Shared and Specific Lung Microbiota with Metabolic Profiles in Bronchoalveolar Lavage Fluid Between Infectious and Inflammatory Respiratory Diseases

Yukun He1
Wenxi Yu1
Pu Ning1,2
Qiongzheng Luo1,3
Lili Zhao1
Yu Xie1
Yan Yu1
Xinqian Ma1
Li Chen1
Yali Zheng1,4
Zhancheng Gao1,4

1Department of Respiratory and Critical Care Medicine, Peking University People’s Hospital, Beijing, People’s Republic of China; 2Department of Respiratory and Critical Care Medicine, Starbucks Affiliated Hospital, Xi’an Jiaotong University, Xi’an, Shaanxi, People’s Republic of China; 3Department of Respiratory & Critical Care Medicine, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, People’s Republic of China; 4Department of Respiratory, Critical Care, and Sleep Medicine, Xiang’an Hospital of Xi’an University, School of Medicine, Xiamen University, Xiamen, People’s Republic of China

Correspondence: Yali Zheng
Department of Respiratory, Critical Care, and Sleep Medicine, Xiang’an Hospital of Xi’an University, School of Medicine, Xi’an University, Xi’an, People’s Republic of China
Email yali_zheng@126.com

Zhancheng Gao
Department of Pulmonary and Critical Care Medicine, Peking University People’s Hospital, Beijing, People’s Republic of China
Email zcgao@bjmu.edu.cn

Background: Infiltration of the lower respiratory tract (LRT) microenvironment could be significantly associated with respiratory diseases. However, alterations in the LRT microbiome and metabolome in infectious and inflammatory respiratory diseases and their correlation with inflammation still need to be explored.

Methods: Bronchoalveolar lavage samples from 44 community-acquired pneumonia (CAP) patients, 29 connective tissue disease-associated interstitial disease (CTD-ILD) patients, and 30 healthy volunteers were used to detect microbiota and metabolites through 16S rRNA gene sequencing and untargeted high-performance liquid chromatography with mass spectrometry.

Results: The composition of the LRT microbial communities and metabolites differed in disease states. CAP patients showed a significantly low abundance and both diseases presented a depletion of some genera of the phylum Bacteroidetes, including Prevotella, Porphyromonas, and health-associated metabolites, such as sphingosine (d16:1), which were negatively correlated with infectious indicators. In contrast, Bacillus and Mycoplasma were both enriched in the disease groups. Streptococcus was specifically increased in CTD-ILD. In addition, co-elevated metabolites such as FA (22:4) and pyruvic acid represented hypoxia and inflammation in the diseases. Significantly increased levels of amino acids and succinate, as well as decreased itaconic acid levels, were observed in CAP patients, whereas CTD-ILD patients showed only a handful of specific metabolic alterations. Functions related to microbial lipid and amino acid metabolism were significantly altered, indicating the possible contributions of microbial metabolism. Dual omics analysis showed a moderate positive correlation between the microbiome and metabolome. The levels of L-isoleucine and L-arginine were negatively correlated with Streptococcus, and itaconic acid positively correlated with Streptococcus.

Conclusion: In the LRT microenvironment, shared and specific alterations occurred in CAP and CTD-ILD patients, which were associated with inflammatory and immune reactions, which may provide a new direction for future studies aiming to elucidate the mechanism, improve the diagnosis, and develop therapies for different respiratory diseases.

Keywords: community-acquired pneumonia, connective tissue disease-associated interstitial disease, microbiome, metabolome, shared and specific alterations

Introduction
Community-acquired pneumonia (CAP) and connective tissue disease-associated interstitial disease (CTD-ILD) are dynamic, heterogeneous, debilitating lung diseases with multiple comorbidities that affect millions of people worldwide.1,2 In
Materials and Methods

Patient Screening, Enrollment, and Clinical Examination

Application of the study was submitted to the Ethical Review Committee of Peking University People’s Hospital (No.2016PHB202-01) and the CAP cohort was registered at ClinicalTrials.gov (NCT03093220). The study was conducted under the principles of the Helsinki Declaration and patient informed consent was obtained from all participants or their next of kin for incapacitated patients or unconscious subjects who were unable to give informed consents.

Forty-four CAP patients admitted to any of the four participating hospitals between March 2017 and August 2017 as part of a multicenter clinical study were enrolled. Twenty-nine patients with CTD-ILD were admitted to Peking University People’s Hospital and thirty healthy volunteers from the medical center were recruited. CAP and CTD-ILD were defined according to the standard published by the America/American Thoracic Society.2,16

BALF Sample Collection, Detection, and Storage

Bronchoscopy was performed within 72 h after hospital admission. Briefly, 100 mL of sterile normal saline was instilled into the diseased region according to CT scan results (for CAP patients), or into the right middle lobe or lingual lobe (for controls). Aliquots were then retrieved by gentle suctioning. Cell counts and protein concentrations of BALF samples were analyzed immediately after collection. The remaining BALF samples underwent centrifugation and then supernatants and precipitates were separated and stored at −80 °C.

16S Ribosomal RNA Gene Sequencing

Total DNA from BALF samples was extracted using the CTAB/SDS method. 16S rRNA genes of the V3-V4 region were amplified with primers (338F and 806R) and the PCR products were detected by agarose gel electrophoresis (2%) and then mixed in equidensity ratios. The mixture of PCR products was purified with the QIAEX II Gel Extraction Kit (QIAGEN, Germany). Sequencing libraries were generated using the TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, USA) and then sequenced on the HiSeq2500 platform and 250 bp paired-end raw reads were generated.
Sequencing of Reagent Controls
To detect contamination introduced by the experimental operations, specimen-free reagent controls for extraction and PCR amplification (no template, ddH2O only) were included in the experiment. Possible contaminants were excluded from the present analysis using the R “Decontam” package.

16S rRNA Gene Sequencing Analysis
Paired-end reads were merged and quality filtering, trimming, and dereplication of the raw tags were performed under specific conditions to obtain the high-quality clean reads using VSEARCH.17 Chimeric sequences were detected by aligning clean reads to the Silva database (v.132)18 with the UCHIME algorithm. The effective reads were denoised to generate amplicon sequence variants (ASVs) using unoise3. Taxonomy assignment was performed on ASVs using USEARCH (v10) against the RDP database (v 11.5)19 and Greengenes database.20 The Abundance-based coverage estimator (ACE) and the Shannon index were calculated to evaluate alpha diversity. Beta diversity was assessed by permutational multivariate analysis of variance (PERMANOVA) and ANOSIM based on Bray-Curtis distances. Differential bacterial taxa among groups were assessed using the “makeContrasts” function within a multifactorial design in the metagenomeSeq v1.35.0 R package. Statistically significant bacterial differences (LDA > 2, P < 0.05) associated with different groups were explored using linear discriminant analysis (LDA) effect size (LEfSe). The microbiome phenotypes were predicted by BugBase21 based on the Greengenes annotation. PICRUSt2 was used to identify predicted associated pathways from the inferred metagenomes of taxa with the “stratified” mode.22 Comparisons of the predicted pathways were obtained with STAMP. Spearman’s rho was calculated using the “corr.test” function within the R package psych v1.8.12105 based on centered log-ratio-transformed genome relative abundance.

Untargeted Metabolic Profiling Analysis
The detailed method of BALF sample pretreatment is described in the Supplement 3. All compounds in BALF were analyzed using a Cortecs C18 column (2.1 × 100 mm, Waters) on an Ultimate 3000 UHPLC (Dionex) system coupled with a Q Exactive (Orbitrap) mass spectrometer (Thermo Fisher, CA). Detailed parameters for the untargeted metabolic and lipidomic analyses were set following the protocols of our previously reported study.23 Data-dependent MS/MS acquisition (DDA) of all samples was analyzed using TraceFinderTM (Thermo, CA). Metabolites and lipids were identified based on matching precursor and characteristic fragment masses and then assigned using in-house databases in “screening” mode. Any metabolite feature with more than 20% missing values was removed from the result.24,25 Missing values were estimated by the Bayesian PCA (BPCA) method. Data were normalized by the QC group and then auto-scaled using MetaboAnalystR 3.0.26 The normalized data were used for downstream analysis. Principal component analysis (PCA) and orthogonal projections to latent structures discriminant analysis (OPLS-DA) were conducted using SIMCA v14.1 (Umetrics, Sweden).

Metabolomic and Metagenomic Data Integration
Correlation analysis between metagenomic and metabolomic data was undertaken using ASVs and metabolites. DIABLO function from the mixOmics v6.6.2 R package27,28 was used to generate integrated metagenomic and metabolomic signatures. The analysis was performed using centered log-ratio-transformed taxa relative abundance and log-transformed auto-scaled metabolite data. The optimum number of components and variables included within the final model was determined using the “tune.block.splsda” function with 50 × 10-fold cross-validation.29

Statistical Analysis
All categorical variables are presented as numbers (percentages), parametric continuous variables are presented as the mean ± SD, and nonparametric continuous variables are presented as median and interquartile ranges (25th and 75th percentiles). Student’s t-test or analysis of variance (ANOVA) with post-hoc Tukey HSD test were used to analyze continuous parametric data, whereas continuous nonparametric data were analyzed using Mann–Whitney U or a Kruskal–Wallis test. All categorical data were analyzed using the chi-square or Fisher’s exact test. All tests were two-sided, the p values were corrected using FDR, and p < 0.05 was considered statistically significant.
Results
Clinical Characteristics of the Study Population
To identify clinical differences among groups, we first separately described the clinical characteristics of the 44 CAP patients, 29 CTD-ILD patients, and 30 healthy controls (Table 1). Lymphocyte counts and albumin (ALB) levels in peripheral blood (PB) were lower in the CAP patients and CTD-ILD patients than in the HC group, while polymorphonuclear cell (PMN) percentages and albumin (ALB) levels in the BALF, and inflammatory markers in the CAP patients were significantly higher than those in the HC group. Creatinine and total triglyceride (TG) contents were significantly different between the CAP and CTD-ILD groups. No significant differences were observed in age or various comorbidities among the groups. Collectively, these data indicated that CAP patients exhibited substantially different clinical characteristics from those of CTD-ILD or HC subjects.

The LRT Microbial Community is Altered in CAP and CTD-ILD Patients
To explore the LRT bacterial community composition in different diseases, we conducted high throughput 16S rRNA gene sequencing. Rarefaction curves of all samples indicated that the sequencing depth was sufficient to capture the full diversity of taxa associated with each microbial community (Figure S1). In total, 7551 ASVs were identified across all 103 BALF samples (Table S1). After filtering for sequence variants present in at least two samples with a minimum relative abundance of 0.05%, we obtained a total of 1782 sequence variants for community analysis.

In terms of alpha diversity, the ACE index showed a significant decrease in species abundance in the CAP subjects (Figure 1A, all \( P < 0.05 \)), although no significant differences were observed in the Shannon index (Figure 1B). The PERMANOVA index showed statistically significant differences among all groups (Figure 1C and E). NMDS based on the Bray–Curtis distances showed tight clustering of samples from the CTD-ILD group, whereas CAP group samples were distributed diffusely (Figure 1D) and the ANOSIM index presented a larger variability within the CAP group (Figure 1F). We compared the subgroups within the samples and found no significant difference in microbiota composition in terms of smoking status (\( P = 0.509 \)) or sex (\( P = 0.085 \)) (Figure S2). Thus, the LRT microbial community was altered in the diseased groups, and the CAP group presented a lower species abundance and higher heterogeneity.

Shared vs Disease-Specific Microbial Alterations in CAP and CTD-ILD Patients
To identify significantly different genera, we employed the metagenomeSeq R package within a multifactorial design. Moreover, correlation analysis to identify bacteria linked with specific clinical indicators was performed. The four phyla, Proteobacteria, Firmicutes, Bacteroidetes, and Fusobacteria, dominated the community composition across all cohorts (Figure 2A). Specifically, CTD-ILD samples had a higher abundance of Proteobacteria and Firmicutes. Prevotella and Alloprevotella were decreased in both CAP and CTD-ILD patients and were negatively correlated with white blood cell counts (WBCs) and C-reactive protein (CRP) levels (Figure 2B, Table S2). Bacillus was increased in diseased groups. The abundances of the genus Pseudomonas, and well-known gut bacteria, such as Enterococcus, were all significantly increased in CAP patients. Pseudomonas was positively correlated with CRP level. Additionally, Streptococcus intermedeus which was enriched in the CTD-ILD group was moderately correlated with the erythrocyte sedimentation rate (ESR) level (\( r = 0.40, P = 0.003 \)). The differentially abundant genera were further confirmed using LEfSe analysis (Figure 2C), revealing that 19 and 2 genera were discriminative for CTD-ILD and CAP patients, respectively. Streptococcus and Haemophilus were preponderant in CTD-ILD patients, whereas CAP patients presented a higher abundance of Mycoplasma.

Predicted Functional Characterization of the LRT Microbiome
Predicted phenotypes based on taxonomic classification were analyzed with BugBase. The results suggested that gram-positive bacteria were more abundant in the CAP patients than in the HC group, while the abundance of gram-negative bacteria was decreased in the CAP group (Figure S3). No significant difference was observed in the CTD-ILD group.

To explore differences in the potential function of microbiota between the groups, we used PICRUSt2 based on the Kyoto Encyclopedia of Genes and Genomes.
Table 1 Demographical and Clinical Features of Included Subjects

<table>
<thead>
<tr>
<th></th>
<th>CAP (n = 44)</th>
<th>CTD-ILD (n = 29)</th>
<th>HC (n = 30)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>51.9 ± 3.07</td>
<td>51.2 ± 2.46</td>
<td>56.1 ± 2.22</td>
<td>0.449</td>
</tr>
<tr>
<td>Gender, n, (% male)</td>
<td>28(63.6)</td>
<td>6(20.7)</td>
<td>11(36.7)</td>
<td>0.002</td>
</tr>
<tr>
<td>Smokers, n (%)</td>
<td>8(18.2)</td>
<td>2(6.9)</td>
<td>0(0)</td>
<td>0.407</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes Mellitus, n (%)</td>
<td>6(13.6)</td>
<td>6(20.7)</td>
<td>3(10.0)</td>
<td>0.087</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>14(31.8)</td>
<td>6(20.7)</td>
<td>11(36.7)</td>
<td>0.522</td>
</tr>
<tr>
<td>Hyperlipidemia, n (%)</td>
<td>9(20.5)</td>
<td>5(17.2)</td>
<td>3(10.0)</td>
<td>0.119</td>
</tr>
<tr>
<td>Coronary Heart Disease, n (%)</td>
<td>5(11.4)</td>
<td>2(6.9)</td>
<td>5(16.7)</td>
<td>0.166</td>
</tr>
<tr>
<td>Laboratory Findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peripheral blood</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WBC (×10^9/L)</td>
<td>6.40(5.13–9.73)</td>
<td>5.74(4.52–7.80)</td>
<td>6.87(5.68–8.31)</td>
<td>0.193</td>
</tr>
<tr>
<td>Neutrophils(×10^9/L)</td>
<td>4.27(3.55–7.13)</td>
<td>3.70(3.05–5.35)</td>
<td>4.03(3.08–4.97)</td>
<td>0.180</td>
</tr>
<tr>
<td>Lymphocytes(×10^9/L)</td>
<td>1.03(0.70–1.50)</td>
<td>1.30(0.80–1.88)</td>
<td>1.91(1.56–2.66)</td>
<td>0.000</td>
</tr>
<tr>
<td>NLR</td>
<td>3.43(2.31–991)</td>
<td>0.06(0.05–0.19)</td>
<td>2.05(1.49–2.63)</td>
<td>0.000</td>
</tr>
<tr>
<td>BUN (mmol/L)</td>
<td>4.18(2.72–5.40)</td>
<td>4.67(3.3–6.1)</td>
<td>4.92(4.07–5.6)</td>
<td>0.243</td>
</tr>
<tr>
<td>Cr (μmol/L)</td>
<td>65.5(55–83.75)</td>
<td>52(39–67)</td>
<td>60(56–72)</td>
<td>0.009</td>
</tr>
<tr>
<td>ALT (μU/L)</td>
<td>34.5(12–60.25)</td>
<td>26(20–35.5)</td>
<td>17.5(12–29.8)</td>
<td>0.029</td>
</tr>
<tr>
<td>AST (μU/L)</td>
<td>28(20.5–73.75)</td>
<td>27(20.5–37)</td>
<td>19(16–22.5)</td>
<td>0.000</td>
</tr>
<tr>
<td>CK (μU/L)</td>
<td>77.5(46.5–146)</td>
<td>64.5(45.5–134)</td>
<td>58(44–85)</td>
<td>0.312</td>
</tr>
<tr>
<td>ALB (g/L)</td>
<td>35(29–39.78)</td>
<td>34.8(31.5–37.9)</td>
<td>44.1(41–45.4)</td>
<td>0.000</td>
</tr>
<tr>
<td>Glucose (mmol/L)</td>
<td>5.18(4.59–6.48)</td>
<td>4.7(4.53–6.35)</td>
<td>4.89(4.62–5.73)</td>
<td>0.323</td>
</tr>
<tr>
<td>TC (mmol/L)</td>
<td>3.51(2.83–4.36)</td>
<td>4.78(4.27–5.44)</td>
<td>4.63(3.82–5.56)</td>
<td>0.000</td>
</tr>
<tr>
<td>TG (mmol/L)</td>
<td>0.92(0.75–1.58)</td>
<td>1.69(1.22–2.13)</td>
<td>1.4(0.89–2.37)</td>
<td>0.01</td>
</tr>
<tr>
<td>BALF related</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMN percentages (%)</td>
<td>18.00(1.75–65.50)</td>
<td>3.00(1.00–9.75)</td>
<td>1.00(0.50–2.00)</td>
<td>0.000</td>
</tr>
<tr>
<td>Lymphocyte percentages (%)</td>
<td>21.00(11.00–43.50)</td>
<td>35.00(14.75–54.75)</td>
<td>12.00(8.00–30.75)</td>
<td>0.059</td>
</tr>
<tr>
<td>Eosinophil percentages (%)</td>
<td>0.00(0.00–1.00)</td>
<td>0.00(0.00–0.75)</td>
<td>0.00(0.00–1.00)</td>
<td>0.537</td>
</tr>
<tr>
<td>Albumen concentration (g/L)</td>
<td>1128.69(312.07–2230.86)</td>
<td>171.42(75.04–323.68)</td>
<td>158.88(75.50–220.13)</td>
<td>0.000</td>
</tr>
<tr>
<td>Inflammatory markers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCT (μg/L)</td>
<td>0.16(0.05–1.02)</td>
<td>0.06(0.05–0.19)</td>
<td>0.05(0.05–0.09)</td>
<td>0.032</td>
</tr>
<tr>
<td>CRP (mg/L)</td>
<td>40.75(12.12–132.75)</td>
<td>2.52(0.67–7.91)</td>
<td>1.37(0.78–2.81)</td>
<td>0.000</td>
</tr>
<tr>
<td>ESR (mm/h)</td>
<td>41.00(20.00–43.00)</td>
<td>16.00(7.00–29.00)</td>
<td>8.50(6.00–13.50)</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Notes: *Statistical significance exists between CAP and healthy controls; †Statistical significance exists between CAP and CTD-ILD ‡Statistical significance exists between CTD-ILD and healthy controls.

Abbreviations: BALF, bronchoalveolar lavage; PMN, polymorphonuclear leukocyte; TC, serum total cholesterol; TG, serum total triglycerides; AST, aspartate aminotransferase; ALT, alanine aminotransferase; CK, creatine kinase; Cr, creatinine; WBC, white blood cell; N, neutrophil; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; PCT, procalcitonin; ALB, albumin.
and Metacyc databases. No significant differences were observed in the overall predicted pathways among all samples (Figure S4). However, several annotated functions were differentially expressed among the groups at the individual pathway level (Tables S3–6).

The functions “glycerolipid metabolism” and “androstenedione degradation” were significantly enriched in the CAP group whereas “D-arginine and D-ornithine metabolism” was significantly enriched in the CTD-ILD group. These results suggested that the microbial metabolic capability

![Figure 1](https://doi.org/10.2147/JIR.S342462)

**Figure 1** Lower respiratory tract microbiome alpha and beta diversity of taxonomic analysis. (A) Comparison of the ACE index based on the genus profile in different groups for assessment of microbiome alpha diversity of three groups. *P* ≤0.05. (B) Comparison of the Shannon index based on the genus profile in different groups. (C) Beta diversity was assessed by PERMANOVA based on Bray-Curtis distances using principal coordinate analysis (PCoA). (D) Beta diversity was assessed by ANOSIM based on Bray-Curtis distances using non-metric multidimensional scaling (NMDS). (E) PERMANOVA index of groups. (F) ANOISM index of groups.
could contribute to a disease-associated disruption of the LRT metabolic profiles.

**Shared vs Disease-Specific Metabolic Alterations in CAP and CTD-ILD Patients**

To explore the metabolic and lipidomic profiles of the LRT among CAP and CTD-ILD patients, we conducted an untargeted HPLC-MS analysis of 103 paired BALF samples. This assay identified a total of 505 compounds that likely originated from either the microbiota or the host. According to the PCA result, we found that, in agreement with NMDS analysis of microbiota, CAP samples were more diffusely grouped (Figure 3A). OPLS-DA score plots showed a clear separation among all groups (Figure S5).

To identify metabolic alterations, we calculated the fold changes (FCs), P values, and OPLS-DA variable importance in the projection (VIP) scores for all metabolic features. Using a linear model with adjustment for covariates (age, sex, smoking, DM2, and hyperlipidemia), we found that 47 compounds showed the same trend in disease groups, and 181 and 36 compounds were identified as specific to CAP and CTD-ILD patients, respectively (P ≤ 0.05, and VIP > 1) (Figure 3B, Tables S7–9).

Eight lipid species were increased in both diseases. Notably, PI (18:0/20:4) was moderately associated with elevated CRP (r = 0.451) and neutrophils in BALF (r = 0.417), and PC (18:0p/20:4) was moderately associated with elevated ESR (r = 0.564) (Figure 3C, Table S10). These two lipids have one of the acyl chains constituted by C20:4, that is, the
structure of arachidonic acid, which can be oxidized to prostanoids, well-known players in the inflammatory response. Pyruvic acid was also elevated in both diseases due to active glycolysis. In contrast, sphingosine \([\text{So}] (\text{d}16:1)\) was decreased in the disease groups and inversely related to ESR level \((r \text{ value} = -0.562)\).

For CAP-specific metabolic alterations, the levels of 12 kinds of amino acid and nucleotide-related metabolites in BALF were significantly increased, suggesting proteolytic activity and DNA damage induced by oxidative stress. Besides, we found significantly lower levels of fumarate and itaconic acid and increased succinate levels in the CAP groups, indicating remodeling of the tricarboxylic acid (TCA) cycle accompanying inflammatory macrophage activation. However, only a handful of metabolites were uniquely altered in CTD-ILD.

Collectively, the co-elevated metabolites represented general disorders in the disease states, such as hypoxia and inflammation. In acute infection, metabolic disorders are more evident in the LRT.

Dual-Omics Analysis Reveals the Relationship Between the LRT Microbiota and Metabolites in CAP and CTD-ILD

We next sought to examine the associations between the microbiome and metabolome. A moderate positive correlation was observed in the dual omics analysis (Figure 4A). Ultimately, two major microbiome/metabolite clusters were defined (Figure 4B). The first cluster showed associations between a group of five species and nine metabolites. All five species were depleted in CAP patients. *Porphyromonas*, and *Streptococcus* were negatively correlated with the amino acids, Phe, Arg, and Ile, and positively correlated with fumarate and itaconic acid. The second network did not contain any differential metabolites and therefore may represent interactions unrelated to disease states.
Discussion

Several studies have suggested that infiltration of the LRT microenvironment can be related to the pathogenesis of respiratory diseases.\(^{32,33}\) However, due to obvious limitations in obtaining lower airway samples and heterogeneity of lung diseases, the LRT microbiome and associated metabolites have remained poorly characterized. In this study, we present the first analysis to our knowledge of the microbiota and metabolome of the human LRT in both CAP and CTD-ILD patients. We revealed that these two diseases had shared and specific alterations in the lung microenvironment, which could be informative for future research exploring disease mechanisms, or could suggest potential therapeutic targets through validation of their effects in model organisms.

Importantly, both diseases showed a depletion of *Prevotella*, *Alloprevotella*, and *Porphyromonas* compared to the HC group, and all three genera belong to the phylum Bacteroidetes. Duncan reported that most Bacteroidetes are pH sensitive and fail to prosper in acidic environments,\(^ {34}\) so this finding may be explained by the decreased local pH of the LRT microenvironment due to inflammation.\(^ {35}\) In the CAP subjects, loss of overall species abundance was observed, which may diminish colonization resistance against pathogens. We found that some gut-associated bacteria were specifically more abundant in CAP patients, and previous studies have indicated that increased lung and gut permeability may induce bacterial migration in various respiratory diseases, such as acute respiratory disease syndrome (ARDS).\(^ {33,36}\) Besides, *Pseudomonas*, which is a frequent pathogenic colonizer in COPD and asthma patients\(^ {37}\) and predispose the host to more severe respiratory viral pneumonia,\(^ {38}\) was increased in the CAP group and correlated with CRP level. Exoproteins from *Pseudomonas* were reported to disrupt the mucosal barrier and induce IL-6 production potentially contributing to mucosal inflammation.\(^ {39}\) Thus, in CAP patients even without ARDS, the overgrowth of certain pathogens under acute infection may lead to the breakdown of the lung barrier and the induction of gut-lung bacterial translocation, which further correlates with more severe inflammation.\(^ {33,37,38,40}\)

In the CTD-ILD group, many putative commensal microbes, such as *Veillonella* and *Streptococcus*, were enriched compared with HC. *Streptococcus intermedius* enrichment was associated with increased ESR. Liu reported that *Streptococcus intermedius* secreted a histone-like DNA binding protein, which induced proinflammatory cytokine production in a macrophage-derived cell line.\(^ {41}\) Therefore, the increased abundance of *Streptococcus intermedius* and its interaction with the...
host immune system could be related to the inflammatory status of CTD-ILD.

Predicted phenotypes based on BugBase suggested that gram-positive bacteria were more abundant in the CAP patients, while the abundance of gram-negative bacteria was decreased in the CAP group. According to guidelines for the diagnosis and treatment of adult community-acquired pneumonia in China, common pathogens of CAP include gram-positive bacteria such as Streptococcus pneumoniae and Staphylococcus aureus, as well as mycoplasma pneumoniae/chlamydia. But for the hospital acquired pneumonia (HAP), common pathogens are gram-negative bacteria. This high level of the analysis may indicate a higher overall level of gram-positive bacteria in the lower respiratory flora in CAP patients, which may be related to the etiology of CAP.

PICRUSt2 analysis also suggested that microbial metabolic function may contribute to the LRT metabolome, and hence respiratory health. The predicted function associated with “D-arginine and D-ornithine metabolism” was significantly enriched in the CTD-ILD group, which plays a vital role in pulmonary fibrosis pathogenesis. Arginine can be converted into ornithine through arginase and proline, as a metabolite from ornithine, participates in fibrosis. Thus, arginase inhibitors might be potential therapeutic agents for CTD-ILD by rectifying immune imbalance and arginine metabolism disorder.

Dual-omics analysis revealed that Streptococcus was negatively correlated with three amino acids but positively correlated with itaconic acid. Streptococcus is one of the most abundant genera inhabiting the respiratory tract of the healthy population. LEfSe analysis showed that Streptococcus was increased in CTD-ILD patients but decreased in CAP patients. Besides, we found that the CAP group was highly enriched for various amino acids, which is consistent with the finding in the inflamed CF airway and had a significantly lower level of itaconic acid. Iaconic acid is derived from activated macrophages and can inhibit inflammation at low doses but promote inflammatory apoptosis at high doses. Thus, in the HC group, the balanced composition of commensal flora may lead to relative amino acid starvation in the airway that could restrict viral invasion and may also inhibit excessive inflammation. Once dysbiosis occurs in the LRT, depletion or abnormal enrichment of commensal flora will cause the host’s original balanced immune state to be disrupted, and metabolic reprogramming of immune cells may also occur, causing invasion of pathogens and then an excessive inflammatory response.

In this context, it is possible that targeting specific metabolites and/or pathways in combination with narrow-spectrum antibiotics for specific microbes, can represent a potential therapeutic strategy to control acute or chronic inflammation in respiratory diseases.

The present study had some limitations. First, in all identified ASVs, the proportion of “unassigned” in the BALF samples reached 35% and varied from 10% to 40% in the previous studies, which may be caused by the imperfect recall (not all sequences or taxa are detected) or imperfect precision (additional false sequences or taxa are detected) of the 16S V3-V4 rRNA sequencing. Meanwhile, the different phenomena of species within genera highlight the interspecies variability that complicates microbiome data interpretation, so it is important to identify organisms at the species level. However, resolution of 16S method at the species level is especially low, so bacterial community analysis using full-length 16S rRNA gene or metagenomics is needed.

Second, the discovered statistical correlations between metabolites and clinical indicators were generally moderate, which could likely be improved with a further larger study. Third, the relationship between bacteria and metabolites was limited and further verification using metatranscriptome analysis and animal models is necessary.

Despite these limitations, our study is the first to our knowledge to compare the LRT environment between acute infections and auto-immune disorders, providing a benchmark for future studies evaluating associations between the LRT microenvironment and the immune homeostasis. Our results provide more nuanced insight into dysbiosis of the LRT microenvironment, revealing shared and specific alterations in CAP and CTD-ILD patients. These findings may provide new insights into the etiologies and pathophysiological mechanisms of different respiratory diseases. However, the causal interactions between the observed metabolic and microbiome changes and the host in disease development remain to be elucidated. Moreover, the alterations in the LRT microenvironment should be interpreted with caution, as many identified alterations may be indicative of general disease states. Thus, microbes or metabolites that are non-specifically associated with multiple diseases would not be useful as diagnostics. On the other hand, bacteria and metabolites that are associated with specific diseases could be developed into diagnosis and therapy targets.
Data Sharing Statement
The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA751994.

Ethics Approval and Informed Consent
Application of the study was submitted and approved by the Ethical Review Committee of Peking University People’s Hospital (No.2016PHB202-01) and the CAP cohort was registered at ClinicalTrials.gov (NCT03093220). The study was conducted under the principles of the Helsinki Declaration and patient informed consent was obtained from patients or their next of kin.

Consent for Publication
All authors agreed to the publication of the article.

Acknowledgments
The authors wish to thank staff members of the cooperating hospitals (Department of Pulmonary and Critical Care Medicine, Shanghai Pulmonary Hospital, Shanghai, China; Department of Pulmonary and Critical Care Medicine, West China Hospital, Sichuan, China; and Department of Pulmonary and Critical Care Medicine, Second Hospital of Jilin University, Jilin, China) for assistance with samples and clinical data collection. They are also grateful to the Metabolomics Facility at the Technology Center for Protein Sciences, Tsinghua University for technical support.

Funding
This work is funded by the National Natural Science Foundation of China (No.82000019, No. 81870010), National Key Research and Development Programme of China (No. 2016YFC0903800).

Disclosure
The authors declare that they have no competing interests.

References


