Alterations to the Lung Microbiome in Idiopathic Pulmonary Fibrosis Patients

Xunliang Tong [1†], Fei Su [2†], Xiaomao Xu [1], Hongtao Xu [3], Ting Yang [4,5], Qixia Xu [6], Huaping Dai [4,5], Kewu Huang [7], Lihui Zou [8], Wenna Zhang [1], Surui Pei [9], Fei Xiao [2,8], Yanming Li [1*] and Chen Wang [4,5,10*]

[1] Department of Respiratory and Critical Care Medicine, National Center of Gerontology, Beijing Hospital, Beijing, China, [2] Clinical Biobank, National Center of Gerontology, Beijing Hospital, Beijing, China, [3] Department of Laboratory Medicine, Beijing Hospital, Beijing, China, [4] National Clinical Research Center for Respiratory Diseases, Center for Respiratory Diseases, China-Japan Friendship Hospital, Peking University Health Science Center, Beijing, China, [5] Department of Pulmonary and Critical Care Medicine, China-Japan Friendship Hospital, Peking University Health Science Center, Beijing, China, [6] Department of Respiratory and Critical Care Medicine, Bengbu University Affiliated Hospital, Bengbu, China, [7] Department of Respiratory and Critical Care Medicine, Beijing Chao-Yang Hospital, Capital Medical University and Beijing Institute of Respiratory Medicine, Beijing, China, [8] The Key Laboratory of Geriatrics, National Center of Gerontology, Beijing Hospital, Beijing, China, [9] Annoroad Gene Technology (Beijing) Co., Ltd., Beijing, China, [10] National Clinical Research Center for Respiratory Diseases, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China

Lung microbiome ecosystem homeostasis in idiopathic pulmonary fibrosis (IPF) remains uncharacterized. The aims of this study were to identify unique microbial signatures of the lung microbiome and analyze microbial gene function in IPF patients. DNA isolated from BALF samples was obtained for high-throughput gene sequencing. Microbial metagenomic data were used for principal component analysis (PCA) and analyzed at different taxonomic levels. Shotgun metagenomic data were annotated using the KEGG database and were analyzed for functional and metabolic pathways. In this study, 17 IPF patients and 38 healthy subjects (smokers and non-smokers) were recruited. For the PCA, the first and the second principal component explained 16.3 and 13.4% of the overall variability, respectively. The β diversity of microbiome was reduced in the IPF group. Signature of IPF's microbes was enriched of *Streptococcus*, *Pseudobutyrivibrio*, *and Anaerorhabdus.* The translocation of lung microbiome was shown that 32.84% of them were from oral. After analysis of gene function, ABC transporter systems, biofilm formation, and two-component regulatory system were enriched in IPF patients' microbiome. Here we shown the microbiology characteristics in IPF patients. The microbiome may participate in altering internal conditions and involving in generating antibiotic resistance in IPF patients.

Keywords: idiopathic pulmonary fibrosis, bronchoalveolar lavage fluid, microbiota, antibiotic resistant gene, virulence factor

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a progressive, fatal disease without a known cause (Molyneaux et al., 2014). IPF is increasing in prevalence and has a median survival of 3 years after diagnosis; the life expectancy for IPF patients is worse than those of some cancers (Maher et al., 2007; Ley et al., 2011; Navaratnam et al., 2011; Molyneaux et al., 2014). Because microorganisms

may disturb the internal environment of the lower airway and cause lung damage, it is crucial to determine the precise composition of the lung microbiota and predict associated gene functions to understand IPF (Han et al., 2014; O'dwyer et al., 2016).

Traditional microbiological culture-based methods do not provide a complete profile of the microbiota in the lower respiratory tract (Garg et al., 2017; Ubags and Marsland, 2017). In contrast, culture-independent approaches, primarily based on gene sequencing, better describe the wide diversity of the microbial inhabitants of the lung microbiota and are able to identify significant differences between healthy subjects and patients with various respiratory diseases (Turnbaugh et al., 2007; Huse et al., 2012; Dickson and Huffnagle, 2015; O'dwyer et al., 2016).

Bronchoalveolar lavage fluid (BALF) samples obtained from the lower respiratory tract are used for routine diagnostic procedures. BALF sampling is suitable for the identification of a distinct microbiome in the lower respiratory tract (Zemanick et al., 2017). The bronchi harbor a specific microbiome associated with healthy individuals as well as a microbiome specific to pulmonary fibrosis patients (Charlson et al., 2011; Zimmermann et al., 2015). In addition, some studies have identified differences in the microbiomes of the upper respiratory tracts of smokers vs. non-smokers (Morris et al., 2013; Faner et al., 2017). Smoking may directly affect the composition of the respiratory microbiome, leading to changes or shifts in microbial community structure (Charlson et al., 2010; Sapkota et al., 2010; Lee et al., 2012). However, few studies have examined differences in the microbiomes of the lower respiratory tract in smokers vs. non-smokers (Garmendia et al., 2012; Morris et al., 2013).

In our current study, we broadly characterized the BALF-associated microbial communities in IPF patients and healthy individuals, including smokers and non-smokers. Additionally, we compared the lung microbiome with other local microbial communities, including those of the upper airway tract (the oropharynx) and gut. We attempted to assess the usefulness of shotgun metagenomics applied directly to DNA extracted from BALF samples to characterize the lung microbiomes in IPF and healthy patients and to reveal the role of the microbiome in IPF pathophysiology.

METHODS

Study Design

A diagnosis of IPF was made after multidisciplinary consultations. Only individuals diagnosed according to the international guidelines of the American Thoracic Society (ATS) and the European Respiratory Society (ERS) were subsequently included in this study (Travis et al., 2013). Healthy control subjects included non-smokers and smokers with normal lung function. Exclusion criteria included: (a) a history of self-reported upper or lower respiratory tract infection in the previous 3 months; (b) antibiotic use in the previous 3 months; (c) acute IPF exacerbation; and (d) other respiratory disorders. Written informed consent was obtained from all subjects, and the study was approved by the Institutional Review Board and

Ethics Committee of Beijing Hospital (Beijing, China). The workflow of this research is shown in **Supplementary Figure 1**.

Bronchoscopy

Fiberoptic bronchoscopies with bronchoalveolar lavage (BAL) were performed according to ATS guidelines via the oropharyngeal route in accordance with a standard operating procedure, during which 20 ml of normal saline was instilled into a designated segment of the lobe showed in **Figure 1A** (Busse et al., 2005). After specimen collection, an aliquot of unfiltered and unprocessed BAL was immediately placed on ice and then frozen at $-80°C$ for further analysis.

DNA Extraction

DNA was extracted from 3.0 ml of each original BALF sample using a QIAamp DNA Microbiome Kit (catalog 51704, Qiagen, Hilden, Germany) in strict accordance with the manufacturer's instructions. The DNA concentration was measured using a Qubit® 2.0 Fluorimeter (Life Technologies, Invitrogen, USA).

DNA Library Construction and Sequencing

DNA libraries were constructed using an Illumina TruSeq DNA kit according to the manufacturer's instructions. The same Illumina workflows were used to perform cluster generation, template hybridization, isothermal amplification, linearization, blocking, denaturation and hybridization of the sequencing primers. We performed paired-end sequencing on 2×100 base pairs (bp) or single-end sequencing on 100 bp. The base-calling pipeline was used to process raw fluorescent images and to call sequences. Reads that mapped to the human genome (hg19) were removed from each sample with BWA MEM using the default parameter (Li and Durbin, 2009). For quality control, we employed Trimmomatic (version 0.36) with the following criteria: (a) scan the read with a 4-base-wide sliding window, cutting when the average quality per base drops below 15; (b) drop reads below 50 bases in length; and (c) remove leading/tailing low quality ($Q < 13$). This produced an average of 8.5 gigabases (Gb) of high-quality sequences for each sample, resulting in a total of 470 Gb of sequence data (**Supplementary Table 1**). The reads were assembled into contigs for all samples using the assembly software IDBA (version 1.1.1) (Peng et al., 2012). IDBA was applied for Illumina short read assembly with the parameters "–pre_correction –mink 30 –maxk 120 –step 10." Then, we removed ambiguous bases from the assembled scaffolds and discarded scaffolds with lengths less than 500 bp. Finally, 73% of the total reads were used to generate 88,761 contigs without ambiguous bases (minimum length of 500 bp). These contigs had a total length of 210,670,314 bp, an average N50 length of 4,151 bp and ranged from 503 to 620,529 bp.

Construction of a Non-redundant Microbiome Gene Set

To predict microbial genes from the assembly, we applied the methodology used in the MetaHIT human gene catalog study (Qin et al., 2010). MetaGeneMark (MetaGeneMark_v1.mod) was used to predict the open reading frames (ORFs) in contigs

FIGURE 1 | Principal component analysis with BALF samples from IPF patients and health subjects. **(A)** Diagram of the bronchoscopy performance. **(B)** Principal component analysis with BALF samples from IPF patients and health subjects (including smokers and non-smokers) was performed based on the taxonomic profiles at species level. Different colors of points were represented for different groups: the blue points represent BALF samples from IPF group, the green points represent BALF samples from smoking control subjects, and the red points represent BALF samples from non-smoking control subjects. BALF samples from IPF patients were with an apparent clustering pattern of microbial composition compared with smoking control subjects, but no significant distribution pattern with the non-smoking control subjects.

without ambiguous bases (Zhu et al., 2010). The program predicted 250,268 ORFs using a 100-bp cut-off. The total length of the predicted ORFs was 183,986,409 bp, representing 87.3% of the total contig length. Among these ORFs, 142,871 (57.1%) were complete genes, while 107,397 (42.9%) were incomplete without a start or stop codon. A non-redundant "gene set" was established by removing redundant ORFs, defined as those sharing 95% identity with VSEARCH (Rognes et al., 2016). Then, genes from the HMP and HOMD databases (Dewhirst et al., 2010) were merged as a non-redundant reference gene set for gene-source analysis.

Organism Abundance Profiling

We followed the approach described in previous study to perform organism abundance profiling (Qin et al., 2014). BWA was applied to align paired-end clean reads against reference genomes using default parameters. Reads with alignments on the same reference genomes were assigned into the following two groups: reads having alignments with only one genome were identified as unique reads (U), and reads having alignments with more than one genome were identified as multiple reads (M).

For species S, if its abundance was Ab (S), and if it potentially aligned with U unique reads and M multiple reads, the following formulae were used:

$$\text{Ab}(S) = Ab(U) + Ab(M)$$
$$\text{Ab}(U) = U/l$$
$$\text{Ab}(M) = (\sum_{i=l}^{M} Co * \{M\})/l$$

Shannon diversity index (H) at the species level was calculated as the following, with p_i denoting the proportion in group k.

$$H = -\sum_{i=1}^{k} p_i \log(p_i)$$

Gene Abundance Profiling

Gene abundance was determined using a method similar to RPKM (reads per kilobases per million reads) used to quantify gene expression based on RNA sequencing data. In brief, high-quality original Illumina reads from each sample were aligned with a reference gene set using BWA. For each gene, G_i, the number of read pairs that aligned to it divided by the length of the gene was calculated as Num_G_i, and the relative abundance, $RNum\_G_i$, of each gene in each sample (n genes) was computed using the following formula:

$$\text{RNum}_{Gi} = Num\_Gi / \sum_{i=1}^{n} Num\_Gi$$

Gene Function Classification and Ortholog Group Abundance Profiling

Protein sequences of the predicted genes were searched using BLASTp (Altschul et al., 1990) against the EggNOG 4.5 database (Huerta-Cepas et al., 2016) (**Supplementary Table 3**) and the KEGG gene database(Kanehisa et al., 2017) with the parameters "-num_descriptions 10,000, -e-value 1e-5." Genes that had alignments with a score slightly higher than 60 were assigned into one or more eggnog or KEGG ortholog groups. We used

the methods introduced in previous study (Li et al., 2014) to calculate the abundances of KEGG ortholog groups; additionally, we added the abundances of proteins assigned into the same KEGG ortholog groups as abundances of KEGG ortholog groups, and KEGG ortholog group profiles were generated.

Quantification of Antibiotic Resistance Genes and Virulence Factors

Antibiotic resistance genes (ARGs) were annotated using the ARDB (Liu and Pop, 2009) database with BLASTp (*e*-value $< 1 \times 10^{-5}$). Then, antibiotic resistance gene abundance was determined as mentioned above. We used the Wilcoxon rank sum test to determine significantly different abundances between conditions. Virulence factors (VFs) were analyzed using a similar strategy. In brief, all VFs were identified using the VFDB database (Chen et al., 2016) with BLASTp (*e*-value $< 1 \times 10^{-5}$). Gene abundances and related statistical analysis were followed the ARGs' method.

Multivariate Analysis

To analyze the beta diversity of various groups based on the Bray-Curtis dissimilarity, ANOSIM was performed to detect significant dissimilarity or similarity in community composition among different groups (IPF/smoking/non-smoking). Pairwise comparisons were performed in ANOSIM, and pairwise *P*-values were obtained to assess significant differences in community composition between any two groups.

Statistical Analysis

PCA was analyzed using QIIME (Version 1.9.1) (Caporaso et al., 2010). Differential gene abundance and KEGG modules were tested with the Wilcoxon rank-sum test. One-way ANOVA test was performed and *P*-values were corrected for multiple testing with the Benjamin & Hochberg method using R software Version 3.1.1 (http://www.r-project.org/).

Availability of Data

The data set supporting the results of this article has been deposited in the NCBI Short Read Archive database under BioProject accession code PRJNA387212.

RESULTS

Patients Characteristics

In total, 17 IPF patients (IPF group) and 38 healthy control subjects (control group, including 23 non-smokers and 15 smokers) were included according to the inclusion criteria. The demographic data for these subjects are summarized in **Table 1**. IPF groups were slightly older than healthy subjects (IPF group vs. control group: [62.71 ± 7.90] years vs. [61.28 ± 9.37] years). As expected, IPF patients exhibited reduced forced expiratory volume in 1 one second (FEV1) and forced vital capacity (FVC) compared with those of healthy subjects. Total lung volume (TLC), residual lung volume (VR), lung diffusing capacity (DLCO) and the transfer coefficient (KCO) were sharply reduced in the IPF group, shown in **Table 1**.

TABLE 1 | Patient characteristics.

	Healthy control (*n* = 38)	IPF patients (*n* = 17)
Age, years	62.71 ± 7.90	61.28 ± 9.37
Gender (M/F)	24/14	11/6
Smoker (Never/current), n	23/15	11/6
Pack-year	14 ± 6	48 ± 32
FEV1% pred	110 ± 10	68 ± 17
FVC% pred	115 ± 14	66 ± 16
FEV1/FVC%	82 ± 5	77 ± 12
TLC, %pred	nd	65 ± 13
RV, %pred	nd	65 ± 30
DLCO% pred	nd	34 ± 12
KCO% pred	nd	54 ± 14

Principal Component Analysis for Microbiome

To identify any differences in the organismal structure of the lung microbiota, principal component anlaysis was performed based on taxonomic profiles at the species level (**Figure 1B**). Different colored points represent different groups: the blue points represent BALF samples from the IPF group, the green points represent BALF samples from smoking control subjects, and the red points represent BALF samples from non-smoking control subjects. The first principal component explained 16.32% of the overall variability among different groups, whereas the second principal component explained 13.42% of variability. As shown in **Figure 1B**, an apparent microbial composition clustering pattern was identified for IPF patients and smoking control subjects. Conversely, non-smoking control subjects, which are marked in red, exhibited no significant distribution pattern.

Microbial Community Structure of BALF

Sequencing reads (average number: 7,242,099) were aligned against 3,096 reference genomes from the National Center for Biotechnology Information and the HMP reference sequence, which contains 131 archaeal strains comprising 97 species, 326 lower eukaryotes comprising 326 species, 3,683 viral strains comprising 1,420 species, and 1,751 bacterial strains comprising 1,253 species (**Supplementary Table 1**). Relative abundances at the phylum, class, order, family, genus, and species levels were compared between the IPF group and control group. Intrinsic differences in lung microbiota composition are shown in **Figure 2** and **Supplementary Table 2**. Species with a median relative abundance larger than 0.01% of the total abundance in either the control group or the IPF group were included for comparison. Furthermore, we imported taxonomic data into QIIME for comparison. At the phylum level, *Firmicutes* (ANOVA, *P*-value = 0.02) and *Fusobacterium* (ANOVA, *P*-value = 0.04) dominated the BALF microbial communities in both groups (**Figure 2**). Compared to the control group, the IPF group had lower levels of *Bacteroidetes* (ANOVA, *P*-value = 0.03) but higher levels of *Proteobacteria* (ANOVA, *P*-value = 0.004) and *Fusobacteria* (ANOVA, *P*-value = 0.022). At the genus

FIGURE 2 | Microbiota composition of BALF samples in IPF and healthy subjects. The microbiota in IPF patients and healthy subjects (including smokers and non-smokers) were shown at the phylum **(A)**, class **(B)**, order **(C)**, family **(D)**, and genus **(E)** levels. Species with a median relative abundance larger than 0.01% of the total abundance in either the control group or the IPF group were included for comparison. P stands for IPF group, N stands for non-smoking normal subjects, S stands for smoking normal subjects. All results are presented as the median, the 25–75 % percentiles and the variation range; results for IPF group (P) are presented as yellow boxes, smoking control subjects (S) are marked with pink boxes and non-smoking control subjects (N) are marked with violet boxes. Blue dots represent the abnormal observations at the corresponding taxonomic levels.

level (**Figure 2E**), *Bacteroides* was the dominant phylotype in both groups but was significantly decreased in the IPF group. Similar to previous studies (Morris et al., 2014), *Streptococcus*, representing 23.0 % of total reads, was the most common genus in subjects with IPF, followed by *Pseudobutyrivibrio*, *Anaerorhabdus*, *Campylobacter*, and *Blautia*, all of which were enriched in the IPF group. In contrast, *Sutterella*, *Coprococcus*, *Parasutterella*, *Paludibacter* and *Dorea* were dominant in the healthy group. The microbial communities of subjects with IPF were by contrast less diverse (Shannon diversity index, 2.81 ± 0.08 vs. 4.01 ± 0.10; P = 0.004) than the healthy subjects.

Biodiversity Analysis of BALF Samples

To measure the diversity of the BALF communities in IPF patients and healthy subjects (including smokers and non-smokers), we employed various methods, such as Adonis and ANOSIM of QIIME, to evaluate differences (beta-diversity) in the lung microbiota among the groups. BALF biodiversity was significantly reduced (P-value = 0.01) in the IPF group compared

with that in the smoking/non-smoking groups according to the ADONIS results. A similar grouping pattern was identified based on the Bray-Curtis matrix using hclust2 (**Figure 3**). We identified the following two groups that were differentiated based on combinations of sex, age, and condition: Group 1, opportunistic pathogenic bacteria primarily derived from the skin and mouth, according to study of Blauwkamp (Blauwkamp et al., 2019); and Group 2, bacteria derived from the gut. These two groups were similar to those identified by ANOSIM. Mantel tests of the vegan package revealed the significant correlation of community structure with a P-value = 0.001 (Mantel Statistic R: 0.3057).

Crosstalk of BALF Microbiom to the Oral and Gut Microbiomes

To explore the origins of the lung microbiome, we constructed a reference gene catalog based on the HOMD and HMP (https://portal.hmpdacc.org) databases. We compared annotated genes with the reference database to determine functional gene sources. According to our results, the majority of genes were shared,

FIGURE 3 | Heatmap of the normalized abundances of all BALF samples. A similar grouping pattern on the combinations of the gender, age, and condition was also identified by Bray-Curtis matrix using the hclust2 and Mantel tests. The significant correlated of community structure of two groups was revealed (P-value = 0.01). The species in the red box are group 1, opportunistic pathogenic bacteria primarily derived from the skin and mouth, as mentioned above. The rest species are group 2, which is mainly from gut. Heatmap is color-coded based on the normalized abundance of species, from black (lower abundance) to red (higher abundance).

32.84% genes were derived from oral, and 1.32% genes were derived from the gut, as determined by V search (see the Methods section for details) and the results were shown in **Figure 4A**. We also analyzed the crosstalk of the different enriched genes between smoking group and IPF group: 38% of them were from oral and rest of them were unique in BALF (**Figure 4B**). None of the different enriched genes were from gut.

Functional Analysis of the Lung Microbiome

To identify functional differences in the BALF microbiome between control subjects (including smokers and non-smokers) and IPF patients, shotgun metagenomics data were annotated using the KEGG and eggNOG database and analyzed for functional and metabolic pathways (**Supplementary Tables 3, 4**). A total of 2,707 KOs were identified either in healthy controls or IPF samples. We further mapped these KOs to

KEGG modules and pathways and calculated the adjusted P-value for the hypergeometric distribution. In particular, ABC transporter systems, the two-component regulatory system, biofilm formation, methane metabolism, aromatic compound degradation, amino acid biosynthesis (including alanine, tyrosine, valine and leucine), vitamin biosynthesis, biotin metabolism, and amino sugar and nucleotide sugar metabolism were enriched in the IPF group (**Supplementary Figure 2**). Among these pathways, the ABC transporter system, which imports amino acids, represented the majority of the enriched genes in the IPF group (**Supplementary Figure 3**). Multidrug efflux genes, which contribute to drug resistance, were enriched in the IPF group. According to ARDB databases, antibiotic resistant genes (ARG) was checked and 48 different types of ARGs was identified, which showed in **Supplementary Table 5**. The relative abundance of identified ARGs in IPF group and healthy control groups (smoking and non-smoking) was shown

FIGURE 4 | Crosstalk of BALF's microbiome to oral and gut microbiome. **(A)** The origin lung microbiome was explored by comparison with HOMD and HMP databases. The genes rescore was shown: 32.84% genes were from oral and 1.32% genes were from the gut, as determined by the Vsearch. **(B)** Crosstalk of the different enriched genes were analyzed in this research: 38% of them were from oral and rest of them were unique in BALF.

in **Figure 5**. Compared with microbial genes from healthy control subjects, genes from IPF group were enriched of ARGS (5/10), including *tet*40, *acr*R, *evg*A, *tol*C, *mdt*O, with significantly abundance changes. These ARGs identified in IPF group are related to multidrug efflux, suggesting that the multidrug efflux pumps maybe the primary mechanism for bacteria to extrude antibiotics and other materials (**Figure 5A**). At the same time, virulence genes were also detected in this research showed in **Supplementary Table 6** and the relative abundance of each group was showed in **Figure 6**.

DISCUSSION

The Human Microbiome Project was launched in 2007 to study the microbial inhabitants of the human body and the microbes that live on human mucosal surfaces with the understanding that human biology may partially depend on interactions with these microorganisms (Turnbaugh et al., 2007). Because the human lung was believed to be sterile, it was not included in the original Human Microbiome Project. However, a growing number of studies have revealed the presence of microbial communities in the lungs. Next generation sequencing technology has dramatically improved lung microbiome identification in healthy subjects and in patients with chronic airway diseases, but humanized genomic DNA contamination remains a major technical challenge (Corless et al., 2000; Zoetendal et al., 2006; Salonen et al., 2010; Wu et al., 2010). As shown in our previous study, the use of pure microbial DNA avoids humanized genomic pollution and facilitates investigation of the lung microbiota (Wen et al., 2016).

In our current study, bacteria and virus microbial genes discriminated between IPF patient BALF and the BALF of healthy individuals with a high specificity; for example, *Acinetobacter* and *Neisseria* were identified at the genus level. The most common pathogenic bacteria in the IPF group were *Acinetobacter* and *Neisseria* (at the genus level), and the most common pathogenic viruses were microvirus, simplex virus and varicella virus. The

same results had been reported with an increased bacterial load and decreased microbial diversity in BAL samples from IPF patients (Morris et al., 2014). At the same time, there are clear differences in the IPF BAL microbiome (increase in *Neisseria, Streptococcus* and some other species) compared to healthy subjects, which was also proven in our study. The bacterial load affects survival in these patients whereas retrospective analysis of the lung microbiome in BAL samples from the COMET-cohort suggested that presence of specific *Streptococcus* operational taxonomic units (OTU) or *Staphylococcus* OTU was associated with worse outcomes of IPF. These studies were showed the differential microbiome and specific microbial genes changes in BALF isolated from IPF patients, which suggested microbiota-targeted biomarkers may be potential tools for disease diagnosis and predicting prognosis. Biodiversity analysis revealed greater microbiome diversity associated with better lung function in the control group; conversely, the IPF group demonstrated decreased microbial diversity in terms of bacterial gene richness. Analysis of BALF microbiome revealed the appearance of common oral and gut inhabitants in BALF. Notably, shifts in the bacterial make-up of the BALF were associated with the human gut and oral microbiotas. In the IPF group, the majority of inhabitants were derived from the oral microbiome, suggesting they shifted from the mouth and throat. Comparison of BALF microbial genes to oral and gut microbiomes demonstrated that majority genes were from oral source, which hint the potential mechanism of oral aspiration as a risk factor in pathogenesis of IPF.

This study provides an opportunity to identify the bacterial functions required for a bacterium to thrive in the context of the BALF from IPF group and healthy subjects. To identify the functions of the BALF genome, gene length and copy number were normalized, and the relative frequencies of different functions were deduced based on the number of genes recruited to different EggNOG clusters. ABC transporter systems, the two-component regulatory system, biofilm formation, amino acid and vitamin biosynthesis, and central carbon metabolism were enriched in the IPF group. Not surprisingly,
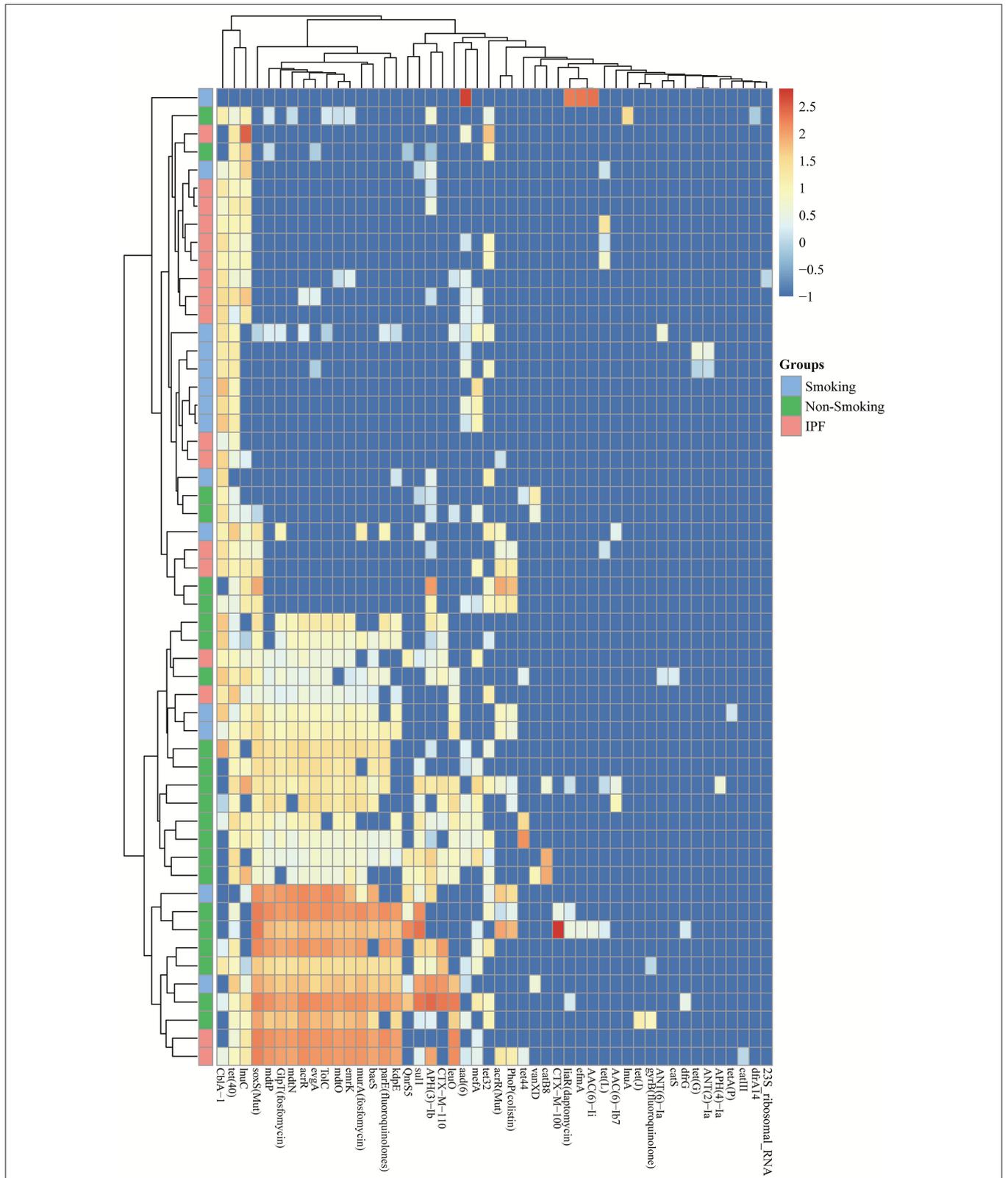
FIGURE 5 | Abundances of antibiotic resistant genes from BALF samples. Each column corresponds to an individual BALF's sample from IPF group (pink color) and healthy control group (smoking: blue color; non-smoking: green color) showed on the left side. The types of antibiotic resistant genes were indicated in the boxes at the bottom. Each row corresponds to a specific abundance of antibiotic resistant genes based on different colors according to different abundance folds.

FIGURE 6 | Abundances of virulence genes from BALF samples. Each column corresponds to an individual BALF's sample from IPF group (pink color) and healthy control group (smoking: blue color; non-smoking: green color) showed on the left side. The types of virulence factors were indicated in the boxes at the bottom. Each row corresponds to a specific abundance of virulence genes based on different colors according to different abundance folds.

projection of the range clusters in KEGG metabolic pathways provided highly explicit directions associated with antimicrobial resistance. Antimicrobial resistance is a natural response of bacteria to antibiotic exposure (Sherrard et al., 2014). Antimicrobial resistance may be intrinsic to a bacterium, arise from spontaneous genetic mutations, or be associated with horizontal gene transfer (Sherrard et al., 2014). The enrichment of ABC transporters and biofilm formation signals indicates the presence of major antimicrobial resistance pathways in the IPF lung.

ABC transporter systems play a large variety of biological roles in processes such as translation, elongation, and DNA repair. In some antibiotic-producing or drug-resistant bacteria, particularly Gram-negative bacteria, ABC systems are responsible for the active efflux of drugs and other harmful compounds across the cell envelope (Seeger and Van Veen, 2009). Relative abundance of multidrug efflux genes were increased in IPF group compared with healthy controls which listed in this research. Furthermore, the virulence factors also evaluated in this research and showed significant differences between IPF group and control group. Biofilm formation is a feature of chronic airway infection. Bacteria growing in biofilms are embedded in a matrix of exopolymeric substances and are much more resistant to antibiotics than organisms growing

planktonically (Stewart and Costerton, 2001; Wu et al., 2015). Together, these findings emphasize antimicrobial resistance as a primary function of the BALF ecosystem.

Our descriptions of the major characteristics of the lung microbiome provide a holistic view to understand IPF patients. In our study, intrinsic changes in the microbiome emphasize the production of strong antimicrobial resistance by the BALF ecosystem, which is important to understand pathophysiological processes. The interesting modules and pathways found in this project suggest the BALF microbiome in IPF patients is related to antibiotic resistance. Therefore, more therapeutic options for IPF may become available through the adaptation of lung environments or the identification of beneficial probiotic microorganisms. Further studies in this area will lead to a deeper understanding of bacterial life in the BALF.

ETHICS STATEMENT

Institutional Review Board and Ethics Committee of Beijing Hospital.

AUTHOR CONTRIBUTIONS

CW, YL, FX, and XT designed the study, interpreted the results, and drafted the manuscript. FS and SP analyzed the data. HD, TY, XX, QX, and KH made substantial contributions to the experimental protocol. HX, LZ, and WZ participated in sample preparation and DNA extraction. CW, YL, XT, and FS critically revised the manuscript for intellectual content. All authors participated in discussing and interpreting the results and gave final approval for manuscript submission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2019.00149/full#supplementary-material

Supplementary Figure 1 | The workflow of this research was shown in this figure.

Supplementary Figure 2 | Functional analysis of lung microbiome. The shotgun metagenomics data were annotated using the KEGG databases and analyzed at the functional and metabolic pathway levels. The KOs to KEGG modules and pathways were mapped to identify the BALF microbial functional differences between control subjects. The blue and red lines denote the enriched KEGG pathways in the control and IPF subjects, representatively.

Supplementary Figure 3 | Enrichment of ABC transporter systems based on the knowledge of KEGG database. The red boxes stand for the enriched genes in BALF samples. The blue boxes stand for the rest genes from KEGG database.

Supplementary Table 1 | Sequencing statistic of all BALF samples.

Supplementary Table 2 | Taxonomy composition at different levels.

Supplementary Table 3 | KEGG functional enrichment analysis.

Supplementary Table 4 | EggNOG functional enrichment analysis.

Supplementary Table 5 | Antibiotic resistant genes abundance comparison between different conditions.

Supplementary Table 6 | Virulence genes abundance comparison between different conditions.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Blauwkamp, T. A., Thair, S., Rosen, M. J., Blair, L., Lindner, M. S., Vilfan, I. D., et al. (2019). Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat. Microbiol.* 4, 663–674. doi: 10.1038/s41564-018-0349-6

Busse, W. W., Wanner, A., Adams, K., Reynolds, H. Y., Castro, M., Chowdhury, B., et al. (2005). Investigative bronchoprovocation and bronchoscopy in airway diseases. *Am. J. Respir. Crit. Care Med.* 172, 807–816. doi: 10.1164/rccm.200407-966WS

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Charlson, E. S., Bittinger, K., Haas, A. R., Fitzgerald, A. S., Frank, I., Yadav, A., et al. (2011). Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am. J. Respir. Crit. Care Med.* 184, 957–963. doi: 10.1164/rccm.201104-0655OC

Charlson, E. S., Chen, J., Custers-Allen, R., Bittinger, K., Li, H., Sinha, R., et al. (2010). Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS ONE* 5:e15216. doi: 10.1371/journal.pone.0015216

Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239

Corless, C. E., Guiver, M., Borrow, R., Edwards-Jones, V., Kaczmarski, E. B., and Fox, A. J. (2000). Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. *J. Clin. Microbiol.* 38, 1747–1752.

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W. H., et al. (2010). The human oral microbiome. *J. Bacteriol.* 192, 5002–5017. doi: 10.1128/JB.00542-10

Dickson, R. P., and Huffnagle, G. B. (2015). The lung microbiome: new principles for respiratory bacteriology in health and disease. *PLoS Pathog.* 11:e1004923. doi: 10.1371/journal.ppat.1004923

Faner, R., Sibila, O., Agusti, A., Bernasconi, E., Chalmers, J. D., Huffnagle, G. B., et al. (2017). The microbiome in respiratory medicine: current challenges and future perspectives. *Eur. Respir. J.* 49:1602086. doi: 10.1183/13993003.02086-2016

Garg, N., Wang, M., Hyde, E., Da Silva, R. R., Melnik, A. V., Protsyuk, I., et al. (2017). Three-dimensional microbiome and metabolome cartography of a diseased human lung. *Cell Host Microbe.* 22, 705–716.e704. doi: 10.1016/j.chom.2017.10.001

Garmendia, J., Morey, P., and Bengoechea, J. A. (2012). Impact of cigarette smoke exposure on host-bacterial pathogen interactions. *Eur. Respir. J.* 39, 467–477. doi: 10.1183/09031936.00061911

Han, M. K., Zhou, Y., Murray, S., Tayob, N., Noth, I., Lama, V. N., et al. (2014). Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir. Med.* 2, 548–556. doi: 10.1016/S2213-2600(14)70069-4

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved

functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–293. doi: 10.1093/nar/gkv1248

Huse, S. M., Ye, Y., Zhou, Y., and Fodor, A. A. (2012). A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS ONE* 7:e34242. doi: 10.1371/journal.pone.0034242

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

Lee, J., Taneja, V., and Vassallo, R. (2012). Cigarette smoking and inflammation: cellular and molecular mechanisms. *J. Dent. Res.* 91, 142–149. doi: 10.1177/0022034511421200

Ley, B., Collard, H. R., and King, T. E. Jr. (2011). Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 183, 431–440. doi: 10.1164/rccm.201006-0894CI

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., et al. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* 32, 834–841. doi: 10.1038/nbt.2942

Liu, B., and Pop, M. (2009). ARDB–antibiotic resistance genes database. *Nucleic Acids Res.* 37, D443–D447. doi: 10.1093/nar/gkn656

Maher, T. M., Wells, A. U., and Laurent, G. J. (2007). Idiopathic pulmonary fibrosis: multiple causes and multiple mechanisms? *Eur. Respir. J.* 30, 835–839. doi: 10.1183/09031936.00069307

Molyneaux, P. L., Cox, M. J., Willis-Owen, S. A., Mallia, P., Russell, K. E., Russell, A. M., et al. (2014). The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 190, 906–913. doi: 10.1164/rccm.201403-0541OC

Morris, A., Beck, J. M., Schloss, P. D., Campbell, T. B., Crothers, K., Curtis, J. L., et al. (2013). Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am. J. Respir. Crit. Care Med.* 187, 1067–1075. doi: 10.1164/rccm.201210-1913OC

Morris, A., Gibson, K., and Collman, R. G. (2014). The lung microbiome in idiopathic pulmonary fibrosis. what does it mean and what should we do about it? *Am. J. Respir. Crit. Care Med.* 190, 850–852. doi: 10.1164/rccm.201409-1626ED

Navaratnam, V., Fleming, K. M., West, J., Smith, C. J., Jenkins, R. G., Fogarty, A., et al. (2011). The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax* 66, 462–467. doi: 10.1136/thx.2010.148031

O'dwyer, D. N., Dickson, R. P., and Moore, B. B. (2016). The lung microbiome, immunity, and the pathogenesis of chronic lung disease. *J. Immunol.* 196, 4839–4847. doi: 10.4049/jimmunol.1600279

Peng, Y., Leung, H. C., Yiu, S. M., and Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821

Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., et al. (2014). Alterations of the human gut microbiome in liver cirrhosis. *Nature* 513, 59–64. doi: 10.1038/nature13568

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahe, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584

Salonen, A., Nikkila, J., Jalanka-Tuovinen, J., Immonen, O., Rajilic-Stojanovic, M., Kekkonen, R. A., et al. (2010). Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J. Microbiol. Methods* 81, 127–134. doi: 10.1016/j.mimet.2010.02.007

Sapkota, A. R., Berger, S., and Vogel, T. M. (2010). Human pathogens abundant in the bacterial metagenome of cigarettes. *Environ. Health Perspect.* 118, 351–356. doi: 10.1289/ehp.0901201

Seeger, M. A., and Van Veen, H. W. (2009). Molecular basis of multidrug transport by ABC transporters. *Biochim. Biophys. Acta* 1794, 725–737. doi: 10.1016/j.bbapap.2008.12.004

Sherrard, L. J., Tunney, M. M., and Elborn, J. S. (2014). Antimicrobial resistance in the respiratory microbiota of people with cystic fibrosis. *Lancet* 384, 703–713. doi: 10.1016/S0140-6736(14)61137-5

Stewart, P. S., and Costerton, J. W. (2001). Antibiotic resistance of bacteria in biofilms. *Lancet* 358, 135–138. doi: 10.1016/S0140-6736(01)05321-1

Travis, W. D., Costabel, U., Hansell, D. M., King, T. E. Jr., Lynch, D. A., Nicholson, A. G., et al. (2013). An official american thoracic society/european respiratory society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am. J. Respir. Crit. Care Med.* 188, 733–748. doi: 10.1164/rccm.201308-1483ST

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature* 449, 804–810. doi: 10.1038/nature06244

Ubags, N. D. J., and Marsland, B. J. (2017). Mechanistic insight into the function of the microbiome in lung diseases. *Eur. Respir. J.* 50:1602467. doi: 10.1183/13993003.02467-2016

Wen, Y., Xiao, F., Wang, C., and Wang, Z. (2016). The impact of different methods of DNA extraction on microbial community measures of BALF samples based on metagenomic data. *Am. J. Transl. Res.* 8, 1412–1425.

Wu, G. D., Lewis, J. D., Hoffmann, C., Chen, Y. Y., Knight, R., Bittinger, K., et al. (2010). Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol.* 10:206. doi: 10.1186/1471-2180-10-206

Wu, H., Moser, C., Wang, H. Z., Hoiby, N., and Song, Z. J. (2015). Strategies for combating bacterial biofilm infections. *Int. J. Oral Sci.* 7, 1–7. doi: 10.1038/ijos.2014.65

Zemanick, E. T., Wagner, B. D., Robertson, C. E., Ahrens, R. C., Chmiel, J. F., Clancy, J. P., et al. (2017). Airway microbiota across age and disease spectrum in cystic fibrosis. *Eur. Respir. J.* 50:1700832. doi: 10.1183/13993003.00832-2017

Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38:e132. doi: 10.1093/nar/gkq275

Zimmermann, A., Zissel, G., Muller-Quernheim, J., Hofmann, S., Schreiber, S., and Fischer, A. (2015). Are bronchoalveolar lavages a good source for microbial profiling? differences between throat and bronchoalveolar lavage microbiomes. *J. Med. Microbiol.* 64, 948–951. doi: 10.1099/jmm.0.000112

Zoetendal, E. G., Heilig, H. G., Klaassens, E. S., Booijink, C. C., Kleerebezem, M., Smidt, H., et al. (2006). Isolation of DNA from bacterial samples of the human gastrointestinal tract. *Nat. Protoc.* 1, 870–873. doi: 10.1038/nprot.2006.142