* and *Neisseria*, and negative correlations with *Streptococcus*, *Megasphaera* and *Veillonella* across all samples (Spearman's rho = 0.33, FDR  $P \leq 0.05$ , Fig. 3, Additional file 1: Table S5). The significant correlation between *Haemophilus* and sputum neutrophil count was further confirmed by qPCR (Spearman's rho = 0.37,  $P = 0.037$ , Additional file 1: Table S6). No bacterial taxa or sputum cell counts were associated with QoL scores, FEV<sub>1</sub> or FVC.

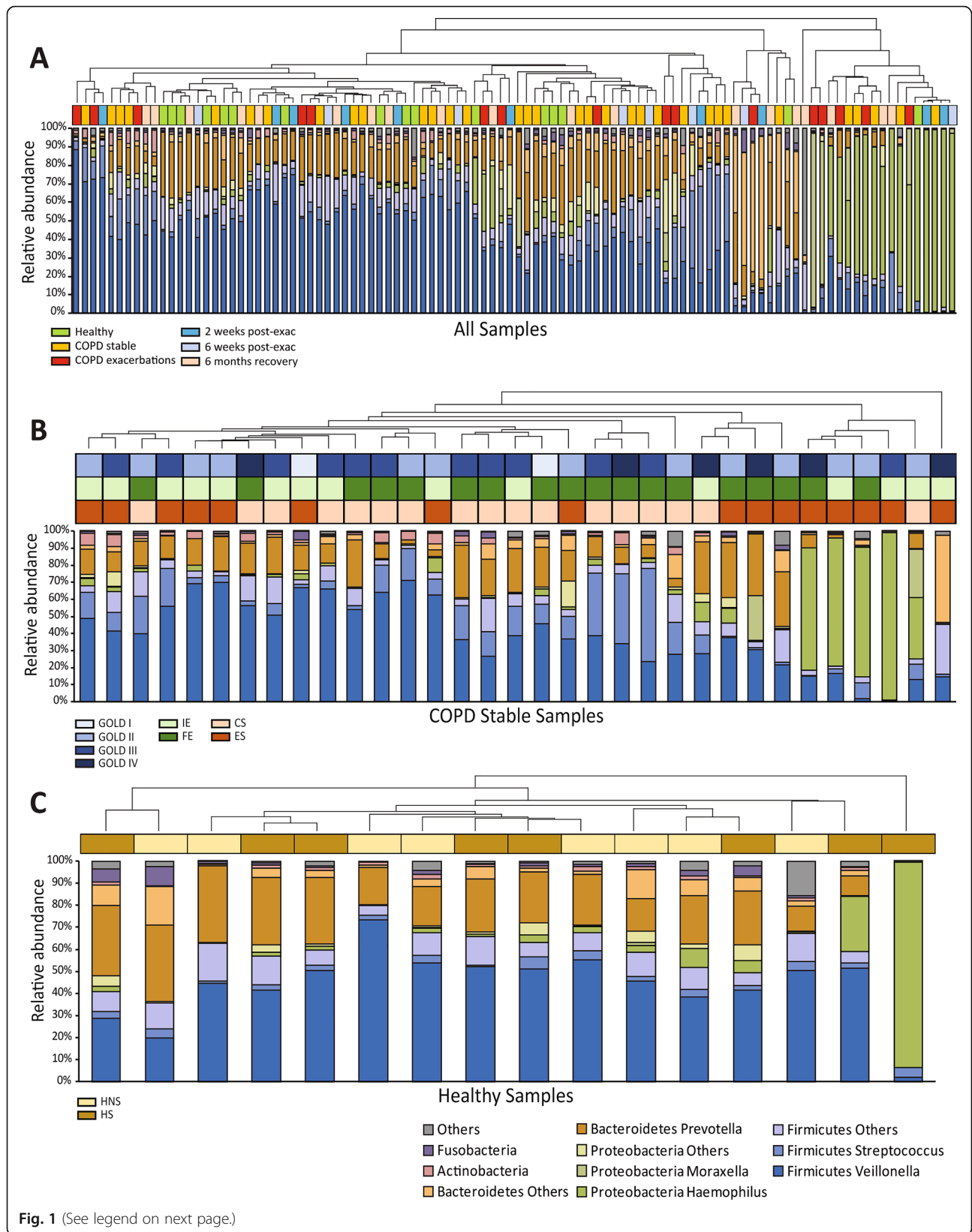
### Host transcriptome and proteome at COPD stable state and exacerbations

We compared host transcriptome differences between COPD stable and exacerbations. A substantial amount of 2453 upregulated and 4814 downregulated differentially expressed genes (DEGs) were identified at exacerbations versus stable state ( $FC \geq 1.5$ , FDR  $P \leq 0.05$ ), in which 239 and 8 MetaCore pathways were significantly enriched respectively (FDR  $P \leq 0.01$ , Additional file 2). A large proportion of the upregulated pathways were involved in immune response with the top pathways being interferon and interleukin-6 signaling pathways. The downregulated pathways included cell cycle, nucleotide metabolism and phagocytosis pathways. No DEGs were found between stable patient subgroups according to clinical characteristics (GOLD stage, smoking status, ICS administration and exacerbation frequency).

For patient proteome data, 790 of the 1310 proteins had significantly higher expression levels in stable COPD ex-smokers compared to current smokers, including multiple pro-inflammatory markers such as interleukin-36, fibrinogen and matrix metalloproteinase 10 ( $FC \geq 1.5$ , FDR  $P \leq 0.05$ , Mann-Whitney-Wilcoxon test, Additional file 2). No differentially expressed proteins were identified for other comparisons.

### *Haemophilus* and *Moraxella* are most significantly associated with host transcriptome and proteome

To gain insights into airway host-microbiome interactions in COPD, we established a multivariate linear model between microbiome, host transcriptome and proteome profiles across all samples (including exacerbations) and within stable samples only. We first performed a power estimation and calculated that given a true covariance of 0.5 between bacterial taxa and gene expression in 40 samples (the number of samples with both transcriptome/proteome and microbiome data), it would be possible to perform a maximum of  $10^2$  pairwise tests (or approximately 10 microbiome and 10 host expression factors) and retain 80% power and an alpha of 0.05 using Bonferroni correction (Additional file 1: Figure S7). As it is impossible for significant associations



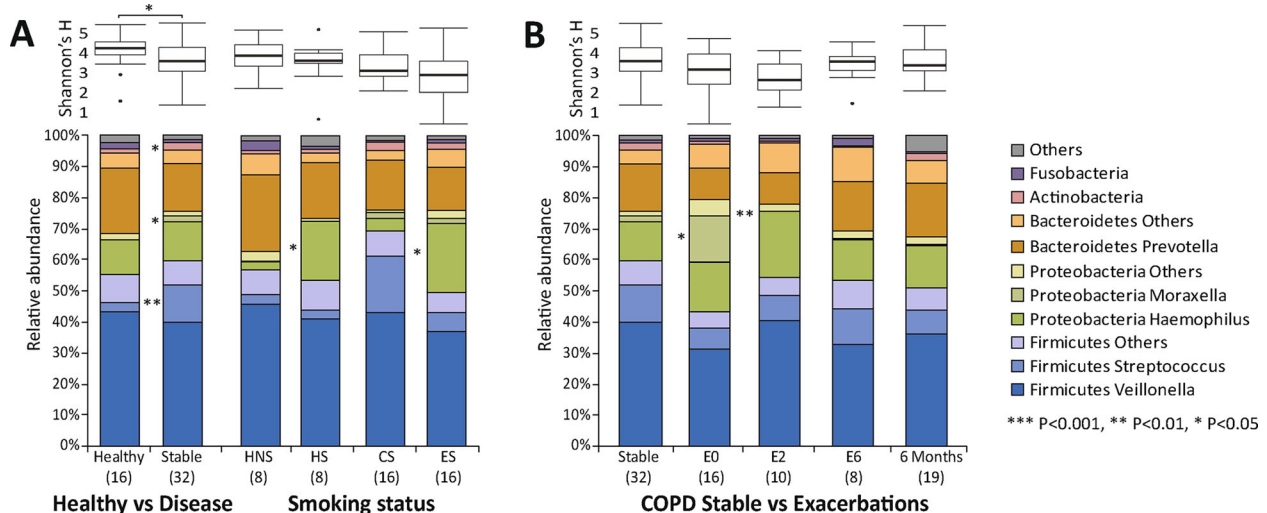
(See figure on previous page.)

**Fig. 1** Overview of the sputum microbiome taxa distributions. **a** Overall study clustering for all 101 samples. **b** Clustering of 32 COPD stable samples. **c** Clustering of 16 healthy samples. Each column represents one sample colored by different subgroups. Y-axis represents relative abundances of major phyla and genera. Samples were clustered by UPGMA clustering based on the weighted UniFrac distances. HNS: healthy non-smokers, HS: healthy smokers, CS: COPD current smokers, ES: COPD ex-smokers, non-ICS: non-ICS exposer, ICS: ICS exposer, IE: infrequent exacerbators, FE: frequent exacerbators

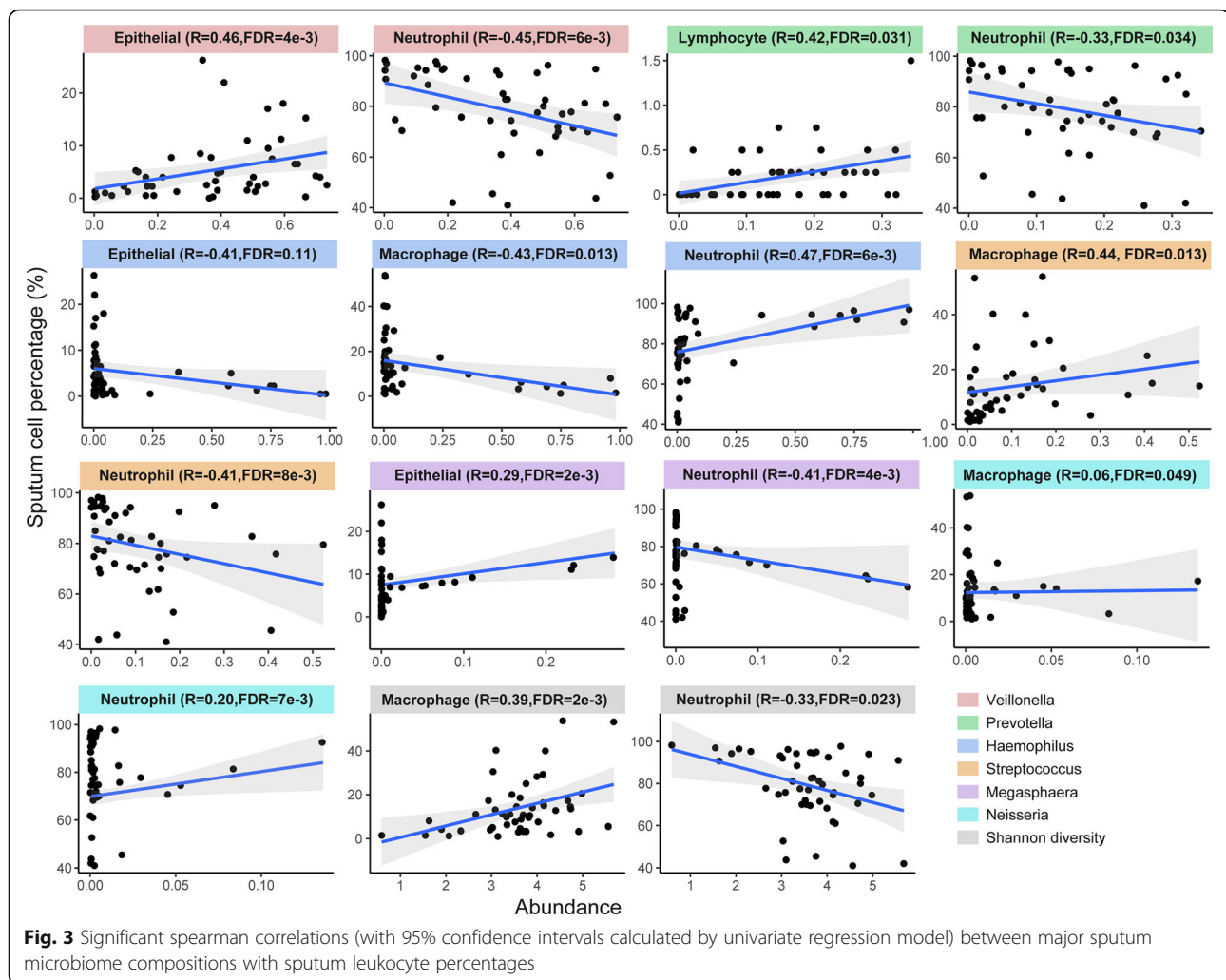
to survive correction for multiple testing of ~ 20,000 human genes, we performed an unsupervised dimensionality reduction on host multi-omics data using PCA. A total of 9 transcriptome and 8 proteome PCs (tPCs and pPCs, respectively) with proportion of variance > 2% were selected, together explaining 72 and 84% of observed variance. Using all samples (including exacerbations), a GLMM was established between each of the 9 major bacterial genera and all tPCs or pPCs, adjusting for different timepoints and patient demographic factors. Among all genera, *Haemophilus* and *Moraxella* were most strongly associated with host factors, in particular strong positive correlations with tPC2 and tPC4, respectively (FDR  $P < 5.0E-4$ , Fig. 4a, Table 2, Additional file 3).

Examining the top loadings of tPC2 and tPC4 revealed that they reflected increased expression of some immune response genes, such as interleukin-1 receptor-associated kinase 1 (*IRAK1*), interleukin-18 binding protein (*IL18BP*), linker for activation of T-cells family member 1 (*LAT*) for tPC2, and several interferon genes for tPC4 (Additional file 1: Figure S8, Additional file 4), indicating that high level of these two tPCs might correspond to increased immune activities. We performed GSEA on the loadings of

each tPC to further understand its functional properties. Both tPC2 and tPC4 were most significantly positively enriched for host immune response pathways. Several T-cell differentiation (i.e. Th1 and Th2 cell differentiation) and pro-inflammatory cytokine (i.e. IL-12) signaling pathways were among the top positive pathways for tPC2, while interferon signaling pathways were the top positive pathways for tPC4 (Fig. 4b, Additional file 4). Individual genes in the top pathways of tPC2 and tPC4 showed consistent correlations with *Haemophilus* and *Moraxella* respectively in both microbiome and qPCR datasets (Additional file 5), further supporting the associations in the multivariate models. Furthermore, both *Haemophilus* and tPC2 exhibited positive correlations with pPC3 (FDR  $P = 3.1E-3$ , Table 3, Additional file 3), together forming an interconnected subnetwork (Fig. 4a). Likewise, both *Moraxella* and tPC4 were positively correlated with pPC1 (FDR  $P = 0.014$ , Additional file 3). Multiple pro-inflammatory markers such as interleukin-1 receptor, matrix metalloproteinase 7, galectin-2 and TNF-related weak inducer of apoptosis were among the top loadings for pPC1 or pPC3 (Fig. 4c). In addition, several known bronchial epithelial cell receptors that respond to bacterial lipopolysaccharide (LPS)



**Fig. 2** Sputum microbiome profiles in healthy subjects and COPD patients. **a** Shannon diversity and relative abundance of major bacterial taxa in healthy controls and stable COPD patients, and in healthy and COPD subgroups in relation to smoking status. **b** Shannon diversity and relative abundances of major bacterial taxa in COPD patients at different visits. The number of samples in each group is indicated in the parenthesis. Significantly differentially represented bacterial taxa were identified using edgeR [27]. For visit, statistical analysis was performed on each adjacent two time points. E0: COPD exacerbations, E2: 2 week post-exacerbations, E6: 6 week post-exacerbations, 6 Months: 6 months from first stable visit, HNS: healthy non-smokers, HS: healthy smokers, CS: COPD current smokers, ES: COPD ex-smokers. \*\*\* FDR  $P < 0.001$ , \*\* FDR  $P < 0.01$ , \* FDR  $P < 0.05$



such as angiotensin-1 receptor [40] and Ephrin-A2 [41] were among the top loadings for the two pPCs (Additional file 4). The correlations of *Haemophilus*-tPC2-pPC3 and *Moraxella*-tPC4-pPC1 were further confirmed by qPCR ( $P \leq 0.1$ , Additional file 1: Table S6). In addition, both tPC4 and pPC1 were increased at exacerbations versus stable (Additional file 1: Figure S9). Both tPC2 and pPC3 were significantly positively correlated with sputum neutrophil counts (FDR  $P \leq 0.05$ , Additional file 1: Table S5).

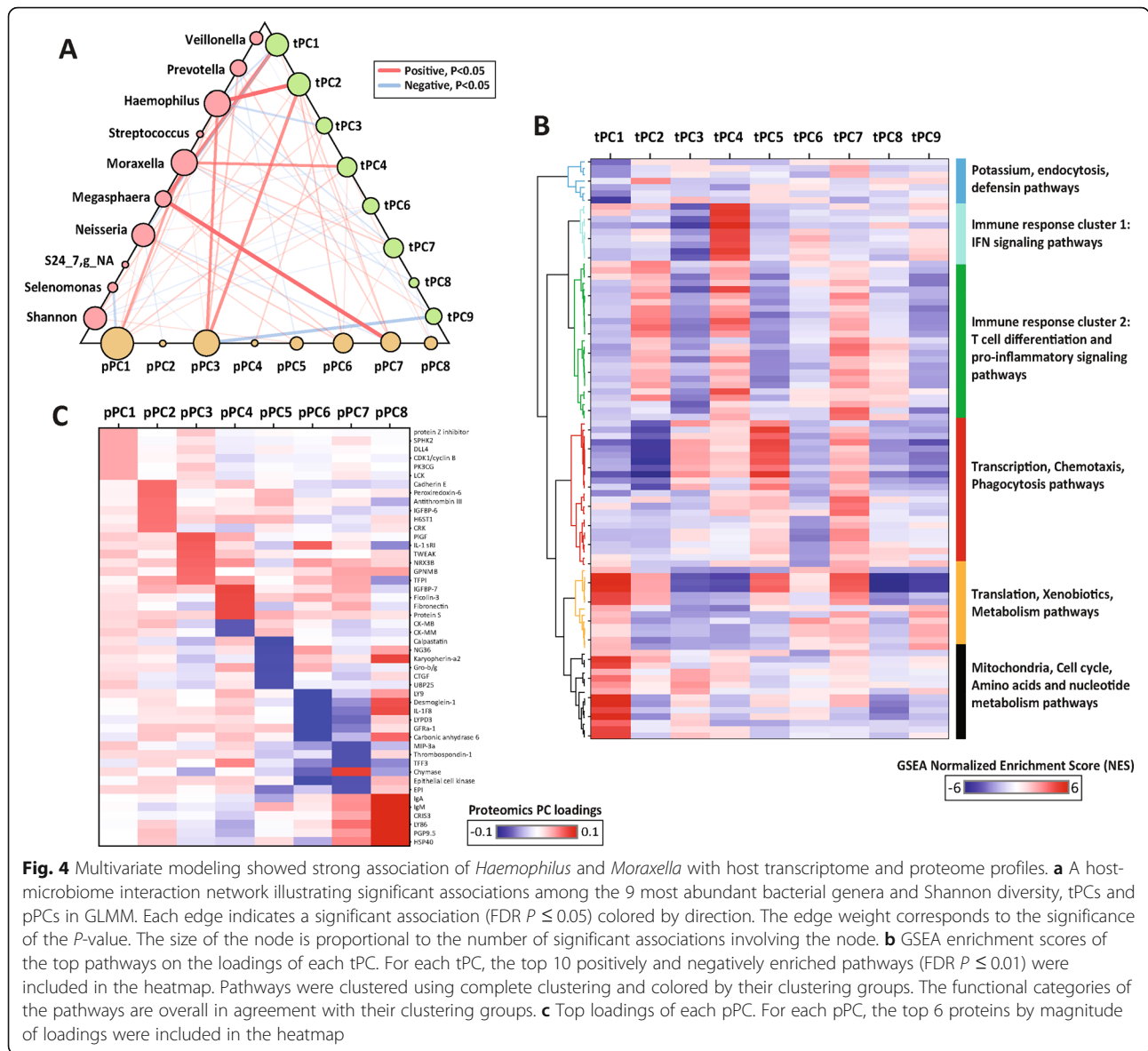
Among other major genera, *Megasphaera* was strongly positively correlated with tPC1 and pPC7 (FDR  $P = 2.0E-4$ ), which were negatively enriched for host immune pathways such as IL-17 and interferon pathways (Fig. 4b) and associated with reduced expression of pro-inflammatory markers such as C-C motif chemokine 20 and interleukin-36 (Fig. 4c, Additional file 4). Therefore, increased abundance of *Megasphaera* could be associated with reduced airway inflammatory responses. In comparison, other major genera such as *Streptococcus* and *Veillonella* were associated with relatively little host response.

Within stable samples only, the *Haemophilus*-tPC2-pPC3 associations persisted, while *Moraxella* was not associated with any host PCs (Additional file 1: Figure S8, Tables 2-3). Within stable, *Streptococcus* showed a significantly negative correlation with tPC6 (FDR  $P = 3E-3$ , Table 2, Additional file 3), in which several phagocytosis and neutrophil migration pathways were most negatively enriched. Thus, increased *Streptococcus* could be associated with greater expression of these pathways at stable. Furthermore, an unknown genus in S24-7 family had significant positive correlations with tPC9 at stable (FDR  $P = 2E-3$ , Table 2, Additional file 3), in which IL-12, IL-23 and T-cell differentiation pathways were most negatively enriched. This genus, despite its low abundance, could be associated with reduced inflammatory response at stable.

## Discussion

Here we present the first comprehensive study characterizing airway host-microbiome interactions in COPD





**Fig. 4** Multivariate modeling showed strong association of *Haemophilus* and *Moraxella* with host transcriptome and proteome profiles. **a** A host-microbiome interaction network illustrating significant associations among the 9 most abundant bacterial genera and Shannon diversity, tPCs and pPCs in GLMM. Each edge indicates a significant association (FDR  $P \leq 0.05$ ) colored by direction. The edge weight corresponds to the significance of the  $P$ -value. The size of the node is proportional to the number of significant associations involving the node. **b** GSEA enrichment scores of the top pathways on the loadings of each tPC. For each tPC, the top 10 positively and negatively enriched pathways (FDR  $P \leq 0.01$ ) were included in the heatmap. Pathways were clustered using complete clustering and colored by their clustering groups. The functional categories of the pathways are overall in agreement with their clustering groups. **c** Top loadings of each pPC. For each pPC, the top 6 proteins by magnitude of loadings were included in the heatmap

integrating lung microbiome and host multi-omics datasets both in stable state and during exacerbations. The systems biology approach revealed a significant airway host-microbiome interplay associated with COPD inflammation and exacerbations. Among all major genera, *Haemophilus* and *Moraxella* were most strongly associated with host gene expression profiles, particularly immunity and inflammation, suggesting the two genera as key players in airway host-bacterial crosstalk in COPD.

Importantly, our results revealed different timing of host responses to these two genera. While *Haemophilus* was associated with host responses both in stable state and during exacerbations, the associations for *Moraxella* were primarily related to exacerbations. This is consistent with a previous study [42] and highlights the role of *Haemophilus* as a stable airway colonizer and *Moraxella*

as an exacerbation-related opportunistic pathogen in COPD. Furthermore, the *Haemophilus*-associated immune responses were correlated with the degree of neutrophilic inflammation, underscoring the interactions between bacterial presence, host immune responses and cellular inflammation. This suggests that chronic airway inflammation in some COPD patients may not respond to anti-inflammatory therapies alone [43] unless the underlying bacterial infection driving the abnormal immune response is addressed.

To achieve statistical power for a genome-wide analysis associating microbiome and host multi-omics datasets in a relatively small sample set, we performed dimensionality reduction on host data and used multivariate modeling to identify significant associations between microbiome composition and overall patterns of

**Table 2** Associations of the 9 bacterial genera and Shannon diversity with tPCs both across all samples and within stable samples in generalized linear mixed models. FDR *P*-values are indicated in the table. Significant associations are highlighted in asterisks. Only significant variables were included in the final model unless otherwise stated

	Microbiome	Transcriptome								
		tPC1	tPC2	tPC3	tPC4	tPC5	tPC6	tPC7	tPC8	tPC9
All	<i>Veillonella</i>	0.57	0.05*	0.91	0.39	0.32	0.53	0.71	0.82	0.70
	<i>Prevotella</i>	0.34	0.42	0.04*	0.25	0.93	0.92	0.59	0.59	0.14
	<i>Haemophilus</i>	0.01*	1E-4*	3E-3*	0.97	0.68	0.41	0.01*	0.33	0.02*
	<i>Streptococcus</i>	0.19	0.17	0.29	0.67	0.59	0.87	0.03*	0.36	0.36
	<i>Moraxella</i>	0.03*	0.45	0.88	5E-4*	0.09	0.30	0.01*	0.02*	0.02*
	<i>Megasphaera</i>	2E-4*	0.95	0.15	0.03*	0.28	0.78	0.91	0.84	0.79
	<i>Neisseria</i>	5E-3*	0.92	0.26	0.29	0.21	0.05*	0.19	0.18	0.23
	<i>S24-7</i>	0.72	0.73	0.38	0.52	0.94	0.84	0.80	0.57	0.46
	<i>Selenomonas</i>	0.29	0.97	0.13	0.24	0.09	0.04*	0.33	0.21	0.48
	Shannon	0.01*	0.10	0.02*	0.91	0.51	0.03*	0.04*	0.06	0.92
Stable	<i>Veillonella</i>	0.27	0.06	0.05	0.25	0.26	0.95	0.08	0.83	0.53
	<i>Prevotella</i>	0.37	0.36	0.27	0.51	0.56	0.67	0.27	0.90	0.27
	<i>Haemophilus</i>	0.10	0.01*	0.04*	0.36	0.76	0.62	0.16	0.31	0.51
	<i>Streptococcus</i>	0.80	0.64	0.73	0.28	0.56	3E-3*	0.08 <sup>a</sup>	0.86	0.51
	<i>Moraxella</i>	0.40	0.59	0.81	0.46	0.57	0.35	0.75	0.13	0.24
	<i>Megasphaera</i>	0.58	0.43	0.02*	0.50	0.39	0.07 <sup>a</sup>	0.45	0.67	0.78
	<i>Neisseria</i>	0.73	0.49	0.95	0.44	0.09 <sup>a</sup>	0.60	0.50	0.83	0.60
	<i>S24-7</i>	0.39	0.73	0.52	0.82	0.42	0.90	0.03*	0.72	2E-3*
	<i>Selenomonas</i>	0.02*	0.59	0.50	0.58	0.90	0.92	0.57	0.88	0.56
	Shannon	0.06	0.10	0.08	0.33	0.26	0.37	0.92	0.07	0.26

<sup>a</sup>Variables not statistically significant but present in the model

host gene expression. Similar approaches were employed by Morgan et al. in associating gut microbiome with host transcriptome in inflammatory bowel disease patients [29]. The strong correlations of *Haemophilus* and *Moraxella* with host immune and inflammation-related tPCs and pPCs highlight the positive links between the two genera and host immune responses that predominated airway host-microbiome interactions. The presence of lipopolysaccharide-induced bronchial epithelial receptors among the top loadings demonstrates that our approach can recapitulate an active host-bacterial cross-talk in COPD. While both *Haemophilus* and *Moraxella* were positively associated with T-cell induced pro-inflammatory signaling, the interferon signaling was more strongly linked to *Moraxella* than *Haemophilus*. This is consistent with one previous study showing that *M. catarrhalis* but not *H. influenzae* induced interferon-beta expression in bronchial epithelial cells [44] and aligns with the different pathogenicity profiles between the two pathogens [45]. Differential involvement of viral co-infection could be another important factor [46]. We have not fully characterized the sputum viral load of this cohort due to limited sputum available. Additional viral load data is key to further resolving this question.

Our multivariate analysis showed that *Megasphaera* and an unknown genus in *S24-7* family were associated with reduced expression of host inflammatory pathways and therefore could potentially reverse airway inflammation (i.e. interferon, IL-12 pathway) induced by *Haemophilus* and *Moraxella*. Furthermore, *Megasphaera* was negatively correlated with sputum neutrophil counts. *Megasphaera* is a known member of human lung microbiome [39] and has beneficial effects on the host through short chain fatty acids (SCFAs) production [47]. In the lung microenvironment, bacterial SCFAs were shown to inhibit cytokine production and inflammation after LPS stimulation of macrophages [48]. Trompette et al. also showed that bacterial SCFAs reduce neutrophil recruitment to the airways and protect against influenza virus infection in mice, suggesting that it has anti-inflammatory effects [49]. Cait et al. demonstrated that diet-derived SCFAs ameliorate allergic inflammation in mice, suggesting its anti-inflammatory effects in the lung [50]. One study on oropharyngeal microbiome of H7N9-infected patients showed that *Megasphaera* increased in patients without secondary bacterial infection, suggesting its potential role in preventing colonization of respiratory pathogens [51]. Further validation on the identity and prevalence of these

**Table 3** Associations of the 9 bacterial genera and Shannon diversity with pPCs both across all samples and within stable samples in generalized linear mixed models. FDR *P*-values are indicated in the table. Significant associations are highlighted in asterisks. Only significant variables were included in the final model unless otherwise stated

	Microbiome	Proteome							
		pPC1	pPC2	pPC3	pPC4	pPC5	pPC6	pPC7	pPC8
All	<i>Veillonella</i>	0.04*	0.82	0.54	0.62	0.40	0.19	0.62	0.04*
	<i>Prevotella</i>	0.52	0.94	0.01*	0.11	0.02*	0.76	0.01*	0.34
	<i>Haemophilus</i>	0.01*	0.06	1E-3*	0.69	0.25	0.47	0.64	0.56
	<i>Streptococcus</i>	0.29	0.35	0.92	0.69	0.45	0.93	0.13	0.61
	<i>Moraxella</i>	1E-3*	0.54	0.05 <sup>a</sup>	0.45	0.13	0.40	0.85	0.52
	<i>Megasphaera</i>	0.31	0.36	0.06 <sup>a</sup>	0.67	0.51	0.01*	1E-4*	0.84
	<i>Neisseria</i>	0.03*	0.06 <sup>a</sup>	0.22	0.26	0.02*	0.01*	0.02*	0.06
	<i>S24-7</i>	0.67	0.12	0.04*	0.24	0.21	0.58	0.49	0.55
	<i>Selenomonas</i>	2E-3*	0.14	0.58	0.12	0.72	0.69	0.13	0.29
	Shannon	0.04*	0.54	0.78	0.10	0.03*	0.56	0.58	0.09
	Stable	<i>Veillonella</i>	0.05 <sup>a</sup>	0.10	0.10	0.23	0.71	0.47	0.99
<i>Prevotella</i>		0.73	0.05 <sup>a</sup>	0.21	0.03*	0.86	0.38	0.70	0.50
<i>Haemophilus</i>		0.03*	0.03*	2E-3*	0.08	0.75	0.29	0.73	0.60
<i>Streptococcus</i>		0.03*	0.93	0.58	0.35	0.76	0.82	0.62	0.24
<i>Moraxella</i>		0.46	0.97	0.26	0.59	0.95	0.96	0.30	0.79
<i>Megasphaera</i>		0.38	0.76	0.80	0.69	0.80	0.27	0.67	0.25
<i>Neisseria</i>		0.23	0.66	0.58	0.70	0.46	0.85	0.55	0.99
<i>S24-7</i>		0.23	0.83	0.89	0.56	0.95	0.42	0.96	0.21
<i>Selenomonas</i>		0.75	0.01*	0.65	0.03*	0.07 <sup>a</sup>	0.70	0.39	0.05*
Shannon		0.25	0.04*	0.13	0.02*	0.33	0.58	0.77	0.41

<sup>a</sup>Variables not statistically significant but present in the model

genera is warranted to explore their functions in the COPD lung.

Our study provides novel insights on the impact of smoking on the lung microbiome, although individual subgroups had small sample sizes and the results need further confirmation in larger cohorts. Our results suggest that the effect of current smoking on the lung microbiome differs between healthy subjects and COPD patients. In healthy subjects, a significantly increased *Haemophilus* was observed in smokers versus non-smokers, suggesting that smoking could be a risk factor for airway dysbiosis in healthy populations. In COPD patients, a significantly increased *Haemophilus* was observed in ex-smokers versus current smokers. The greater dysbiosis in COPD ex-smokers was further associated with their greater airway inflammatory states, as evident by significantly higher expression of sputum pro-inflammatory markers. Our findings further support the view that smoking likely had resulted in an irreversible airway inflammation in COPD, which persisted despite smoking cessation [52].

We observed significant increase of *Moraxella* in stable COPD patients versus healthy subjects, and in COPD exacerbations versus stable, in agreement with

previous observations [5, 8, 9, 36]. The reversal trends of microbiome diversity and composition prior and post exacerbations further support the lung microbiome dysbiosis during exacerbations. Increased *Haemophilus* and *Moraxella* were found in stable ICS versus non-ICS exposed patients, consistent with earlier observations [6, 9], with the caveat being the small sample size of non-ICS users. At stable, the microbiome was comparable between frequent and infrequent exacerbators, suggesting that the baseline microbiome does not effectively predict exacerbation frequency. Identifying markers that predict the exacerbation frequency is of great importance for COPD management [53]. Differences in baseline respiratory microbiota composition were hypothesized to explain the different exacerbation frequency in COPD patients [13]. However, neither this study nor earlier reports support this hypothesis [54, 11]. Instead, previous longitudinal studies showed that there is an association between temporal variability of the airway microbiome and patient exacerbation frequency [11, 12], suggesting that the frequent exacerbator phenotype might be more relevant to the destabilization of the microbiome over time but not the microbiome composition at baseline per se. We observed no significant association between microbiome or

sputum cell count changes with CAT score, FEV1 or FVC, which suggests that different patient inflammatory profiles (i.e. neutrophilic or eosinophilic inflammation) and their associated airway microbiome changes are likely independent of disease severity and cannot be distinguished clinically [55].

There are several caveats to our study. First, the sample size was relatively small particularly for the subgroup analysis, and the longitudinal profiling was limited due to the limited amount of sputum produced in some visits and the technical difficulty in extracting sufficient material from sputum for the various aspects of downstream experiments (i.e. microbiome, transcriptome, proteome, cell counting). We performed a power estimation to ensure adequate statistical sensitivity could be achieved after dimensionality reduction. Nevertheless, the associations observed in our study need to be validated in larger independent patient cohorts. Second, host transcriptome and proteome were not profiled for healthy subjects, which is important to understand to what extent the observed host-microbiome associations are disease specific. Our study provides a method for profiling airway host-microbiome interactions that should catalyze future efforts on characterizing lung microbiome and host multi-omics in larger healthy and disease populations.

## Conclusions

To our knowledge, this is the first study that depicts airway host-microbiome interactions in COPD and highlights the differential role of *Haemophilus* and *Moraxella* in terms of host interactions. Our study provides support for novel therapies targeting both genera and their associated host pathways to overcome the abnormal immune response in COPD.

## Additional files

**Additional file 1:** Supplementary figures and tables (DOCX 856 kb)

**Additional file 2:** a. MetaCore pathways significantly enriched for differential expressed genes (DEGs) between stable and exacerbations (FDR  $P \leq 0.01$ ). b. Significantly differentially expressed proteins between sputum samples of COPD ex-smokers and current smokers (Wilcoxon, FDR  $P \leq 0.05$ ). (XLSX 68 kb)

**Additional file 3:** Cross associations among microbiome (9 microbiome genera and Shannon diversity), host transcriptome (9 tPCs) and proteome (8 pPCs) data both across all samples and within stable samples in the multivariate analysis. FDR  $P$ -values are indicated in the table. Significant positive correlations are highlighted in red. Significant negative correlations are highlighted in blue. (XLSX 15 kb)

**Additional file 4:** a. Top 25 loadings by magnitude for each tPC. b. Top 25 loadings by magnitude for each pPC. c. Top 10 positively and negatively enriched pathways and their enrichment scores as shown in the heatmap in Fig. 4b. d. Top 25 significantly enriched positive and negative pathways (FDR  $P < 0.01$ ) for the loadings of tPC1–9. (XLSX 103 kb)

**Additional file 5:** Spearman correlation of individual genes in the top two pathways. a. tPC2 and b. tPC4 with *Haemophilus* and *Moraxella* in both microbiome and qPCR datasets. (XLSX 32 kb)

## Abbreviations

FDR: False discovery rate; GLMM: Generalized linear mixed model; GSEA: Gene set enrichment analysis; ICS: Inhaled corticosteroid; LPS: Lipopolysaccharide; OTUs: Operational taxonomic units; PCA: Principal component analysis; QoL: Quality of life; qPCR: Quantitative polymerase chain reaction; SCFA: Short chain fatty acid

## Acknowledgments

We thank all the volunteers for their participations in the study. Dave Singh is supported by the National Institute for Health Research (NIHR) Manchester Biomedical Research Centre (BRC).

## Author contributions

ZW had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. ZW, BM, SL, UK, DM, JRB, EMH and DS contributed to the study conception and design. SL, UK and DS coordinated the collection of sputum samples and clinical data. SVH and CT performed microbiome DNA purification, sequencing and qPCR experiments. ZW performed all data analyses and interpretation. ZW wrote the initial draft of the manuscript with additional contents and critical revisions from all authors. All authors read and approved the final version of the manuscript.

## Funding

This work was supported by GlaxoSmithKline.

## Competing interests

Z. W., B. M., D. M., S. V. H., C. T., J. R. B. and E. M. H. were employees and shareholders in GlaxoSmithKline PLC at the time of this study. Other authors have no competing interest to declare.

## Author details

<sup>1</sup>Computational Biology, Human Genetics, Research and Development (R&D), GlaxoSmithKline (GSK), 1250 S. Collegeville Road, Collegeville, PA 19426-0989, USA. <sup>2</sup>Present Address: School of Life Sciences, South China Normal University, Guangzhou 510631, People's Republic of China. <sup>3</sup>Refractory Respiratory Inflammation Discovery Performance Unit, Respiratory Therapy Area, R&D, GSK, Stevenage SG1 2NY, UK. <sup>4</sup>University of Manchester and University Hospital of South Manchester, Manchester M23 9QZ, UK. <sup>5</sup>Functional Genomics, Medicinal Science and Technology, R&D, GSK, Collegeville, PA 19426, USA.

Received: 1 January 2019 Accepted: 28 May 2019

Published online: 06 June 2019

## References

- Ball P. Epidemiology and treatment of chronic bronchitis and its exacerbations. *Chest*. 1995;108(2 Suppl):435–525.
- Miravittles M, Espinosa C, Fernandez-Laso E, Martos JA, Maldonado JA, Gallego M. Relationship between bacterial flora in sputum and functional impairment in patients with acute exacerbations of COPD. *Study Group of Bacterial Infection in COPD*. *Chest*. 1999;116(1):40–6.
- Monso E, Ruiz J, Rosell A, Manterola J, Fiz J, Morera J, et al. Bacterial infection in chronic obstructive pulmonary disease. A study of stable and exacerbated outpatients using the protected specimen brush. *Am J Respir Crit Care Med*. 1995;152(4 Pt 1):1316–20.
- Soler N, Torres A, Ewig S, Gonzalez J, Celis R, El-Ebiary M, et al. Bronchial microbial patterns in severe exacerbations of chronic obstructive pulmonary disease (COPD) requiring mechanical ventilation. *Am J Respir Crit Care Med*. 1998;157(5 Pt 1):1498–505.
- Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, et al. Disordered microbial communities in asthmatic airways. *PLoS One*. 2010;5(1):e8578.
- Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PLoS One*. 2012;7(10):e47305.

7. 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.
8. Huang YJ, Sethi S, Murphy T, Nariya S, Boushey HA, Lynch SV. Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. *J Clin Microbiol*. 2014;52(8):2813–23.
9. Wang Z, Bafadhel M, Haldar K, Spivak A, Mayhew D, Miller BE, et al. Lung microbiome dynamics in chronic obstructive pulmonary disease exacerbations. *The European respiratory journal*. 2016.
10. Millares L, Ferrari R, Gallego M, Garcia-Nunez M, Perez-Brocal V, Espasa M, et al. Bronchial microbiome of severe COPD patients colonised by *Pseudomonas aeruginosa*. *European journal of clinical microbiology & infectious diseases* : official publication of the European Society of Clinical Microbiology. 2014;33(7):1101–11.
11. Wang Z, Singh R, Miller BE, Tal-Singer R, Van Horn S, Tomsho L, Mackay A, Allinson JP, Webb AJ, Brookes AJ et al: Sputum microbiome temporal variability and dysbiosis in chronic obstructive pulmonary disease exacerbations: an analysis of the COPDMap study. *Thorax* 2018, 73(4):331–338.
12. Mayhew D, Devos N, Lambert C, Brown JR, Clarke SC, Kim VL, Magid-Slav M, Miller BE, Ostridge KK, Patel R et al: Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. *Thorax* 2018, 73(5):422–430
13. Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. *Lancet*. 2014;384(9944):691–702.
14. Sze MA, Dimitriu PA, Suzuki M, McDonough JE, Campbell JD, Brothers JF, et al. Host response to the lung microbiome in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2015;192(4):438–45.
15. Molyneaux PL, Willis-Owen SAG, Cox MJ, James P, Cowman S, Loebinger M, et al. Host-microbial interactions in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2017;195(12):1640–50.
16. General Assembly of the World Medical A. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *J Am Coll Dent*. 2014;81(3):14–8.
17. International Conference on. Harmonisation of technical requirements for registration of pharmaceuticals for human u. ICH harmonized tripartite guideline: guideline for good clinical practice. *J Postgrad Med*. 2001;47(1):45–50.
18. Anthonisen NR, Manfreda J, Warren CP, Hershfield ES, Harding GK, Nelson NA. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. *Ann Intern Med*. 1987;106(2):196–204.
19. Rodriguez-Roisin R. Toward a consensus definition for COPD exacerbations. *Chest*. 2000;117(5 Suppl 2):398S–401S.
20. Bafadhel M, McCormick M, Saha S, McKenna S, Shelley M, Hargadon B, et al. Profiling of sputum inflammatory mediators in asthma and chronic obstructive pulmonary disease. *Respiration*. 2012;83(1):36–44.
21. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335–6.
22. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12:87.
23. Wang X, Lin Y, Song C, Sibille E, Tseng GC. Detecting disease-associated genes with confounding variable adjustment and the impact on genomic meta-analysis: with application to major depressive disorder. *BMC Bioinf*. 2012;13:52.
24. Menni C, Kiddle SJ, Mangino M, Vinuela A, Psatha M, Steves C, et al. Circulating proteomic signatures of chronological age. *J Gerontol A Biol Sci Med Sci*. 2015;70(7):809–16.
25. Mehan MR, Williams SA, Siegfried JM, Bigbee WL, Weissfeld JL, Wilson DO, et al. Validation of a blood protein signature for non-small cell lung cancer. *Clin Proteomics*. 2014;11(1):32.
26. Hathout Y, Brody E, Clemens PR, Cripe L, DeLisle RK, Furlong P, et al. Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. *Proc Natl Acad Sci U S A*. 2015;112(23):7153–8.
27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
28. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
29. Morgan XC, Kabakchiev B, Waldron L, Tyler AD, Tickle TL, Milgrom R, et al. Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biol*. 2015;16:67.
30. Bates DMM, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1):1–48.
31. BPaCR KA. lmerTest: tests in linear mixed effects models; 2014.
32. Brindefalk B, Ettema TJ, Viklund J, Tholleson M, Andersson SG. A phylometagenomic exploration of oceanic alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade. *PLoS One*. 2011;6(9):e24457.
33. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
34. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet*. 2006;38(5):500–1.
35. Benjamini Y, Hochberg Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57:289–300.
36. Molyneaux PL, Mallia P, Cox MJ, Footitt J, Willis-Owen SA, Homola D, et al. Outgrowth of the bacterial airway microbiome after rhinovirus exacerbation of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2013;188(10):1224–31.
37. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, et al. Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PLoS One*. 2011;6(2):e16384.
38. Huang YJ, Kim E, Cox MJ, Brodie EL, Brown R, Wiener-Kronish JP, et al. A persistent and diverse airway microbiota present during chronic obstructive pulmonary disease exacerbations. *Omic*. 2010;14(1):9–59.
39. Zakharkina T, Heinzel E, Koczulla RA, Greulich T, Rentz K, Pauling JK, et al. Analysis of the airway microbiota of healthy individuals and patients with chronic obstructive pulmonary disease by T-RFLP and clone sequencing. *PLoS One*. 2013;8(7):e68302.
40. Mofarrah M, Noh T, Qureshi S, Guillot L, Mayaki D, Hussain SN. Regulation of angiopoietin expression by bacterial lipopolysaccharide. *Am J Physiol Lung Cell Mol Physiol*. 2008;294(5):L955–63.
41. Ivanov AI, Romanovsky AA. Putative dual role of ephrin-Eph receptor interactions in inflammation. *IUBMB Life*. 2006;58(7):389–94.
42. Barker BL, Haldar K, Patel H, Pavord ID, Barer MR, Brightling CE, et al. Association between pathogens detected using quantitative polymerase chain reaction with airway inflammation in COPD at stable state and exacerbations. *Chest*. 2015;147(1):46–55.
43. King PT. Inflammation in chronic obstructive pulmonary disease and its role in cardiovascular disease and lung cancer. *Clin Transl Med*. 2015;4(1):68.
44. Klaile E, Klassert TE, Scheffrahn I, Muller MM, Heinrich A, Heyl KA, et al. Carcinoembryonic antigen (CEA)-related cell adhesion molecules are co-expressed in the human lung and their expression can be modulated in bronchial epithelial cells by non-typable *Haemophilus influenzae*, *Moraxella catarrhalis*, TLR3, and type I and II interferons. *Respir Res*. 2013;14:85.
45. Sethi S, Murphy TF. Bacterial infection in chronic obstructive pulmonary disease in 2000: a state-of-the-art review. *Clin Microbiol Rev*. 2001;14(2):336–63.
46. DeMuri GP, Gern JE, Eickhoff JC, Lynch SV, Wald ER. Dynamics of bacterial colonization with *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Moraxella catarrhalis* during symptomatic and asymptomatic viral upper respiratory infection. In: *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*; 2017.
47. Shetty SA, Marathe NP, Lanjekar V, Ranade D, Shouche YS. Comparative genome analysis of *Megasphaera* sp. reveals niche specialization and its potential role in the human gut. *PLoS One*. 2013;8(11):e79353.
48. Chang PV, Hao L, Offermanns S, Medzhitov R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc Natl Acad Sci U S A*. 2014;111(6):2247–52.
49. Trompette A, Gollwitzer ES, Pattaroni C, Lopez-Mejia IC, Riva E, Pernot J, et al. Dietary Fiber confers protection against flu by shaping Ly6c(–) patrolling monocyte hematopoiesis and CD8(+) T cell metabolism. *Immunity*. 2018; 48(5):992–1005 e8.
50. Cait A, Hughes MR, Antignano F, Cait J, Dimitriu PA, Maas KR, et al. Microbiome-driven allergic lung inflammation is ameliorated by short-chain fatty acids. *Mucosal Immunol*. 2018;11(3):785–95.
51. Lu HF, Li A, Zhang T, Ren ZG, He KX, Zhang H, et al. Disordered oropharyngeal microbial communities in H7N9 patients with or without secondary bacterial lung infection. *Emerg Microbes Infect*. 2017;6(12):e112.

52. Laniado-Laborin R. Smoking and chronic obstructive pulmonary disease (COPD). Parallel epidemics of the 21 century. *Int J Environ Res Public Health*. 2009;6(1):209–24.
53. Wedzicha JA, Brill SE, Allinson JP, Donaldson GC. Mechanisms and impact of the frequent exacerbator phenotype in chronic obstructive pulmonary disease. *BMC Med*. 2013;11:181.
54. Wang Z, Bafadhel M, Haldar K, Spivak A, Mayhew D, Miller BE, et al. Lung microbiome dynamics in COPD exacerbations. *Eur Respir J*. 2016; 47(4):1082–92.
55. Bafadhel M, McKenna S, Terry S, Mistry V, Reid C, Haldar P, et al. Acute exacerbations of chronic obstructive pulmonary disease: identification of biologic clusters and their biomarkers. *Am J Respir Crit Care Med*. 2011; 184(6):662–71.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

