The Microbiome in Lung Explants of Idiopathic Pulmonary Fibrosis (MiLEs-IPF): a case-control study in patients with end-stage fibrosis.

Supplementary Material

Contents:

EXTENSIVE METHODS

Case definitions:

We diagnosed Idiopathic Pulmonary Fibrosis (IPF) according to standard American Thoracic Society-European Respiratory Society clinical criteria.[1] We included patients with “end-stage” IPF, as our patients either underwent lung transplantation or died from refractory hypoxemia due to IPF from 2012 to 2016. After reviewing 46 lung explant cases with pulmonary fibrosis, 40 patients were included with consensus diagnosis of IPF according to established criteria. Excluded patients had evidence of other forms of pulmonary fibrosis (hypersensitivity pneumonitis, mixed connective tissue disease and scleroderma). We recorded pathology results from the lung explants (or prior available surgical lung biopsies performed for the initial diagnosis of IPF) as reported by specialized thoracic pathologists at the University of Pittsburgh Medical Center (UPMC) and included only patients with a confirmed usual interstitial pneumonia (UIP) pattern of fibrosis. We classified patients with IPF further as those with acute exacerbations of IPF (AEIPF) versus chronic IPF according to recently proposed criteria.[2] Among patients with AEIPF, we examined for receipt of rescue therapies (investigational auto-antibody targeted therapies[3] or extracorporeal membrane oxygenation therapy[4]) and the presence of diffuse alveolar damage in pathology specimens of the explanted lungs. We recorded clinical variables for demographics, smoking history, spirometry, diffusion capacity, oxygen requirement, functional status, immunosuppressive medications, presence of gastroesophageal reflux disease, recent antibiotic use, and available microbiological cultures within one week of lung explantation (Supplementary Table S1). Clinical cultures were negative in all but three (7.5%) cases.

We obtained informed consent for conducting research utilizing the lung explants by the patients or their designated representatives. The University of Pittsburgh Institutional Review Board and Committee for Oversight of Research and Clinical Training Involving

Decedents approved the study.

Control definitions:

Donor lungs: We included specimens from explanted lungs procured from organ donation candidates that were deemed unsuitable for lung transplantation. These organ grafts were rejected for reasons that included donor age, smoking history, hypoxemia, and mechanical complications. We excluded cases with positive clinical cultures with respiratory pathogens or with macroscopic appearance of infection. We were able to identify 32 control lungs, and we used a total of 37 tissue samples (we used 2 basilar samples from each of 5 explants, in order to obtain close to 1:1 matching of controls for the 40 IPF basilar samples). Consent for utilizing these lung explants for research purposes was provided by the patients' representatives to the Pennsylvania Center for Organ Recovery and Education (CORE). All patients met brain death criteria, and these samples were anonymized by CORE investigators.

Diseased Control lungs: We included explants from 5 cystic fibrosis (CF) patients who underwent lung transplantation at our institution. Three patients were known to be colonized by *Burkholderia* and two patients by *Pseudomonas*, and these pathogens were also identified clinically by microbiologic cultures of airway samples obtained at timing of transplantation.

Study Procedures:

Please also refer to Figure 1 in the main manuscript.

Procedures in cases: We performed tissue sampling of the explanted IPF lung(s) in two different settings: first, in the operating room for patients who underwent lung transplantation, and second, in the morgue for patients who died from respiratory failure and had consented to participate in a Rapid Tissue Donation Program at UPMC.[5] For

patients who underwent lung transplantation, a researcher from our team received the lung explant in the operating room immediately after transfer of the lung away from the surgical field. Without any break in sterility, we resected 3-4 pieces of basilar lung tissue from a lower lobe (either left or right, whichever was removed first during surgery). Tissue samples were flash frozen in liquid nitrogen and stored at -80^{o}C. The remaining lung explant(s) were then available to the treating physicians for standard pathology studies. For patients included in the Rapid Tissue Donation Program, a trained pathologist procured the lungs in the morgue and provided the explants to our research team immediately after. We performed tissue sampling, processing and storage in a similar fashion as described above. In three IPF cases, we also performed bronchial wash sampling prior to tissue sample collection, with 30cc of phosphate-buffered saline instilled into a bronchial segment using a sterile tube. For another three IPF cases, we also resected additional apical samples (apart from the basilar subpleural samples obtained from all patients) to make comparisons between apical and basilar samples.

Procedures in controls: For CORE lungs, we obtained tissue samples in the operating room under sterile conditions as above. We made every effort to sample macroscopically normal appearing lung tissue from the basilar portions of the lower lobes, in order to match the corresponding regions sampled from the IPF lungs. For CF lungs, we obtained tissue samples in the operating room at time of transplantation similarly to the IPF cases. Samples were flash frozen and stored as above.

DNA extraction, 16S rRNA gene sequencing and quantitative PCR:

We extracted bacterial DNA from ~85 milligrams (mg) of whole lung explant tissue using the UltraClean Tissue & Cell DNA Kit (Mo Bio) following the manufacturer's instructions with slight modification.[6] Briefly, we added frozen lung tissue and 700ul TD1 buffer to a Dry Bead tube. We then added Proteinase K (15 microliters (µL)) and

Lysozyme (2 µL) and placed the tube at 65 °C for 20 minutes. We added the tubes to the Mo Bio vortex adapter and lysed for 10 minutes at top speed. Post-lysis, we spun the tubes at > 13,000 x g for 1 minute, and transferred the supernatant to a spin filter. We then centrifuged filters at 10,000 x g for 1 minute to bind DNA with the eluate discarded. Next, we washed the filters with 400ul TD2, spun at 10,000 x g for 1 minute, and again discarded the eluate. We performed a second spin to remove all residual TD2. We transferred the filters to a new tube and added 80ul TD3 buffer, followed by a 5-minute incubation at room temperature. Finally, we eluted DNA by centrifugation at 10,000 x g for 1 minute. For DNA extractions from bronchial washings samples, we utilized the Power Soil (Mo Bio) as previously described.[7]

We amplified extracted DNA by PCR using the method of Caporaso et al.[8] and the Q5 HS High-Fidelity polymerase (NEB) targeting the V4 hypervariable region of the 16S rRNA gene. We utilized reagent controls for each step of the process (DNA extraction and PCR amplification). As PCR amplification positive controls, we utilized the ZymoBIOMICS Microbial Community DNA Standard (Zymo Research, Irvine, CA), a mock microbial community consisting of genomic DNA of eight bacterial strains. We utilized 1 µL of the genomic mixture with concentration of 10ng/µl for each reaction. We amplified 4 µL per reaction of each sample with a single barcode in triplicate 25 µL reactions. Cycle conditions were 98°C for 30s, then 33 cycles of 98°C for 10s, 57°C for 30s, 72°C for 30s, with a final extension step of 72°C for 2 min. We combined triplicates and purified with the AMPure XP beads (Beckman) at a 0.8:1 ratio (beads:DNA) to remove primer-dimers. We quantitated eluted DNA on a Qubit fluorimeter (Life Technologies). We performed sample pooling on ice by combining 20ng of each purified band. For negative controls and poorly performing samples, we used 20 µL of each sample. We purified the sample pool with the MinElute PCR purification kit. The final sample pool underwent two more purifications – AMPure XP beads to 0.8:1 to remove

all traces of primer dimers and a final cleanup in Purelink PCR Purification Kit (Life Technologies). We quantitated the purified pool in triplicate on the Qubit fluorimeter prior to preparing for sequencing.

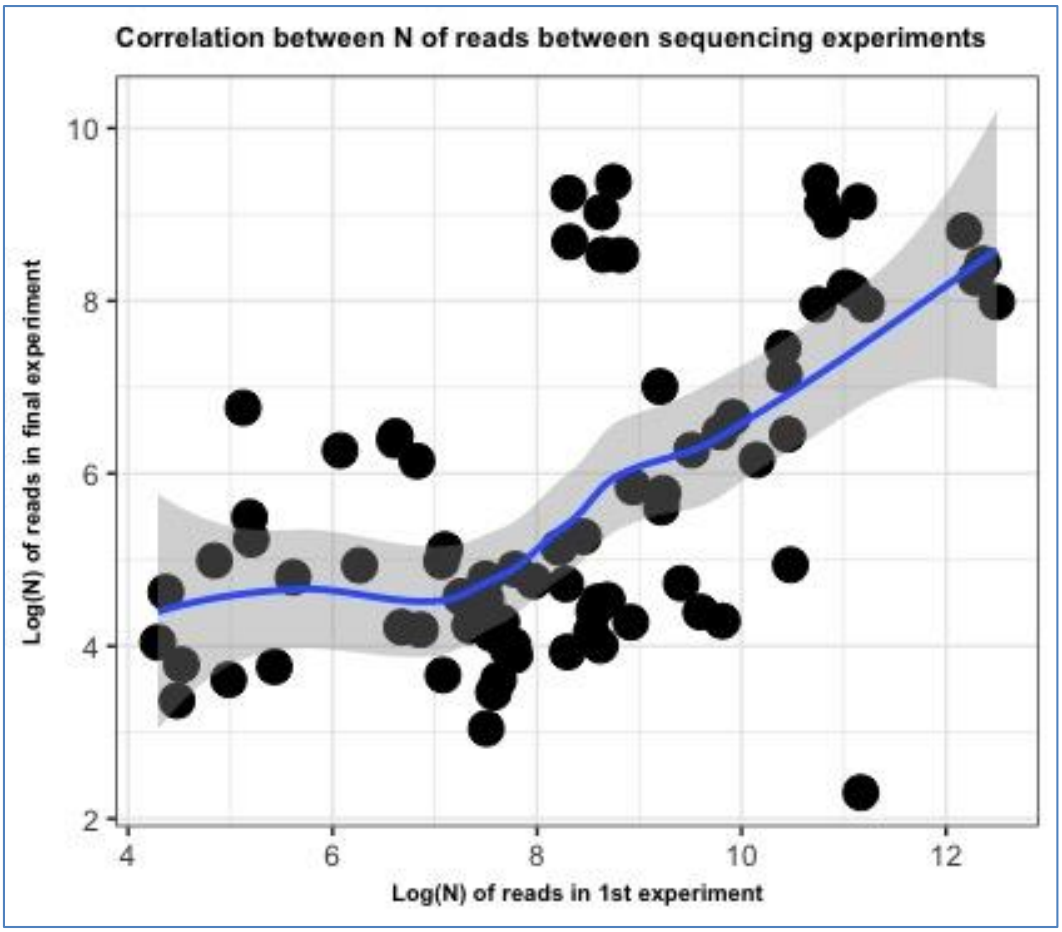
We prepared the sequencing pool according to instructions by Illumina, with an added incubation at 95^{o}C for 2 minutes immediately following the initial dilution to 20 picomolar. We then diluted the sequencing pool to a final concentration of 7 pM + 15% PhiX control. Amplicons were sequenced on the Miseq platform. We trimmed sequence reads and assigned them to operational taxonomic units (OTUs) with the QIIME software (ver 1.91) package by using a closed-reference approach and QIIME's default reference sequences, Greengenes 13_8, 97_otus.

We amplified DNA templates encoding 16S rRNA by using a quantitative PCR (qPCR) as previously described.[9] The amplification reaction of 20 µL consisted of 2 µL of PCR buffer (100 mmol/L Tris-HCl, pH 8.3; and 500 mmol/L KCl [Invitrogen]), 3.5 mmol/L $MgCl_2$, 0.2 mmol/L deoxynucleoside triphosphate, 0.5 µmol/L forward and reverse primers, 0.32 µmol/L probe, 0.75 U of Platinum Taq polymerase (Invitrogen), and 5 µL of the template DNA. The DNA was amplified in duplicate, and mean values were calculated. A standard curve was created from serial dilutions of plasmid DNA containing known copy numbers of the template. The reaction conditions for amplification of DNA were 95°C for 5 min, followed by 45 cycles at 95°C for 15 s and at 60°C for 1 min.

Internal validity and replication of microbiome experiments: We initially conducted DNA extraction, PCR amplification and sequencing in two different batches of experiments, which included random subsets of case and control samples in random processing order. We used the same type of DNA extraction kit for all experiments. As our initial results demonstrated extremely low bacterial signal for IPF lungs, we then

repeated all experiments of PCR amplification and sequencing in a single MiSeq run to ensure reproducibility among experiments. The pattern of results was consistent among experiments, and we noted statistically significant correlation between N of 16S reads (sequences) between 1st experiment and final sequencing run (Figure S1).

Figure S1. Correlation between Log of 16S reads between first and final sequencing experiments for all samples included.
Adjusted R-squared: 0.31, p-value = 4.8 x10^{-08}.



To minimize risks of contamination during the tissue DNA extraction process, we made every effort to minimize manipulation of the tissue samples to prevent procedural contamination, and thus we accepted a distribution of weights of tissue samples (mean 85mg, standard deviation 21.6mg) instead of aiming for narrow-range accuracy on

actual weight, which would have required multiple measurements on scale and dissections with sterile scalpels. There were no statistically significant differences in tissue weights of samples between the IPF, CORE and CF cohorts, and we found no significant association between tissue sample weights and 16S rRNA gene qPCR signal or number of reads (sequences) from MiSeq.

Statistical Analyses:

We calculated descriptive statistics and performed comparisons with non-parametric statistical tests using the R platform.[10] We performed taxonomic descriptions and beta diversity analyses (Bray Curtis dissimilarity index) with QIIME and compared distance matrices between groups with the anosim method.[11] Due to the very low number of reads produced by IPF lungs (nearly all samples <1000 reads), we did not perform any formal alpha diversity analyses. To identify operational taxonomic units (OTUs or taxa) with differential abundance between groups, we applied linear discriminant analysis (LDA) effect size (LEfSE) analysis, with thresholds of effect size by LDA scores >3 and $p<0.05$ adjusted for multiple comparisons with the Benjamini & Hochberg method (false discovery rate (FDR) probability).[12,13] To assess the origin of microbiota in the tissue communities and explore for potential sources of contamination, we applied the Bayesian approach of the SourceTracker software,[14] utilizing reference communities from the Human Microbiome Project (HMP).[15]

Statistical comparisons:

We considered the following two main comparisons of interest: 1) All patients with IPF versus controls from CORE (referred to as CORE lungs) and CF lungs, 2) AEIPF patients versus those with chronic IPF.

We performed additional secondary analyses for subsets of matched apical and basilar samples or tissue and bronchial washing samples. To contextualize our findings with the published evidence in the field, we also compared our IPF sequencing data with available published sequencing data for IPF patients from BAL samples (from the Molyneaux et al. study).[16]

ROLE OF THE FUNDING SOURCE:

The study sponsor had no role in the analysis or interpretation of the data and was not involved in the writing of the manuscript or decision to submit the paper for publication. The corresponding author had full access to the data and final responsibility for the decision to submit for publication.

RESULTS:

Table S1. Clinical data for patients with IPF.
Patients stratified by presentation with acute exacerbation of IPF (AEIPF, n=10) or chronic IPF (n=30). Data presented as mean (standard deviation) or N (%).

Variable	All IPF patients	Patients with AEIPF	Patients with chronic IPF
N	40	10	30
Age (years), mean (SD)	63.1 (8.1)	59.5 (10.5)	64.3 (6.9)
Male, n (%)	31 (77.5%)	9 (90.0%)	22 (73.3%)
Ever-smokers, n (%)	28 (70.0%)	9 (90.0%)	19 (63.3%)
Total pack-years, mean (SD)	18.2 (16.1)	20.5 (11.7)	17.4 (17.5)
FEV1 (L), mean (SD)	1.54 (0.56)	1.70 (0.72)	1.48 (0.49)
FEV1 predicted %, mean (SD)	53.1 (16.9)	53.9 (24.3)	52.8 (14.2)
FVC (L), mean (SD)	1.86 (0.72)	2.04 (0.89)	1.79 (0.67)
FVC predicted %, mean (SD)	45.5 (14.4)	44.7 (18.9)	45.7 (12.9)
DLCO (ml), mean (SD)	7.8 (4.9)	10.7 (8.7)	6.8 (2.5)
DLCO predicted %, mean (SD)	31.4 (14.5)	35.8 (24.6)	29.9 (9.7)
02 requirement at rest (L/min), mean (SD)	6.9 (3.4)	8.1 (4.4)	6.5 (2.9)
Mean PAP (mmHg), mean (SD)	25.5 (6.5)	27.7 (6.8)	24.8 (6.4)
GERD[#], n (%)	27 (67.5%)	8 (80.0%)	19 (63.3%)
Systemic steroids, n (%)	12 (30.0%)	4 (40.0%)	8 (26.7%)
Prednisone dose equivalent (mg/day), mean (SD)	6.8 (17.5)	18.5 (32.1)	3 (5.2)
Inhaled corticosteroids, n (%)	6 (15.0%)	2 (20.0%)	4 (13.3%)
Other immunomodulator, n (%)	15 (37.5%)	5 (50.0%)**	10 (33.3%)[##]
BAL + culture or recipient tissue culture*, n (%)	3 (7.5%)[$]	2 (22.2%) (*Pseudomonas, Candida Albicans)*	1 (4.3%)[$] (*Aspergillus fumigatus*)
Years with IPF diagnosis, mean (SD)	4.3 (2.6)	2.9 (2.5)	4.7 (2.5)
Diffuse alveolar damage on explant pathology, n (%)	10 (25.0%)[$]	6 (60.0%)[$]	4 (13.3%)[$]
Antibiotics in past 3 months, n (%)	12 (30.0%)	8 (80.0%)	4 (13.3%)
Lung transplant, n (%)	35 (87.5%)	6 (60.0%)	29 (96.7%)

[#] GERD definition: based on clinical history, available data from esophagogram or upper endoscopy, or prescription for proton pump inhibitor or histamine receptor 2 blocker.
* All patients who underwent lung transplantation had a tissue culture of their resected main stem bronchial tissue, as per our transplant protocol.
[$] Percentage calculated with denominator of all patients with available data (patients with unavailable data for each variable were removed from calculations).
** Intravenous immunoglobulin, plasma exchange and Rituximab (n=3), (Simtuzumab or placebo; unknown treatment assignment as this patient was enrolled in a double blind clinical trial) (n=1), Mycophenolate (n=1).
[##] Pirfenidone (n=5), Nintedanib (n=1), Azathioprine (n=1), anti-IL13 (n=1), (Simtuzumab or placebo; unknown treatment assignment as this patient was enrolled in a double blind clinical trial) (n=1), Hydroxychloroquine (n=1).
Abbreviations: FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity, 6MWD: 6-

minute walk distance; DLCO: diffusion capacity of the lungs for carbon monoxide; GERD: gastroesophageal reflux disease; PAP: pulmonary artery pressure.

Table S2:
Diagnostic criteria for cases with AEIPF (suspected and confirmed), according to the revised diagnostic criteria.[2]

	Criteria					
Patient	1. Diagnosis of IPF	2. Acute worsening or development of dyspnea typically less than one month duration	3. Computed tomography with new bilateral ground-glass opacity and/or consolidation superimposed on a background pattern consistent with usual interstitial pneumonia pattern	4. Deterioration not fully explained by cardiac failure or fluid overload	Confirmed AEIPF (4/4 diagnostic criteria met)	Suspected AEIPF (3/4 diagnostic criteria met)
2	Yes	Yes	Yes	Yes	Yes	No
4	Yes	Yes	NA	Yes	No	Yes
5	Yes	Yes	Yes	Yes	Yes	No
8	Yes	Yes	NA	Yes	No	Yes
10	Yes	Yes	Yes	Yes	Yes	No
25	Yes	Yes	NA	Yes	No	Yes
29	Yes	Yes	Yes	Yes	Yes	No
30	Yes	Yes	Yes	Yes	Yes	No
31	Yes	Yes	NA	Yes	No	Yes
35	Yes	Yes	Yes	Yes	Yes	No
Total	**10/10**	**10/10**	**6/10**	**10/10**	**6/10**	**4/10**

Figure S2: Log DNA concentration of post-PCR amplification product by each sample type. Concentration measured with the Qubit method. Mean concentration in IPF lungs was statistically significantly lower than in CORE lungs ($p<0.0005$). This finding indicates that while we are able to extract DNA from IPF lungs, this is likely of low concentration of 16S template that fails to produce a detectable signal by 16S sequencing.

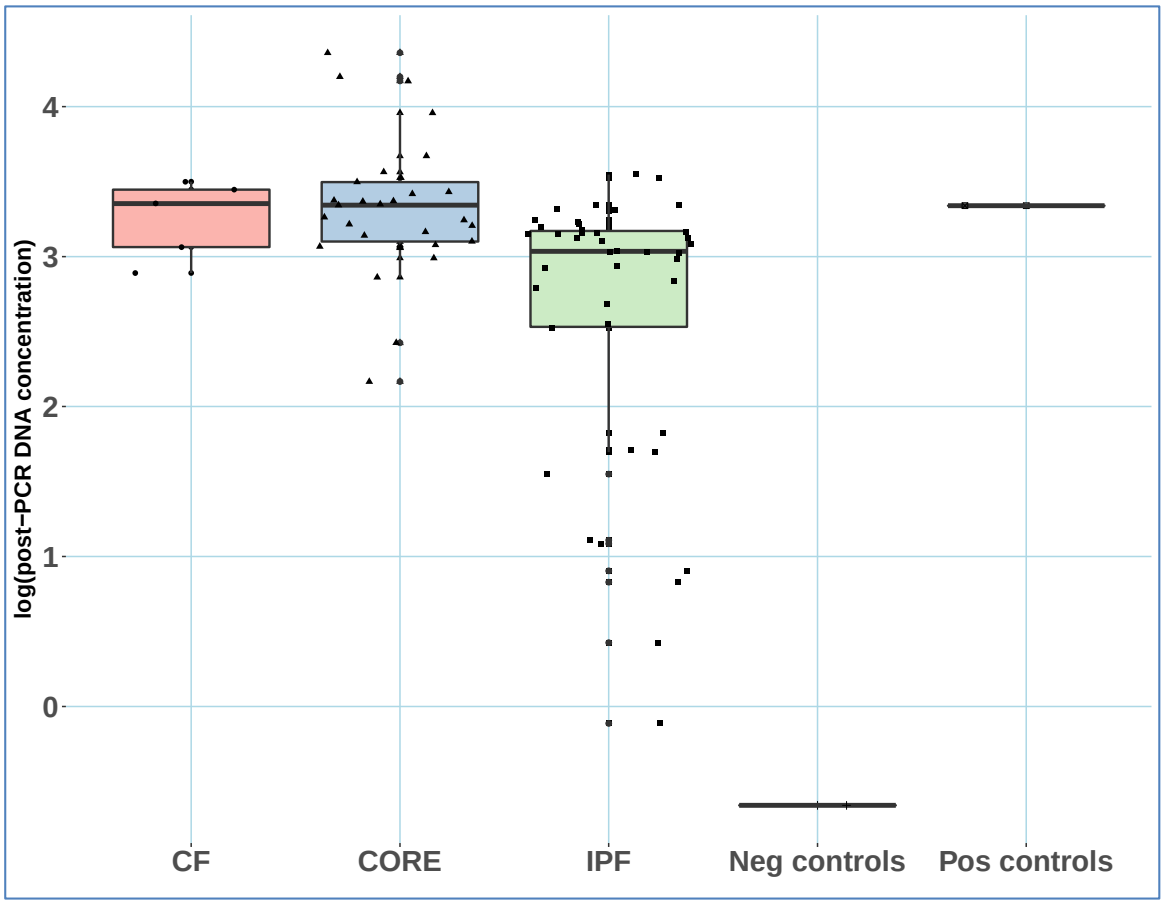


Figure S3: Linear Discriminant Analysis (LDA) scores for different relative abundances between IPF and CORE lungs. Taxonomic information for the 33 differentially abundant taxa are shown in the y-axis. Comparisons were adjusted for multiple testing (False Discovery Rate p-values <0.05). Analyses performed with the Linear discriminant analysis Effect Size (LEfSE) software. Comparisons between IPF and negative (reagent) controls revealed no statistically significant taxonomic abundance differences adjusted for multiple testing (False Discovery Rate p-values <0.05).

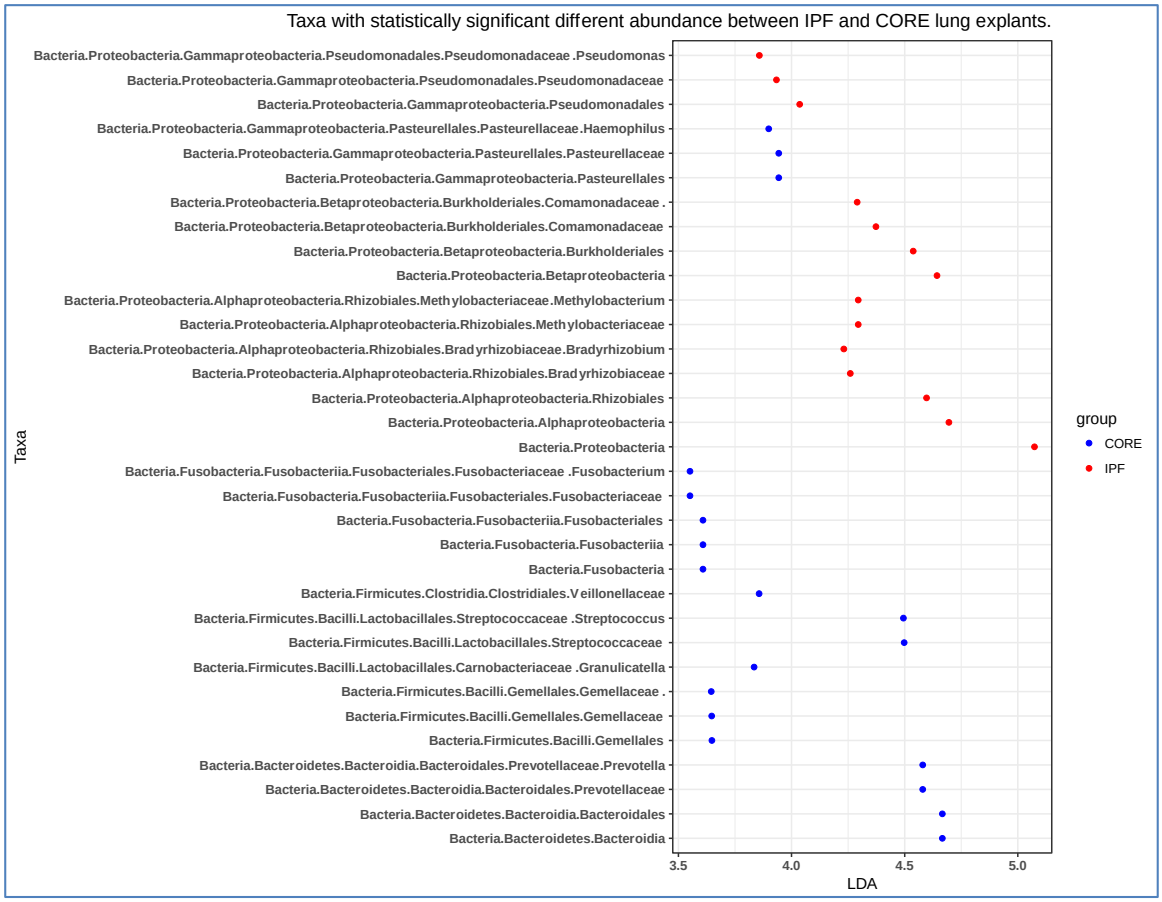


Figure S4: Sourcetracker analysis results. We utilized publicly available data from the Human Microbiome Project (HMP) to identify potential origin and sources of bacteria in our tissue samples. We utilized 16S rRNA gene sequencing data (V1-3 regions) from skin (right and left antecubital fossa), oral (throat, dorsum of tongue, saliva and tonsils), stool and nares. We applied the Sourcetracker analysis algorithm with default settings.[14] Mean proportions of skin and oral origin microbiota in the communities of IPF and CORE lung tissue and reagent controls are shown as bars with associated standard errors. Statistically significant differences between IPF and CORE lungs for oral and skin origin bacteria were found (non-parametric p-values for comparisons of mean oral and skin bacteria abundance were 4.87×10^{-05} and 0.01, respectively)

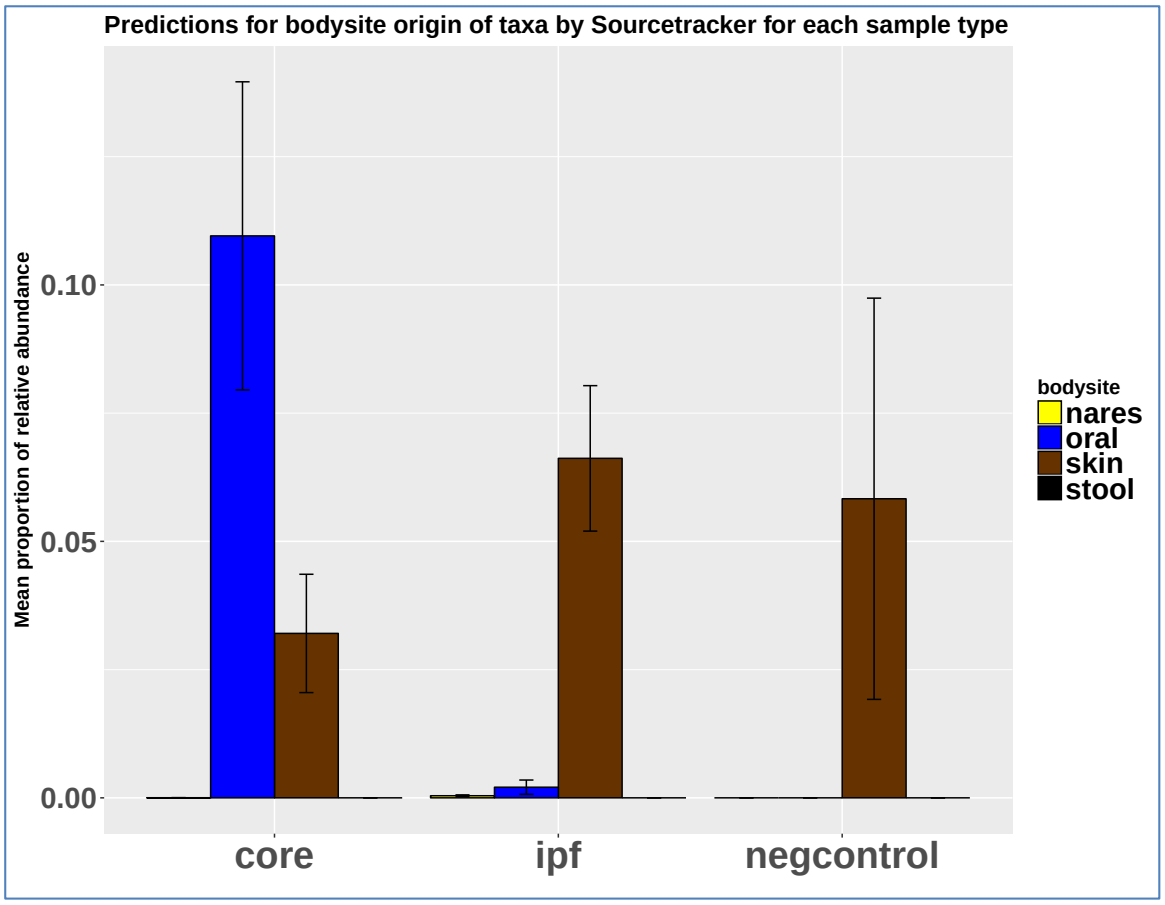


Figure S5: Comparisons between acute exacerbations of IPF and chronic IPF by A. N of reads, B. Principal coordinates analysis of Bray-Curtis dissimilarity distances and C. Taxonomic composition. While AEIPF lungs had higher median number of reads compared to chronic IPF (non-parametric p-value p=0.005), both sample types produced an exceedingly low number of reads (<1000) and inferences for a high microbial load in AEIPF lungs cannot be made. By comparison of Bray-Curtis dissimilarity distances with anosim testing, AEIPF samples were not statistically significant different from chronic IPF samples in terms of their taxonomic composition (p>0.05).

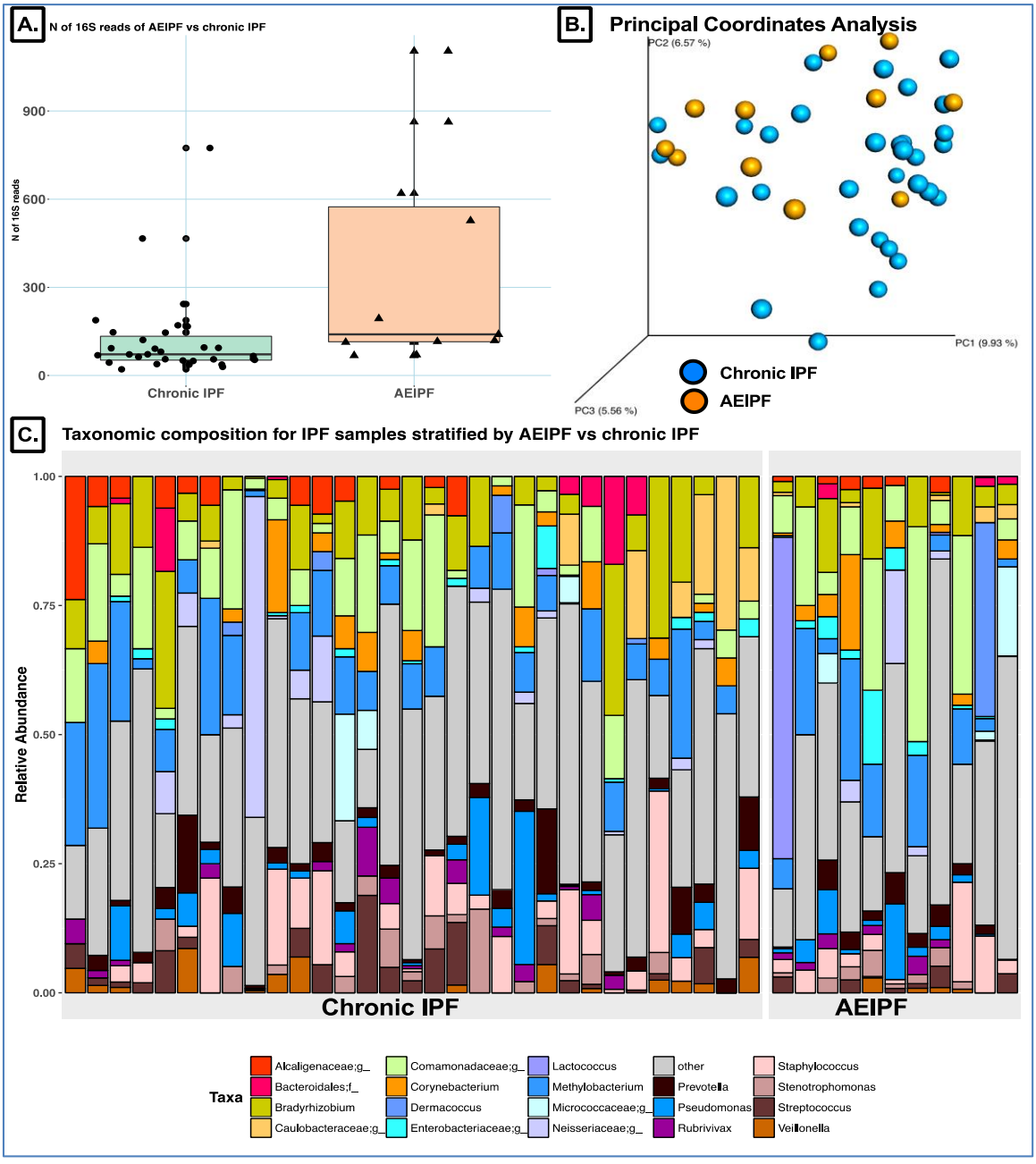


Figure S6. Heatmap of summary relative abundance by groups of sample types. CF communities had strong signal by *Burkholderia* or *Pseudomonas* taxa (corresponding to pathogens identified in explant cultures) and CORE samples had overall higher abundance of typical oral taxa (e.g. *Prevotella*, *Streptococcus*) as expected for the supraglottic pneumotype of the healthy lung microbiome. In contrast, IPF samples had high abundance of typical skin microbiota (e.g. *Comamonadaceae, Methylobacterium)* also seen in negative controls.

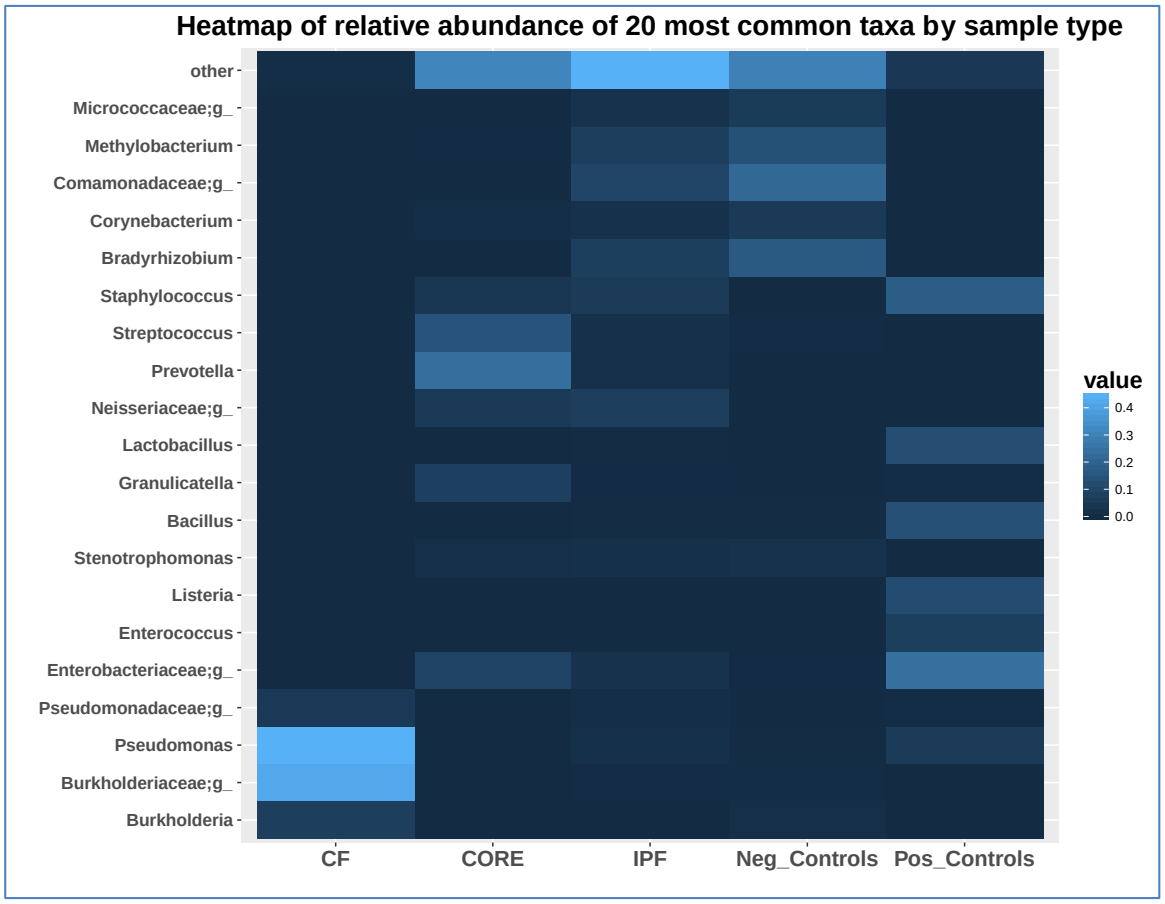


Figure S7: Comparison between apical and basilar samples in 3 IPF patients with matched sample types. A. N of reads, B. Principal coordinates analysis of Bray-Curtis dissimilarity distances and C. Taxonomic composition. In one patient (patient B), we also had an available right middle lobe sample, thus making a total of 4 non-basilar samples from 3 patients for comparisons with 3 basilar samples.

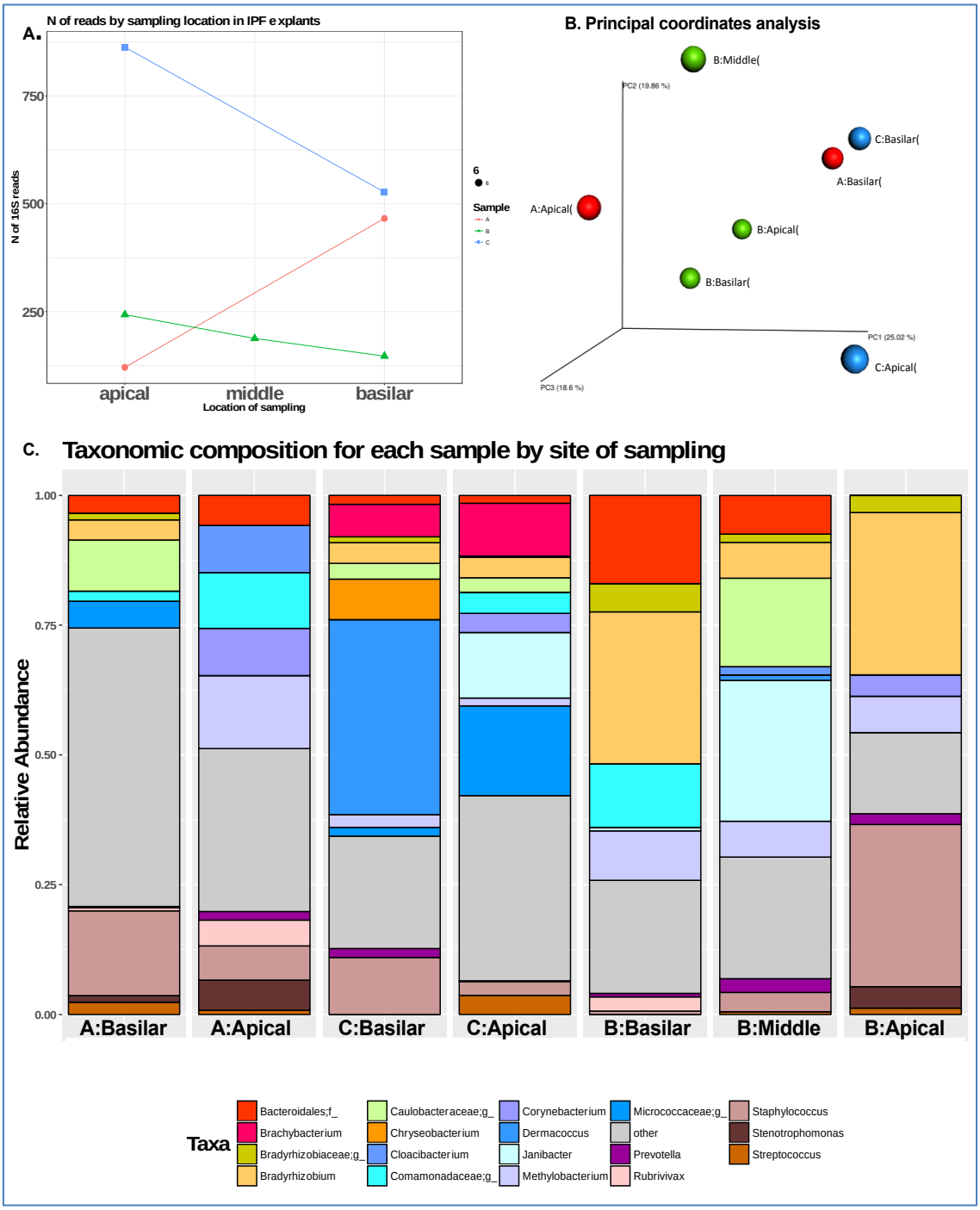


Figure S8: Comparison between bronchial washings and basilar tissue samples in 3 IPF patients with matched sample types. A. N of reads, B. Taxonomic composition. Taxonomic composition comparisons indicated limited taxonomic overlap between bronchial washings and tissue samples (range of overlapping taxa 29-53%).

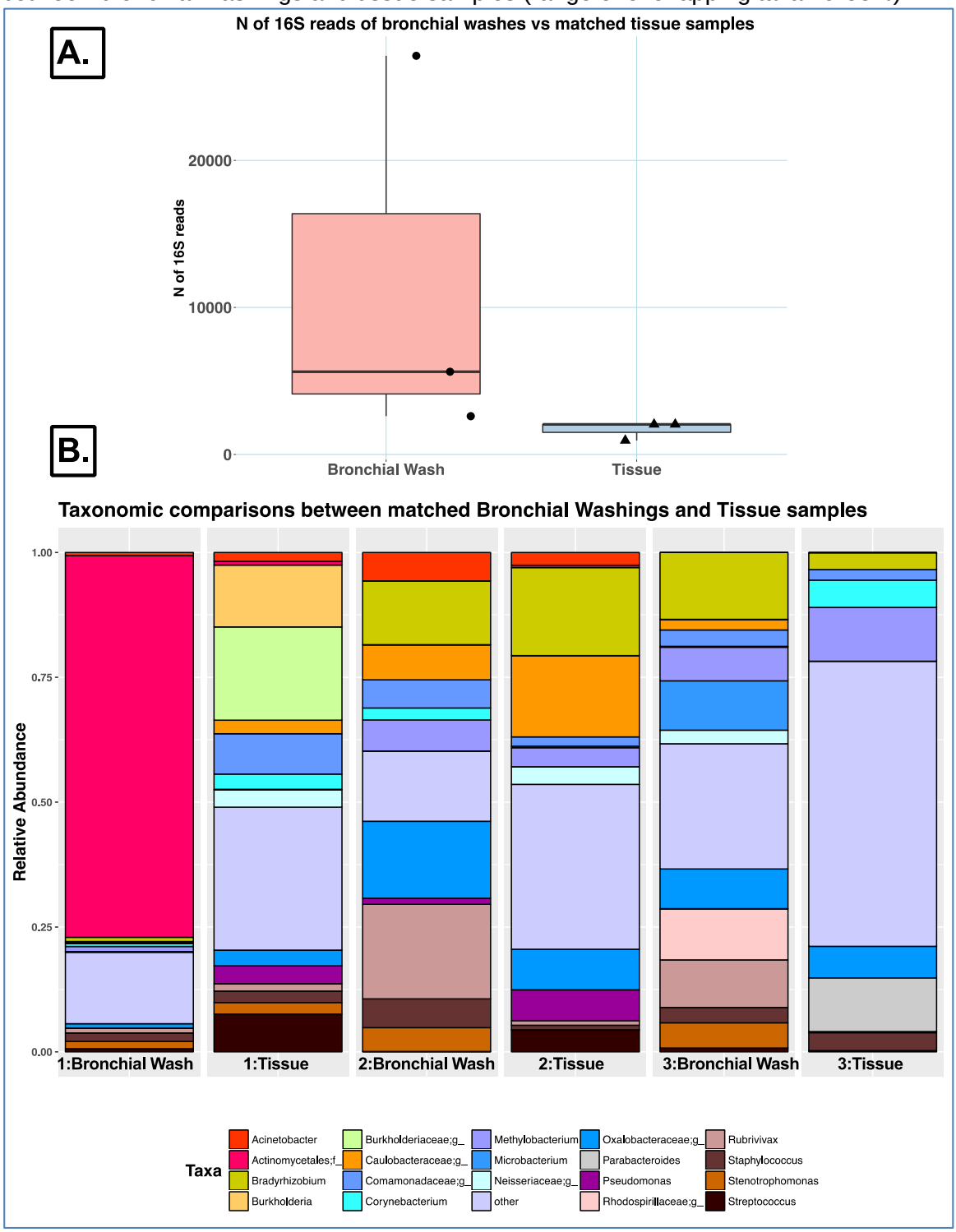
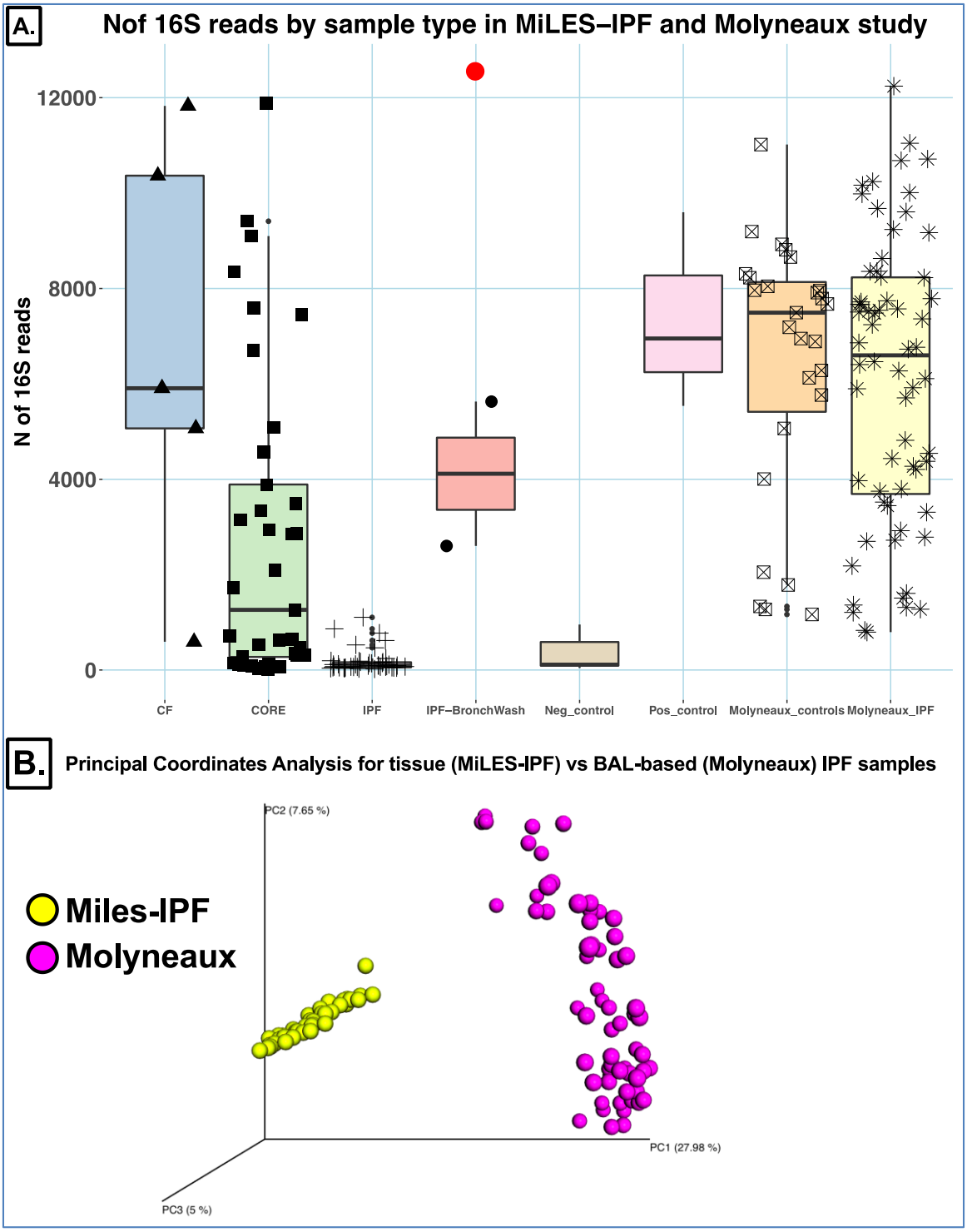


Figure S9: Comparisons between tissue-based Samples in MiLES-IPF and BAL-based samples in Molyneaux et al. study. A. N of reads by sample type (CF, CORE, IPF tissue, IPF bronchial washings, Negative controls and Positive Controls in MiLES-IPF study; IPF BAL and Control BAL samples in the Molyneaux et al. study). B. Principal coordinates analysis for Bray Curtis dissimilarity distances between MiLES-IPF and Molyneaux BAL-based IPF samples, showing striking compositional differences of the microbial communities identified in IPF lungs between the two studies (Anosim test statistic 0.998, p-value <0.001). The tissue samples in our study cluster closely together and are markedly dissimilar from the published BAL-based data, because the tissue samples largely represent background signal. In graph A, an outlying observation of number of reads (n= 27,118) from one bronchial wash IPF sample is highlighted with a red sphere. The y-axis is limited to 12,500 reads to improve readability of available sample distributions.


A.
Nof 16S reads by sample type in MiLES–IPF and Molyneaux study
N of 16S reads
12000
8000
4000
0
CF
CORE
IPF
IPF–BronchWash
Neg_control
Pos_control
Molyneaux_controls
Molyneaux_IPF
B.
Principal Coordinates Analysis for tissue (MiLES-IPF) vs BAL-based (Molyneaux) IPF samples
Miles-IPF
Molyneaux
PC2 (7.65 %)
PC1 (27.98 %)
PC3 (5 %)

Bibliography

1 Raghu G, Collard HR, Egan JJ, *et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;**183**:788–824. doi:10.1164/rccm.2009-040GL

2 Collard HR, Ryerson CJ, Corte TJ, *et al.* Acute exacerbation of idiopathic pulmonary fibrosis. an international working group report. *Am J Respir Crit Care Med* 2016;**194**:265–75. doi:10.1164/rccm.201604-0801CI

3 Donahoe M, Valentine VG, Chien N, *et al.* Autoantibody-Targeted Treatments for Acute Exacerbations of Idiopathic Pulmonary Fibrosis. *PLoS ONE* 2015;**10**:e0127771. doi:10.1371/journal.pone.0127771

4 Bermudez CA, Rocha RV, Zaldonis D, *et al.* Extracorporeal membrane oxygenation as a bridge to lung transplant: midterm outcomes. *Ann Thorac Surg* 2011;**92**:1226–31; discussion 1231. doi:10.1016/j.athoracsur.2011.04.122

5 Lindell KO, Erlen JA, Kaminski N. Lessons from our patients: development of a warm autopsy program. *PLoS Med* 2006;**3**:e234. doi:10.1371/journal.pmed.0030234

6 Mazhar K, Gunawardana M, Webster P, *et al.* Bacterial biofilms and increased bacterial counts are associated with airway stenosis. *Otolaryngol Head Neck Surg* 2014;**150**:834–40. doi:10.1177/0194599814522765

7 Morris A, Beck JM, Schloss PD, *et al.* Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med* 2013;**187**:1067–75. doi:10.1164/rccm.201210-1913OC

8 Caporaso JG, Lauber CL, Walters WA, *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 2012;**6**:1621–4. doi:10.1038/ismej.2012.8

9 Jiang W, Lederman MM, Hunt P, *et al.* Plasma levels of bacterial DNA correlate with immune activation and the magnitude of immune restoration in persons with antiretroviral-treated HIV infection. *J Infect Dis* 2009;**199**:1177–85. doi:10.1086/597476

10 R Foundation for Statistical Computing RCT. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: 2016. https://www.R-project.org (accessed 11 Oct2016).

11 Caporaso JG, Kuczynski J, Stombaugh J, *et al.* QIIME allows analysis of high-

throughput community sequencing data. *Nat Methods* 2010;**7**:335–6. doi:10.1038/nmeth.f.303

12 Segata N, Izard J, Waldron L, *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol* 2011;**12**:R60. doi:10.1186/gb-2011-12-6-r60

13 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* 1995;**57**:289–300.http://www.jstor.org/stable/2346101

14 Knights D, Kuczynski J, Charlson ES, *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nat Methods* 2011;**8**:761–3. doi:10.1038/nmeth.1650

15 Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;**486**:207–14. doi:10.1038/nature11234

16 Molyneaux PL, Cox MJ, Willis-Owen SAG, *et al.* The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014;**190**:906–13. doi:10.1164/rccm.201403-0541OC