

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

3  
4 Eleanor B Valenzi<sup>1</sup>, Haopu Yang<sup>2,3,4</sup>, John C Sembrat<sup>1</sup>, Libing Yang<sup>2,4</sup>, Spencer Winters<sup>1,5</sup>, Rachel Nettles<sup>2</sup>, Daniel J Kass<sup>1</sup>,  
5 Shulin Qin<sup>1,2</sup>, Xiaohong Wang<sup>1</sup>, Michael M Myerburg<sup>1</sup>, Barbara Methé<sup>1</sup>, Adam Fitch<sup>1,2</sup>, Jonathan K Alder<sup>1</sup>, Panayiotis V  
6 Benos<sup>3</sup>, Bryan J McVerry<sup>1</sup>, Mauricio Rojas<sup>1</sup>, Alison Morris<sup>1,2</sup>, Georgios D Kitsios<sup>1,2</sup>

7  
8 <sup>1</sup>Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh School of  
9 Medicine and University of Pittsburgh Medical Center, Pittsburgh, PA, USA; <sup>2</sup>Center for Medicine and the Microbiome,  
10 University of Pittsburgh, Pittsburgh, PA, USA; <sup>3</sup>Department of Computational and Systems Biology, University of  
11 Pittsburgh, Pittsburgh, PA, USA; <sup>4</sup>School of Medicine, Tsinghua University, Beijing, China; <sup>5</sup>Bronson Adulte Critical  
12 Care, Kalamazoo, MI, USA

13  
14  
15 Corresponding Author:

16 Georgios D. Kitsios, MD, PhD

17 Assistant Professor of Medicine

18 Division of Pulmonary, Allergy and Critical Care Medicine

19 University of Pittsburgh Medical Center

20 Address: UPMC Montefiore Hospital, NW628, 3459 Fifth Avenue, Pittsburgh, PA 15213

21 Email: [kitsiosg@upmc.edu](mailto:kitsiosg@upmc.edu)

22 Phone: +1-412-383-0127

23  
24 Word Count: 3180

27 **Abstract:**

28  
29 **Background:** Lung microbiota profiles in patients with early idiopathic pulmonary fibrosis (IPF) have been associated  
30 with disease progression; however, the topographic heterogeneity of lung microbiota and their roles in advanced IPF are  
31 unknown.

32  
33 **Methods:** We sampled subpleural tissue from up to three lobes as well as airway-based specimens (bronchial washings  
34 and airway tissue) in patients with IPF, connective tissue disease-associated interstitial lung disease (CTD-ILD), cystic  
35 fibrosis (CF), and chronic obstructive pulmonary disease (COPD) and donor lungs deemed unsuitable for transplant  
36 (controls). We quantified bacterial load and profiled communities by polymerase chain reaction (PCR) amplification and  
37 sequencing of the 16S rRNA gene.

38  
39 **Findings:** Explants from 62 IPF, 15 CTD-ILD, 20 CF, 20 COPD and 20 control patients were included. Airway-based  
40 samples had higher bacterial load compared to distal parenchymal tissue across all patient groups. IPF basilar tissue had  
41 much lower bacterial load compared to CF and control lungs ( $p < 0.001$ ). Among patients with IPF, no differences in  
42 microbial community profiles were found between parenchymal tissue samples from different lobes. With Dirichlet  
43 multinomial models, a cluster of IPF patients (29%) with distinct composition, high bacterial load and low alpha diversity  
44 was identified, exhibiting higher odds for acute exacerbation of IPF or death.

45  
46 **Interpretation:** IPF explants exhibited low biomass in the distal parenchyma of all three lobes with higher bacterial load  
47 in the airways. The discovery of a distinct subgroup of IPF patients with higher bacterial load and worse clinical outcomes  
48 supports investigation of personalized medicine approaches for microbiome-targeted interventions.

49  
50  
51 **Keywords: Idiopathic pulmonary fibrosis, lung microbiome**  
52  
53

## 54 **Key Messages**

- 55 • What is the key question?

56 Bronchoalveolar lavage microbiome profiles in early idiopathic pulmonary fibrosis (IPF) have been associated  
57 with disease progression, but the regional heterogeneity of resident microbiota in end-stage IPF has not been  
58 defined.

- 59 • What is the bottom line?

60 IPF explants demonstrate higher bacterial load in airway compared to parenchymal samples, but no differences in  
61 between apical or basilar parenchymal samples. A subgroup of patients with higher bacterial load and respiratory  
62 pathogen abundance was associated with worse clinical outcomes.

- 63 • Why read on?

64 Patient-specific heterogeneity in the lung microbiome of IPF supports the need for personalized microbiome-  
65 targeted interventions in IPF.

## 58 INTRODUCTION

59 Idiopathic pulmonary fibrosis (IPF) is a devastating age-associated disease, occurring more frequently in smokers  
60 and carriers of host-defense gene mutations.[1, 2] While it is theorized that alveolar injury in a genetically-susceptible  
61 host propagates aberrant repair mechanisms resulting in fibrosis, the exact environmental factors provoking lung injury  
62 have not been defined.[3] Dysbiosis in the respiratory tract has been proposed as a potential mechanism for precipitation  
63 and/or perpetuation of lung injury. The hypothesis that lung microbiota contributes to IPF progression emerged from  
64 epidemiologic observations suggestive of host-microbiome interactions in the respiratory tract, such as the finding that  
65 immunosuppression increases mortality in IPF, whereas antibiotics may offer survival benefit in treatment-tolerant and  
66 adherent IPF patients.[4-6] Three prospective cohort studies in patients with early IPF provided direct evidence for the  
67 lung microbiome hypothesis with the use of culture-independent, bacterial DNA sequencing techniques. Bronchoalveolar  
68 lavage (BAL) fluid from patients with IPF had higher bacterial burden compared to chronic obstructive pulmonary disease  
69 (COPD) patients and healthy controls,[7] and those IPF patients with the highest bacterial burden exhibited worse  
70 outcomes.[7-9] Lung microbial profiles were associated with distinct host transcriptome responses.[10, 11] Furthermore,  
71 in murine model of fibrosis, lung dysbiosis preceded peak lung injury and was associated with worse survival.[8]  
72 Therefore, lung microbiota manipulation with antibiotic therapies has become an attractive target for intervention,  
73 currently assessed by ongoing clinical trials.[12, 13]

74 The role of lung microbiota in later stages of IPF remains unknown. Furthermore, the involvement of microbes  
75 across the respiratory tract in such a disease hallmarked by spatial heterogeneity (subpleural predominance and an  
76 apicobasal gradient of fibrosis) is poorly understood. The BAL samples used by prior studies capture microbiota from  
77 lower generations of the tracheobronchial tree and up to 5% of the alveolar space, but do not provide detailed regional  
78 characterization of microbiota. In a previous study from our group (Microbiome in Lung Explants - MiLEs study), we  
79 examined distal parenchymal tissue from lung explants of patients with end-stage usual interstitial pneumonia (UIP) at the  
80 time of lung transplant or death.[14] Surprisingly, we found exceedingly low bacterial signals by 16S rRNA gene  
81 sequencing, in contrast to explant tissue from cystic fibrosis (CF) or donor lungs. These findings suggest that advanced  
82 honeycombing may represent not only a physiologic dead-space, but also an ecological dead-space (“*microbiome desert*”),  
83 whereas microbiota may primarily colonize the airways and areas of traction bronchiectasis.[14] The discordance between  
84 IPF bacterial burden in BAL versus in distal parenchymal tissue suggests that fibrosis and the resulting honeycombing  
85 could be a maladaptive fibrotic response to airway microbiota.[15]

To gain further understanding of the spatial heterogeneity of lung microbiota in IPF, our current study (MiLEs-2) characterized the regional tissue microbiome across the apicobasal axis of UIP as well as the airways in lung explants from patients with IPF. In particular, we sampled up to three different lobes and airway-based samples in IPF explants, as well as similar samples from explants with connective tissue disease-associated interstitial lung disease (CTD-ILD), CF, COPD, and lungs that had been donated but rejected for transplant (control samples).

## **METHODS**

### **Study design**

MiLEs-2 is a retrospective, case-control study of explanted lung tissue obtained at the time of lung transplantation or rapid autopsy from patients with IPF and diseased controls (CTD-ILD, CF, and COPD). We also obtained control tissue samples from lung donation candidates deemed unsuitable for transplant, via the Center for Organ Recovery and Education (CORE). IPF diagnoses were confirmed according to 2018 clinical practice guidelines, with histopathology results from lung explants (and prior surgical lung biopsies for some cases) reviewed by specialized thoracic pathologists at the University of Pittsburgh Medical Center.[16] We classified patients with IPF as those with an acute exacerbation of IPF (AE-IPF) versus chronic IPF per established criteria.[14, 17] Informed consents for conducting research utilizing the explanted lung specimens were obtained from patients or their designated representatives. The University of Pittsburgh Institutional Review Board and Committee for Oversight of Research and Clinical Training Involving Decedents approved this study.

### **Sample acquisition and processing**

We obtained lung tissue and airway samples in the operating room or autopsy suite per established protocols.[14] We resected a subpleural lower lobe tissue segment, which was further dissected into pieces weighing an average of 45mg under sterile conditions (Figure 1). For a random subset of explants based on logistical feasibility, 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 also collected a bronchial wash specimen (by aspiration of 30mL 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.

24

## 25 **DNA extraction, 16S rRNA qPCR and pyrosequencing**

26 We extracted genomic DNA and performed polymerase chain reaction (PCR) amplification of the V4 hypervariable  
27 region of the 16S rRNA gene.[18] Amplicons of the V4 rRNA bacterial gene subunit were sequenced on the Illumina  
28 MiSeq platform and quantified by qPCR.[14] We also quantified the human genomic DNA present in each sample by  
29 qPCR of the human Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene. We conducted a series of experiments  
30 with lung explant tissue to identify optimal sample type (whole tissue vs. swab) and ruled out the presence of PCR  
31 inhibitors (Supplement).

32

## 33 **Statistical analysis**

34 From derived 16S sequences, we applied a custom pipeline for Operational Taxonomic Units (OTUs-taxa)  
35 classification (Supplement) and performed analyses at genus level. We calculated descriptive statistics of clinical  
36 characteristics and performed nonparametric comparisons using the R software (v.3.5.1). Ecological analyses of alpha  
37 diversity (Shannon index) and beta diversity (Bray-Curtis index with permutational analysis of variance [permanova] at  
38 1000 permutations) were conducted using the R *vegan* package and visualized with principal coordinates analyses (PCoA)  
39 plots. For comparisons with lower respiratory communities in healthy controls, we utilized 16S sequencing data from a  
40 previous study that had analyzed BAL specimens from healthy volunteers.[19] For 30 IPF patients who had survived to  
41 lung transplantation and had whole genome sequencing performed in genomic DNA extracted from blood samples, we  
42 obtained genotypes for the promoter single nucleotide polymorphism (SNP) rs35705950 of the *MUC5B* gene. To  
43 agnostically examine for distinct clusters of microbial composition (“meta-communities”) in our IPF basilar tissue  
44 samples (n=62), we applied unsupervised Dirichlet Multinomial Models (DMM) with Laplace approximations to define  
45 the optimal number of clusters in our dataset.[20] We then examined for associations of microbiome variables (bacterial  
46 load [log-transformed end-point fluorescence of qPCR assay], alpha diversity, beta-diversity and DMM clusters) with  
47 clinical variables (disease classification, diagnosis of AE-IPF, lung transplant vs. death outcome and *MUC5B* genotypes).

48

## 49 **RESULTS**

### 50 **Study population**

51 We analyzed basilar lung explant tissue specimens from 62 IPF patients, 15 CTD-ILD patients, 20 CF patients, 20  
52 COPD patients and 20 control lungs (CORE). Additional specimens (middle/upper lobe tissue or airway-based samples)  
53 were available from a subset of explants (Figure 1). Comparisons of clinical characteristics showed that a higher  
54 percentage of IPF patients were male (81%) (Table 1), and that IPF patients had severely decreased forced vital capacity  
55 (FVC) (median 41%) and diffusing capacity for carbon monoxide (DLCO) (median 30%), reflecting their end-stage status  
56 at the time of lung transplantation (76%) or rapid autopsy (24%). AE-IPF was clinically diagnosed in 35% of patients with  
57 IPF and was strongly associated with the finding of diffuse alveolar damage on explant histopathology (odds ratio [OR] =  
58 9.8, 95% confidence interval [CI]: 2.5-44.3,  $p < 0.0001$ ).

**Table 1. Clinical data for patients included in the MiLEs-2 cohort.** Data presented as medians (Interquartile range [IQR]) or N (%).

Variable	IPF	CF	COPD	CTD-ILD
<b>N</b>	62	20	20	15
<b>Age (years), median (IQR)</b>	64.5 [60.0, 69.0]	29.0 [23.8, 35.0]	62.0 [59.0, 65.0]	63.0 [54.0, 68.5]
<b>Male, n (%)</b>	50 (81%)	7 (35%)	9 (45%)	7 (47%)
<b>Ever-smokers, n (%)</b>	42 (68%)	0 (0%)	20 (100%)	10 (67%)
<b>Total pack-years, median (IQR)</b>	10.2 [0.0, 30.0]	0.0 [0.0, 0.0]	57.5 [40.0, 65.5]	10.0 [0.0, 26.8]
<b>FEV1 (L), median (IQR)</b>	1.5 [1.2, 2.0]	0.7 [0.6, 0.9]	0.6 [0.5, 0.7]	1.4 [1.2, 1.6]
<b>FEV1 predicted %, median (IQR)</b>	53.0 [43.0, 67.0]	22.0 [18.8, 24.2]	25.0 [22.0, 30.2]	52.0 [45.5, 68.0]
<b>FVC (L), median (IQR)</b>	1.8 [1.4, 2.5]	1.4 [1.1, 1.8]	2.0 [1.8, 2.3]	1.7 [1.5, 2.1]
<b>FVC predicted %, median (IQR)</b>	41.0 [37.0, 58.0]	32.5 [28.0, 41.2]	58.5 [48.0, 66.0]	47.0 [39.2, 60.2]
<b>DLCO (ml), median (IQR)</b>	6.4 [4.7, 9.3]	12.8 [11.8, 14.6]	6.8 [5.1, 9.9]	5.1 [4.6, 7.5]
<b>DLCO predicted %, median (IQR)</b>	30.0 [22.0, 36.0]	58.0 [49.0, 67.0]	32.5 [26.8, 51.0]	25.0 [20.5, 28.5]
<b>Mean PAP (mmHg), median (IQR)</b>	24.0 [20.0, 28.0]	25.0 [21.0, 33.0]	21.5 [19.0, 27.0]	24.5 [21.0, 38.8]
<b>GERD<sup>#</sup>, n (%)</b>	48 (77%)	19 (95%)	12 (60%)	15 (100%)
<b>Systemic steroids, n (%)</b>	15 (24%)	9 (45%)	6 (30%)	9 (60%)
<b>Inhaled corticosteroids, n (%)</b>	11 (18%)	8 (40%)	19 (95%)	5 (33%)
<b>Specific therapies:</b>				
<b>Pirfenidone, n (%)</b>	10 (16%)	0 (0%)	0 (0%)	1 (7%)
<b>Nintedanib, n (%)</b>	9 (15%)	0 (0%)	0 (0%)	0 (0%)
<b>Other immunomodulator, n (%)</b>	15 (24%)	0 (0%)	1 (5.0%)	9 (60%)
<b>BAL + culture or recipient tissue culture*, n (%)</b>	7 (12%) <sup>§</sup>	20 (100%)	7 (35%)	2 (13%) <sup>§</sup>
<b>Diffuse alveolar damage on explant pathology, n (%)</b>	21 (36%) <sup>§</sup>	0 (0%)	0 (0%)	5 (33%)
<b>Antibiotics in past 3 months, n (%)</b>	25 (40%)	20 (100%)	9 (45%)	9 (60%)
<b>Lung transplant, n (%)</b>	47 (76%)	20 (100%)	20 (100%)	10 (67%)

<sup>#</sup> 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 the institutional transplant protocol.

<sup>§</sup> Percentage calculated with denominator of all patients with available data (patients with unavailable data for each variable were removed from calculations).

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.



## 59 **Airway-based samples have higher bacterial load than corresponding parenchymal tissue**

70 By 16S qPCR across all available samples, bronchial washings had higher bacterial load than corresponding  
71 airway tissue, which in turn had higher bacterial load than their counterparts in basilar parenchymal tissue (Figure 2A).  
72 This bacterial load gradient from the airways to the distal parenchyma was present within both IPF and COPD lungs  
73 (Figure S1). As for the amount of human DNA available in each sample (quantified by *GAPDH* qPCR), a reverse gradient  
74 was observed compared to bacterial DNA load, with tissue samples having much higher human DNA content compared to  
75 bronchial washings ( $p < 0.0001$ ) (Figure S2). Sample type was further associated with significant difference in alpha  
76 diversity (Shannon index, Figure 2B), with airway tissue samples having the lowest alpha diversity compared to the other  
77 types of samples, as well as significantly different taxonomic composition by beta-diversity comparisons (Figure 2C).  
78 Overall, airway-based samples captured higher microbial biomass with lower human DNA abundance, whereas distal  
79 parenchymal tissue (mainly from IPF and COPD explants) had lower biomass and higher amount of human DNA present.

## 31 **IPF basilar tissue samples have low bacterial load**

32 After demonstration of the airway-parenchyma gradient of bacterial load, we then examined for differences in  
33 bacterial load and community profiles in basilar tissue parenchymal specimens across the different disease states. By  
34 qPCR, IPF and COPD basilar tissue samples had much lower bacterial load (approximately 40-fold less bacterial DNA)  
35 compared to basilar parenchymal tissue from CF patients or CORE lungs ( $p < 0.0001$ , Figure 3A). Notably, parenchymal  
36 samples in IPF and COPD have extensive anatomic destruction and physiologic dead-spaces (due to advanced  
37 honeycombing and emphysema, respectively), suggesting the presence of “ecological dead-spaces” for respiratory  
38 microbiota. These observations agree with our prior ones in a smaller cohort of end-stage IPF explants.[14] Examination  
39 of alpha diversity provided a reciprocal image of bacterial load: IPF samples had much higher Shannon index compared to  
40 CF samples ( $p < 0.0001$ , Figure 3B). The pattern of low bacterial burden with high alpha diversity in IPF samples is  
41 strongly suggestive of experimental contamination due to low signal/noise ratio.[21] Significant taxonomic composition  
42 differences by disease were detected with beta-diversity comparisons (Figure 3C).

43 To further interpret the low bacterial signal from IPF basilar tissue samples, we examined the specific taxonomic  
44 composition of all samples available (Figure S3). We first focused on CF samples, which consisted of low diversity  
45 communities with a high abundance of typical pathogenic taxa (e.g. *Pseudomonas* or *Burkholderia* genera). Among 21  
46 samples from CF patients, 16S sequencing of tissue samples demonstrated dominance by one or two genera that

corresponded to the clinically isolated pathogens identified by airway cultures in 80% of samples (Figure S4). The high concordance with clinical isolates established that whenever there is high bacterial load, even basilar tissue samples can reliably capture the bacteria present, despite their smaller biomass compared to corresponding airway-based samples. Importantly, for the majority of IPF (and COPD) parenchymal samples, typical respiratory bacteria (commensal or pathogenic) were conspicuously absent (Figure S3).

### **Bacterial load and composition in IPF tissue samples is associated with clinical outcomes.**

We noted a distinct subgroup of IPF samples with high bacterial load by qPCR, in the range of bacterial loads observed for CF samples. IPF samples within the highest bacterial load tertile had the lowest alpha diversity (Figure 4A) and were taxonomically distinct from samples from the other two tertiles (Figure 4B). Upon retrospective review of associated clinical variables, we found that patients receiving systemic antibiotics within the last three months had higher bacterial load compared to those not receiving antibiotics ( $p=0.04$ ) (Figure 4C), possibly signaling a recent clinical deterioration that prompted an antibiotic prescription. Patients diagnosed with AE-IPF also had higher bacterial load ( $p=0.03$ ) and were less likely to have survived to undergo lung transplantation ( $p=0.02$ ) (Figure 4C).

We then agnostically examined for the presence of distinct microbial composition clusters in basilar IPF tissue samples by DMM. Two clusters offered the best model fit. Cluster 1 (71% of samples) consisted of communities with low bacterial load, high alpha diversity and abundance of bacteria that are not typical members of the respiratory microbiome (e.g. *Bradyrhizobium* and *Methylobacterium*), which likely represent experimental contamination (Figure 5). Cluster 2 (29%) demonstrated high abundance of typical members of the microbiome of respiratory tract (*Streptococcus*, *Veillonella* or *Prevotella* genera), and communities with higher bacterial load ( $p<0.001$ ) and lower alpha diversity ( $p<0.05$ ) compared to cluster 1 (Figure 5). Membership in cluster 2 was associated with higher odds ratio for diagnosis of AE-IPF (OR=3.3 [0.9-12.2],  $p=0.04$ ) and recent antibiotic prescription (OR=6.6 [1.8-29.5],  $p=0.002$ ) and lower odds ratio for survival to lung transplantation (OR=0.16 [0.04-0.66],  $p=0.007$ ). We did not find any association between bacterial load or DMM clusters by *MUC5B* genotypes in the subset of IPF patients who underwent lung transplantation.

### **Bacterial load and diversity do not differ along the apicobasal gradient of fibrosis**

Examination of bacterial communities in IPF lungs along the typical apicobasal gradient of fibrosis seen in IPF revealed no difference in bacterial load by qPCR between apical and basilar tissue (Figure S5). Upon closer inspection of

25 bacterial distribution, the majority of patient samples exhibited consistently low bacterial signals across all three lobes.  
26 Additionally, the taxonomic composition of communities was indistinguishable between each lobe (Permanova p-value  
27 for lobe non-significant). However, a distinct subgroup of six IPF lungs with high bacterial load (above 75<sup>th</sup> percentile)  
28 was identified not only in the basilar but also in the middle/lingula and upper lobe samples (Figure S5). In these six cases,  
29 similar taxa were highly abundant in samples from all two or three lobes available, further underlying the notion for the  
30 presence of a patient-specific rather than a lobe-specific microbiome in IPF (Figure S6).

## 32 DISCUSSION

33 In this study, we identified that basilar lung parenchyma samples have consistently decreased bacterial load  
34 compared to airway-based samples. End-stage parenchymal destruction as seen with advanced IPF and COPD represents  
35 a “microbiome desert” with the absence of identifiable respiratory bacterial communities in most cases. We also  
36 determined that while bacterial communities in IPF do not differ across the apicobasal gradient of fibrosis, a distinct  
37 subgroup of IPF patients demonstrated high bacterial load with abundance of typical respiratory microbiota, as well as  
38 worse clinical outcomes including acute exacerbations and death. Amongst the subgroup of IPF patients receiving a lung  
39 transplant, no association was detected between *MUC5B* genotypes and bacterial load, nor with DMM cluster assignment.  
40 These novel data generate further hypotheses regarding the potential of personalized microbiome-targeted therapies in IPF.

41 The spatial and temporal heterogeneity of UIP in IPF classically develop along an apicobasal gradient with  
42 fibrosis, with end-stage honeycombing appearing mostly in the subpleural lower lobes compared to the upper lobes.[16]  
43 Recent studies examining whether BAL-sampled microbiota vary among IPF patients with different radiographic features  
44 (e.g. honeycombing or traction bronchiectasis) did not reveal any significant associations.[9, 22] Similarly, we did not  
45 identify significant differences in the microbial communities of apical vs. basilar parenchymal samples, with the latter  
46 having greater extent of radiographic honeycombing. Instead, we found evidence of patient-specific, rather than lobe-  
47 specific microbiome in IPF. If local replication within the lower respiratory tract was a predominant source of origin for  
48 the lung microbiome, one might expect to see greater diversity across lobes. Our findings may indicate that  
49 microaspiration and dispersion of microbiota in the lower airways and parenchyma is the primary shaping force of  
50 microbial communities in IPF.

51 Our study utilized lung explants for excision of tissue specimens and acquisition of bronchial washings rather  
52 than obtaining trans-oral bronchoscopic samples as in prior investigations. Explanted tissue not only allowed us to

53 examine the bacterial burden and communities in patients with advanced IPF that were unable to undergo research  
54 bronchoscopy, but also obviated the concern for bronchoscopic contamination by upper airway bacteria. Previous  
55 prospective analyses of IPF lung microbiota have focused on BAL within early-stage IPF patients, which may not be  
56 reflective of the microbiota present in those with the most advanced disease. Unlike previous studies,[7, 23] we did not  
57 identify an association between *Streptococcus* or *Staphylococcus* genus abundance and disease progression within our  
58 advanced IPF population. Instead, we utilized an unsupervised clustering approach to identify whether distinct meta-  
59 communities of lung microbiota existed in the IPF cohort. With agnostic DMM, we identified that 71% of samples had  
60 low bacterial burden and high alpha diversity, with several genera not typically associated with the respiratory tract,  
61 suggesting their possible origin from experimental contamination rather than true biologic presence. The remaining 29%  
62 of samples exhibited higher bacterial loads with typical respiratory microbiota abundance (*Streptococcus*, *Veillonella*,  
63 *Prevotella* genera). Notably, this cluster of patients with high loads of respiratory microbiota was associated with worse  
64 clinical outcomes, suggesting that high bacterial burden may impact clinical decline not only in early disease shown in  
65 previous studies,[7-9] but also in end-stage IPF.

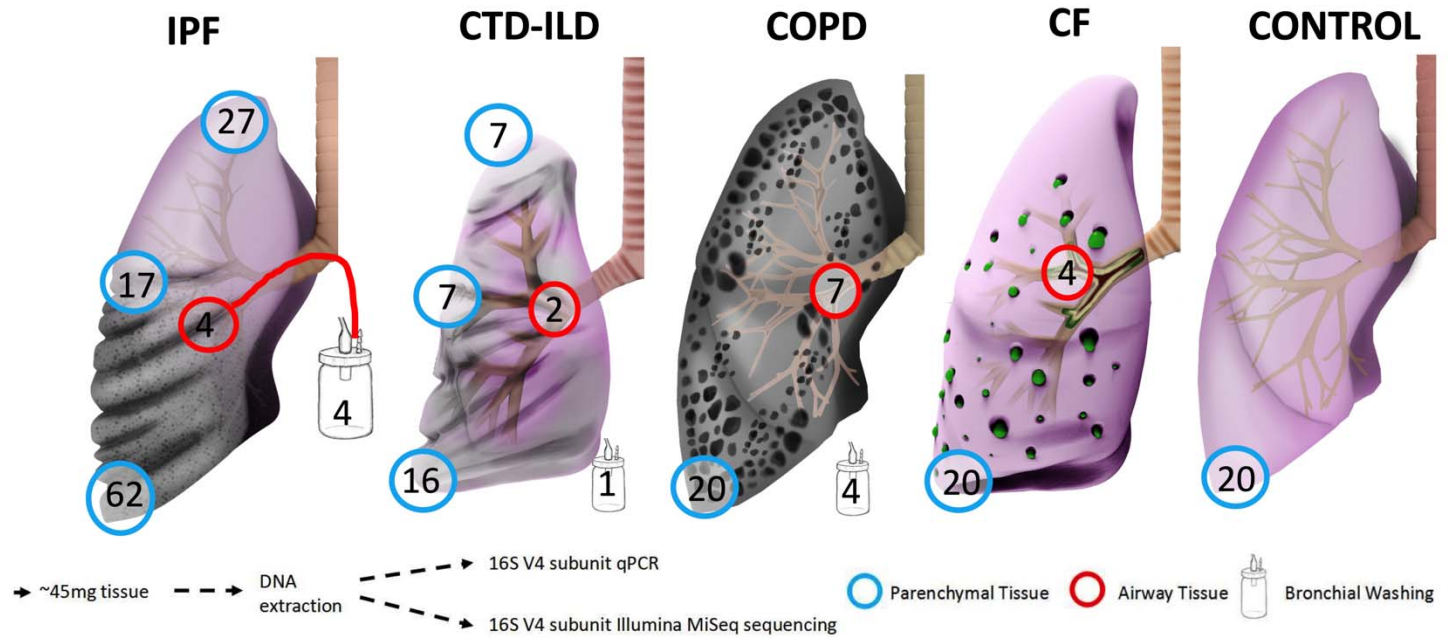
66 Lung bacterial communities resemble those of the oral cavity in healthy individuals, though certain bacteria are  
67 significantly more abundant in the lung.[19] As oral samples were not collected from study patients, we were unable to  
68 ascertain the correlation between oral and lung microbiota in these patients. Given the high prevalence of  
69 gastroesophageal reflux disease (GERD) in IPF and the subsequent probable repeated micro-aspirations, higher  
70 correlation of lung and gut microbiome might be expected in IPF.[24, 25] While much of the association between GERD  
71 and IPF may be related to smoking, GERD has been implicated in many other respiratory diseases as well including  
72 asthma, COPD, bronchiolitis obliterans, and organizing pneumonia.[26-30] GERD was clinically present in 77% of our  
73 IPF patients (as defined by clinical history, esophagram and upper endoscopy results, and prescription for proton-pump  
74 inhibitor or H2 blocker), though the rate of silent micro-aspiration could be even higher. Indeed, increased micro-  
75 aspiration could account for the higher bacterial burden of BAL fluid in IPF compared to healthy controls previously  
76 identified. Further efforts to treat GERD, and more importantly prevent micro-aspiration events in IPF (rather than acid  
77 controlled focused therapy alone), may have a vital role in preventing the dysbiosis associated with worse clinical  
78 outcomes. However, our analyses by clinically-defined GERD did not reveal significant differences in microbial  
79 communities, similar to previous observations.[9]

30 Our study was limited by its retrospective design and utilization of a convenience dataset rather than a prospective  
31 cohort. All IPF samples are from patients with end-stage disease receiving care at a tertiary medical center, and thus may  
32 not reflect the general IPF population. Lung tissue remains a challenging biospecimen for microbiome work due to its low  
33 biomass relative to the amount of human DNA present, contamination risks, and readouts at levels near the detection limit  
34 of assays used. While we examined the bacterial composition of more anatomical regions compared to previous studies,  
35 the current lack of viral and fungal data limits our study from presenting a comprehensive representation of the  
36 microbiome in IPF. Noted associations with clinical outcomes, including the decreased likelihood of receiving a lung  
37 transplant and increased incidence of AE-IPF in patients with higher bacterial load, are hypothesis-generating only, given  
38 their retrospective nature. How to identify such patients while balancing the potential benefits of targeted antimicrobials  
39 with the risks of overtreatment and diagnostic procedures in patients with advanced lung disease remains unclear.

40 In summary, our analysis utilizes the distinct capacity of culture-independent sequencing techniques to discern  
41 that end-stage IPF lungs have limited biomass bacterial communities, without evidence of spatial heterogeneity across the  
42 apicobasal gradient of fibrosis. A subpopulation of patients with higher bacterial load and taxonomic species dominated  
43 by typical respiratory pathogens are more likely to die than undergo lung transplantation and may represent a crucial  
44 population towards whom microbiome-targeted interventions ought to be considered. The ongoing development of rapid,  
45 culture-independent methods for profiling microbiota holds the promise for personalized medicine approaches in IPF.

08 **Figure Legends:**

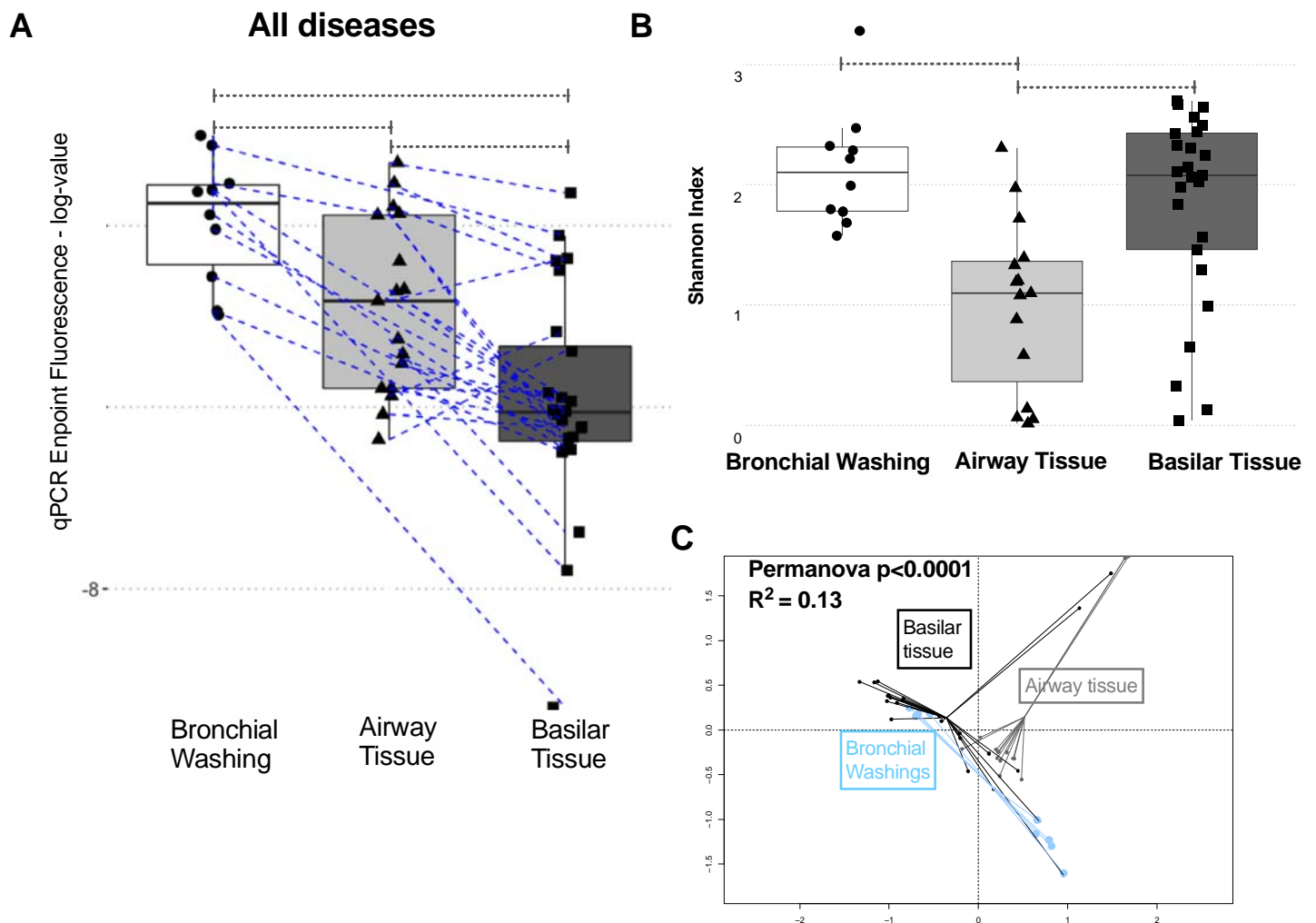
09 **Figure 1.** Explanted lung tissue and bronchial washing samples included in the study, depicted by disease and tissue  
10 anatomic location. Numbers in the circles indicate available number of samples from each anatomic site or type. IPF,  
11 idiopathic pulmonary fibrosis; CTD-ILD, connective tissue disease-associated interstitial lung disease; COPD, chronic  
12 obstructive pulmonary disease; CF, cystic fibrosis.



16

17 **Figure 2. Airway-based samples have higher bacterial load compared to parenchymal tissue samples.** (A) Bacterial  
18 load by qPCR endpoint fluorescence (log-transformed) in bronchial washings, airway tissue and basilar parenchyma tissue  
19 samples for all diseased samples available (IPF, CTD-ILD, COPD and CF). (B) Airway tissue samples had the lowest  
20 alpha diversity (Shannon Index) compared to bronchial washings or basilar tissue samples. (C) Significant compositional  
21 differences by beta diversity (Bray-Curtis dissimilarity index) comparisons visualized with Principal Coordinates Analysis  
22 plot between bronchial washings, airway tissue, and basilar parenchyma tissue samples. Pairwise p-values obtained from  
23 Wilcoxon tests. \*= p<0.05, \*\*\*= p<0.001, \*\*\*\*= p<0.0001.

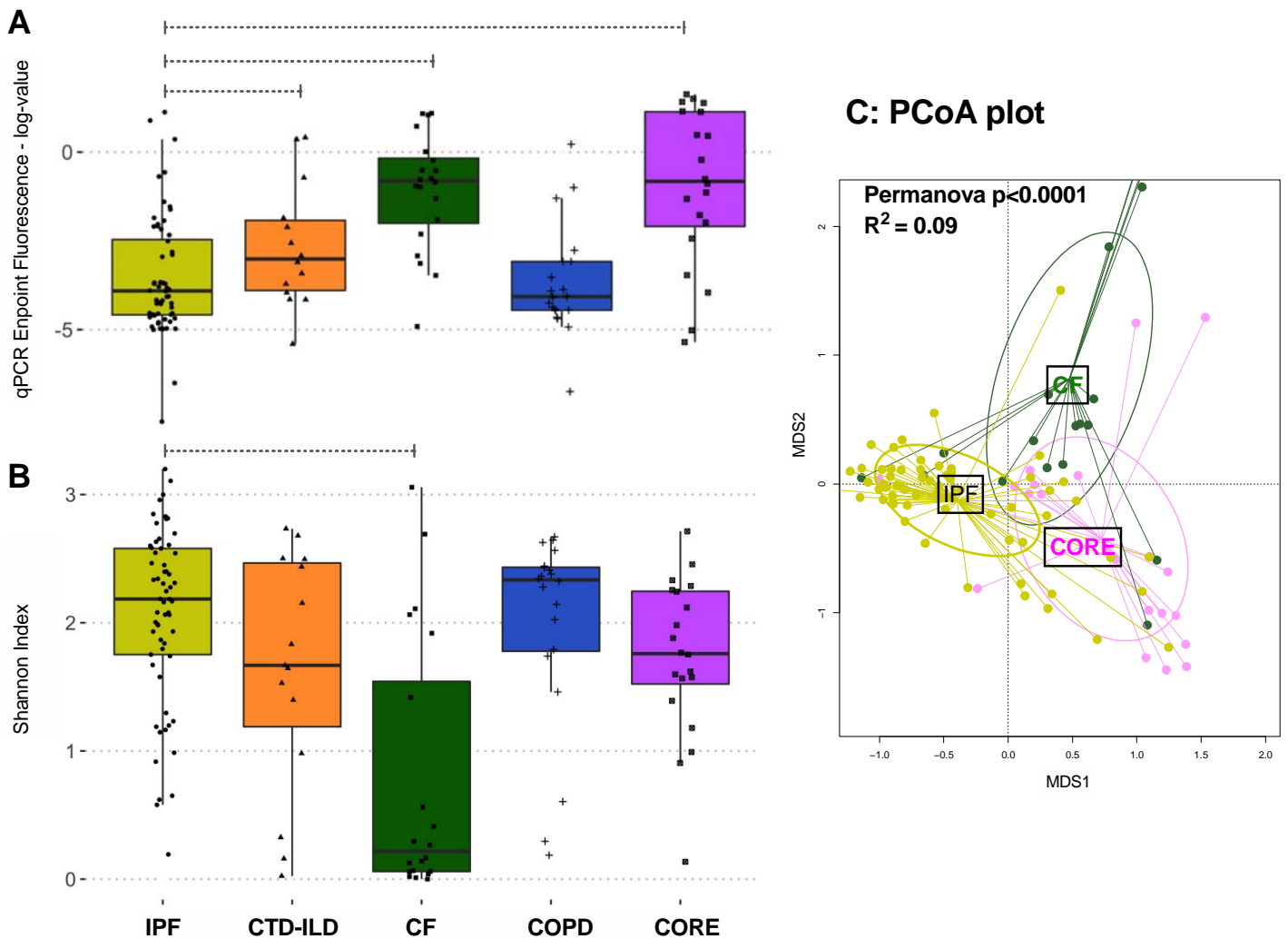
14



15



16 **Figure 3.** (A) Bacterial burden by qPCR endpoint fluorescence in basilar parenchyma tissue across all diseases and  
17 controls. (B) Alpha diversity, as measured by Shannon index, of basilar parenchyma tissue across all diseases and controls.  
18 (C) Beta diversity, as measured by Bray-Curtis dissimilarity in basilar parenchyma tissue of IPF, CF, and control lungs.  
19 P-values are obtained by Wilcoxon tests. \*= p<0.05, \*\*\*=p<0.001, \*\*\*\*=p<0.0001. IPF, idiopathic pulmonary fibrosis;  
20 CTD-ILD, connective tissue disease-associated interstitial lung disease; COPD, chronic obstructive pulmonary disease;  
21 CF, cystic fibrosis; MDS1, primary NMDS axis; MDS2, secondary NMDS axis; NMDS, non-metric multidimensional  
22 scaling.

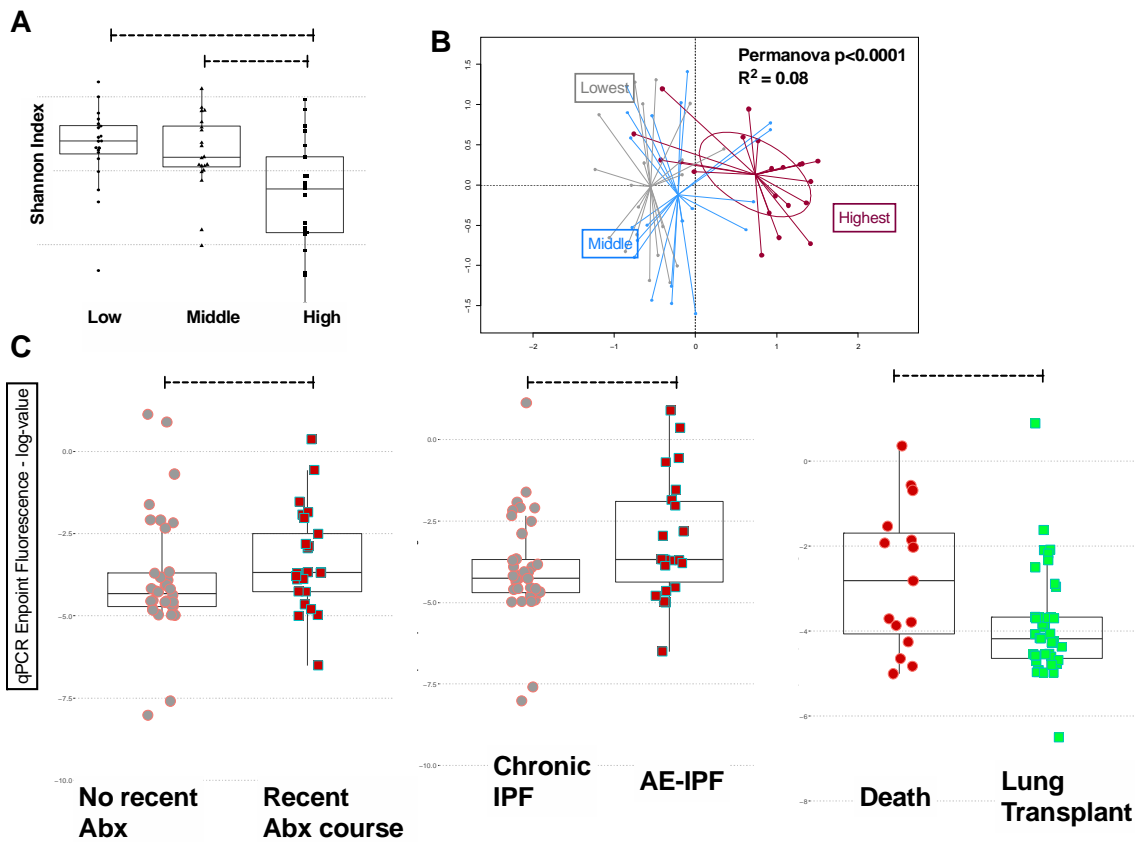




28

29 **Figure 4.** (A) Alpha diversity, as measured by Shannon index, of IPF basilar parenchymal tissue stratified by tertiles of  
30 bacterial load. (B) Beta diversity, as measured by Bray-Curtis dissimilarity, in IPF basilar parenchymal tissue stratified by  
31 tertiles of bacterial load. (C) Mean bacterial burden by qPCR endpoint fluorescence in basilar parenchyma IPF tissue  
32 stratified by clinical outcomes: by whether patients received a recent (within the preceding 90 days) antibiotic prescription,  
33 by whether patients had an AE-IPF at the time of lung explantation (transplant or death) and by whether patients died or  
34 received a lung transplant. P-values are obtained by Wilcoxon tests. \*=p<0.05, \*\*\*=p<0.001, \*\*\*\*=p<0.0001. Abx,  
35 antibiotics; IPF, idiopathic pulmonary fibrosis; AE-IPF, acute exacerbation of IPF; MDS1, primary NMDS axis; MDS2,  
36 secondary NMDS axis; NMDS, non-metric multidimensional scaling.

37

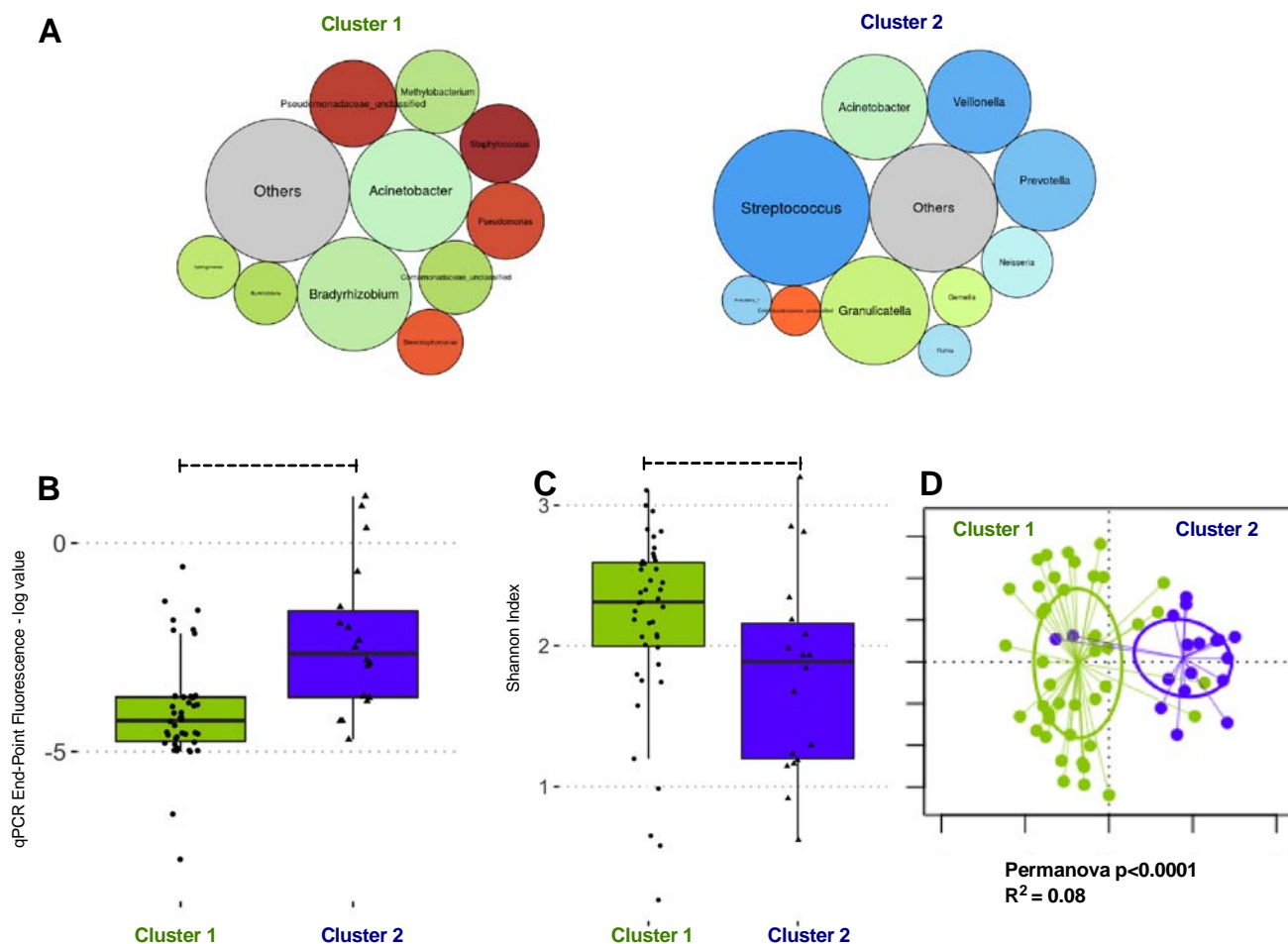


38

39

**Figure 5: Dirichlet Multinomial Modeling clustering of IPF basilar tissue sample communities reveals two distinct**

**taxonomic clusters.** (A) Summary relative abundance for top 10 genera in each cluster, visualized as bubble plot (diameter of each circle corresponds to relative abundance of each taxon across all samples in the cluster). Cluster 1 (71% of samples) had high abundance for several genera that are not typical members of the respiratory microbiome and may represent procedural contamination (e.g. *Bradyrhizobium*, *Methylobacterium*, *Comamonadaceae*). Cluster 2 (29% of samples) had high abundance of typical members of the respiratory microbiome (*Streptococcus*, *Prevotella* and *Veillonella* genera). Genera beyond the top 10 genera demonstrated in these bubble plots were summarized to their overall relative abundance as a single bubble in grey and annotated “Others”. (B) Cluster 2 had significantly higher bacterial load compared to cluster 1 (log-transformed end-point fluorescence of 16S rRNA gene quantitative polymerase chain reaction). (C) Cluster 2 had lower alpha diversity (Shannon index) compared to cluster 1. (D) Principal coordinates analysis for visualization of beta-diversity differences (Bray-Curtis dissimilarity) between Cluster 1 and Cluster 2 samples. P-values are obtained by Wilcoxon tests. \*= $p < 0.05$ ; \*\*\*= $p$ -value  $< 0.001$ .





55 Acknowledgements

56 The authors would like to acknowledge The Center for Organ Recovery & Education (CORE) as well as organ donors and  
57 their families for the generous donation of tissue used in this study.

58

59 Competing interests

60 Dr. Bryan J. McVerry is a consultant for The VeraMedica Institute, LLC and receives research funding from Bayer  
61 Pharmaceuticals, Inc. Dr. Georgios D. Kitsios receives research funding from Karius, Inc. The other authors have no  
62 conflicts of interest to declare.

63 Funding

64 National Institutes of Health [K23 HL139987 (GDK); U01 HL098962 (AM); U01 HL137159 (PVB); R01 HL127349  
65 (PVB); K24 HL123342 (AM); R01 HL123766 (MR), RO1 HL126990 (DK), T32 HL007563-31 (EV), CFF RDP to the  
66 University of Pittsburgh (MM)], Breathe Pennsylvania Lung Health Research Grant (GDK).

67

68 Author contributions

69 Conception and design: GDK, MR, AM, BJM

70 Acquisition, analysis or interpretation of data: EV, HY, JCS, LY, SW, RN, DJK, SQ, XW, MMM, BM, AF, JKA, PVB,  
71 BJM, MR, AM, GDK

72 Clinical cohort characterization: GDK, EV, SW, MMM, DJK

73 Drafting of work and/or revising for important intellectual content: EV, HY, JCS, LY, SW, RN, DJK, SQ, XW, MMM,  
74 BM, AF, JKA, PVB, BJM, MR, AM, GDK

75 Final approval of version to be published; agreement to be accountable for all aspects of the work in ensuring that  
76 questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: EV, HY,  
77 JCS, LY, SW, RN, DJK, SQ, XW, MMM, BM, AF, JKA, PVB, BJM, MR, AM, GDK

78

79 Data sharing

80 All de-identified sequencing data have been submitted to Sequence Read Archive (SRA) database, with BioSample  
81 accession numbers of SAMN13906474-13906711.

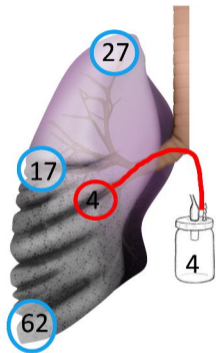
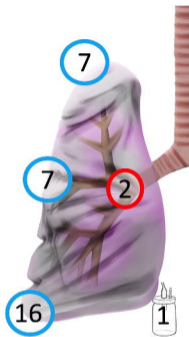
82

33 References

- 34 1. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM, et al. Genetic variants associated  
35 with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association  
36 study. *Lancet Respir Med*. 2017 Nov; 5(11):869-880.
- 37 2. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, et al. Genetic variants associated with  
38 idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir*  
39 *Med*. 2013 Jun; 1(4):309-317.
- 40 3. Maher TM, Wells AU, Laurent GJ. Idiopathic pulmonary fibrosis: multiple causes and multiple  
41 mechanisms? *Eur Respir J*. 2007 Nov; 30(5):835-839.
- 42 4. Idiopathic Pulmonary Fibrosis Clinical Research N, Raghu G, Anstrom KJ, King TE, Jr., Lasky JA,  
43 Martinez FJ. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med*. 2012 May  
44 24; 366(21):1968-1977.
- 45 5. Kreuter M, Wuyts W, Renzoni E, Koschel D, Maher TM, Kolb M, et al. Antacid therapy and disease  
46 outcomes in idiopathic pulmonary fibrosis: a pooled analysis. *Lancet Respir Med*. 2016 May; 4(5):381-389.
- 47 6. Shulgina L, Cahn AP, Chilvers ER, Parfrey H, Clark AB, Wilson EC, et al. Treating idiopathic  
48 pulmonary fibrosis with the addition of co-trimoxazole: a randomised controlled trial. *Thorax*. 2013 Feb;  
49 68(2):155-162.
- 50 7. Molyneaux PL, Cox MJ, Willis-Owen SA, Mallia P, Russell KE, Russell AM, et al. The role of bacteria  
51 in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2014 Oct 15;  
52 190(8):906-913.
- 53 8. O'Dwyer DN, Ashley SL, Gurczynski SJ, Xia M, Wilke C, Falkowski NR, et al. Lung Microbiota  
54 Contribute to Pulmonary Inflammation and Disease Progression in Pulmonary Fibrosis. *Am J Respir Crit Care*  
55 *Med*. 2019 May 1; 199(9):1127-1138.
- 56 9. Invernizzi R, Barnett J, Rawal B, Nair A, Ghai P, Kingston S, et al. Bacterial burden in the lower  
57 airways predicts disease progression in idiopathic pulmonary fibrosis and is independent of radiological disease  
58 extent. *Eur Respir J*. 2020 Jan 24.
- 59 10. Huang Y, Ma SF, Espindola MS, Vij R, Oldham JM, Huffnagle GB, et al. Microbes Are Associated  
60 with Host Innate Immune Response in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2017 Jul 15;  
196(2):208-219.
- 61 11. Molyneaux PL, Willis-Owen SAG, Cox MJ, James P, Cowman S, Loebinger M, et al. Host-Microbial  
62 Interactions in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2017 Jun 15; 195(12):1640-1650.
- 63 12. Hammond M, Clark AB, Cahn AP, Chilvers ER, Fraser WD, Livermore DM, et al. The Efficacy and  
64 Mechanism Evaluation of Treating Idiopathic Pulmonary fibrosis with the Addition of Co-trimoxazole (EME-  
65 TIPAC): study protocol for a randomised controlled trial. *Trials*. 2018 Feb 5; 19(1):89.
- 66 13. Brownell R, Kaminski N, Woodruff PG, Bradford WZ, Richeldi L, Martinez FJ, et al. Precision  
67 Medicine: The New Frontier in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2016 Jun 1;  
68 193(11):1213-1218.
- 69 14. Kitsios GD, Rojas M, Kass DJ, Fitch A, Sembrat JC, Qin S, et al. Microbiome in lung explants of  
70 idiopathic pulmonary fibrosis: a case-control study in patients with end-stage fibrosis. *Thorax*. 2018 May;  
71 73(5):481-484.
- 72 15. Jenkins G. A big beautiful wall against infection. *Thorax*. 2018 May; 73(5):485.
- 73 16. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al. Diagnosis of Idiopathic  
74 Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med*.  
75 2018 Sep 1; 198(5):e44-e68.
- 76 17. Collard HR, Moore BB, Flaherty KR, Brown KK, Kaner RJ, King TE, Jr., et al. Acute exacerbations of  
77 idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2007 Oct 1; 176(7):636-643.
- 78 18. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput  
79 microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012 Aug; 6(8):1621-1624.
- 80

- 31 19. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the  
32 respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med*. 2013 May 15;  
33 187(10):1067-1075.
- 34 20. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial  
35 metagenomics. *PLoS One*. 2012; 7(2):e30126.
- 36 21. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory  
37 contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014 Nov 12; 12:87.
- 38 22. Dickson RP, Huffnagle GB, Flaherty KR, White ES, Martinez FJ, Erb-Downward JR, et al.  
39 Radiographic Honeycombing and Altered Lung Microbiota in Patients with Idiopathic Pulmonary Fibrosis. *Am*  
40 *J Respir Crit Care Med*. 2019 Dec 15; 200(12):1544-1547.
- 41 23. Han MK, Zhou Y, Murray S, Tayob N, Noth I, Lama VN, et al. Lung microbiome and disease  
42 progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med*. 2014 Jul;  
43 2(7):548-556.
- 44 24. Bedard Methot D, Leblanc E, Lacasse Y. Meta-analysis of Gastroesophageal Reflux Disease and  
45 Idiopathic Pulmonary Fibrosis. *Chest*. 2019 Jan; 155(1):33-43.
- 46 25. Wang Z, Bonella F, Li W, Boerner EB, Guo Q, Kong X, et al. Gastroesophageal Reflux Disease in  
47 Idiopathic Pulmonary Fibrosis: Uncertainties and Controversies. *Respiration*. 2018; 96(6):571-587.
- 48 26. Harding SM, Guzzo MR, Richter JE. 24-h esophageal pH testing in asthmatics: respiratory symptom  
49 correlation with esophageal acid events. *Chest*. 1999 Mar; 115(3):654-659.
- 50 27. Rascon-Aguilar IE, Pamer M, Wludyka P, Cury J, Coultas D, Lambiase LR, et al. Role of  
51 gastroesophageal reflux symptoms in exacerbations of COPD. *Chest*. 2006 Oct; 130(4):1096-1101.
- 52 28. King BJ, Iyer H, Leidi AA, Carby MR. Gastroesophageal reflux in bronchiolitis obliterans syndrome: a  
53 new perspective. *J Heart Lung Transplant*. 2009 Sep; 28(9):870-875.
- 54 29. Gaillet G, Favelle O, Guilleminault L, de Muret A, Lemarie E, Lecomte T, et al. Gastroesophageal  
55 reflux disease is a risk factor for severity of organizing pneumonia. *Respiration*. 2015; 89(2):119-126.
- 56 30. Sedgewick AJ, Buschur K, Shi I, Ramsey JD, Raghu VK, Manatakis DV, et al. Mixed graphical models  
57 for integrative causal analysis with application to chronic lung disease diagnosis and prognosis. *Bioinformatics*.  
58 2019 Apr 1; 35(7):1204-1212.
- 59



**IPF****CTD-ILD****COPD****CF****CONTROL**

→ ~45mg tissue

- - - - -

DNA  
extraction

→ 16S V4 subunit qPCR

→ 16S V4 subunit Illumina MiSeq sequencing



Parenchymal Tissue

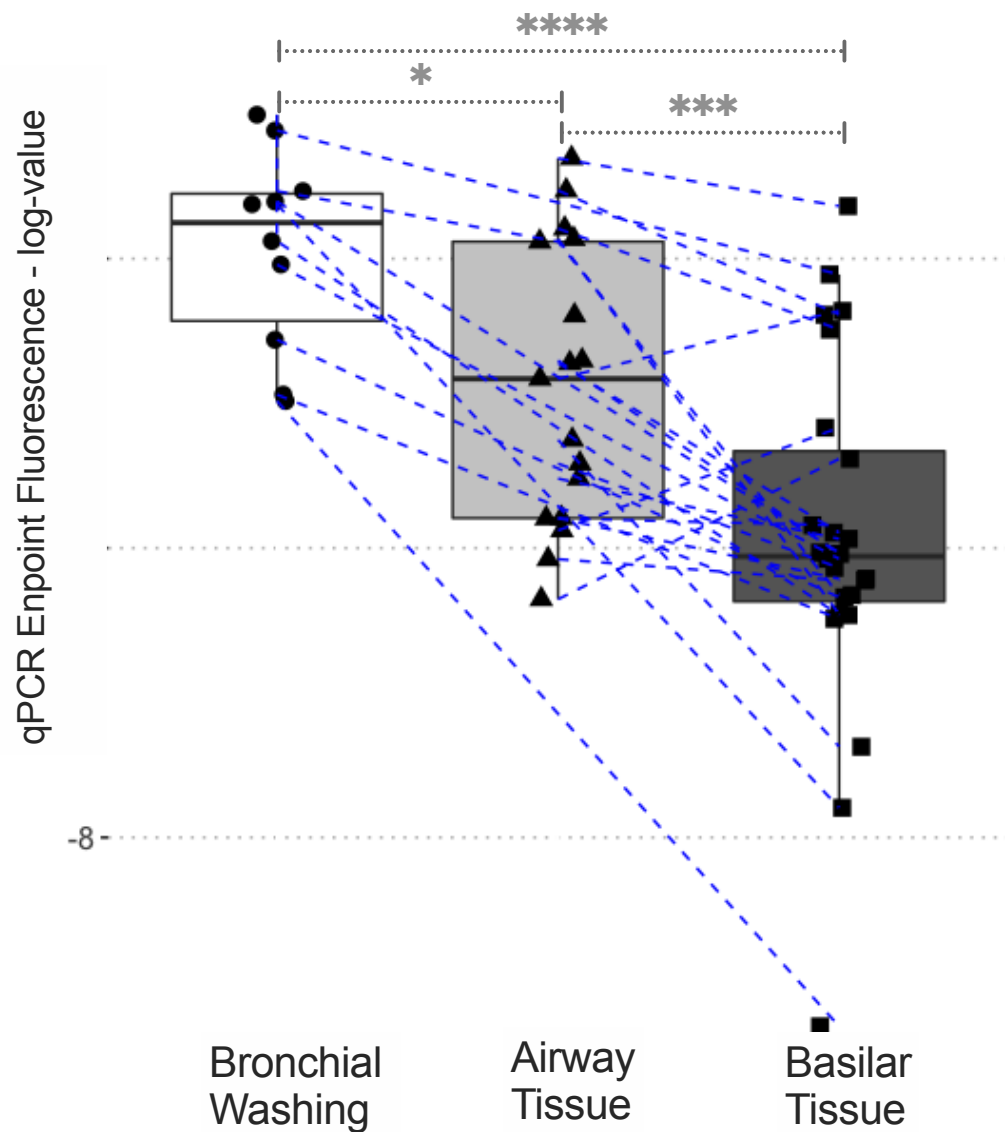
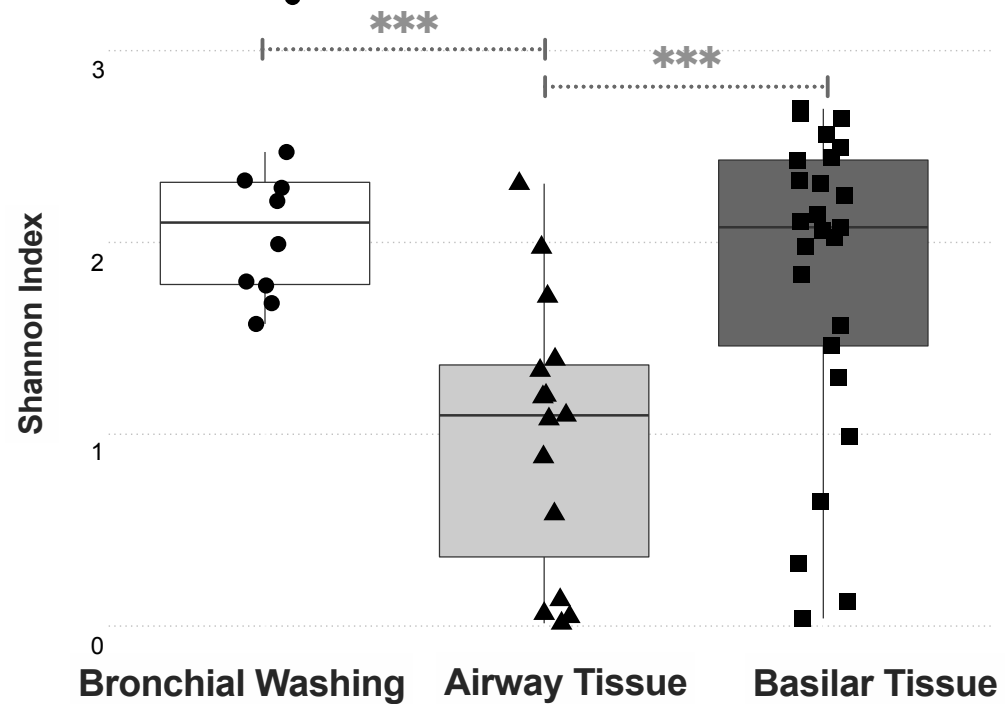
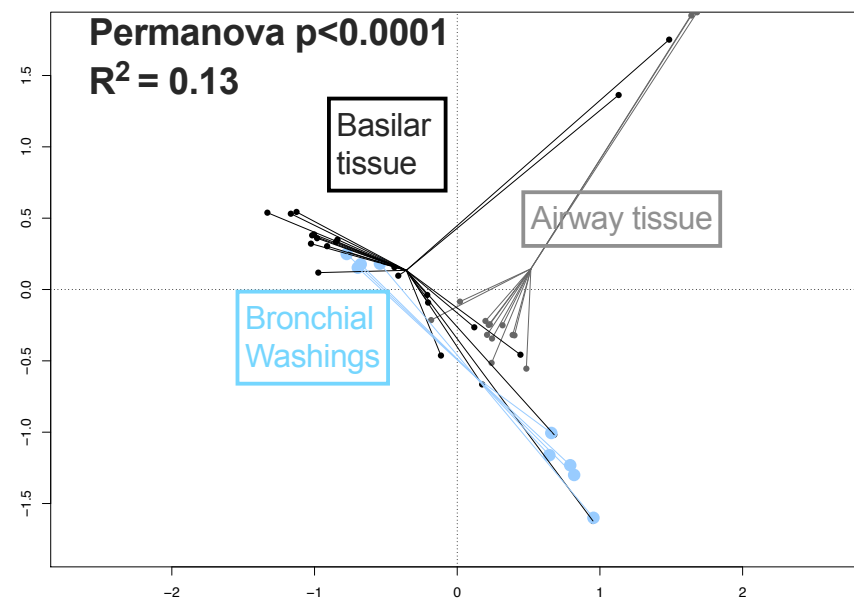


Airway Tissue



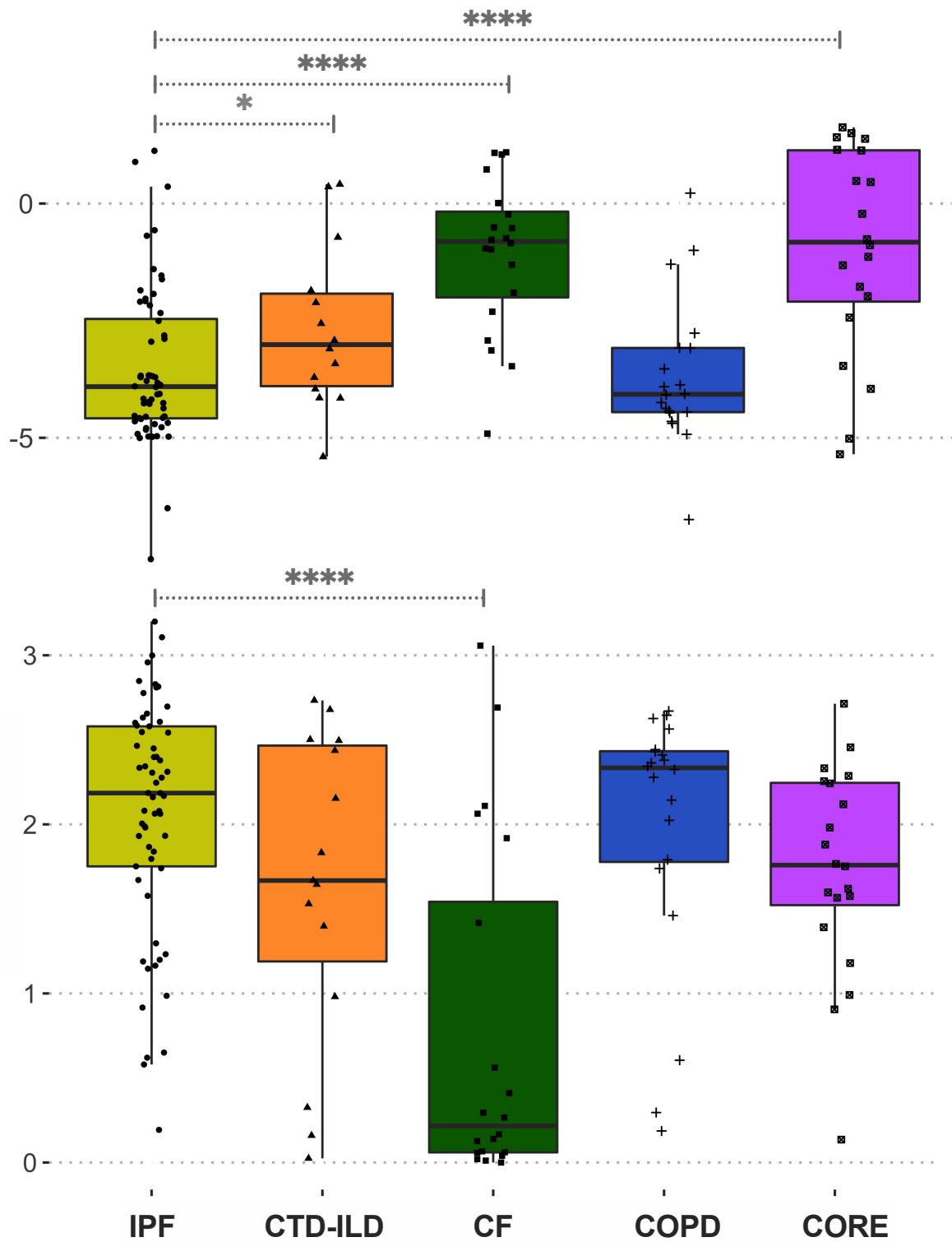
Bronchial Washing



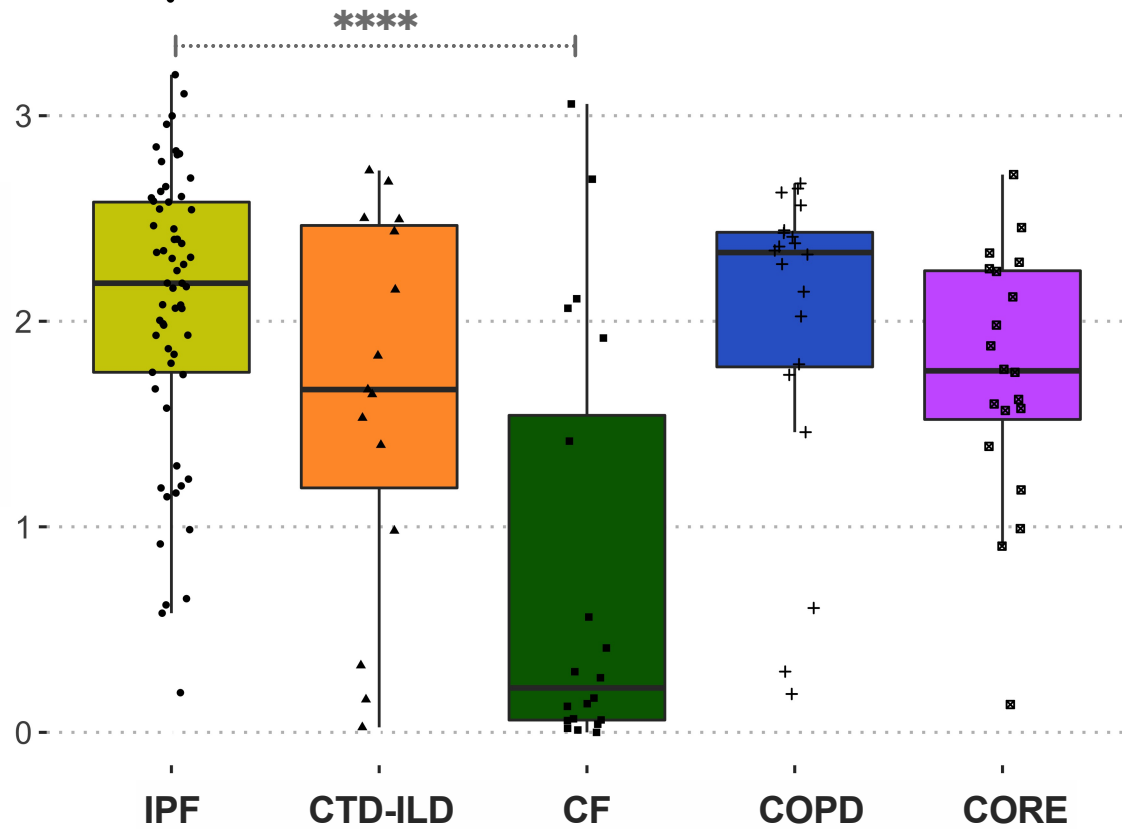
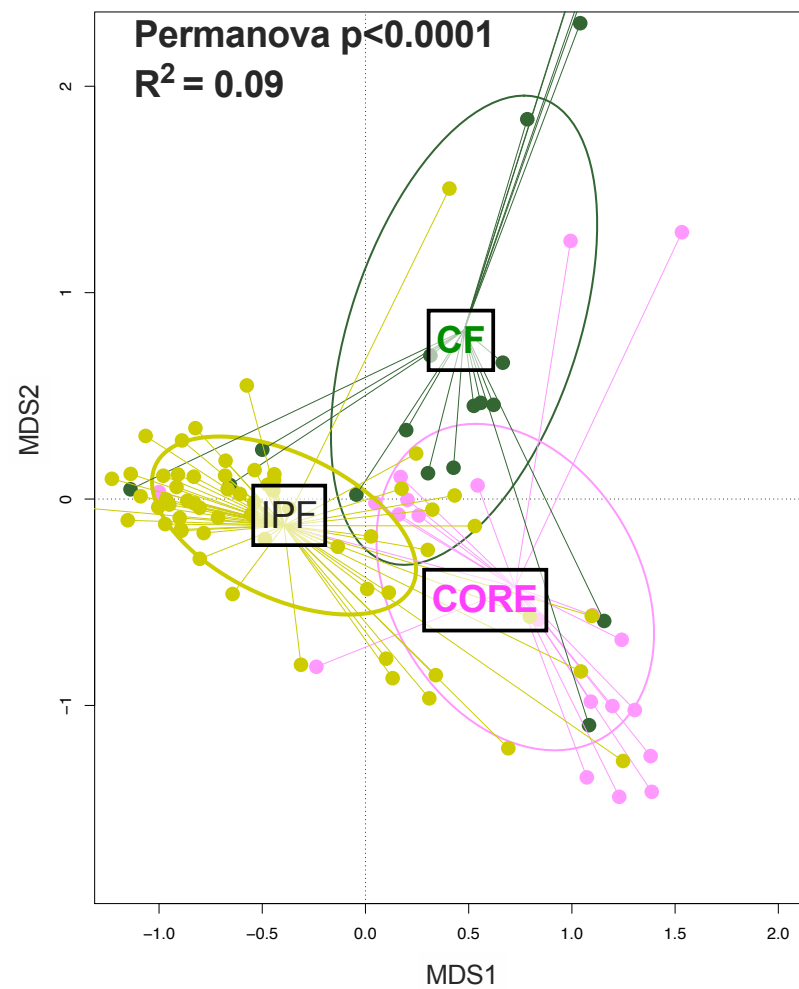
**A****All diseases****B****C**

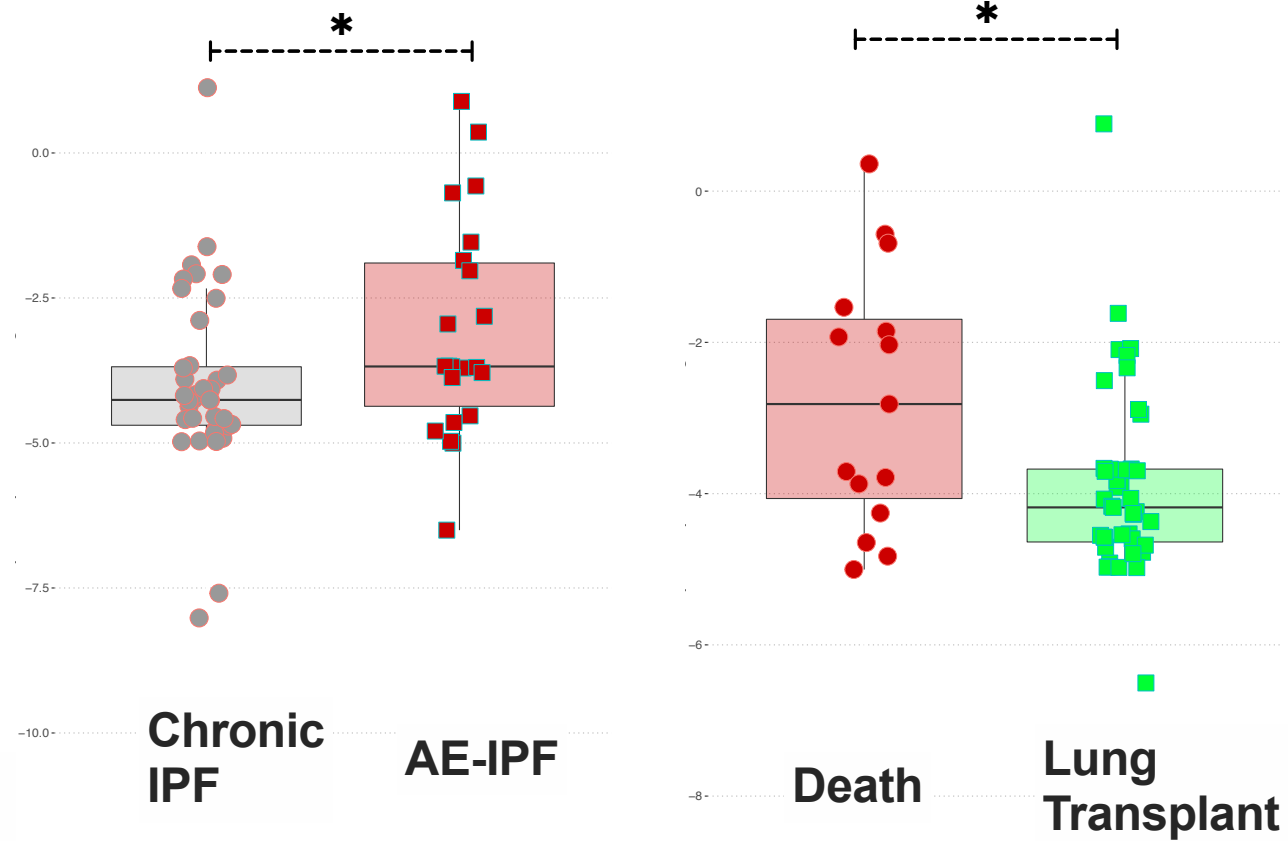
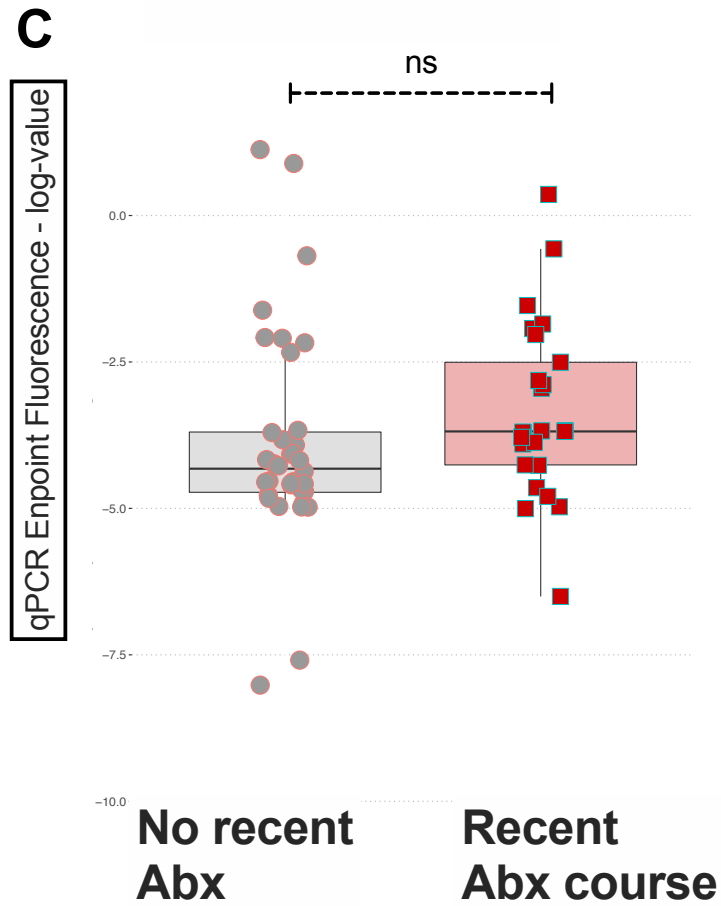
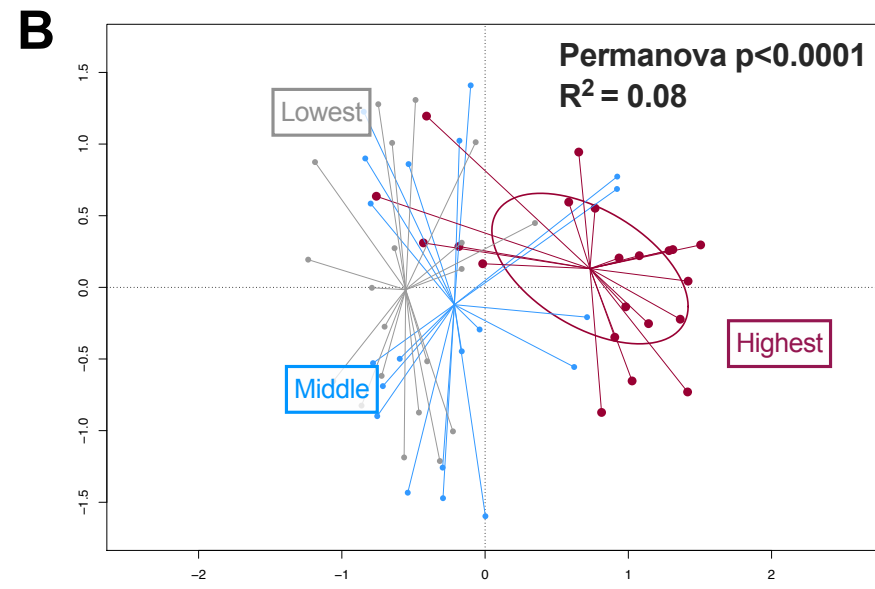
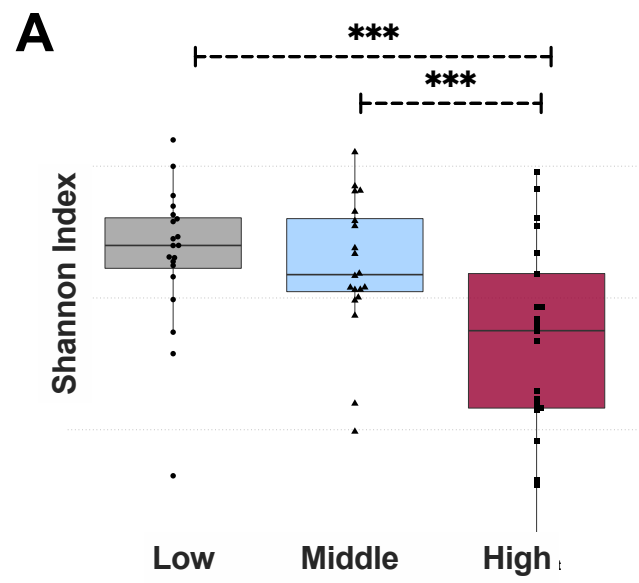
**A**

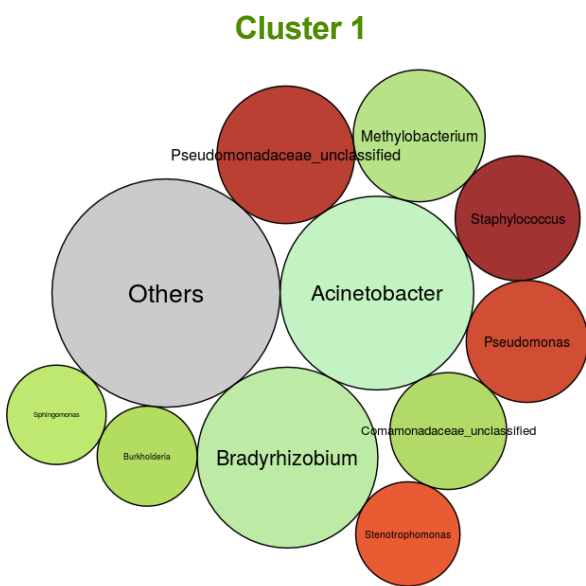
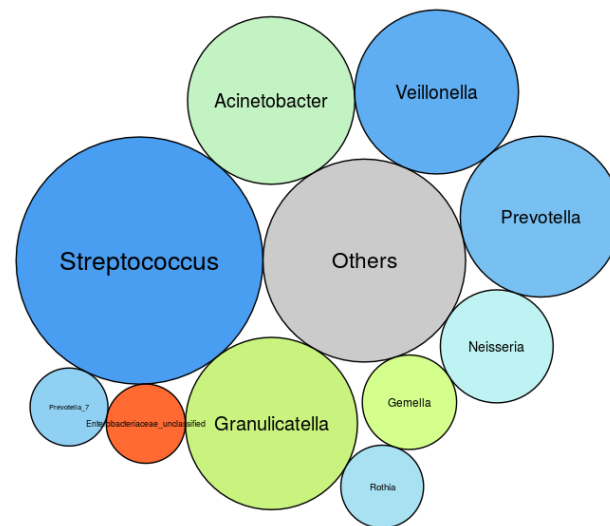
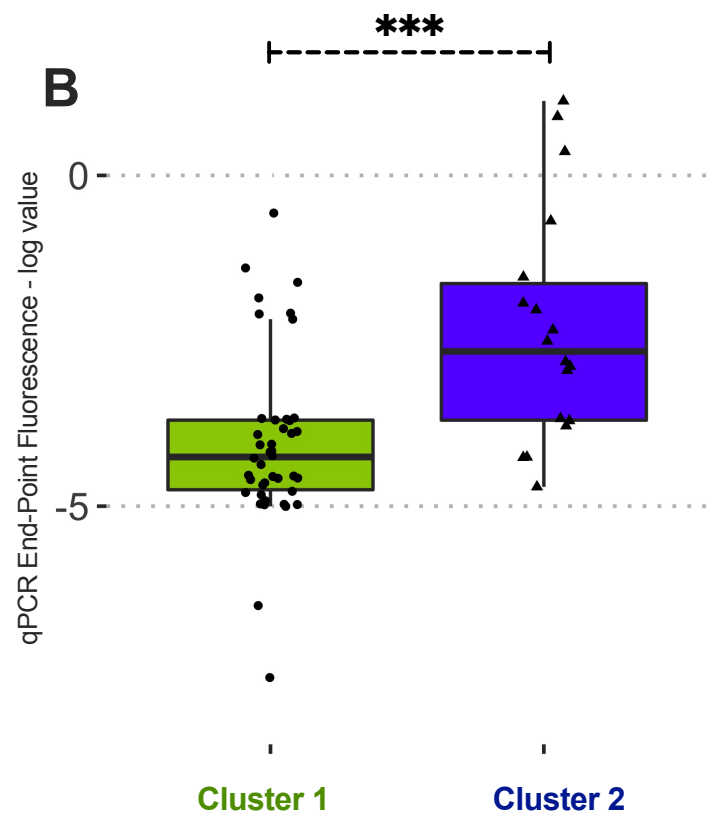
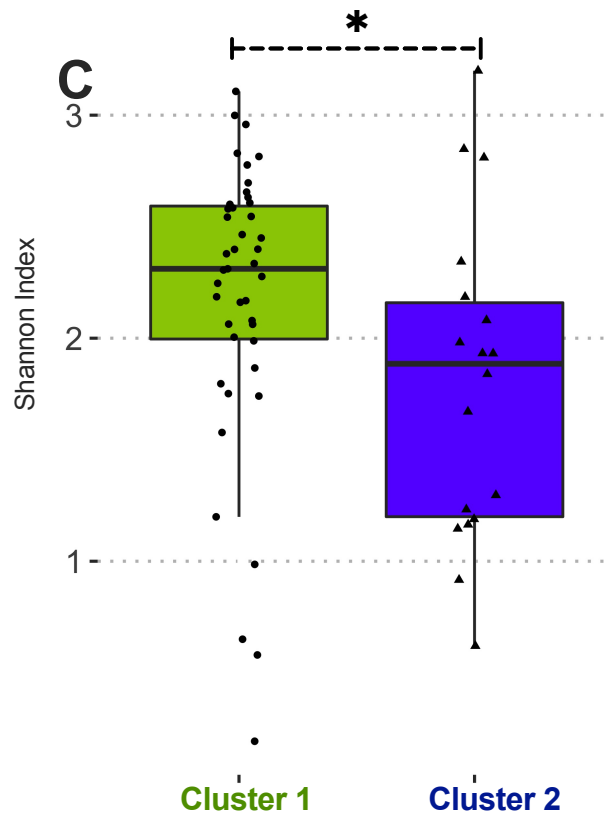
qPCR Endpoint Fluorescence - log-value

**B**

Shannon Index

**C: PCoA plot**



**A****Cluster 2****B****C****D**