Topographic Heterogeneity of Lung Microbiota in End-Stage Idiopathic Pulmonary Fibrosis: The Microbiome in Lung Explants-2 (MiLEs-2) Study.

Online Supplement

Extended Methods: pages 2-11

Results: pages 12-27

References: page 28

Valenzi E, *et al. Thorax* 2020;0:1–9. doi: 10.1136/thoraxjnl-2020-214770

EXTENDED METHODS

Study procedures:

We performed sampling of the explanted lung(s) in two different pneumonectomy settings, as previously described: 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 the University of Pittsburgh Medical Center (UPMC).[1] 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) from all explants included in the study. From a subset of explants, we also resected subpleural tissue from the right middle lobe or lingula (for right or left lung explants respectively) and the upper lobe. For a smaller subset of explants, we were also able to collect a bronchial wash specimen (by aspiration of 30 mL of phosphate-buffered saline instilled into a bronchial segment using a sterile tube) as well as an airway tissue specimen (from a segmental bronchus) prior to parenchymal tissue sample collection. Samples were frozen in liquid nitrogen and stored at -80°C until processing. 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 within six hours of death and provided the explants to our research team immediately after. We performed tissue sampling, processing and storage in a similar fashion as described above. The University of Pittsburgh's rapid autopsy program is currently limited to patients with interstitial lung disease or pulmonary arterial hypertension; hence no COPD or CF lungs were collected in this manner.

Our study utilized a retrospective cohort of samples previously collected by a Lung Tissue biospecimen core repository for use by multiple investigators at our institution. The varied amount of sampling for each lung is primarily a result of feasibility at the time of procurement, including but not limited to: staff availability, timing of the transplant, and time that can be dedicated to sample collection prior to time-sensitive experiments on fresh tissue. Additionally, patients undergoing a single lung transplant (rather than a double lung transplant) only had basilar parenchyma tissue available for sampling. No bronchial washings or airway tissue was sampled from CORE lungs and no bronchial washings from CF lungs, as this was not previously part of the protocol for these specimens. We included specimens from explanted lungs procured from organ donation candidates that were deemed unsuitable for lung transplantation, obtained from the Pennsylvania Center for

Organ Recovery and Education (CORE). All patients met brain death criteria, and these samples were anonymized by CORE investigators. These organ grafts were rejected for reasons that included donor age, hypoxemia, and mechanical complications among others. We excluded cases with clinical or microbiologic evidence of pulmonary infection prior to procurement. Details for the reasons of rejection of these organs for lung transplantation are provided in Supplementary Table 1:

Supplementary Table 1: Reasons of rejection of explanted donor lungs for transplantation:

Reason for rejection	**N**
Age	2
Persistent hypoxemia	5
Poor respiratory mechanics	1
Ex vivo pulmonary edema	4
Pulmonary nodules	1
Pulmonary contusions from chest trauma	3
Size	1
Aspiration	1
Purulence on bronchoscopy	1
Bilateral pulmonary embolism	1
Total CORE subjects	20

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 idiopathic pulmonary fibrosis (IPF) lungs.

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

DNA extraction: We extracted bacterial DNA from ~45 milligrams (mg) of whole lung explant tissue using DNeasy Powersoil Kits (Qiagen, Germantown, MD), following the manufacturer's instructions. For each DNA extraction experiment, we included negative controls (sterile water instilled in extraction columns) to examine for contamination prior and during the DNA extraction phase.

16S rRNA gene sequencing: We amplified extracted DNA by PCR using the method of Caporaso et al. and the Q5 HS High-Fidelity polymerase (NEB) targeting the V4 hypervariable region of the 16S rRNA gene.[2] We utilized reagent controls for each PCR amplification plate to examine for contamination post-DNA extraction. We amplified four microliters

per reaction of each sample with a single barcode in triplicate 25 microliter 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 20 ng of each purified band. For negative controls and poorly performing samples, we used 20 microliters 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˚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.

Quantitative PCR (qPCR): We performed qPCR reactions of the V3-V4 region of the 16S rRNA gene to quantify the bacterial load in each sample, with the BactQuant protocol.[2] PCR reactions were prepared in a total volume of 20µL, which consisted of 2µL of 10XPCR buffer, 3.5 mmol/L $MgCl_2$, 0.2 mmol/L deoxynucleoside triphosphate, 0.5 µmol/L forward and reverse primers, 0.225 µmol/L probe, 0.75 U of Platinum Taq polymerase (Invitrogen), and 2µL of each DNA. The forward and reverse primers and the probe sequences that were used to amplify DNA templates encoding the V3-V4 region of the 16S rRNA gene were identical to those previously described.[2] 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 assays were performed on the LightCyler System (Roche) using the following PCR conditions: 95°C for 5 min, followed by 50 cycles at 95°C for 15 s and at 60°C for 1 min. SYBR GREEN-based qPCR of human Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene was performed to detect and quantify human DNA in available samples on LightCyler System (Roche, Basel, Switzerland).

Protocol optimization:

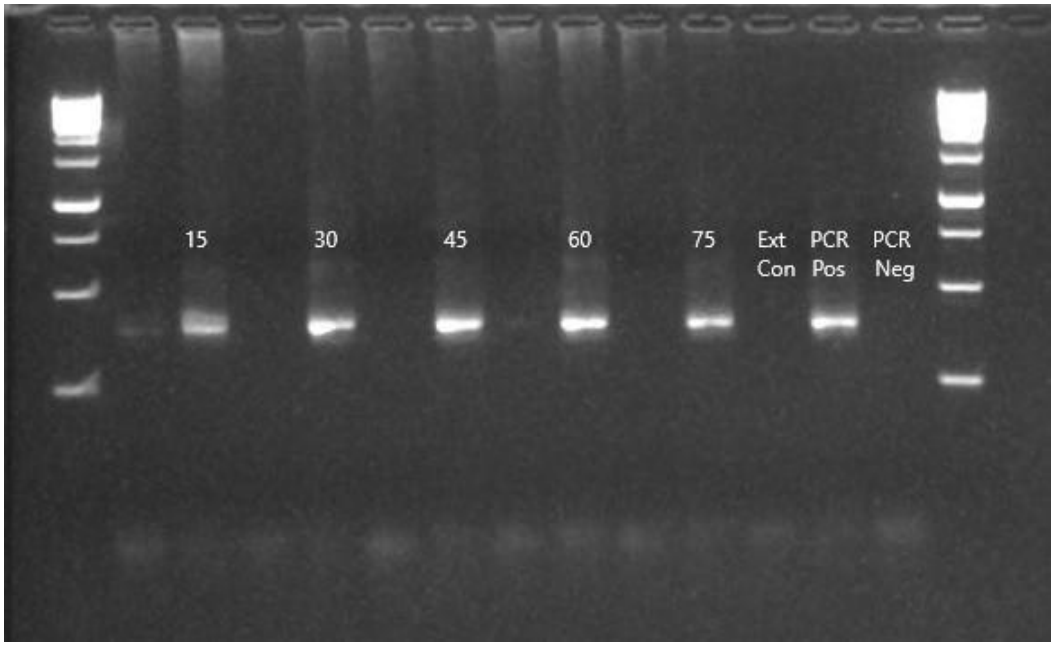
In previous work from our group,[1] we demonstrated extremely low bacterial DNA amplification from IPF explants, raising concerns about the feasibility of studying low biomass communities from explanted lung tissue.[3, 4] To examine for potential methodological issues and define optimal methods for studying low biomass communities in lung tissue

specimens, we conducted a series of methodological experiments prior to defining the final protocol for the MiLEs-2 study processes described above. Specifically, we examined for the presence of PCR inhibitors of bacterial DNA amplification in human lung tissue (CORE lungs or IPF lungs) as well as for the optimal method of sampling tissue specimens (tissue swabs as performed in the study by Pragman et al vs. whole tissue DNA extracts).[5] In these experiments, we utilized our standard PCR amplification protocol of the V4 hypervariable region of the 16S rRNA gene [2], and examined for the detection of corresponding band of the amplified PCR product in electrophoresis for each experimental condition.

a. Examination for PCR inhibitors in CORE lung tissue

To determine whether lung tissue specimens contained inhibitors that interfered with the amplification of the 16S rRNA gene by PCR (PCR inhibitors), we conducted the following experiments. First, we spiked CORE lung tissue specimens with bacterial cells by using varying amounts of ZymoBIOMICS Microbial Community Standards (Zymo, USA). Tissue specimens were spiked with incremental amounts of the Zymo community standard solution: 15µl, 30µl, 45µl, 60µl, and 75µl. Then, these spiked lung tissue specimens underwent whole genomic DNA (gDNA) extraction with Qiagen's DNeasy PowerSoil Kit (Qiagen, USA). After the addition of C1, samples were heated for 15 minutes at 65°C, and vortexed for 20 minutes. All other steps were performed according to the kit's protocol. Extracted gDNA was amplified using 4µl of template for a total reaction volume of 25µl. The thermal cycling program was as follows: initial activation at 98°C for 30 seconds, followed by 25 cycles of 98°C for 10s, 57°C for 30s, and 72°C for 30s, with final elongation of 72°C for 2 minutes. To test the success of the PCR we used gel electrophoresis. Every spiked tissue sample displayed bright bands of amplified DNA (e-Methods Figure 1), thus ruling out the presence of PCR inhibitors at concentrations sufficient to inhibit the amplification of bacterial DNA in the amounts available in the Zymo community standard.

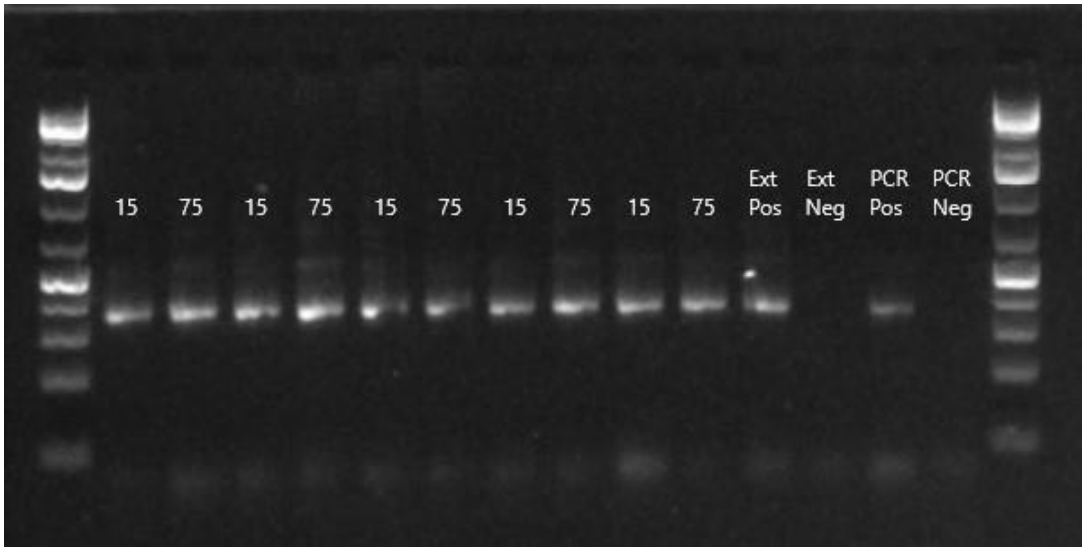
e-Methods Figure 1. Gel electrophoresis of the PCR amplified V4 subunit of the 16S rRNA gene for CORE tissue specimens spiked with Zymo bacterial cells. Lane 1: Ladder. Lanes 2, 4, 6, 8, 10: CORE tissue extractions. Lanes 3, 5, 7, 9, 11: spiked CORE tissues ranging from 15ul Zymo cells, to 75ul Zymo cells. Lane 12: Extraction negative control. Lane 13: PCR positive control. Lane 14: PCR negative control. Lane 15: Ladder



b. Examination for PCR inhibitors in IPF lung tissue

To test for PCR inhibitors in end-stage IPF lung tissues, specimens were spiked with either 15µl or 75µl of bacterial cells from the Microbial Community Standard (Zymo, USA). DNA extractions were completed as described above. PCR was performed using 4µl of template and 25 cycles of the above described thermal cycling program. All tissues spiked with Zymo cells yielded visible bands using gel electrophoresis (e-Methods Figure 2). Therefore, we concluded that IPF lung tissues do not contain sufficient amounts of PCR inhibitors to inhibit 16S rRNA gene amplification at the concentrations used.

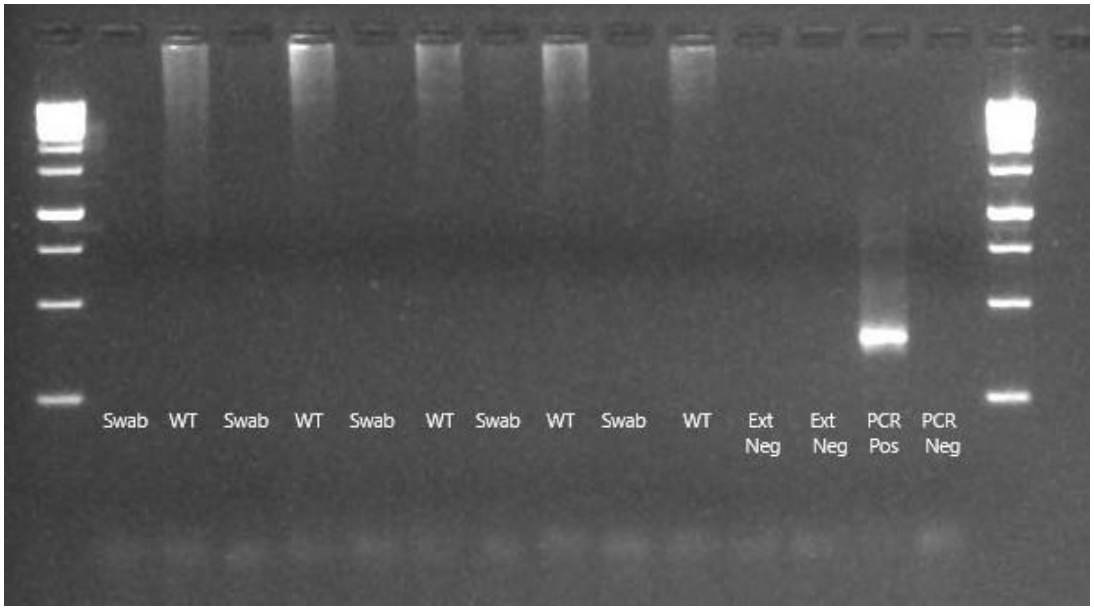
e-Methods Figure 2. Gel electrophoresis of the PCR amplified V4 subunit of the 16S rRNA gene for IPF tissues spiked with Zymo bacterial cells. Lane 1: Ladder. Lanes 2-11: spiked IPF tissues alternating between 15µl and 75µl spike-in samples. Lane 12: Extraction positive control (75ul of Zymo cells). Lane 13: Extraction negative control. Lane 14: PCR positive control. Lane 15: PCR negative control. Lane 16: Ladder.



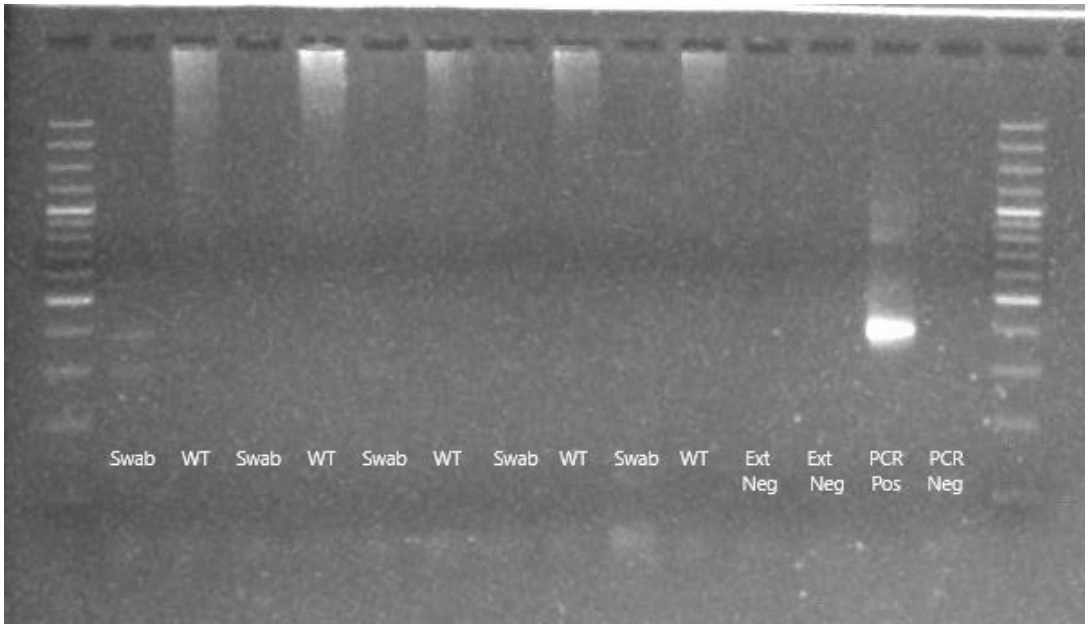
c. Comparison of PCR amplification yield between tissue swabs and whole tissue extractions – derivation of optimal number of PCR cycles:

A plausible obstacle in successfully amplifying bacterial DNA from whole human tissue extractions is the high ratio of human to microbial DNA in gDNA extracts. To address this issue, tissue swabs have been proposed in an effort to sample retrievable bacteria from the tissue surface and avoid the large amounts of human DNA in the gDNA extracted pool. To compare the yield of tissue swabs vs. whole tissue extractions, we swabbed the surface area of five CORE lung tissue samples. These swabs were extracted in addition to the resected sections of lung tissues weighing approximately 45mg each. PCR was performed using 8µl of template for a total reaction volume of 25µl. The thermal cycling program used is as described above. Using gel electrophoresis, no samples yielded visible bands of amplified DNA using 25 cycles (e-Methods Figure 3) or 30 cycles of the above-mentioned thermal cycling program (e-Methods Figure 4). Using 33 cycles, all samples yielded visible bands of amplified DNA (e-Methods Figure 5). Therefore, we proceed with whole tissue sample extractions and 33 cycles of PCR to maximize the yield of amplified bacterial DNA from each sample.

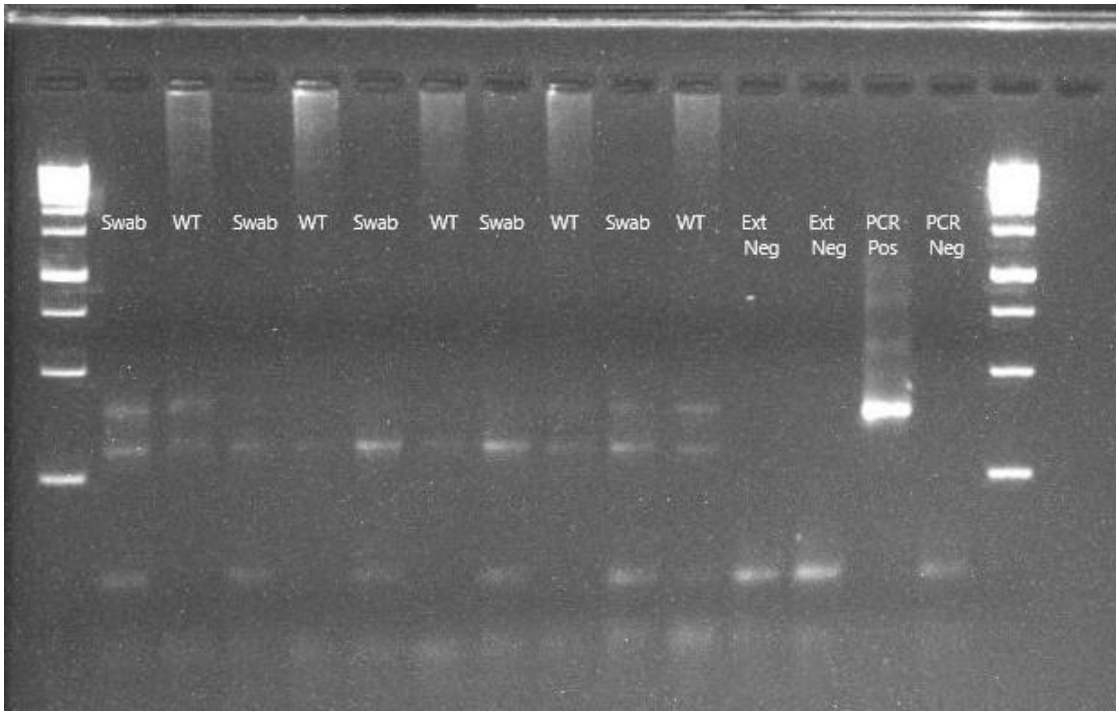
e-Methods Figure 3. Gel electrophoresis of the PCR amplified V4 subunit of the 16S rRNA gene for CORE tissues and swabs using 25 cycles and 4µl of template. Lane 1: Ladder. Lanes 2-11: CORE lung extractions alternating between swabs and whole tissue (WT) extractions. Lane 12-13: Extraction negative controls. Lane 14: PCR positive control. Lane 15: PCR negative control. Lane 16: Ladder.



e-Methods Figure 4. Gel electrophoresis of the PCR amplified V4 subunit of the 16S rRNA gene for CORE tissues and swabs using 30 cycles and 8µl of template. Lane 1: Ladder. Lanes 2-11: CORE lung extractions alternating between swabs and whole tissue (WT) extractions. Lane 12-13: Extraction negative controls. Lane 14: PCR positive control. Lane 15: PCR negative control. Lane 16: Ladder.



e-Methods Figure 5. Gel electrophoresis of the PCR amplified V4 subunit of the 16S rRNA gene for CORE tissues and swabs using 33 cycles and 8µl of template. Lane 1: Ladder. Lanes 2-11: CORE lung extractions alternating between swabs and whole tissue (WT) extractions. Lane 12-13: Extraction negative controls. Lane 14: PCR positive control. Lane 15: PCR negative control. Lane 16: Ladder.



Analytics:

a. 16S annotation and the making of operational taxonomic unit (OTU) table

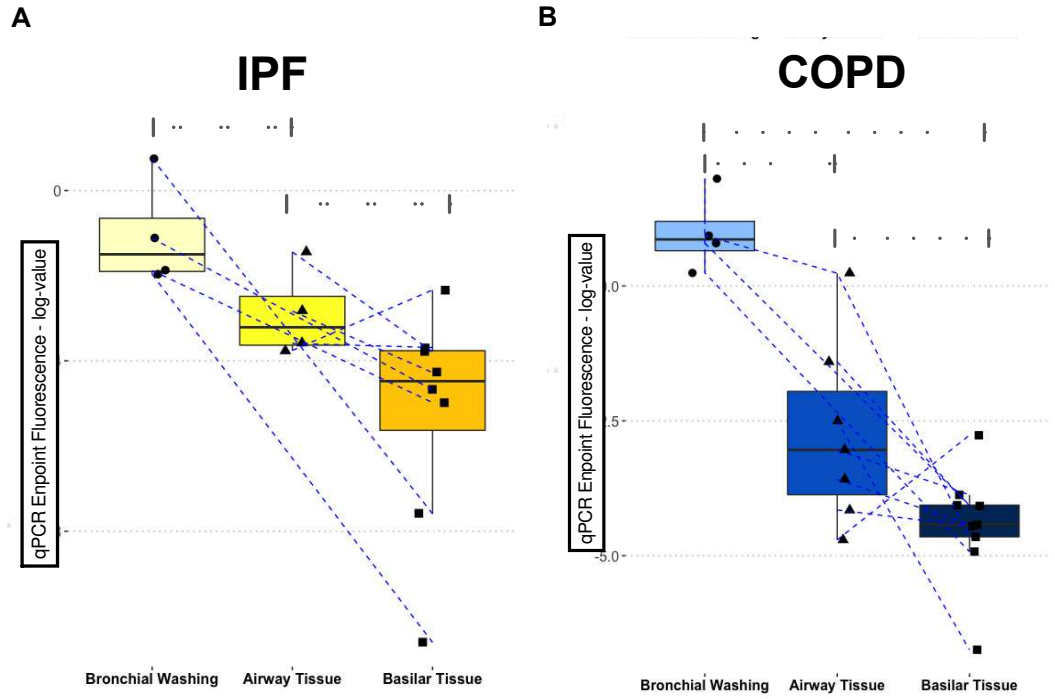
Sequences from the pooled sequencing run were demultiplexed into individual sample/replicate fastq files. Each fastq file was then processed through the Center for Medicine and the Microbiome (CMM) custom modular read QC pipeline that was configured to perform the following steps: low complexity filtering, QV trimming, Illumina sequencing adapter trimming, and 16S primer trimming. Mated pair forward and reversed reads were joined into contigs with the make.contigs function in Mothur, followed by additional screening with the following criterion: minimum required overlap = 25 bp, proportion overlap mismatch > 20%, maximum N's allowed = 4, and a read length minimum of 125 bp. Then, consensus sequences passing screening were then passed through CMM's 16S clustering and annotation pipeline, a Mothur-dependent wrapper designed to streamline and automate the execution of the following Mothur steps in version v.1.39.1: unique.seqs, align.seqs, screen.seqs, filter.seqs, second uniq.seqs, pre.cluster, chimera.uchime, remove.seqs, classify.seqs, dist.seqs, cluster, make.shared, and classify.otu. *Mothur* output files were then reformatted to sample x category (taxonomic levels or OTU at 97% sequence similarity). For downstream statistical modeling and analyses, we utilized taxonomic tables at the genus level, with taxonomic assignments performed using a naïve Bayes k-mer classifier in conjunction with the Ribosomal Database Project (RDP) 16S rRNA gene sequences. Details of the custom analytical pipeline used are provided in https://github.com/MedicineAndTheMicrobiome/AnalysisTools

b. Cluster derivation

In order to agnostically examine for distinct clusters of microbial composition ("meta-communities") in our samples, we applied unsupervised Dirichlet Multinomial Models (DMM) with Laplace approximations to define the optimal number of clusters in our dataset. DMM clusters were derived in a trimmed dataset following exclusion of low abundance taxa, which could represent experimental noise and not true biological signal. We therefore excluded taxa whose relative abundance did not exceed 20% in any of the 62 IPF samples, which resulted in a trimmed dataset with 76% of the total reads and a sum of 40 distinct genera for derivation of the DMM clusters. Application of the DMM models resulted into an interpretable classification of our dataset in two distinct clusters, which were then examined further for associations with clinical variables.

184 **RESULTS**

185 **Supplemental Figure 1.** Bacterial load by qPCR endpoint fluorescence (log transformed) in bronchial washings, airway
186 tissue and basilar parenchyma tissue in IPF (A) and COPD (B) samples. Higher qPCR endpoint fluorescence indicates
187 higher bacterial load. Dashed lines connect samples obtained from the same patient. Pairwise p-values obtained from
188 Wilcoxon tests. *= p<0.05,***= p<0.001 ,****= p<0.0001. IPF, idiopathic pulmonary fibrosis; COPD, chronic obstructive
189 pulmonary disease.

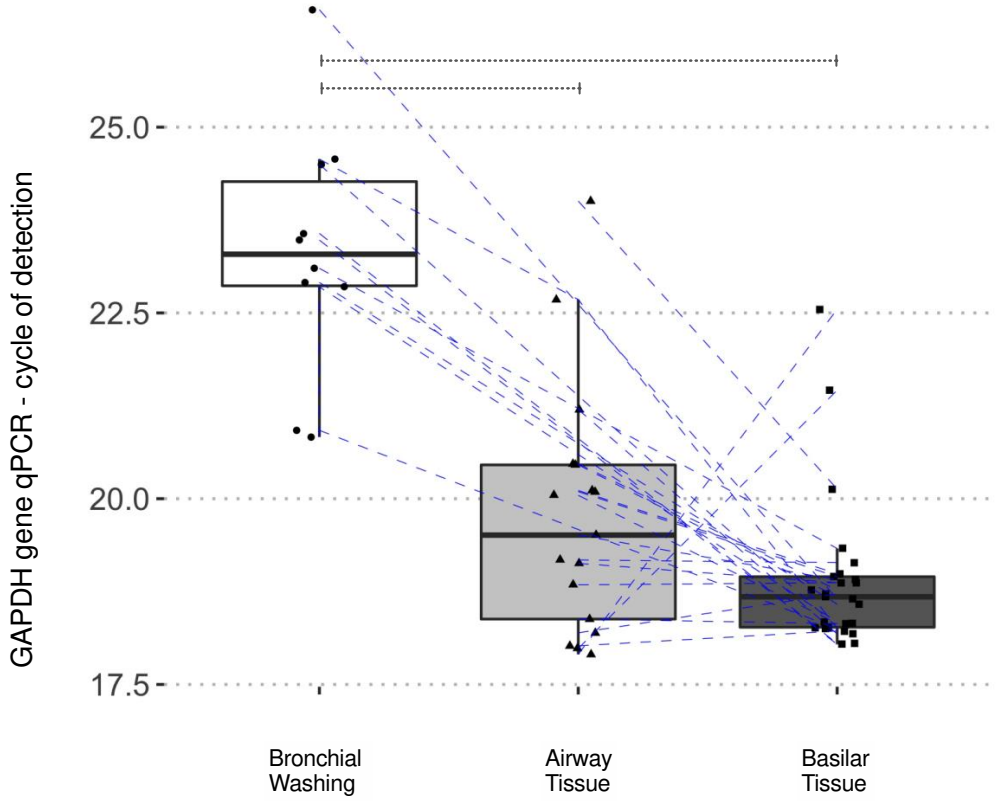


190

191

192

193 **Supplemental Figure 2:** Bronchial washings samples have lower amounts of human DNA (higher cycle of detection for
194 qPCR of the human *GAPDH* gene) compared to airway and parenchymal tissue samples (Wilcoxon p-values < 0.0001).
195 Dashed lines connect samples obtained from the same patient. Higher cycle of detection during qPCR on the y-axis indicates
196 lower amount of target DNA.

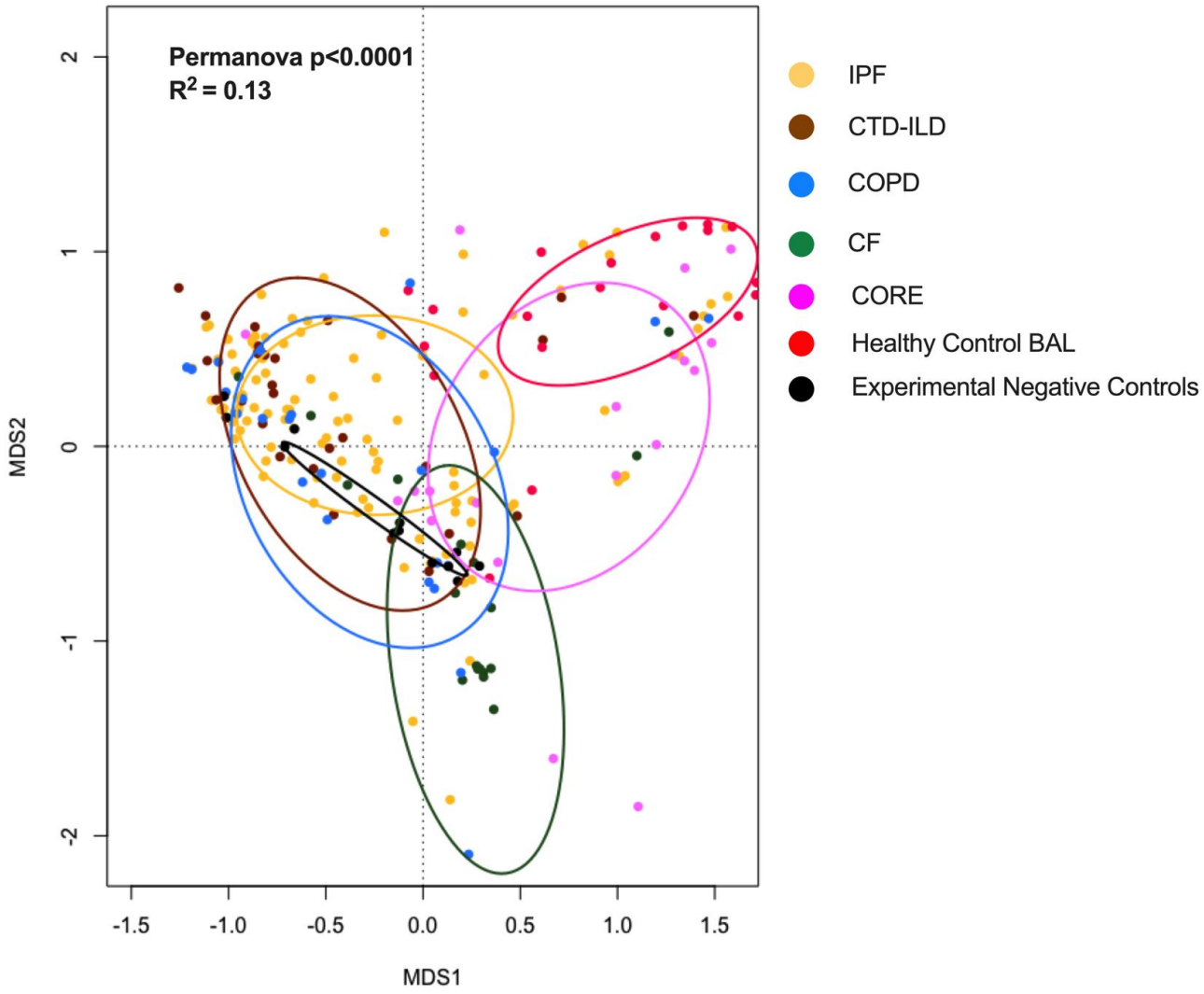


197

198

199

Supplemental Figure 3. Significant taxonomic composition differences by sample and disease type. Beta diversity, as measured by Bray-Curtis dissimilarity in basilar parenchyma tissue of IPF, CTD-ILD, COPD, CF, and CORE lungs, BAL samples from healthy controls, and reagent control samples during DNA extraction and PCR amplification steps. Reagent controls (experimental negative controls) appeared more similar with parenchymal tissue samples of IPF, CTD-ILD and COPD. Permanova; permutation analysis of variance; IPF, idiopathic pulmonary fibrosis; CTD-ILD, connective tissue disease-associated interstitial lung disease; COPD, chronic obstructive pulmonary disease; CF, cystic fibrosis; CORE, Center for Organ Recovery and Education; MDS1, primary NMDS axis; MDS2, secondary NMDS axis; NMDS, non-metric multidimensional scaling.

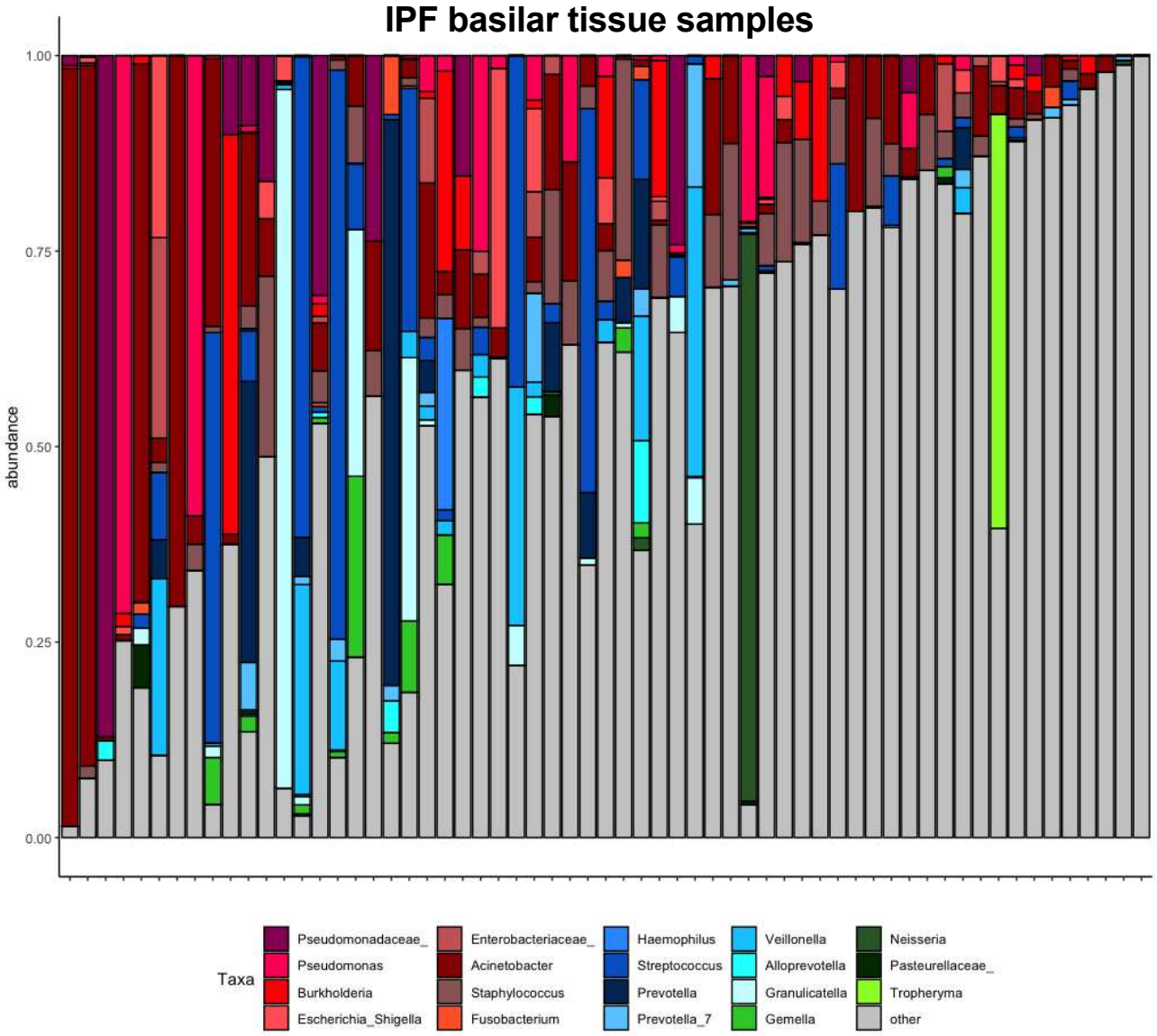


208

209

210 **Supplemental Figure 4.** Taxonomic composition of all basilar parenchyma samples for IPF subjects. IPF taxonomic

211 composition revealed communities with low abundance of typical respiratory microbiota microbes (depicted in blue color).

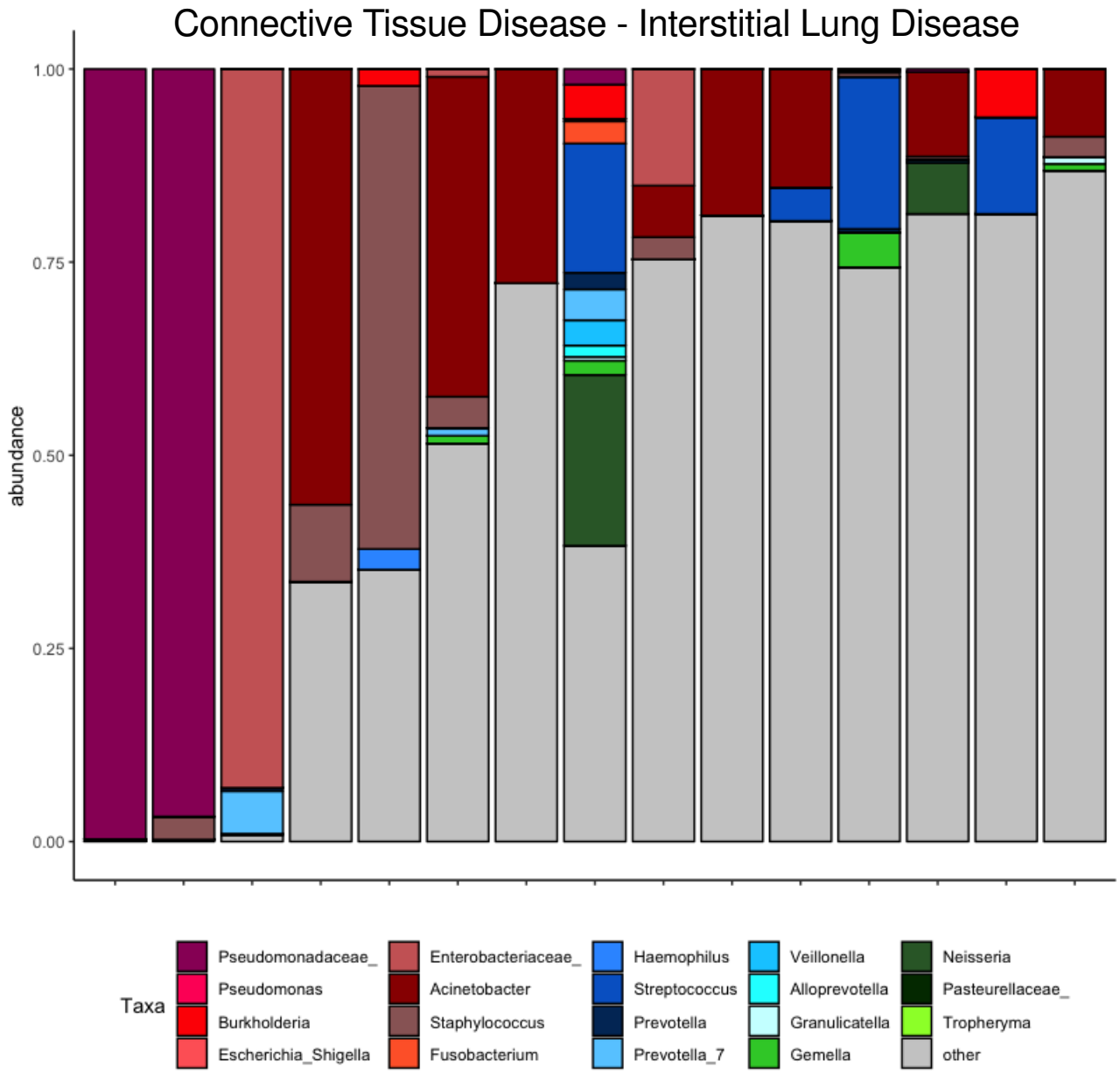

IPF basilar tissue samples
abundance
1.00
0.75
0.50
0.25
0.00
Taxa
Pseudomonadaceae_
Pseudomonas
Burkholderia
Escherichia_Shigella
Enterobacteriaceae_
Acinetobacter
Staphylococcus
Fusobacterium
Haemophilus
Streptococcus
Prevotella
Prevotella_7
Veillonella
Alloprevotella
Granulicatella
Gemella
Neisseria
Pasteurellaceae_
Tropheryma
other

212

213

214

215 **Supplemental Figure 5**. Taxonomic composition of all basilar parenchyma samples for connective tissue disease-
216 associated interstitial lung disease subjects.

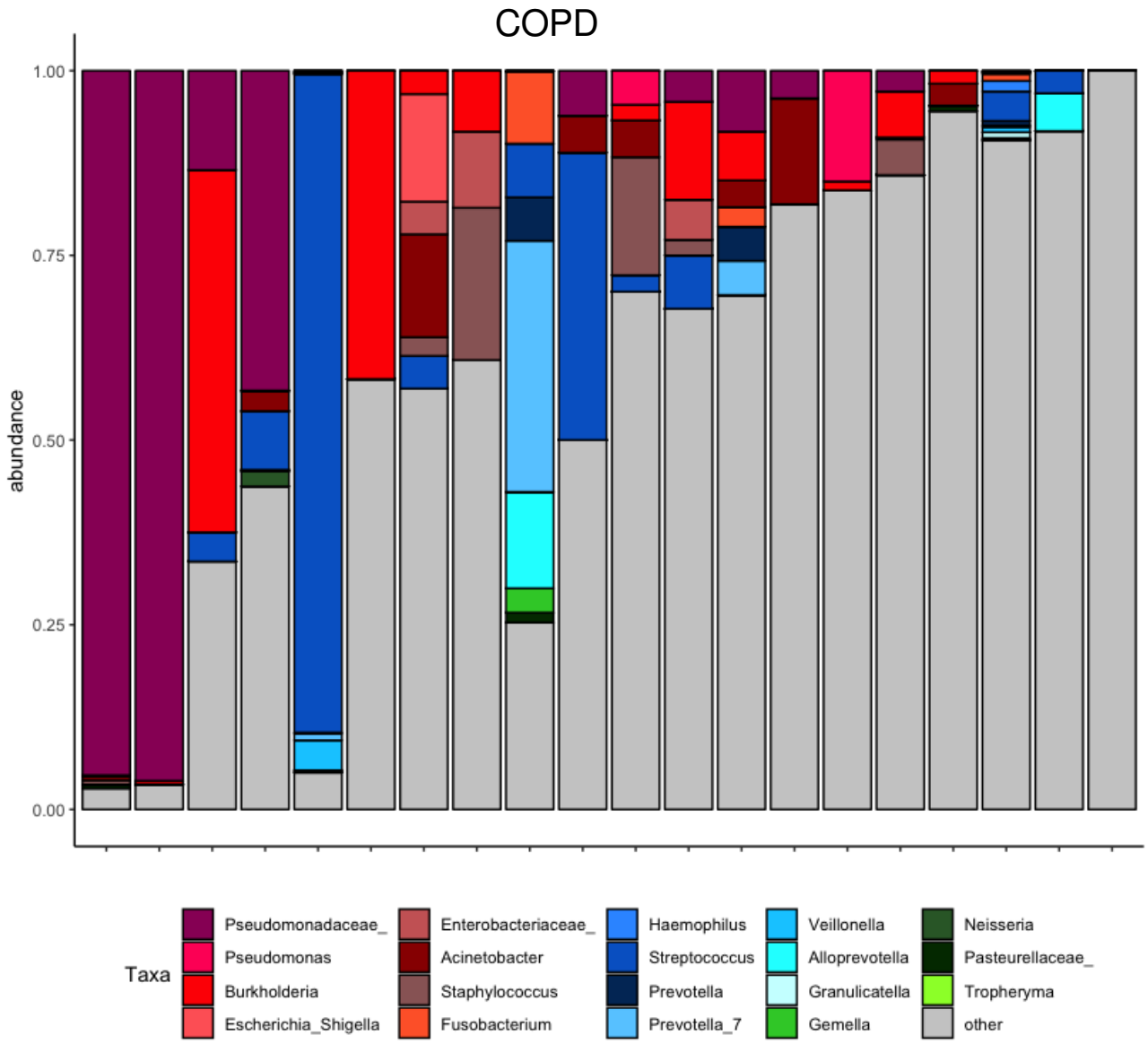


217

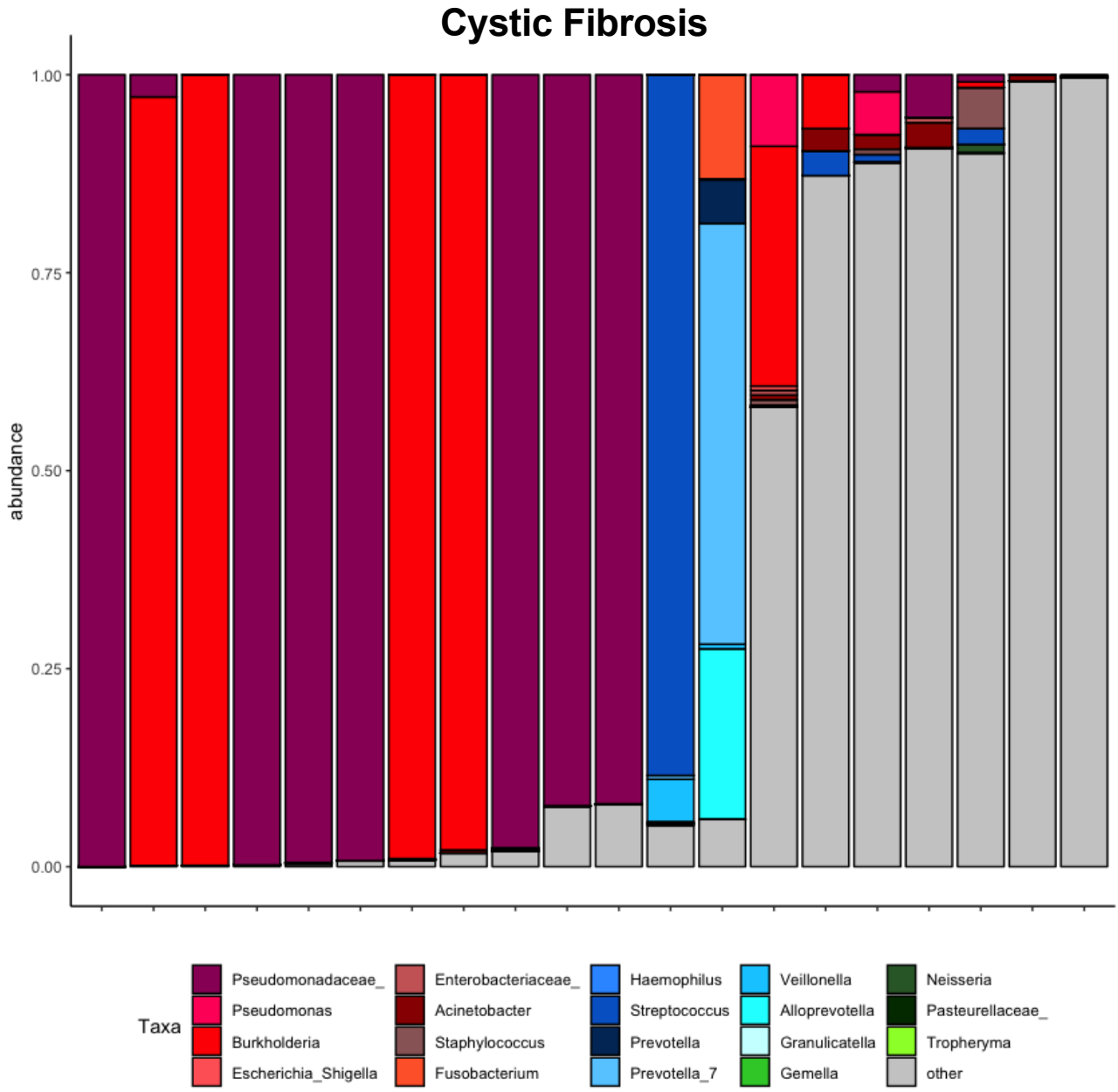
218

219

Supplemental Figure 6. Taxonomic composition of all basilar parenchyma samples for chronic obstructive pulmonary disease subjects.



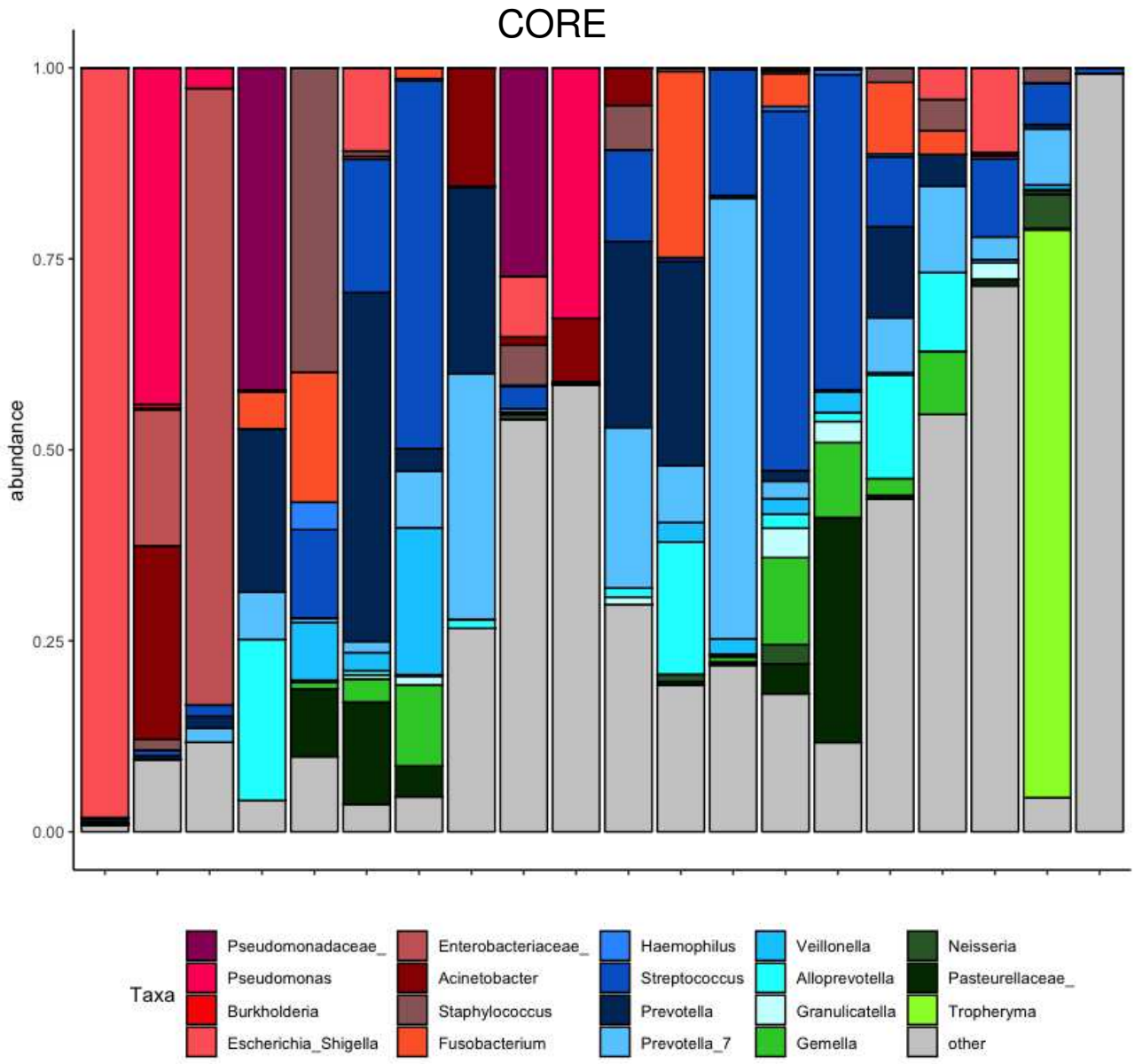
226 **Supplemental Figure 7.** Taxonomic composition of all basilar parenchyma samples for cystic fibrosis subjects.



227

228

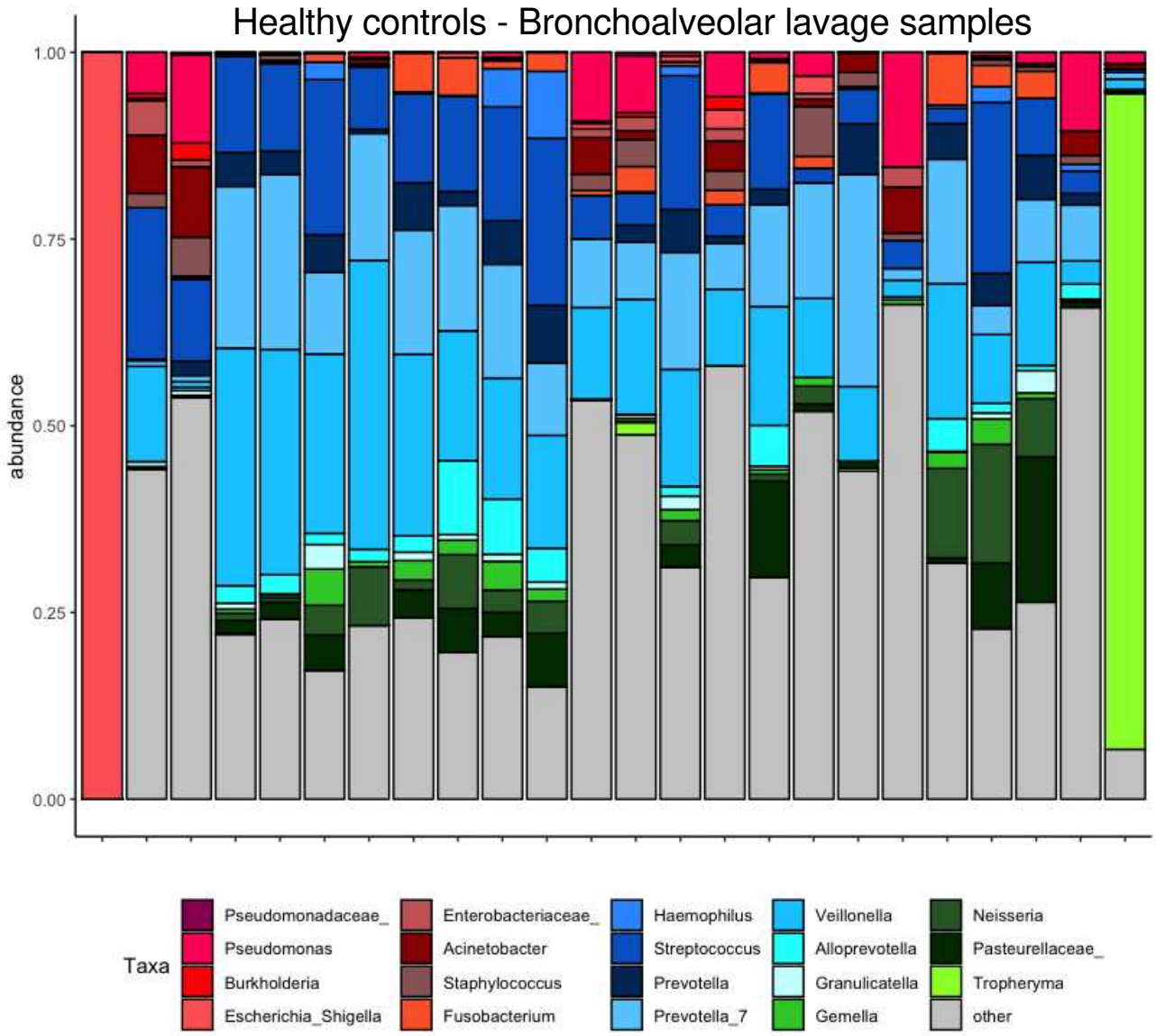
229 **Supplemental Figure 8.** Taxonomic composition of all basilar parenchyma samples for CORE subjects.



230

231

232 **Supplemental Figure 9.** Taxonomic composition of all bronchoalveolar lavage samples for healthy controls.

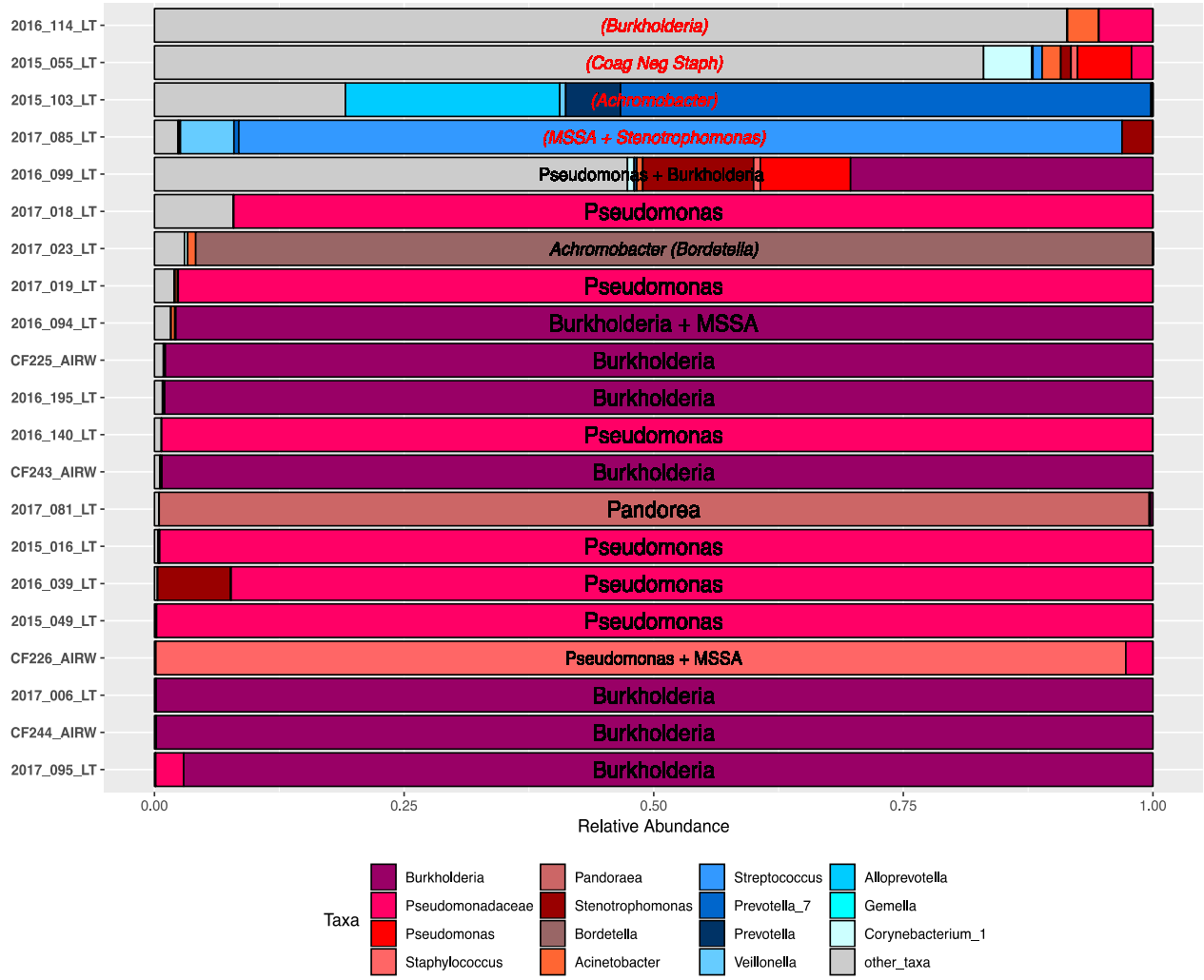


233

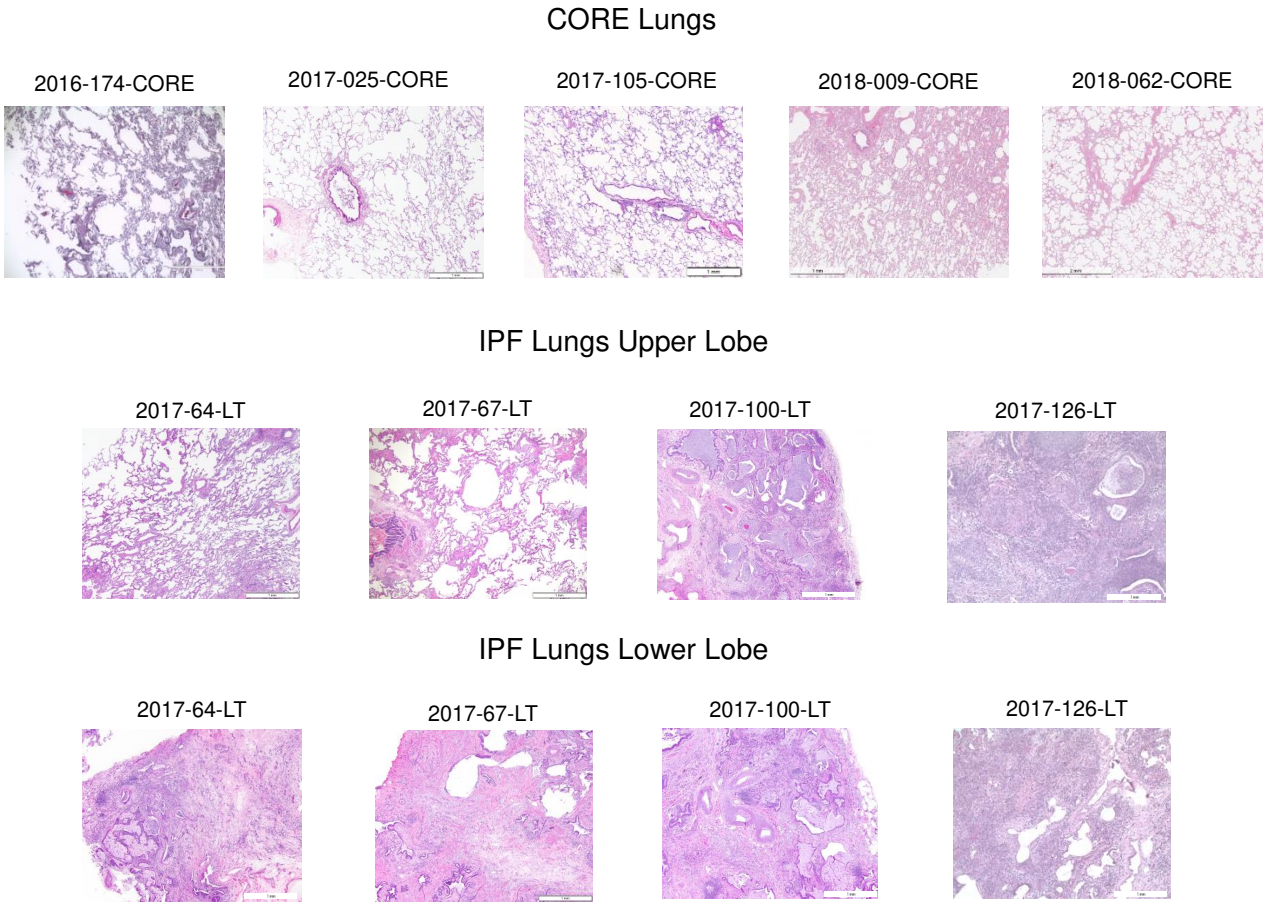
234

235

Supplemental Figure 10. Taxonomic composition of 17 cystic fibrosis basilar parenchyma and 4 cystic fibrosis airway tissue samples, demonstrating very low diversity communities in most samples. In 80% of these cases the dominant taxon by sequencing corresponded to the clinically isolated pathogen identified by airway cultures at the time of transplantation.



Supplemental Figure 11. Explant histopathology in five CORE samples and four matched apical and basal IPF samples.



Thoracic Pathology reports on IPF samples:

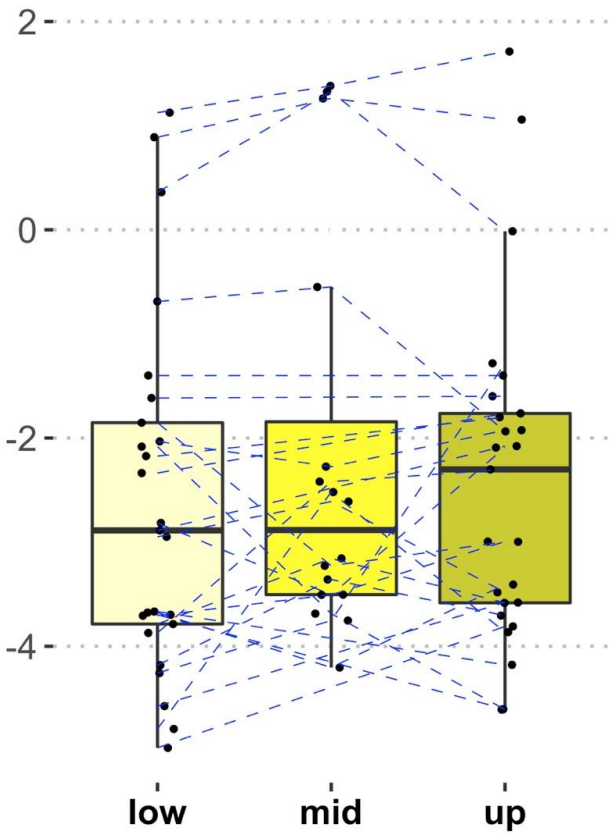
2017-64-LT: usual interstitial pneumonia (UIP), more diffuse in lower lobe than upper lobe. Diffuse alveolar damage in lower lobe suggestive of acute exacerbation.

2017-67-LT: upper lobe with scarring around the airways in a former smoker. No UIP or fibrosis in upper lobe. Lower lobe with advanced diffuse usual interstitial pneumonia, honeycomb cysts, bronchial metaplastic cells, and some smooth muscle metaplasia.

2017-100-LT: complete replacement of the lung architecture by honeycomb cysts and end stage lung disease

2017-126-LT: usual interstitial pneumonia in upper and lower lobes with bronchial metaplasia and mild lymphocytic infiltrate

255 **Supplemental Figure 12.** Bacterial burden by qPCR endpoint fluorescence in upper, middle (or lingula), and lower lobe
256 parenchyma IPF samples. Dashed lines connect lobar samples obtained from the same patient. IPF, idiopathic pulmonary
257 fibrosis.

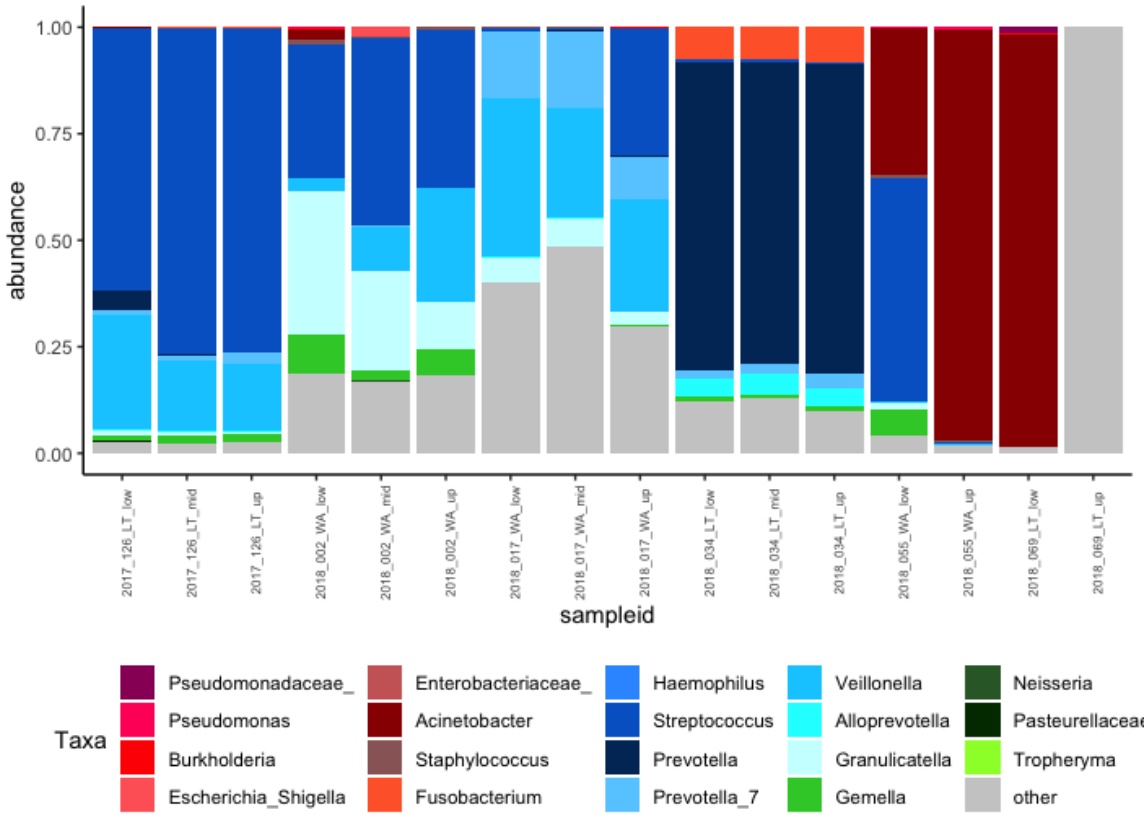


258

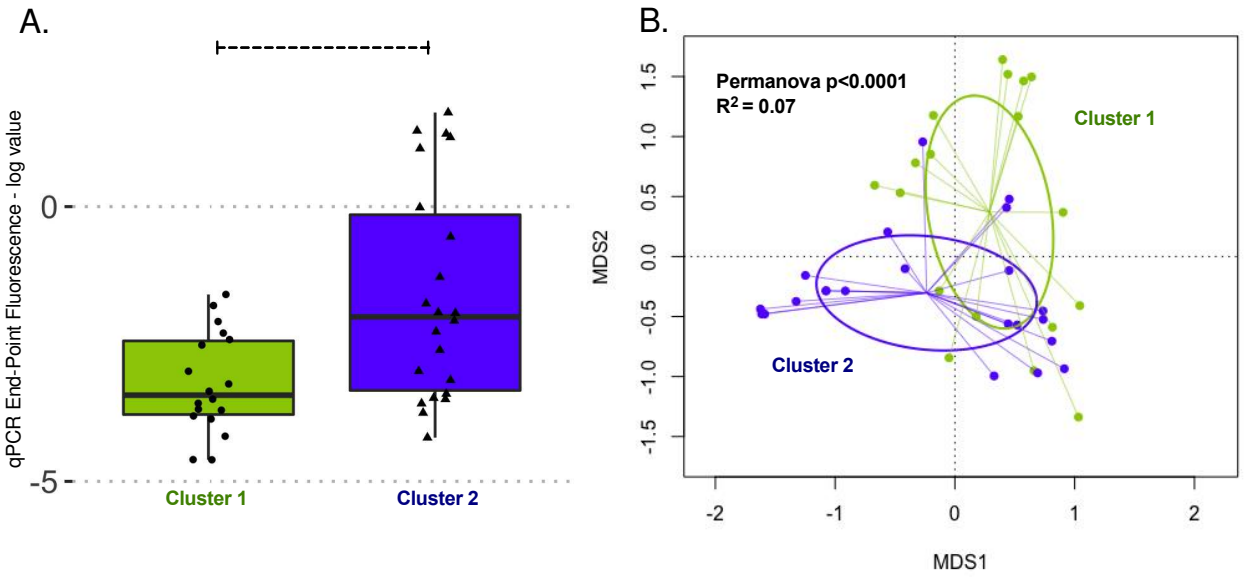
259

260

Supplemental Figure 13. Taxonomic composition by anatomic lobe of select high bacterial load IPF samples. IPF, idiopathic pulmonary fibrosis.



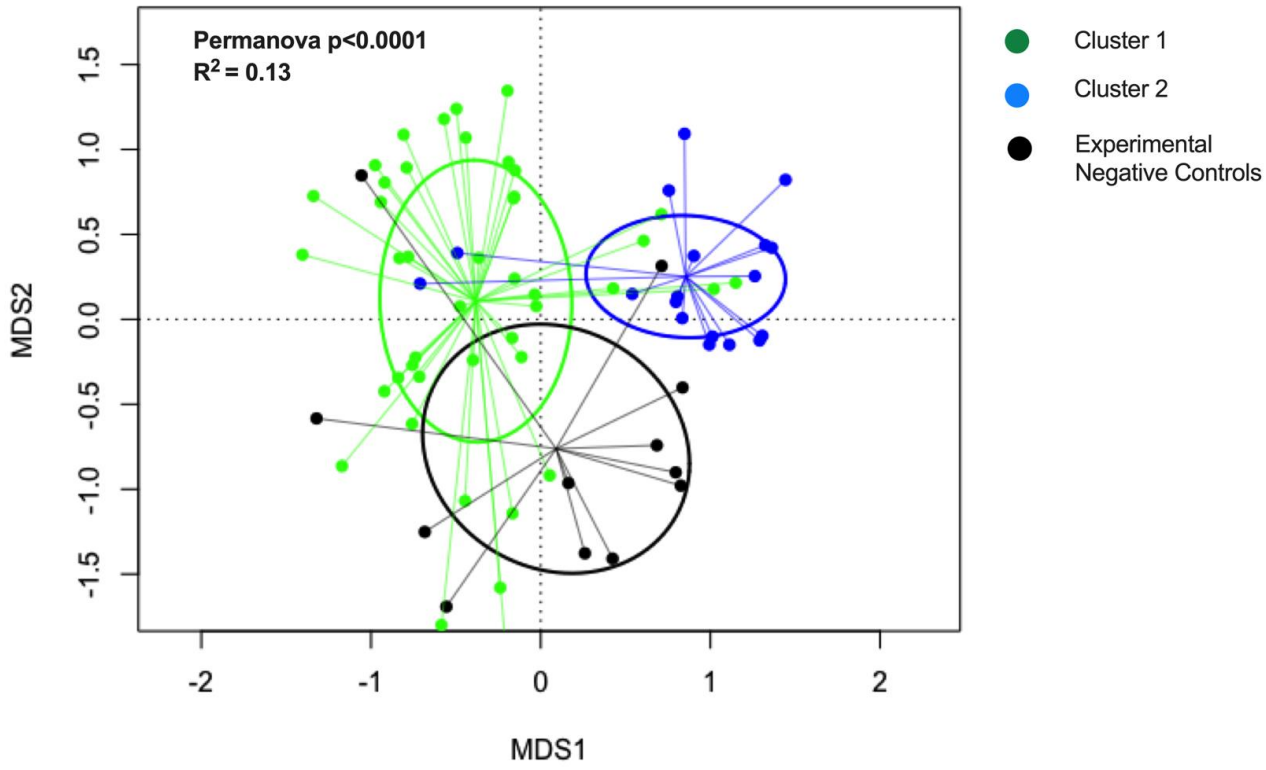
267 **Supplemental Figure 14.** Cluster 2 specimens from upper lobe or middle/lingular lobes had higher bacterial load by
268 qPCR (Panel A) and differential taxonomic composition compared to corresponding cluster 1 specimens in patients with
269 IPF.



270

271 **Supplemental Figure 15.** Experimental negative controls had different taxonomic composition compared to IPF patient

272 DMM clusters 1 and 2. Cluster 1 IPF samples had closer similarity to reagent controls.



273

REFERENCES

1. Kitsios GD, Rojas M, Kass DJ, Fitch A, Sembrat JC, Qin S, et al. Microbiome in lung explants of idiopathic pulmonary fibrosis: a case-control study in patients with end-stage fibrosis. Thorax. 2018 May; 73(5):481-484.
2. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012 Aug; 6(8):1621-1624.
3. Dickson RP, Huffnagle GB, Flaherty KR, White ES, Martinez FJ, Erb-Downward JR, et al. Radiographic Honeycombing and Altered Lung Microbiota in Patients with Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2019 Dec 15; 200(12):1544-1547.
4. Spagnolo P, Molyneaux PL, Bernardinello N, Cocconcelli E, Biondini D, Fracasso F, et al. The Role of the Lung's Microbiome in the Pathogenesis and Progression of Idiopathic Pulmonary Fibrosis. Int J Mol Sci. 2019 Nov 10; 20(22).
5. Pragman AA, Lyu T, Baller JA, Gould TJ, Kelly RF, Reilly CS, et al. The lung tissue microbiota of mild and moderate chronic obstructive pulmonary disease. Microbiome. 2018 Jan 9; 6(1):7.