Transcriptionally Active Lung Microbiome and Its Association with Bacterial Biomass and Host Inflammatory Status

Lili Ren,[a] Rongbao Zhang,[b] Jian Rao,[a] Yan Xiao,[a] Zhao Zhang,[b] Bin Yang,[a] Depan Cao,[a] Hui Zhong,[a] Pu Ning,[b] Ying Shang,[b] Mingkun Li,[c,d,e] Zhancheng Gao,[b] Jianwei Wang[a]

[a]MOH Key Laboratory of Systems Biology of Pathogens and Christophe Mérieux Laboratory, Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[b]Department of Respiratory and Critical Care Medicine, Peking University People's Hospital, Beijing, China

[c]Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

[d]Fondation Mérieux, Lyon, France

[e]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

ABSTRACT Alteration of the lung microbiome has been observed in several respiratory tract diseases. However, most previous studies were based on 16S ribosomal RNA and shotgun metagenome sequencing; the viability and functional activity of the microbiome, as well as its interaction with host immune systems, have not been well studied. To characterize the active lung microbiome and its associations with host immune response and clinical features, we applied metatranscriptome sequencing to bronchoalveolar lavage fluid (BALF) samples from 25 patients with chronic obstructive pulmonary disease (COPD) and from nine control cases without known pulmonary disease. Community structure analyses revealed three distinct microbial compositions, which were significantly correlated with bacterial biomass, human Th17 immune response, and COPD exacerbation frequency. Specifically, samples with transcriptionally active *Streptococcus*, *Rothia*, or *Pseudomonas* had bacterial loads 16 times higher than samples enriched for *Escherichia* and *Ralstonia*. These high-bacterial-load samples also tended to undergo a stronger Th17 immune response. Furthermore, an increased proportion of lymphocytes was found in samples with active *Pseudomonas*. In addition, COPD patients with active *Streptococcus* or *Rothia* infections tended to have lower rates of exacerbations than patients with active *Pseudomonas* and patients with lower bacterial biomass. Our results support the idea of a stratified structure of the active lung microbiome and a significant host-microbe interaction. We speculate that diverse lung microbiomes exist in the population and that their presence and activities could either influence or reflect different aspects of lung health.

IMPORTANCE Recent studies of the microbiome proposed that resident microbes play a beneficial role in maintaining human health. Although lower respiratory tract disease is a leading cause of sickness and mortality, how the lung microbiome interacts with human health remains largely unknown. Here we assessed the association between the lung microbiome and host gene expression, cytokine concentration, and over 20 clinical features. Intriguingly, we found a stratified structure of the active lung microbiome which was significantly associated with bacterial biomass, lymphocyte proportion, human Th17 immune response, and COPD exacerbation frequency. These observations suggest that the microbiome plays a significant role in lung homeostasis. Not only microbial composition but also active functional elements and host immunity characteristics differed among different individuals. Such diversity may partially account for the variation in susceptibility to particular diseases.

Investigation of the lung microbiome is a relatively young field; however, there has been remarkable progress in understanding the composition and function of the lung microbiome in the last few years (1). Dickson and colleagues have described an adapted island model for the lung microbiome (2), in which lung microbes are largely derived from the upper respiratory tract and oral cavity (UO) through microaspiration and mucosal dispersion. Simultaneously, microbes in the lung are eliminated by mucociliary clearance, cough, and host immune defenses, while reproduction and growth of the microbes are determined by regional conditions in the lung. Alterations of the lung microbiome have been observed in many respiratory diseases, including chronic obstructive pulmonary disease (COPD), asthma, and cystic fibrosis, but associations with clinical features and interactions with host genes are largely unknown (3–5).

Sequencing of 16S rRNA gene amplicons is a convenient method to characterize the structure of the microbial community (6, 7). However, information provided by this method is limited due to its narrow detection spectrum (bacteria), low resolution (genus), and lack of direct insights into the viability and functional activity of the microbiome. An alternative approach is shotgun metagenome sequencing, which has been used to reveal the composition of the lung microbiome with a higher resolution and wider detection spectrum (8, 9); however, recent studies proposed that expression of microbial genes could significantly change without large alterations in overall community structure (10, 11), which emphasize the importance of investigating the functional activity of the microbiome. Metatranscriptome sequencing can provide not only active microbiome profiles at high resolution but also the details of functional elements (10, 12). Moreover, it enables analysis of interactions between the microbiome and the host, as both microbial and human transcripts can be analyzed (13). Such studies on the lung microbiome are very limited thus far, and there is an urgent need to explore the active host-microbe interaction in the lung.

COPD, a chronic inflammatory disorder characterized by long-term poor airflow, has been predicted by the World Health Organization to become the third leading cause of death by 2030 (14). The lung microbiome changes dramatically in COPD patients during the exacerbation period (15–17). However, microbes associated with COPD and exacerbation are inconsistent among different studies (15, 17–22). In this study, we aimed to provide a comprehensive description of transcriptionally active microbes and their associations with host gene expression, cytosine concentration, and different clinical features, on the basis of metatranscriptome data.

RESULTS

Overview of the active lung microbiome. The transcriptionally active microbiome was examined in the bronchoalveolar lavage fluid (BALF) samples of 25 COPD patients (during their stable period) and 9 non-COPD controls by metatranscriptome sequencing (demographic and clinical information are described in Table S1 in the supplemental material). After stringent quality control was performed, 76.7% ($\pm$11.5%) of the reads were mapped to the human genome. Among the reads mapping to archaea, bacteria, fungi, and viruses (ABFV), 92% could be assigned to a specific genus, and 60% could be assigned to a specific species or subspecies. Rarefaction analysis showed that the current sequencing depth (30 million reads per sample) enabled us to study most genera/species with read abundance (proportion of reads among all ABFV reads, which reflects the transcriptional activity of the genera/species) greater than 0.1% but not those with lower read abundance. Thus, our study focused only on microbes with read abundance of at least 1% in at least one sample.

Three phyla (*Proteobacteria*, *Firmicutes*, and *Actinobacteria*) were detected in all 34 samples and accounted for 91% of the total ABFV reads (see Fig. S1 in the supplemental material). At the genus level, *Streptococcus*, *Pseudomonas*, *Ralstonia*, *Escherichia*, and
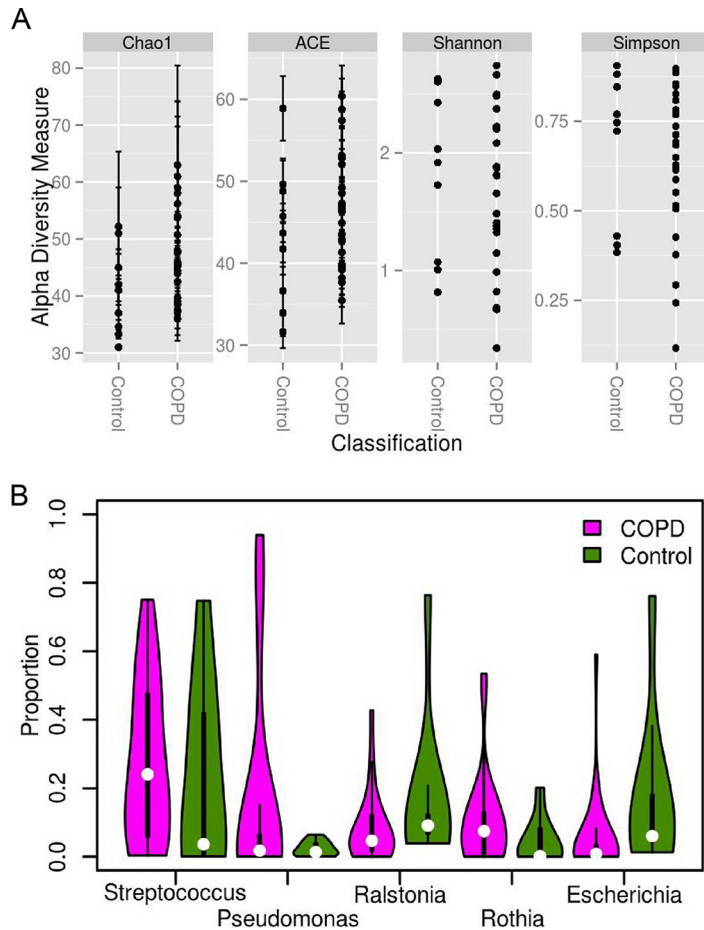
FIG 1 Diversity and composition of the lung microbiome at the genus level. (A) Alpha diversity values for COPD patients and non-COPD controls. (B) Violin plot of the active lung microbiome composition; only genera with a mean read abundance of at least 5% are shown; thickness indicates the density of the value, and each white dot indicates the median value.

Rothia were the most abundant; those genera accounted for 66% of all ABFV reads and explained more than 70% of the variance among samples (Fig. 1). Eighty-seven species from 49 genera were identified. Among all genera, *Streptococcus* and *Pseudomonas* were most diverse, with 14 species and 11 species detected, respectively. Of note, the number of species identified was partially determined by the richness of diversity of sequences included in the reference database; thus, well-studied genera (such as *Streptococcus* and *Pseudomonas*) could be more diverse at the species level. However, only a limited number of species belonging to the same genus could be identified in any one sample.

No significant difference was found between COPD and non-COPD samples in terms of alpha diversities and microbial composition with our data ($P > 0.05$ and $P > 0.05$, respectively) (Fig. 1; see also Fig. S1). Several microbes showed marginal read abundance differences between the two groups. At the genus level, *Gemella* was observed in only 10 COPD samples and at very low read abundance (median = 0.03%; $P < 0.05$, false-discovery rate [q] > 0.1). At the species level, four low-read-abundance species (*Prevotella enoeca*, *Neisseria gonorrhoeae*, *Bifidobacterium dentium*, and *Enterococcus cecorum*) were enriched in COPD samples ($P < 0.05$, $q > 0.1$), and the first three species are known to inhabit the oral cavity and upper respiratory tract. Notably, the sample size in the study was relatively small, which provided us very limited power to detect differences; thus, we aimed to identify the features that are most closely associated with the active lung microbiome.
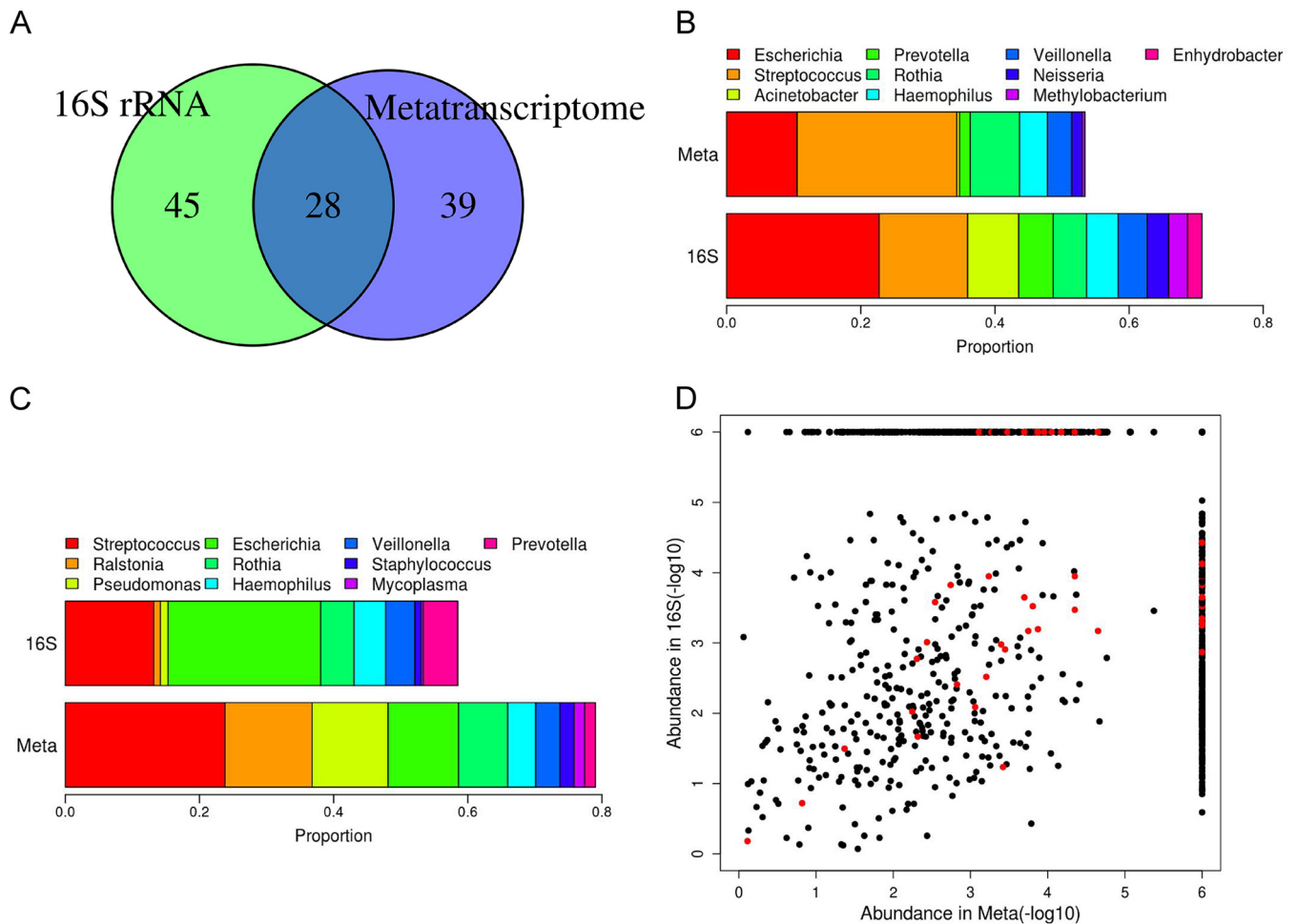
A

B

C

D

FIG 2 Comparison of microbiome composition between metatranscriptome data and 16S rRNA data. (A) Overlap of identified genera between two data sets. (B) The read abundance of the top 10 most abundant genera in 16S rRNA data in two data sets. (C) The read abundance of the top 10 most abundant genera in metatranscriptome data in two data sets. (D) Correlation of the read abundance of each genus between the two methods at the individual level. Red dots denote the genus detected in patient COPD38, who had the highest metatranscriptome-versus-16S rRNA correlation (rho = 0.99, $P < 0.001$). For display, an abundance of 0 was converted to 10^{-6}.

Consistency between metatranscriptome sequencing results and 16S rRNA sequencing results. 16S rRNA sequencing data were successfully obtained for twenty samples (V3-V4 region, with at least 10,000 reads); 73 genera were identified, 28 of which were also discovered in the metatranscriptome data (Fig. 2A). Those 28 genera accounted for 83% of the 16S rRNA reads and 89% of the ABFV metatranscriptome reads, suggesting that the high-abundance genera could be faithfully identified by both methods. However, the abundances determined were not always comparable between the two methods (Fig. 2B and C); for example, *Acinetobacter* had a higher abundance in the 16S rRNA data, while *Ralstonia* and *Pseudomonas* were more highly enriched in the metatranscriptome data. The overall correlation coefficient for the read abundance of each genus estimated on the basis of these two methods was 0.326 ($P < 0.001$), with the highest correlation coefficient being 0.99 in patient COPD38 ($P < 0.001$) (Fig. 2D). Discrepancies between the two methods might reflect the different states of the microbes, which could be either active/viable (overrepresented in metatranscriptome data) or resting/suppressed (underrepresented in the metatranscriptome data). However, an alternative explanation could be that samples with low correlation coefficients had lower bacterial biomass and, hence, that their lung microbiomes were more likely to have been contaminated by reagents and/or the bronchoscope. This hypothesis was supported by the fact that correlation coefficients were
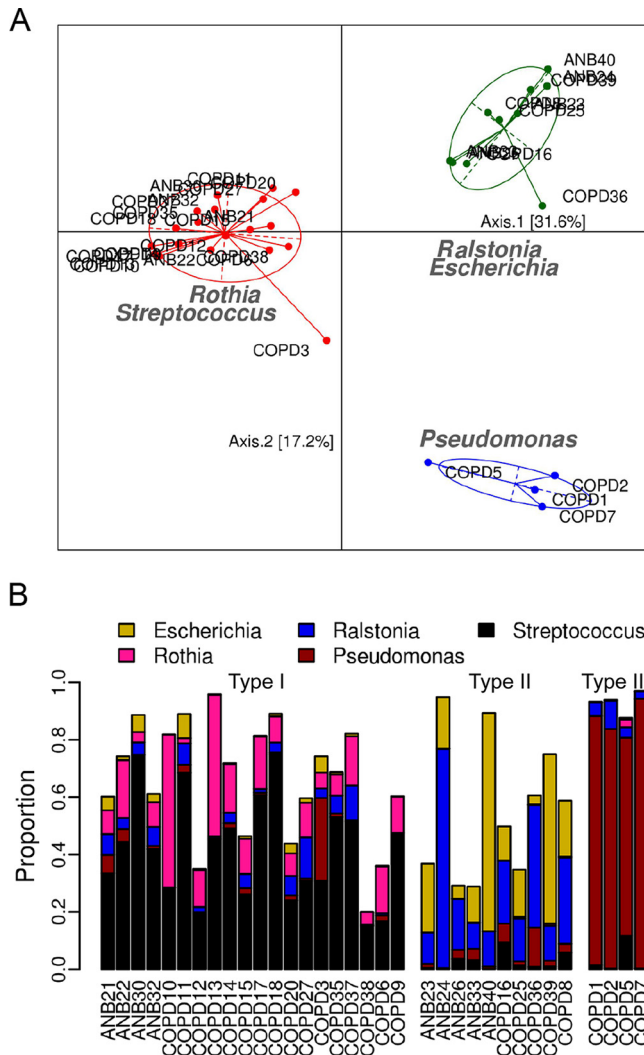
FIG 3 Structure of the lung microbiome at the genus level. (A) Principal-coordinate-analysis (PCoA) plot of the active lung microbiome inferred from metatranscriptome data. Core microbes are labeled on the plot, and the pairwise distance is represented by the Jensen-Shannon divergence (JSD) value. (B) Read abundance of the core microbes in different individuals; samples are ordered by the subgroups to which they belong.

positively correlated with bacterial biomass quantified by a 16S rRNA assay ($P = 0.02$, rho = 0.560).

Structure of the active lung microbiome. Possible lung microbiome subgroups were investigated using two statistical methods (Dirichlet multinomial mixtures and partitioning around medoid clustering coupled with the heuristic Calinski-Harabasz index; see details in Text S1 in the supplemental material), which gave very similar results (only one sample was assigned differently at the phylum level). At the phylum level, samples were clustered into two subgroups, which were dominated by either *Proteobacteria* or *Actinobacteria* and *Firmicutes* (Fig. S2A and C). We found that the read abundances of the dominant (core) microbes (the definition and algorithm are described in Text S1) in the two subgroups were negatively correlated ($P < 0.01$), whereas the read abundances of microbes within each subgroup were positively correlated ($P < 0.01$). At the genus level, samples could be further classified into three subgroups (Fig. 3). Twenty samples were assigned to subgroup I, which was enriched for *Streptococcus* and *Rothia*; 10 samples were assigned to subgroup II, which was enriched for *Ralstonia and Escherichia*; and only 4 samples were assigned to subgroup III, whose

TABLE 1 Tests of the association between the structure of the active lung microbiome and clinical features[a]

Phenotype	Range or results[k]	P value
COPD	{Yes, no}	0.129[h]
Smoking category	{Smoker, quit, never}	0.338[h]
Smoking amount, range[b]	[0, 60]	0.229[i]
Inflammation[c]	{Yes, no, unclear}	0.487[h]
Gender	{Male, female}	0.378[h]
Location	{Left lower lobe, left lingular lobe, right middle lobe}	0.631[h]
Age range (yrs)	[28, 83]	0.355[i]
Smear test	{Positive, negative}	0.019[h]
Inhaled corticosteroids	{Yes, no}	0.731[h]
Bronchodilators	{Yes, no}	0.553[h]
Exacerbation time[d]	[0, 3]	0.819[i]
Macrophage[e] (%)	[0, 100%]	0.153[i]
Lymphocyte (%)	[0, 100%]	0.021[i]
Neutrophil (%)	[0, 100%]	0.896[i]
FEV1	[28.3, 99.3]	0.593[i]
FEV1FVC	[34.4, 70.13]	0.650[i]
RV/TLC	[5.2, 88.6]	0.476[i]
CAT	[2, 23]	0.057[i]
mMRC	{0, 1, 2, 3}	0.904[h]
Severity score (GOLD)[f]	{1, 2, 3, 4}	0.3628[h]
Exacerbation frequency[g]	{0.4, 2, 3.5}	1.3e−5[j]

[a]Use of inhaled corticosteroids and bronchodilators in the previous 3 months prior to the bronchoscope examination was considered. Antibiotics were not used at least 8 weeks preceding the bronchoscopy. CAT, COPD assessment test; FEV1, median forced expiratory volume in 1 s; FVC, forced vital capacity; mMRC, modified Medical Research Council dyspnea scale; RV, residual volume; TLC, total lung capacity;
[b]Data represent numbers of packs of cigarettes smoked per year.
[c]Inflammation status was judged by clinician during bronchoscopy.
[d]Data represent numbers of exacerbations during the year preceding the bronchoscopy.
[e]Cells in the BALF were collected and stained with Wright Giemsa's stain, and cells were counted under a microscope.
[f]GOLD, Global Initiative for Obstructive Lung Disease criteria.
[g]Data represent frequencies of exacerbations for COPD patients in the previous 4 years (2014 to 2018) after the collection of BALF samples.
[h]For discrete data, the contingency table was created and the Fisher exact test was used for the significance test; thus, we were testing whether a specific classification (e.g., male or female) was associated with one of the three active lung microbiome subgroups.
[i]For continuous data, the Kruskal-Wallis rank sum test was applied; thus, we were testing whether a given feature was different among three different active lung microbiome subgroups.
[j]For frequency data, the chi-square test was used for the significance test; thus, we were testing whether the events were randomly distributed in different active lung microbiome subgroups.
[k]Braces mean all possible elements are given here (discrete variable). Square brackets mean a range is given here, e.g., from 0 to 60 (including 0 and 60)(continuous variable).

microbiome was dominated by *Pseudomonas*. Core microbes of subgroup I were negatively correlated with those in subgroup II and subgroup III ($P < 0.01$). Clustering at the species level was identical to that seen at the genus level (Fig. S2B and D).

The significance of the group classification was further evaluated by three methods (average silhouette width, predictive strength, and simulation; see details in Text S1). Clustering at the phylum and genus levels was supported by all statistical metrics, while clustering at the species level had a relatively low predictive strength (0.679) (Fig. S3), suggesting that the clustering scheme is reliable.

Associations between active lung microbiome and clinical features. Associations between the structure of the active lung microbiome and 21 clinical features were investigated (Table 1). First, all samples in subgroup III were COPD patients, suggesting that this might be a COPD-specific group, though this hypothesis needs to be confirmed with a larger sample size. Second, all samples in subgroup II were negative in the bacteria smear test, which differs significantly from the results from the samples in subgroups I and III ($P < 0.05$) (Fig. 4A). We hypothesized that this difference could be an indication of lower bacterial biomass for subgroup II, an interpretation that was supported by the observation that samples in subgroup II had the lowest ratio of bacterial reads to human reads ($P < 0.01$, Fig. 4B). As the proportion of reads could have
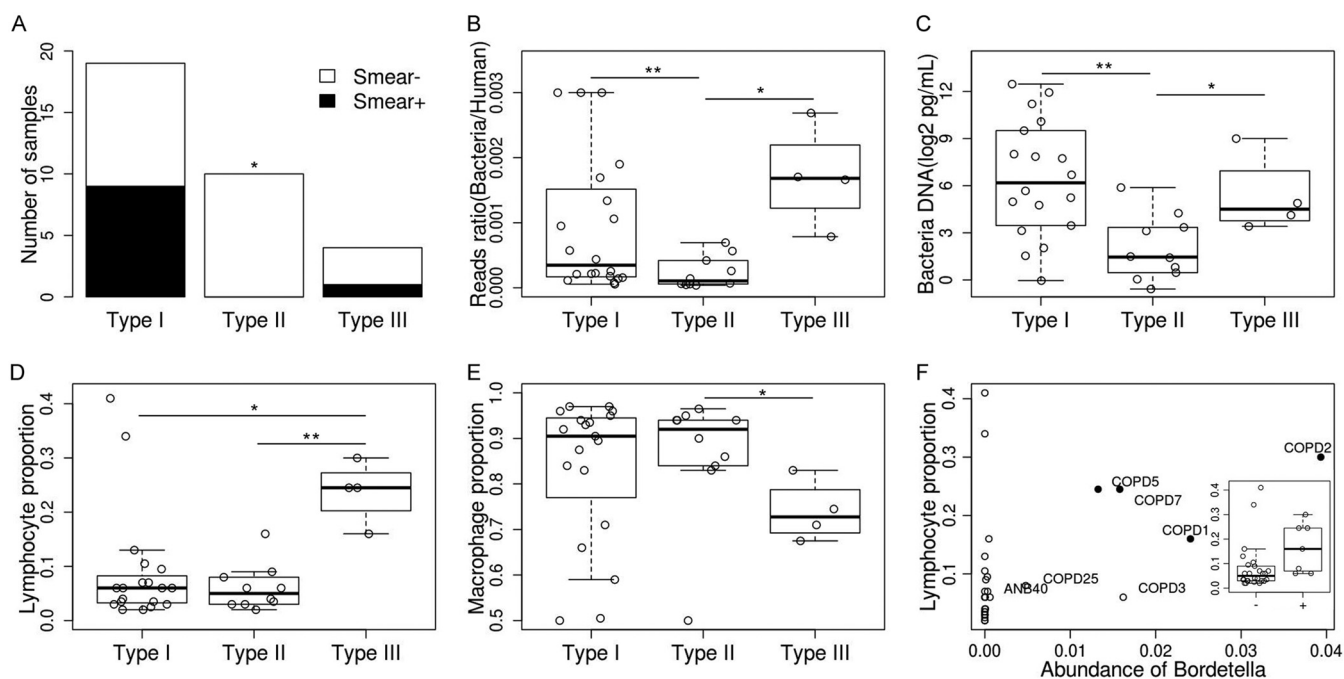
FIG 4 Association between the lung microbiome and clinical features. (A) Bacterial smear test results for different microbiome subgroup samples. (B) Ratio of bacteria reads to human reads. (C) Quantification of bacteria DNA. (D) Proportion of lymphocytes in BALF samples. (E) Proportion of macrophages in BALF samples. (F) Correlation between the proportion of lymphocytes and the read abundance of bacterial genus *Bordetella*; black dots denote samples in subgroup III. The box plot shows the lymphocyte proportion in *Bordetella*-positive samples and *Bordetella*-negative samples.

been biased by the amplification process during library preparation, the actual amount of bacterial DNA was further quantified by a 16S rRNA assay. The median amounts of bacteria DNA in the subgroup I and III samples were 28-fold and 9-fold higher, respectively, than the bacterial DNA amounts in subgroup II samples (77.2 pg/ml and 25.6 pg/ml versus 2.7 pg/ml) ($P < 0.05$) (Fig. 4C). Of note, Salter and colleagues proposed that contamination from laboratory reagents critically impacted results obtained from low-microbial-biomass samples, and both *Escherichia* and *Ralstonia* were on their list of contaminant genera (23). To examine the possibility of contamination, we collected two saline samples (washing through different bronchoscopes before real samples were collected) and prepared the sequencing library following the protocol that had been used with the negative controls. We found that the major compositions were similar, and both *Escherichia* and *Ralstonia* were identified in two negative controls (with read abundances of 35% and 0.1%, respectively) (Fig. S4). Thus, the possibility of contamination is high for this subgroup, and the high read abundance of *Escherichia* and *Ralstonia* may reflect only the background noise introduced from reagents and/or the bronchoscope. In addition, the lymphocyte proportion for the subgroup III samples was significantly higher than that for other subgroups ($P < 0.05$), while the macrophage proportion was lower in this subgroup ($P < 0.05$) (Fig. 4D and E). We further found that the lymphocyte proportion was positively correlated with the relative read abundance of *Bordetella* (mostly *Bordetella pertussis*) ($rho = 0.501$, $P < 0.01$, Fig. 4F). No correlation was found between the subgroups and the severity of COPD, smoking, or use of inhaled corticosteroids, bronchodilators, or other factors (Table 1).

Although COPD is a chronic disease, some patients suffer from exacerbations. Recurrent exacerbation in COPD patients could lead to a faster decline in lung function and could increase their mortality risks. We have obtained the number of exacerbations for 21 COPD patients in the past 4 years (2014 to 2018) (after the bronchoscopy) (Fig. 5). In total, 29 exacerbation events were recorded; 14 of them occurred in 3 patients belonging to subgroup III, 10 of them occurred in 3 patients belonging to subgroup II,
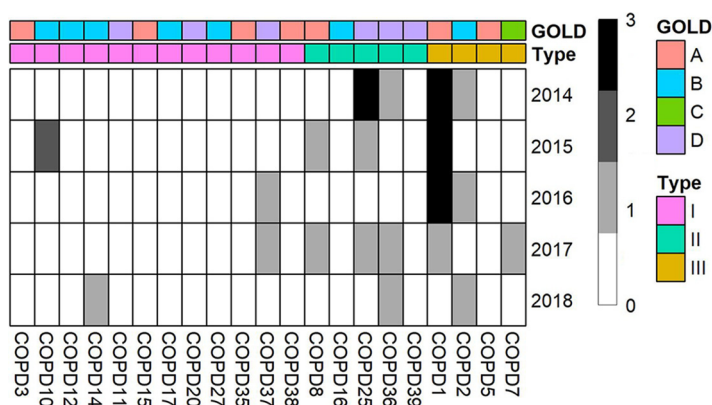
FIG 5 Exacerbation frequency in 21 COPD patients during 2014 to 2018. GOLD (Global Initiative for Obstructive Lung Disease) criteria were used to assess disease severity. A score of "A" represents the mild stage, and a score of "D" represents the most severe stage. Types were defined by microbial composition.

and 5 of them occurred in 3 patients belonging to subgroup I. The exacerbation frequency was significantly higher in patients belonging to subgroups II and III than in those belonging to subgroup I (2 and 3.5 versus 0.4, $P = 1.3e-5$). This observation suggests that colonization of the lung by bacteria of some specific genera (e.g., *Streptococcus*, *Rothia*) might be protective against exacerbation whereas colonization by other bacteria (e.g., *Pseudomonas*) could be harmful. Interestingly, multiple studies have proposed that *Pseudomonas* could be a risk factor for exacerbation in COPD patients (24–26). Associations between individual microbes and clinical features are described in Text S1.

The highly transcribed microbial genes in subgroup I and III samples were similar and enriched for functional catalogs related to metabolism, biosynthesis, replication and repair, and membrane transport (Fig. 6).

Interaction between active lung microbiome and host gene expression. A large (>50%) proportion of the RNA reads in the BALF samples were actually derived from human cells (including macrophages, lymphocytes, and neutrophils) (27), enabling us to investigate the host-microbe interaction in 34 samples. We found that the expression levels of 10 genes were strongly correlated with the read abundance of specific microbes at the genus level (21 genes at the species level) ($P < 0.01$ and $q < 0.01$) (Table S2); however, no specific biological or signaling pathway was enriched in the gene list.

Thousands of genes were differentially expressed in the three microbiome subgroups (adjusted P value [*padj*], <0.01). Interestingly, one gene-enriched pathway ("differential regulation of cytokine production in macrophages and T helper cells by interleukin-17A [IL-17A] and IL-17F") seems to be involved in the immune response to lung microbes, as CD4$^+$ T helper (Th) cells can regulate the adaptive immune response against pathogens and their differentiation has been proposed to be associated with the lung microbiome (28–30). To further investigate the differentiation of Th cells, we compared the expression levels of 36 key genes in this pathway among the three subgroups (Fig. S5A). Overall, samples in subgroup I tended to have a higher expression level of all of these genes than samples in subgroup II (16 of them with $P < 0.05$). In particular, the expression levels of the most critical molecules for Th17 cell differentiation (including IL-6, transforming growth factor β [TGFβ], STAT3, RORC, and IL-17) were all significantly increased in subgroup I samples, and these expression levels were highly synchronized (Fig. 7). The cytokine assay results further confirmed the increased levels of inflammatory cytokines (IL-6, IL-8, and IL-1α) in subgroup I samples ($P < 0.05$, Fig. S5B) compared to those of subgroup II. Subgroup III is not discussed here due to the small sample size.

We then looked for microbes that could potentially associate with the differentiation of Th17 cells. At the genus level, only *Gemella* was positively correlated with the
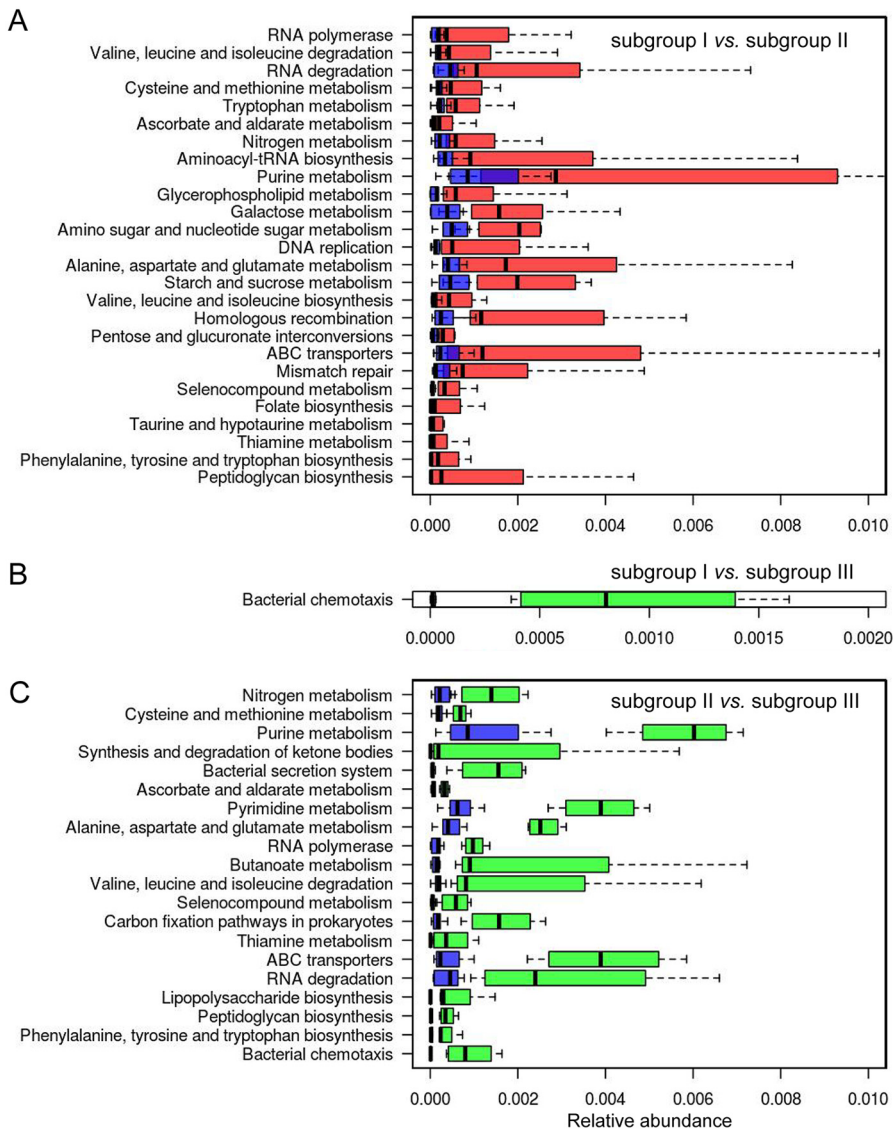
FIG 6 Enrichment of KEGG pathways in microbial genes in different samples. (A) Comparison between subgroup I and subgroup II. (B) Comparison between subgroup I and subgroup III. (C) Comparison between subgroup II and subgroup III. Only pathways with P values of <0.01 and q values of <0.1 (Mann-Whitney U test) are shown. The pathways were sorted by their fold changes in different subgroups (increasing from top to bottom). Red boxes represent subgroup I samples, blue boxes represent subgroup II samples, and green boxes represent subgroup III samples.

expression of *IL-6* (Table S3). At the species level, seven species could potentially stimulate this process by upregulating key genes (Table S3).

DISCUSSION

Although differences in lung microbiome between COPD and non-COPD samples have been found in several studies (15, 19–21), COPD is not the feature that explains most of the variance in lung microbiome in our study. Interestingly, we found that microbes enriched in COPD samples were mostly upper respiratory tract and oral (UO) microbes. They were found exclusively in 15 COPD samples (but were not correlated with severity of COPD), 12 of which had at least two UO microbes codetected ($P < 0.05$). UO microbes can enter the lung through microaspiration but normally are quickly removed by the mucociliary clearance system in the lung. Such clearance is impaired in COPD patients (20, 31), and therefore enrichment of UO microbes is likely to be a
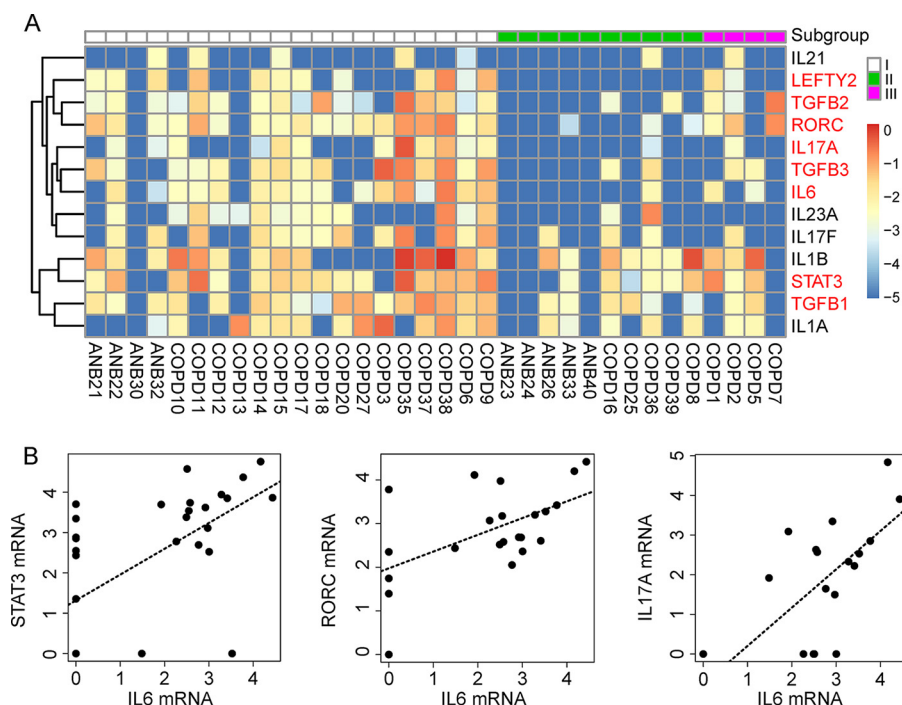
FIG 7 Activation of the Th17 cell differentiation pathway in humans. (A) Expression pattern of 13 key genes involved in Th17 cell differentiation. Differentially expressed genes are indicated in red ($P < 0.05$). (B) Correlation between the expression of *IL-6* and downstream genes in the pathway (*STAT3*, *RORC*, and *IL17A*). The correlation coefficients (*rho*) were 0.641, 0.578, and 0.681, respectively ($P < 0.001$). The gene expression level was calculated as $\log_2$(normalized number of transcripts per million [TPM] + 0.00001).

consequence of COPD. This phenomenon is not specific to COPD but is also observed in other diseases, e.g., mechanically ventilated and pneumonia patients (31, 32).

The composition of the active lung microbiome observed in our study is similar to that found by other studies (2, 8, 9, 12, 33), except that two frequently observed high-abundance microbes, *Prevotella* and *Veillonella*, had relatively low read abundances in our study (1.4% and 3.6%). However, their abundances were higher in the 16S rRNA data (5.1% and 4.3%) and also much higher in the upper respiratory tract (throat swabs were available for seven samples) (see Fig. S6 in the supplemental material). We speculate that these two microbial genera in the lung were acquired by microaspiration from the upper respiratory tract and that their growth was likely to be suppressed by the regional conditions in the lung, although more upper respiratory tract samples were needed to address this issue.

Since the discovery of enterotypes in the gut, similar microbial structures have been identified in other organs (34–36). The concept of a pneumotype was initially proposed by Segal and colleagues in healthy individuals; the researchers defined two pneumotypes according to the abundance of the oral microbes *Veillonella* and *Prevotella* in BALF samples (29, 33). Recently, Shenoy and colleagues also identified two pneumotypes (microbial community states) in HIV and pneumonia patients (30) but with different core microbes. Pneumotypes in both studies were identified on the basis of 16S rRNA data. In our study, three subgroups were identified from metatranscriptome data that may be more functionally relevant as they were inferred from the actively transcribed microbiome.

The subgroup I microbiome was dominated by *Streptococcus* and *Rothia* and was associated with high bacterial biomass, highly expressed microbial genes involved in metabolism and biosynthesis, and activation of the Th17 immune response. These features seem relevant, as bacterial growth requires an abundant nutrient supply and activation of microbial genes that absorb nutrients and synthesize proteins. Actively growing bacteria might activate the host defense system, including recruitment and

differentiation of Th17 cells that mediate host defenses against microbes (37). Coincidentally, Vandeputte and colleagues have recently reported a similar association between microbial load and enterotype in gut (38). We speculate that the correlation could reflect the growth rate variation among different microbes under certain regional conditions. Moreover, COPD patients in this subgroup tended to experience less-frequent exacerbations, suggesting that bacterial colonization could be a crucial stimulus to airway inflammation and could thereby be a risk factor and represent a potential predictor of exacerbations in COPD patients.

Subgroup II had active *Escherichia* and *Ralstonia*, which have been discovered in the respiratory tract and mostly associate with pulmonary inflammation and cystic fibrosis (12, 39, 40). However, they could also have caused the contamination from reagents and bronchoscopes used in our study, as *Escherichia* was the most abundant component in the negative controls. Nevertheless, this subgroup is distinctive in terms of low bacterial biomass and should be considered separately.

Bacterial biomass and microbial gene function enrichment in subgroup III were similar to those in subgroup I, but the proportion of lymphocytes in subgroup III was much higher. This could indicate a more severe inflammation in this subgroup as lymphocytes are normally recruited into the alveolar space during inflammation (41); this may also associated with the fact that COPD patients in this subgroup experienced the most frequent exacerbations. Interestingly, an increased proportion of lymphocytes was found to be associated with the abundance of *Veillonella* (33). However, this correlation was not observed in our data ($P > 0.05$). Rather, the proportion of lymphocytes was found to be associated with the activity of *Bordetella pertussis* (Fig. 4F); infection with *Bordetella pertussis* is not rare in either the healthy population or COPD patients (42, 43), which could be related to the release of pertussis toxin (PT). PT could inhibit the recruitment of neutrophils and macrophages and could impede the movement of lymphocytes into lymph nodes (44). However, there are still samples that have a high proportion of lymphocytes but that have no *Bordetella pertussis*, suggesting that other mechanisms might be involved.

One limitation of the study was the relatively small number of microbial reads used in the analysis (median number = 8,276). Although a host rRNA depletion protocol was applied, 80% of the reads were derived from humans. On the one hand, this enabled us to investigate interactions between the microbiome and the host. On the other hand, less-active microbes could not be detected; however, rarefaction and variance analysis suggested that the data enabled us to identify the most active microbes and capture most of the variance among samples and thus should not have influenced the main conclusion of our study. Another limitation of our study was the relatively small sample size due to the difficulty encountered in collecting lower respiratory tract samples (as the procedure is invasive), which limited the statistical power of the study to detect differences between subgroups; thus, we focused only on the features that most closely associate with lung microbiome (bacterial biomass, Th17 immune response, COPD exacerbation frequency, etc.). Meanwhile, this potentially restricted the study to identification of only a subset of possible microbiome types. Actual lung microbiomes might be more diverse, and their association with clinical features could be more complicated. Nonetheless, data corresponding to the stratified structure of the transcriptionally active lung microbiome and its association with various features are all statistically significant and support the idea of an active host-microbe interaction. Together with previous studies on 16S rRNA and metagenome data, it is tempting to speculate that the lung microbiome variation is stratified in different dimensions in both healthy cases and some disease states. This stratification might represent differences in homeostasis states between the host and the microbiome. Critical follow-up studies should address to what extent the structures exist in different populations, how they are established and persist, and how they interact with the host immunity.

MATERIALS AND METHODS

Subjects and clinical samples. Twenty-five COPD cases and nine non-COPD controls (not paired) were enrolled in this study. All COPD subjects were in a stable state (at least 8 weeks without exacerbation or use of antibiotics). The exclusion conditions included known cardiovascular diseases, renal or liver insufficiency, bronchiectasis, active pulmonary tuberculosis, bronchial asthma, pulmonary fibrosis, and lung cancer. Non-COPD controls had had no respiratory tract infection symptoms in the three months before submitting to bronchoscope examination. Clinical information was obtained for each enrolled patient (Table S1 in the supplemental material).

BALF samples were collected from each subject using a bronchoscope as part of normal clinical management. Two aliquots of 50 ml sterile isotonic saline solution were instilled, with 50% of the volume recovered on average. The BALF samples were immediately placed on ice and processed within 30 min. Bacterial culturing was performed on the BALF samples using an ATB Expression automatic bacterial identification instrument (bioMérieux, Marcy l'Etoile, France). The remnant samples were aliquoted and stored at −80°C before processing. Two negative controls (saline solution passed through a new bronchoscope and a reused sterilized bronchoscope) were collected and processed following the same library preparation protocol.

Metatranscriptome sequencing. A 1-ml aliquot of each whole-BALF sample was pretreated with Turbo DNase (Life Technologies, USA) to decrease the host genome background, according to the manufacturer's instructions. RNA was extracted using a QIAamp UCP pathogen minikit (Qiagen, Valencia, CA, USA), reverse transcribed, and amplified using an Ovation RNA-Seq system (NuGEN, CA, USA). Following fragmentation, the library was constructed using Ovation Ultralow System V2 (NuGEN, CA, USA) and was sequenced on an Illumina HiSeq 2500/4000 platform (Illumina, United Kingdom) (125-bp read length, paired-end protocol).

Metatranscriptome data processing. The raw data were first filtered by base quality score and read length using Trimmomatic (v0.35; SLIDINGINDOW:4:10 MINLEN:70) (45). All filtered reads that could be properly mapped to the human reference genome (GRCh38) or to human cDNA sequences (Ensembl release 83) by Bowtie2 (v2.2.6 – end-to-end, –sensitive) were suspected to represent host contamination and were discarded from further analysis (46). The remaining nonhuman reads were then searched against the ribosome RNA database using SortMeRNA (v2.1, –paired_out) (47), and the nonmapping reads were used for *de novo* assembly. Five assemblers were applied to the data, and the results were compared, "–pre_correction" was used for IDBA_UD and IDBA_Tran (48), "-k 31" was set for Ray(v2.3.1) (49), and default parameters were used for Trinity (v2.1.1) and SOAPdenovo2 (50, 51). Of note, none of these *de novo* assemblers performed well (see Text S1 in the supplemental material); thus, unassembled reads were used directly.

Taxonomy assignment. Unassembled reads were mapped against the NCBI nt database using BLASTN (v2.3.0, -task megablast, -evalue 1e-10, -max_target_seqs 10, -max_hsp 1 – qcov_hsp_perc 60) (52). The results were then used as the input for MEGAN 6 (Min Score 100, Top Percent: 10) (53), and the taxonomic assignment for each read was inferred using the lowest common ancestor (LCA) method. Meanwhile, nonhuman non-rRNA reads were also mapped to the NCBI nr database using Diamond (v0.7.11, –sensitive – c 1) (54), with the thresholds used in MEGAN6 modified accordingly (Min Score: 40, Max Expected 0.001). The conversion file from Gi number to KEGG was used to annotate the function of microbial reads (55). Unless stated otherwise, microbes with a read abundance of at least 1% (among all ABFV reads) in at least 1 sample were regarded as true positives and included in the analysis.

16S rRNA sequencing. The V3-V4 hypervariable region of the bacterial 16S rRNA gene was amplified with barcoded primer set 341F (CCTAYGGGRBGCASCAG) and 806R (GGACTACNNGGGTATCTAAT) with an expected amplicon length of 466 bp. Sequencing of the amplicons was performed using an Illumina HiSeq 2500 instrument (Illumina, United Kingdom) (250-bp read length, paired-end protocol). Reads were analyzed by Mothur (v1.31.2) using the SILVA database (56, 57). Due to the low concentration of microbial DNA in the BALF, enough reads (>10,000) were obtained for only 20 samples with one repeat.

Statistical analysis. Pearson's chi-square test or Fisher's exact test was used for categorical variables, and the Mann-Whitney U test or Kruskal-Wallis rank sum test was used for continuous variables that do not follow a normal distribution. For multiple-test correction, the q value was calculated and a threshold value of 0.1 was applied (58). Benjamini and Hochberg's adjusted P value ($padj$) was given by an integrated pathway analysis (IPA; Ingenuity Systems, Inc.) in the gene enrichment analysis (59), and a threshold value of 0.05 was applied.

More details of the methods employed are provided in the supplemental material.

Ethics statement. The study was approved by the Institutional Review Board of the Peking University People's Hospital. All steps were carried out in accordance with relevant guidelines and regulations. Written informed consent was obtained from each participant.

Data availability. The metatranscriptome and 16S rRNA data have been submitted to NCBI's Sequence Read Archive (SRA) database under project number PRJNA390194.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00199-18.

TEXT S1, DOCX file, 0.04 MB.
FIG S1, TIF file, 0.8 MB.
FIG S2, TIF file, 0.8 MB.

FIG S3, TIF file, 0.4 MB.
FIG S4, TIF file, 0.1 MB.
FIG S5, TIF file, 0.9 MB.
FIG S6, TIF file, 0.1 MB.
TABLE S1, XLSX file, 0.02 MB.
TABLE S2, XLSX file, 0.02 MB.
TABLE S3, XLSX file, 0.02 MB.

REFERENCES

1. Man WH, de Steenhuijsen Piters WA, Bogaert D. 2017. The microbiota of the respiratory tract: gatekeeper to respiratory health. Nat Rev Microbiol 15:259–270. https://doi.org/10.1038/nrmicro.2017.14.

2. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Beck JM, Huffnagle GB, Curtis JL. 2015. Spatial variation in the healthy human lung microbiome and the adapted island model of lung biogeography. Ann Am Thorac Soc 12:821–830. https://doi.org/10.1513/AnnalsATS.201501-029OC.

3. Park H, Shin JW, Park SG, Kim W. 2014. Microbial communities in the upper respiratory tract of patients with asthma and chronic obstructive pulmonary disease. PLoS One 9:e109710. https://doi.org/10.1371/journal.pone.0109710.

4. Huang YJ, Nariya S, Harris JM, Lynch SV, Choy DF, Arron JR, Boushey H. 2015. The airway microbiome in patients with severe asthma: associations with disease features and severity. J Allergy Clin Immunol 136:874–884. https://doi.org/10.1016/j.jaci.2015.05.044.

5. Carmody LA, Zhao J, Schloss PD, Petrosino JF, Murray S, Young VB, Li JZ, LiPuma JJ. 2013. Changes in cystic fibrosis airway microbiota at pulmonary exacerbation. Ann Am Thorac Soc 10:179–187. https://doi.org/10.1513/AnnalsATS.201211-107OC.

6. Maurice CF, Haiser HJ, Turnbaugh PJ. 2013. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell 152:39–50. https://doi.org/10.1016/j.cell.2012.10.052.

7. McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, Henrissat B, Oozeer R, Cools-Portier S, Gobert G, Chervaux C, Knights D, Lozupone CA, Knight R, Duncan AE, Bain JR, Muehlbauer MJ, Newgard CB, Heath AC, Gordon JI. 2011. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci Transl Med 3:106ra106. https://doi.org/10.1126/scitranslmed.3002701.

8. Cameron SJ, Lewis KE, Huws SA, Lin W, Hegarty MJ, Lewis PD, Mur LA, Pachebat JA. 2016. Metagenomic sequencing of the chronic obstructive pulmonary disease upper bronchial tract microbiome reveals functional changes associated with disease severity. PLoS One 11:e0149095. https://doi.org/10.1371/journal.pone.0149095.

9. Millares L, Pérez-Brocal V, Ferrari R, Gallego M, Pomares X, García-Núñez M, Montón C, Capilla S, Monsó E, Moya A. 2015. Functional metagenomics of the bronchial microbiome in COPD. PLoS One 10:e0144448. https://doi.org/10.1371/journal.pone.0144448.

10. Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, Giannoukos G, Boylan MR, Ciulla D, Gevers D, Izard J, Garrett WS, Chan AT, Huttenhower C. 2014. Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci U S A 111:E2329–E2338. https://doi.org/10.1073/pnas.1319284111.

11. Abu-Ali GS, Mehta RS, Lloyd-Price J, Mallick H, Branck T, Ivey KL, Drew DA, DuLong C, Rimm E, Izard J, Chan AT, Huttenhower C. 2018. Metatranscriptome of human faecal microbial communities in a cohort of adult men. Nat Microbiol 3:356–366. https://doi.org/10.1038/s41564-017-0084-4.

12. Lee SW, Kuan CS, Wu LS, Weng JT. 2016. Metagenome and metatranscriptome profiling of moderate and severe COPD sputum in Taiwanese Han males. PLoS One 11:e0159066. https://doi.org/10.1371/journal.pone.0159066.

13. Westermann AJ, Barquist L, Vogel J. 2017. Resolving host-pathogen interactions by dual RNA-seq. PLoS Pathog 13:e1006033. https://doi.org/10.1371/journal.ppat.1006033.

14. World Health Organization (WHO). 2008. The global burden of disease: 2004 update. WHO Press, Geneva, Switzerland.

15. Wang Z, Bafadhel M, Haldar K, Spivak A, Mayhew D, Miller BE, Tal-Singer R, Johnston SL, Ramsheh MY, Barer MR, Brightling CE, Brown JR. 2016. Lung microbiome dynamics in COPD exacerbations. Eur Respir J 47:1082–1092. https://doi.org/10.1183/13993003.01406-2015.

16. Huang YJ, Erb-Downward JR, Dickson RP, Curtis JL, Huffnagle GB, Han MK. 2017. Understanding the role of the microbiome in chronic obstructive pulmonary disease: principles, challenges, and future directions. Transl Res 179:71–83. https://doi.org/10.1016/j.trsl.2016.06.007.

17. Mayhew D, Devos N, Lambert C, Brown JR, Clarke SC, Kim VL, Magid-Slav M, Miller BE, Ostridge KK, Patel R, Sathe G, Simola DF, Staples KJ, Sung R, Tal-Singer R, Tuck AC, Van Horn S, Weynants V, Williams NP, Devaster JM, Wilkinson TMA, AERIS Study Group. 31 January 2018. Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. Thorax https://doi.org/10.1136/thoraxjnl-2017-210408.

18. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, Young VB, Toews GB, Curtis JL, Sundaram B, Martinez FJ, Huffnagle GB. 2011. Analysis of the lung microbiome in the "healthy" smoker and in COPD. PLoS One 6:e16384. https://doi.org/10.1371/journal.pone.0016384.

19. Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, Cooper J, Sin DD, Mohn WW, Hogg JC. 2012. The lung tissue microbiome in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 185:1073–1080. https://doi.org/10.1164/rccm.201111-2075OC.

20. Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE. 2012. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. PLoS One 7:e47305. https://doi.org/10.1371/journal.pone.0047305.

21. Huang YJ, Sethi S, Murphy T, Nariya S, Boushey HA, Lynch SV. 2014. Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. J Clin Microbiol 52:2813–2823. https://doi.org/10.1128/JCM.00035-14.

22. Einarsson GG, Comer DM, McIlreavey L, Parkhill J, Ennis M, Tunney MM, Elborn JS. 2016. Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. Thorax 71:795–803. https://doi.org/10.1136/thoraxjnl-2015-207235.

23. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12:87. https://doi.org/10.1186/s12915-014-0087-z.

24. Garcia-Vidal C, Almagro P, Romani V, Rodriguez-Carballeira M, Cuchi E, Canales L, Blasco D, Heredia JL, Garau J. 2009. Pseudomonas aeruginosa in patients hospitalised for COPD exacerbation: a prospective study. Eur Respir J 34:1072–1078. https://doi.org/10.1183/09031936.00003309.

25. Murphy TF. 2008. The many faces of Pseudomonas aeruginosa in chronic obstructive pulmonary disease. Clin Infect Dis 47:1534–1536. https://doi.org/10.1086/593187.

26. Martinez-Solano L, Macia MD, Fajardo A, Oliver A, Martinez JL. 2008. Chronic Pseudomonas aeruginosa infection in chronic obstructive pulmonary disease. Clin Infect Dis 47:1526–1533. https://doi.org/10.1086/593186.

27. Collins AM, Rylance J, Wootton DG, Wright AD, Wright AK, Fullerton DG, Gordon SB. 24 March 2014. Bronchoalveolar lavage (BAL) for research; obtaining adequate sample yield. J Vis Exp https://doi.org/10.3791/4345.

28. Masopust D, Schenkel JM. 2013. The integration of T cell migration, differentiation and function. Nat Rev Immunol 13:309–320. https://doi.org/10.1038/nri3442.

29. Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG, Li Y, Shen N, Ghedin E, Morris A, Diaz P, Huang L, Wikoff WR, Ubeda C, Artacho A, Rom WN, Sterman DH, Collman RG, Blaser MJ, Weiden MD. 2016. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. Nat Microbiol 1:16031. https://doi.org/10.1038/nmicrobiol.2016.31.

30. Shenoy MK, Iwai S, Lin DL, Worodria W, Ayakaka I, Byanyima P, Kaswabuli S, Fong S, Stone S, Chang E, Davis JL, Faruqi AA, Segal MR, Huang L, Lynch SV. 2017. Immune response and mortality risk relate to distinct lung microbiomes in patients with HIV and pneumonia. Am J Respir Crit Care Med 195:104–114. https://doi.org/10.1164/rccm.201603-0523OC.

31. Sze MA, Hogg JC, Sin DD. 2014. Bacterial microbiome of lungs in COPD. Int J Chron Obstruct Pulmon Dis 9:229–238.

32. Paju S, Scannapieco FA. 2007. Oral biofilms, periodontitis, and pulmonary infections. Oral Dis 13:508–512. https://doi.org/10.1111/j.1601-0825.2007.01410a.x.

33. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z, Chen H, Berger KI, Goldring RM, Rom WN, Blaser MJ, Weiden MD. 2013. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. Microbiome 1:19. https://doi.org/10.1186/2049-2618-1-19.

34. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, Antolín M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, et al. 2011. Enterotypes of the human gut microbiome. Nature 473:174–180. https://doi.org/10.1038/nature09944.

35. Oh J, Byrd AL, Deming C, Conlan S, NISC Comparative Sequencing Program, Kong HH, Segre JA. 2014. Biogeography and individuality shape function in the human skin metagenome. Nature 514:59–64. https://doi.org/10.1038/nature13786.

36. Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UM, Zhong X, Koenig SS, Fu L, Ma ZS, Zhou X, Abdo Z, Forney LJ, Ravel J. 2012. Temporal dynamics of the human vaginal microbiota. Sci Transl Med 4:132ra52. https://doi.org/10.1126/scitranslmed.3003605.

37. Curtis MM, Way SS. 2009. Interleukin-17 in host defence against bacterial, mycobacterial and fungal pathogens. Immunology 126:177–185. https://doi.org/10.1111/j.1365-2567.2008.03017.x.

38. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, Wang J, Tito RY, De Commer L, Darzi Y, Vermeire S, Falony G, Raes J. 2017. Quantitative microbiome profiling links gut community variation to microbial load. Nature 551:507–511. https://doi.org/10.1038/nature24460.

39. Mourani PM, Harris JK, Sontag MK, Robertson CE, Abman SH. 2011. Molecular identification of bacteria in tracheal aspirate fluid from mechanically ventilated preterm infants. PLoS One 6:e25959. https://doi.org/10.1371/journal.pone.0025959.

40. Coenye T, Vandamme P, LiPuma JJ. 2002. Infection by Ralstonia species in cystic fibrosis patients: identification of R. pickettii and R. mannitolilytica by polymerase chain reaction. Emerg Infect Dis 8:692–696. https://doi.org/10.3201/eid0807.010472.

41. Berman JS, Beer DJ, Theodore AC, Kornfeld H, Bernardo J, Center DM. 1990. Lymphocyte recruitment to the lung. Am Rev Respir Dis 142:238–257. https://doi.org/10.1164/ajrccm/142.1.238.

42. He Q, Viljanen MK, Arvilommi H, Aittanen B, Mertsola J. 1998. Whooping cough caused by Bordetella pertussis and Bordetella parapertussis in an immunized population. JAMA 280:635–637. https://doi.org/10.1001/jama.280.7.635.

43. Hashemi SH, Nadi E, Hajilooi M, Seif-Rabiei MA, Samaei A. 2015. High seroprevalence of Bordetella pertussis in patients with chronic obstructive pulmonary disease: a case-control study. Tanaffos 14:172–176.

44. Carbonetti NH, Artamonova GV, Mays RM, Worthington ZE. 2003. Pertussis toxin plays an early role in respiratory tract colonization by Bordetella pertussis. Infect Immun 71:6358–6366. https://doi.org/10.1128/IAI.71.11.6358-6366.2003.

45. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

46. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

47. Kopylova E, Noe L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217. https://doi.org/10.1093/bioinformatics/bts611.

48. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174.

49. Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. 2012. Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biol 13:R122. https://doi.org/10.1186/gb-2012-13-12-r122.

50. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. https://doi.org/10.1038/nprot.2013.084.

51. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18.

52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

53. Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S, Ruscheweyh HJ, Tappu R. 2016. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol 12:e1004957. https://doi.org/10.1371/journal.pcbi.1004957.

54. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

55. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361. https://doi.org/10.1093/nar/gkw1092.

56. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM.01541-09.

57. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41: D590–D596. https://doi.org/10.1093/nar/gks1219.

58. Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100:9440–9445. https://doi.org/10.1073/pnas.1530509100.

59. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300. https://www.jstor.org/stable/2346101.