

# UCLA

## UCLA Previously Published Works

### Title

Lung microbiota associations with clinical features of COPD in the SPIROMICS cohort.

### Permalink

<https://escholarship.org/uc/item/1jc9w4sd>

### Journal

npj Biofilms and Microbiomes, 7(1)

### Authors

Opron, Kristopher

Begley, Lesa

Erb-Downward, John

et al.

### Publication Date

2021-02-05

### DOI

10.1038/s41522-021-00185-9

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

## ARTICLE OPEN



# Lung microbiota associations with clinical features of COPD in the SPIROMICS cohort

Kristopher Opron<sup>1,18</sup>, Lesa A. Begley<sup>1,18</sup>, John R. Erb-Downward<sup>1</sup>, Christine Freeman<sup>1,2</sup>, Siddharth Madapoosi<sup>1</sup>, Neil E. Alexis<sup>3</sup>, Igor Barjaktarevic<sup>4</sup>, R. Graham Barr<sup>5</sup>, Eugene R. Bleecker<sup>6</sup>, Russell P. Bowler<sup>7</sup>, Stephanie A. Christenson<sup>8</sup>, Alejandro P. Comellas<sup>9</sup>, Christopher B. Cooper<sup>4</sup>, David J. Couper<sup>10</sup>, Claire M. Doerschuk<sup>3</sup>, Mark T. Dransfield<sup>10</sup>, Meilan K. Han<sup>1</sup>, Nadia N. Hansel<sup>11</sup>, Annette T. Hastie<sup>12</sup>, Eric A. Hoffman<sup>9</sup>, Robert J. Kaner<sup>13</sup>, Jerry Krishnan<sup>14</sup>, Wanda K. O'Neal<sup>3</sup>, Victor E. Ortega<sup>12</sup>, Robert Paine III<sup>15</sup>, Stephen P. Peters<sup>12</sup>, J. Michael Wells<sup>10</sup>, Prescott G. Woodruff<sup>8</sup>, Fernando J. Martinez<sup>13</sup>, Jeffrey L. Curtis<sup>16</sup>, Gary B. Huffnagle<sup>1,17</sup> and Yvonne J. Huang<sup>18</sup>✉

Chronic obstructive pulmonary disease (COPD) is heterogeneous in development, progression, and phenotypes. Little is known about the lung microbiome, sampled by bronchoscopy, in milder COPD and its relationships to clinical features that reflect disease heterogeneity (lung function, symptom burden, and functional impairment). Using bronchoalveolar lavage fluid collected from 181 never-smokers and ever-smokers with or without COPD (GOLD 0-2) enrolled in the SubPopulations and Intermediate Outcome Measures In COPD Study (SPIROMICS), we find that lung bacterial composition associates with several clinical features, in particular bronchodilator responsiveness, peak expiratory flow rate, and forced expiratory flow rate between 25 and 75% of FVC (FEF<sub>25-75</sub>). Measures of symptom burden (COPD Assessment Test) and functional impairment (six-minute walk distance) also associate with disparate lung microbiota composition. Drivers of these relationships include members of the *Streptococcus*, *Prevotella*, *Veillonella*, *Staphylococcus*, and *Pseudomonas* genera. Thus, lung microbiota differences may contribute to airway dysfunction and airway disease in milder COPD.

npj Biofilms and Microbiomes (2021)7:14; <https://doi.org/10.1038/s41522-021-00185-9>

## INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a predominantly smoking-related lung disease, characterized by airflow obstruction, local and systemic inflammatory responses, and phenotypic heterogeneity. Inflammation can persist after smoking cessation for reasons not fully understood, and patients vary in disease development, progression and other outcomes. Bacterial colonization of the airways in COPD has long been recognized from culture-based studies, but its role in the pathogenesis or progression of COPD remains unclear.

Recent studies, using culture-independent sequencing to profile microbial communities more comprehensively, have reported alterations in airway bacterial composition mostly in more severe COPD<sup>1-5</sup>. Less is known about the lung microbiome in milder COPD or in smokers without evidence of airflow obstruction and, in particular, whether differences in lung microbiota associate with clinical or biological features in earlier stages of COPD development. In the National Heart, Lung and Blood Institute (NHLBI) Subpopulations and Intermediate Outcome Measures in COPD study (SPIROMICS)<sup>6</sup>, 50% of current or former smokers without airflow obstruction reported significant respiratory symptoms, coupled with activity limitation and more exacerbation-like events<sup>7</sup>. These findings indicate a need to determine whether changes in the lung microbiome may contribute to clinical

outcomes in milder COPD or in those who do not meet current spirometric criteria for the diagnosis.

Few studies of COPD have examined the lung microbiome sampled by bronchoscopy<sup>8-13</sup>, an invasive approach that may provide unique insights into the potential role of lung microbiota in COPD pathogenesis. Using bronchoalveolar lavage fluid (BAL) collected from subjects in the well-characterized SPIROMICS cohort<sup>6,14</sup>, we explored the hypotheses that differences in lung bacterial composition are associated with a diagnosis of COPD and/or with clinical features that reflect COPD pathophysiology, including measures of lung function and symptom burden.

## RESULTS

### Cohort characteristics

Characteristics of the 181 subjects in this study are summarized in Table 1. Those with COPD were older and, as expected, had lower lung function and greater symptom burden [higher scores on COPD Assessment Test (CAT), and St. George's Respiratory Questionnaire (SGRQ)]. Subjects with COPD also displayed greater lung function response to an inhaled bronchodilator, as measured by both change in FVC (forced vital capacity) or FEV<sub>1</sub> (forced expiratory volume in one second); these and other lung function measurements, except for FEF<sub>25-75</sub>, did not differ significantly from measurements at the baseline SPIROMICS visit

<sup>1</sup>Division of Pulmonary/Critical Care Medicine, University of Michigan, Ann Arbor, MI, USA. <sup>2</sup>Research Service, VA Ann Arbor Healthcare System, Ann Arbor, MI, USA. <sup>3</sup>University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. <sup>4</sup>University of California at Los Angeles, Los Angeles, CA, USA. <sup>5</sup>Columbia University, New York, NY, USA. <sup>6</sup>University of Arizona College of Medicine, Tucson, AZ, USA. <sup>7</sup>National Jewish Health, Denver, CO, USA. <sup>8</sup>University of California at San Francisco, San Francisco, CA, USA. <sup>9</sup>University of Iowa, Iowa City, IA, USA. <sup>10</sup>University of Alabama at Birmingham, Birmingham, AL, USA. <sup>11</sup>John Hopkins University, Baltimore, MD, USA. <sup>12</sup>Wake Forest School of Medicine, Winston-Salem, NC, USA. <sup>13</sup>Weill Cornell Medical College, New York, NY, USA. <sup>14</sup>University of Illinois, Chicago, IL, USA. <sup>15</sup>University of Utah, Salt Lake City, UT, USA. <sup>16</sup>Medical Service, VA Ann Arbor Healthcare System, Ann Arbor, MI, USA. <sup>17</sup>Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI, USA. <sup>18</sup>These authors contributed equally: Kristopher Opron, Lesa A. Begley. ✉email: [yvhuang@med.umich.edu](mailto:yvhuang@med.umich.edu)

**Table 1.** Baseline characteristics of subjects.

	All (n = 181)	Never-smokers (n = 24)	Smokers, no COPD (FEV <sub>1</sub> /FVC > 0.70) (n = 79)	Mild/Moderate COPD (50% < FEV <sub>1</sub> < 80%) (n = 71)	Severe COPD (FEV <sub>1</sub> < 50%) (n = 7)	p-value <sup>a</sup>
Age (yrs)	61	55 (42, 71)	57 (41, 75)	65 (46, 77)	65 (42, 74)	<0.001
Sex (male %)	52%	33%	49%	63%	50%	0.06
Caucasian (%)	69%	67%	61%	70%	100%	0.05
Current smoker (%)	34%	N/A	43%	38%	17%	<0.001
Pack-yrs	–	0	35 (20, 113)	45 (23, 270)	43 (25, 90)	<0.001
FEV <sub>1</sub> /FVC	0.72	0.81 (0.72, 0.91)	0.77 (0.57, 0.89)	0.60 (0.34, 0.77)	0.52 (0.46, 0.67)	<0.001
FEV <sub>1</sub> (post-BD, L)	2.6	2.9 (1.9, 2.2)	2.9 (1.4, 4.8)	2.2 (0.9, 3.9)	1.7 (1.3, 2.9)	<0.001
FEV <sub>1</sub> (post-BD, %pred)	92.9	101.9 (72.5, 118.1)	98.9 (65.7, 133)	75.5 (41.1, 111.4)	54.9 (46.6, 86.6)	<0.001
FVC (post-BD, L)	3.7	3.4 (2.3, 5.4)	3.7 (2.4, 6.0)	3.8 (2.4, 6.9)	3.4 (2.7, 4.3)	0.73
FVC (post BD, %pred)	99.6	99.6 (71.6, 113.3)	99.9 (73.7, 129.6)	100.9 (67.1, 132.8)	80.6 (73.4, 94.9)	0.027
FEF <sub>25-75</sub> (L/sec)	1.8	3.3 (1.6, 5.4)	2.6 (0.5, 5.2)	0.9 (0.3, 2.9)	0.6 (0.4, 1.5)	<0.001
Peak expiratory flow rate (L/sec)	7.2	8.8 (4.3, 11.3)	8.0 (2.8, 13.3)	6.3 (3.2, 12.4)	5.0 (4.1, 7.8)	<0.001
Bronchodilator response						
FEV <sub>1</sub> (% change)	7.3	5.0 (–6.8, 13.6)	6.2 (–13.6, 33.5)	10.6 (–10.6, 34.2)	5.1 (2.1, 19.0)	<0.001
FVC (% change)	1.8	–0.6 (–11.9, 5.9)	0.5 (–17.7, 19.0)	6.4 (–8.0, 54.2)	9.9 (4.6, 31.2)	<0.001
Six-minute hall walk distance (meters)	462	490 (357, 645)	464 (114, 626)	450 (126, 740)	451 (360, 576)	0.35#
CAT score	8	2 (0, 16)	7 (0, 32)	10 (1, 35)	14 (2, 29)	<0.001#
SGRQ total score	16	4 (3, 35)	13 (3, 85)	25 (3, 83)	32 (18, 67)	<0.001#

Data are median or % (min, max). Data shown are from the annual visit prior to bronchoscopy except for 6-min walk (baseline visit only).

FEV<sub>1</sub> forced expiratory volume in 1 sec, FVC forced vital capacity, FEF<sub>25-75</sub> maximum mid-expiratory flow, CAT COPD Assessment Test, SGRQ St. George's Respiratory Questionnaire.

#p < 0.05 by current smoking status.

<sup>a</sup>Kruskal–Wallis or Fisher's exact test across all four groups.

(Supplementary Table 1). Current smoking was associated with differences in CAT and SGRQ scores, but not lung function measures.

### Lung bacterial composition is associated with specific measures of lung function, symptom burden and functional impairment

First, to examine whether variation between subjects in their overall lung bacterial composition ( $\beta$ -diversity) associated with COPD status and related clinical characteristics, we performed distance-based PERMANOVA using Bray-Curtis and weighted Unifrac distance measures, using data from all 181 subjects. This analysis identified significant associations between variation in lung microbiota composition and specific measures of lung function, symptom burden, and functional impairment (PERMANOVA  $p < 0.05$ ; Table 2 and Supplementary Table 2). Although pre-bronchodilator measures of FEV<sub>1</sub> and FVC were not associated with lung microbiota variation (not shown), measures of post-bronchodilator response (BDR), in particular change in FVC, were. FEF<sub>25-75</sub> (forced expiratory flow rate between 25 and 75% of FVC) and PEFR (peak expiratory flow rate) also were associated with lung microbiota variation, as were CAT and SGRQ scores and baseline six-minute walk distance. This analysis did not identify associations between lung microbiota variation and a diagnosis of COPD, age, sex, current smoking, or inhaled corticosteroid use, nor with study center or season in which the bronchoscopy was performed. Collectively, these initial findings suggested that differences between subjects in their overall lung bacterial composition were more strongly linked to specific measures of airways dysfunction and related symptom burden, rather than a diagnosis of COPD per se.

We next applied principal component analysis (PCA) to examine further the relationships between these associated clinical features and lung microbiota variation between subjects. This analysis demonstrated contrasting gradients of bacterial compositional difference related to BDR and CAT scores versus FEF<sub>25-75</sub> and PEFR (Fig. 1A; R function *envfit*). CAT scores followed the same compositional gradient as BDR, and independently CAT score and FVC\_BDR were correlated (Spearman R;  $r_s = 0.35$ ,  $p < 0.001$ ).

We also determined within-sample measures of bacterial diversity (inverse Simpson index). These correlated negatively with FVC\_BDR ( $r_s = -0.26$ ,  $p = 0.005$ ) but positively with FEF<sub>25-75</sub> ( $r_s = 0.17$ ,  $p = 0.02$ ; Fig. 1B, C), which support the converse relationships of these variables observed by PCA. This diversity measure did not differ by current smoking status, consistent with our above reported finding that current smoking was not associated with inter-subject variation in lung bacterial community structure.

### Analysis of ever-smokers with or without mild-moderate COPD

A crucial longstanding question is why only some smokers develop COPD. Exploiting the advantage that our cohort consisted primarily of ever-smokers (current or former) without COPD or with mild-moderate COPD, we next focused on these two groups to explore the potential role of differences in the lung microbiota in the disease process. We again observed associations between variation in lung bacterial composition and measures of BDR and 6-min walk distance (Unifrac distance-based PERMANOVA;  $r^2 = 0.02$  for both measures,  $p = 0.019$  and  $p = 0.007$ , respectively), but not with FEV<sub>1</sub>/FVC ratio, FEV<sub>1</sub> % predicted or COPD status itself. BDR measures were greater in the mild-moderate COPD group compared to ever-smokers without COPD (Wilcoxon rank-

**Table 2.** Clinical features associated with differences in lung bacterial community structure (N = 181 subjects).

Lung function and symptom-related outcomes	Bray–Curtis $p$	Weighted Unifrac $p$
FVC bronchodilator response (% change)	0.005	0.003
FVC bronchodilator response (mL change)	0.011	0.004
FEV1 bronchodilator response (% change)	0.013	0.032
FEV1 bronchodilator response (mL change)	0.044	0.085
Peak expiratory flow rate, pre-bronchodilator (L/sec)	0.033	0.079
Peak expiratory flow rate, post-bronchodilator (L/sec)	0.036	0.024
FEF25–75, pre-bronchodilator (L/sec)	0.051	0.043
FEF25–75, post-bronchodilator (L/sec)	0.061	0.048
COPD Assessment Test (CAT) score	0.038	0.020
SGRQ total score	0.014	0.017
SGRQ- impact domain	0.009	0.006
SGRQ- activity domain	0.052	0.039
6-min walk distance (meters)	0.003	0.003
6-min walk distance (% predicted)	0.011	0.006

$p$ -values for distance-based PERMANOVA based on 1000 permutations. Variables demonstrating a significant association ( $p < 0.05$ ) by either distance measure are shown. See also Supplemental Table E3. Lung function and symptom measures are from the annual visit preceding bronchoscopy; 6-min walk distance is from the baseline visit.

sum test,  $p < 0.001$  for both change in FEV<sub>1</sub> or FVC), while CAT scores were marginally higher in the COPD group (Wilcoxon rank sum  $p = 0.04$ ). PCA again demonstrated contrasting gradients of lung bacterial composition associated with BDR versus other lung function measures, including PEFR and related parameters that reflect airflow obstruction (Fig. 2, blue arrows; Unifrac distance;  $R_{envfit}$ ). As before, similar directional gradients of bacterial composition were observed for FVC\_BDR and CAT score, which remained correlated with each other ( $r_s = 0.30$ ,  $p < 0.001$ ). This directionality contrasted from the gradients associated with FEV<sub>1</sub>/FVC, FEV<sub>1</sub>, FEF<sub>25–75</sub>, and PEFR. Biplot analysis indicated that members of the *Streptococcus*, *Staphylococcus*, *Pseudomonas*, *Prevotella*, and others were key contributors to the variation in lung bacterial composition (Fig. 2, red arrows) among these ever-smokers without or with mild-moderate COPD.

We next examined each principal component axis to dissect further which particular bacteria and clinical measurements together contributed to the variation in bacterial community structure amongst subjects. For axis 1 (23.5% of the total variation) principal component (PC) scores were significantly correlated with two measures of FVC\_BDR (% change in FVC;  $R = 0.21$ ,  $p = 0.01$ ; FVC volume response in milliliters,  $R = 0.19$ ,  $p = 0.019$ ; Fig. 3A), and with six-minute walk distance ( $R = -0.24$ ,  $p = 0.003$ ; Fig. 3B). Bacterial taxa contributing to PC1 scores included members of *Pseudomonas* (OTU0007  $R = 0.33$ ,  $q < 0.001$ ; OTU0025  $R = 0.22$ ,  $q = 0.055$ ), *Streptococcus* (OTU0005  $R = 0.21$ ,  $q = 0.06$ ; OTU0016  $R = 0.20$ ,  $q = 0.097$ ), *Staphylococcus* (OTU0012  $R = 0.21$ ,  $q < 0.001$ ), and multiple *Prevotella* (OTU0003, OTU0004, OTU0014;  $R = -0.46$  to  $-0.59$ ;  $q < 0.001$ ).

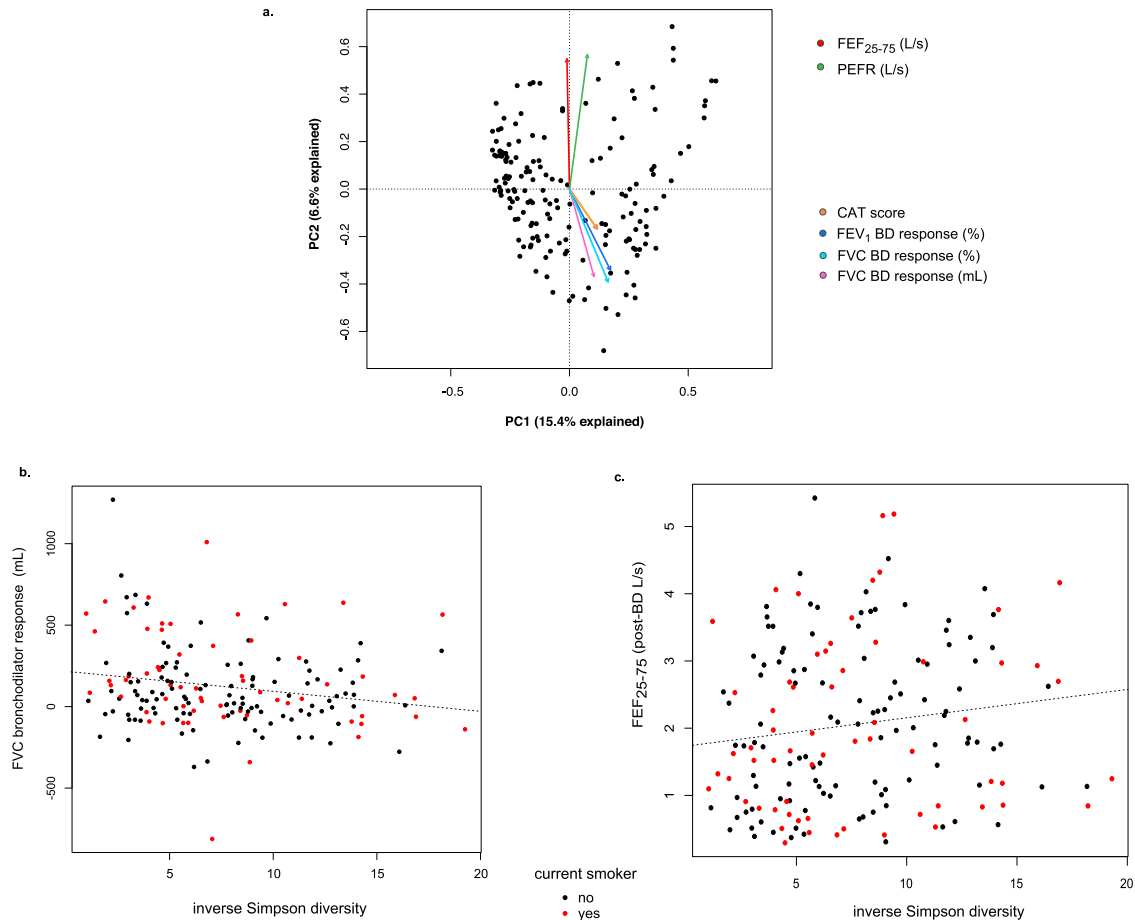
Axis 2, which explained an additional 11.7% of lung bacterial community variation, was associated primarily with measures of PEFR ( $R = 0.20$ ,  $p = 0.014$ ). Five taxa were significantly correlated with PC2 scores ( $q$ -value  $< 0.10$ ). All were negative correlations, the strongest being with three *Streptococcus* members (OTU0005,  $R = -0.73$ ,  $q < 0.001$ ; OTU0016,  $R = -0.33$ ,  $q = 0.002$ ; OTU0019,  $R = -0.25$ ,  $q = 0.049$ ). *Streptococcus* relative abundance also displayed correlative relationships with PEFR, CAT score, as well as FEV<sub>1</sub> and FEV<sub>1</sub>/FVC. OTU0005 had the strongest negative relationship with PC2, associated with lower PEFR, and trended weakly with higher CAT score (Fig. 4) and lower FEV<sub>1</sub> ( $R = -0.14$ ;  $p = 0.095$ ). When the combined relative abundance of these *Streptococci* (OTUs 0005, 0016 and 0019) was considered, a weak negative trend with FEV<sub>1</sub>/FVC also was noted ( $R = -0.16$ ,  $p = 0.081$ ). Altogether these findings suggest that a collection of lung microbiota members contribute to the dysbiosis associated with airways dysfunction and symptoms observed in these ever-smokers.

To complement these analyses, we examined if individual bacterial taxa were differentially abundant between the two groups of ever-smokers and/or associated with lung function measures within each group. Three specific taxa were differentially abundant by DESeq analysis between the ever-smokers without COPD and those with mild-moderate COPD: *Streptococcus* (OTU0005;  $\log_2FC = 2.05$ ,  $q < 0.001$ ), an unclassified *Lactobacillales* (OTU0042;  $\log_2FC = 1.47$ ,  $q = 0.035$ ), and *Veillonella* (OTU0024;  $\log_2FC = -1.41$ ,  $q = 0.079$ ; Fig. 5 and Supplementary Table 3). Among ever-smokers without COPD, the abundances of *Staphylococcus* (OTU0012), *Prevotella* (OTU0014), *Streptococcus* (OTU0016, OTU0019), and *Veillonella* (OTU0024) were negatively associated with lung function by FEV<sub>1</sub>/FVC ratio, FEV<sub>1</sub> % predicted and/or FEF<sub>25–75</sub>. Of these taxa, *Streptococcus* OTU0016 and OTU0019 also negatively associated with lung function in the mild-moderate COPD group (FEV<sub>1</sub>/FVC ratio), along with *Streptococcus* OTU0005 (FEV<sub>1</sub>) and *Gemella* (OTU0044; FEV<sub>1</sub>/FVC ratio, FEV<sub>1</sub> and FEF<sub>25–75</sub>). Taxa associated with greater bronchodilator response in either group included three *Streptococcus* taxa (OTU0005, OTU0016, and OTU0019) and a *Staphylococcus* (OTU0012). We also explored differential abundance relationships with leukocyte populations in the BAL (Supplementary Table 4), which were determined by flow cytometry as previously described<sup>15</sup>. Significant relationships were observed predominantly with the percentages of BAL neutrophils, wherein only *Streptococcus* OTU0019 and an unclassified member of the *Pasteurellaceae* family (OTU0006) were positively associated with BAL % neutrophils in the mild-moderate COPD group.

### Lung bacterial burden and cultivation experiments

Finally, we examined relationships between lung bacterial burden (log-16S rRNA copy numbers as proxy) and lung bacterial composition in BAL samples from all subjects in this study. Bacterial burden positively correlated with phylogenetic diversity of samples (Faith\_PD;  $r_s = 0.30$ ,  $p < 0.001$ ). Microbiota contributing to this relationship included members of *Veillonella* ( $r_s = 0.48$ ), *Prevotella* ( $r_s = 0.28$ – $0.37$ ) and *Streptococcus* ( $r_s = 0.37$ ) (all  $p_{adj} < 0.001$ ). In contrast, the relative abundances of two *Pseudomonas* members and a *Staphylococcus* negatively correlated with bacterial burden ( $r_s = -0.21$  to  $-0.26$ ,  $p_{adj} < 0.001$ ). We noted slightly higher bacterial burden in current smokers compared to subjects not currently smoking (median log-16S copies  $4.6 \pm 0.6$  vs.  $4.4 \pm 0.7$ ; Wilcoxon rank-sum  $p = 0.045$ ). Bacterial burden and within-sample diversity measures did not differ between the groups.

Because of biological interest in several of the identified bacteria in this study (e.g. *Pseudomonas*, *Streptococcus* and *Staphylococcus* taxa), further efforts to identify the particular species were pursued. For *Pseudomonas* (OTU0007) which was associated with community variation along PC1, the



**Fig. 1** Variation in lung bacterial community composition is associated with specific measures of lung function and symptoms. **a** Principal component analysis (PCA) of the overall variation in BAL bacterial composition and relationships to the associated clinical measures for the entire cohort ( $N = 181$ ). Contrasting relationships exist between lung microbiota and measures of airflow limitation ( $FEF_{25-75}$  and peak expiratory flow rate, PEFR) compared to that for measures of bronchodilator response and symptoms (COPD Assessment Test, CAT). PCA based on Hellinger transformation of OTU abundance counts. Vectors indicate direction and magnitude of the linear relationship between a clinical variable and the gradient of bacterial composition shown by PCA (R function *envfit*). **b**, **c** Correlations between within-sample bacterial diversity (inverse Simpson index) to FVC bronchodilator response and  $FEF_{25-75}$  (Spearman  $r = -0.26$ ,  $p = 0.005$  and  $r = 0.17$ ,  $p = 0.02$ , respectively).

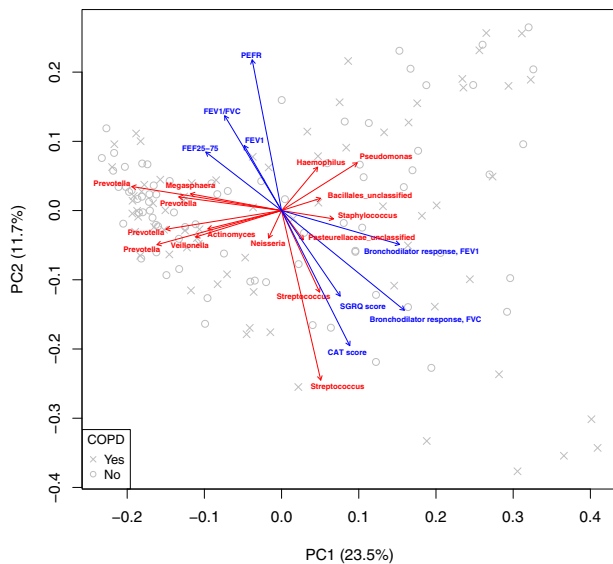
representative sequence demonstrated little similarity to the *P. aeruginosa* represented in the mock community DNA. BLAST analysis of the representative sequence for OTU0007 also did not match *P. aeruginosa*, but rather a range of other species including *P. fluorescens* complex.

We also attempted to isolate by culture and identify the most prevalent *Staphylococcus* and *Streptococcus* species in our dataset (e.g. OTUs 0005, 0012, 0019). From unprocessed frozen ( $-80^{\circ}\text{C}$ ) bronchial wash fluid, single colony isolates of Gram-positive cocci were sub-cultured and sequenced using primers for the full-length 16S rRNA gene, plus the *mnpB* locus (RNA subunit of the bacterial endoribonuclease P complex), as the latter can better differentiate *Streptococcus* species<sup>16</sup>. BLAST analysis of these sequences indicated these cultured isolates to be *Staphylococcus hominis* and *Streptococcus salivarius* (>97% similarity), which by CLUSTAL alignment were identified as OTU0012 and OTU0019, respectively.

## DISCUSSION

In this sub-cohort from SPIROMICS, enriched in ever-smokers without a clinical diagnosis of COPD or with mild-moderate COPD, we observed previously unrecognized relationships between the compositional structure of lung microbiota, the relative abundances of specific bacterial members, and pathophysiological

attributes of COPD, in particular measures reflective of airways dysfunction. Variation in overall lung bacterial community structure did not associate with a categorical diagnosis of COPD, and only two bacterial taxa were significantly positively associated with COPD status (*Streptococcus* and *Lactobacillales*). Instead, we observed more associations between lung bacterial composition and measures of lung function, which provide more granular readouts of airway physiology. Given the milder disease in this cohort, the relationships between these indicators of airways dysfunction and lung bacterial composition hint at interactions in the lung environment that may contribute to airway disease in COPD. To identify specific bacteria that may be involved, we performed taxon-level differential abundance analyses. Although this entailed many comparisons with the lung function measures captured in SPIROMICS, results adjusted for false discovery identified several bacteria negatively associated with  $FEV_1$ /FVC,  $FEV_1$  or  $FEF_{25-75}$  that have previously been implicated in chronic inflammatory airway diseases (e.g. *Streptococcus*, *Staphylococcus*, *Prevotella*, *Gemella*)<sup>17,18</sup>. We also noted relationships between lung bacterial composition and patient-centered measures of respiratory symptom burden (CAT score), suggesting possible clinical consequences related to lung dysbiosis in earlier stages of COPD. Together these results, from one of the largest bronchoscopy-based multicenter investigations focused on understanding the



**Fig. 2 Differential relationships between lung bacterial composition, specific lung function measures, and symptoms are maintained among ever-smokers without or with mild-moderate COPD.** Principal coordinate analysis (PCoA) of lung bacterial community variation (weighted Unifrac distance) among ever-smokers without (gray circles) or with (gray Xs) mild-moderate COPD and relationships to the clinical variables shown (blue arrows;  $R$  envfit). An overlaid biplot analysis shows the microbiota members (red arrows) driving the observed variation in lung bacterial community structure amongst these ever-smokers.

pathogenesis of COPD<sup>14</sup>, suggest that the lung microbiota present in those at risk for COPD may play a role in the pathogenesis of airways dysfunction.

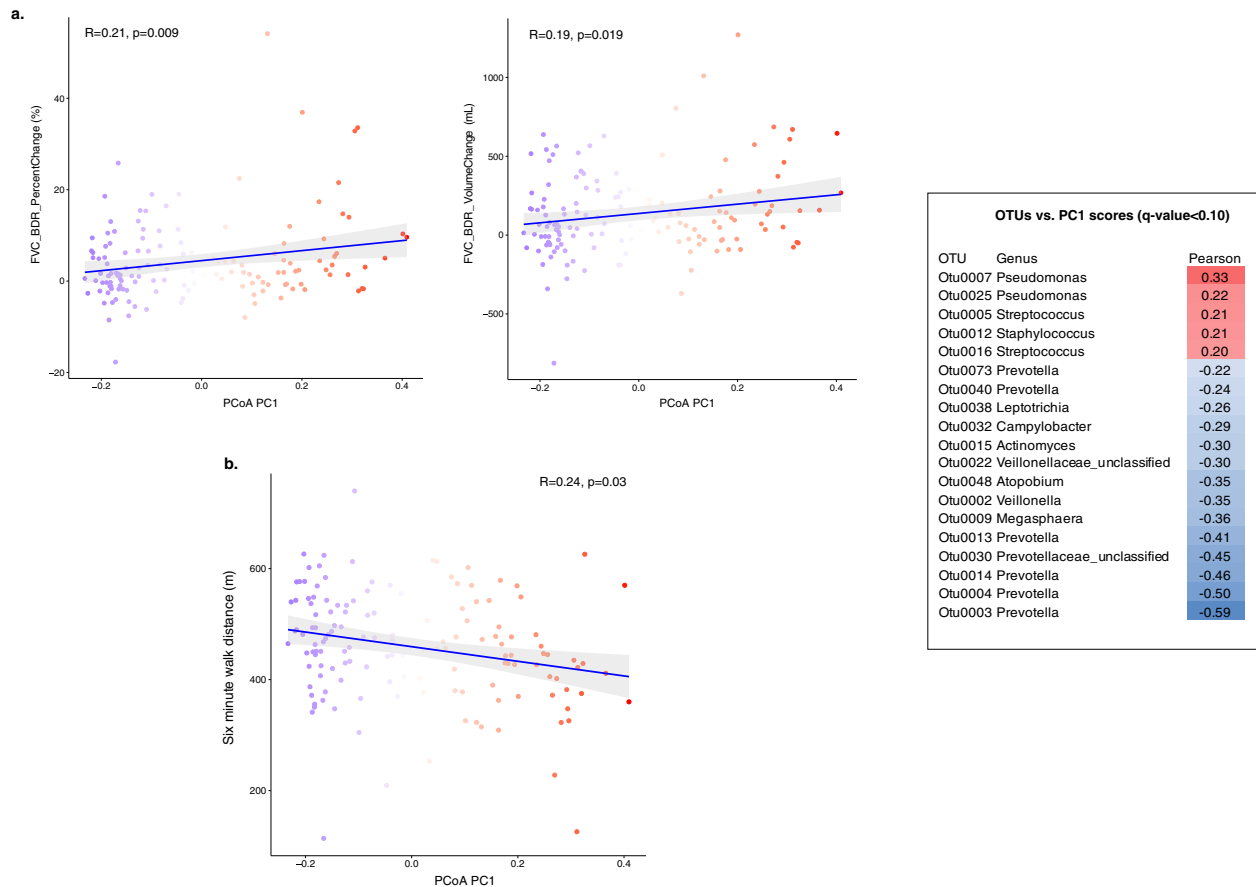
An important strength of our study is the bronchoscopy-based and multicenter nature of our investigation, performed at clinical sites representing the U.S. geographic spectrum. Our study also contrasts from many other studies of the airway microbiome in COPD, where more often sputum has been analyzed to infer insights into the lung microbiome and in patients with more severe COPD or frequent exacerbation history<sup>19–22</sup>. These clinical features were neither the focus of this study nor characteristic of this cohort. We also analyzed cellular BAL which has been shown to capture greater representation of the lung bacterial community and differ in lung bacterial composition compared to acellular samples<sup>23</sup>. It is noteworthy that our data did not reveal differences in lung bacterial composition attributable to geographic locale, season in which bronchoscopy was performed, or current smoking status. An earlier multicenter study of the lung microbiome also did not find an association between current smoking and lung bacterial community structure<sup>24</sup>. We did observe slightly higher lung bacterial burden among current smokers, and it has been reported that BAL from smokers may promote the growth of specific bacterial species<sup>25</sup>. Our study was unique also in our attempt to culture primary isolates of organisms of interest identified in our analyses (e.g. *Staphylococcus*, *Streptococcus*). This led to the identification of two species whose sequences aligned to specific OTUs that may play a role in airways dysfunction. Although *Staphylococcus* and *Streptococcus* are known inhabitants of the upper airways, that these species were viably recovered from BAL samples and not identified as contaminants in our analysis, suggests they have the potential to interact actively in the lung environment.

Our study has a number of limitations. First, it is a descriptive analysis of the lung microbiota in COPD. Our focus on earlier stage disease, however, renders our study somewhat unique. Second, we did not confirm previously reported relationships between

inhaled corticosteroid use and lung bacterial community structure in COPD<sup>2</sup>. This may be due to very few subjects identified as taking an inhaled steroid around the time of bronchoscopy, potentially limiting this evaluation. Similarly, robust examination of lung microbiota relationships to prospective exacerbation events in SPIROMICS subjects, while of interest, was not possible given the limited number of events in the timeframe of available data. Third, due to logistical challenges we could not directly characterize functional features of the microbiota, nor did we examine the identity of non-bacterial members of the lung microbiome. These topics are of interest to advance understanding of interactions between microbial kingdoms and with host immunity. We had limited paired immunological data available for this analysis but were able to explore relationships to BAL leukocyte numbers, which revealed a significant relationship between *Streptococcus* and BAL neutrophils. Future studies incorporating such data from the same biological compartment could be insightful, potentially supplementing the modest strength of the associations we observed with individual bacterial taxa. Indeed, efforts are underway in SPIROMICS to examine these aspects including possible functional interactions between the lung microbiota and host. Lastly, the current study was solely a cross-sectional analysis. While logistically challenging, longitudinal evaluations underway in the current phase of SPIROMICS will permit deeper evaluation of lung microbiota-immune interactions over time and how they relate to COPD development or progression.

The relationships we observed between the lung microbiota and measures of lung function, in particular BDR, FEF<sub>25–75</sub> and PEFr, will also require further study. As with analogous findings in idiopathic pulmonary fibrosis<sup>26,27</sup>, a different lung disease linked in part to smoking, longitudinal studies of COPD could provide stronger evidence of the directionality of these associations and whether particular implicated microbiota precede rather than result from loss of lung function. Molecular mechanisms for the microbial associations with BDR are at present unclear, but this consistent finding in our analyses regardless of the inclusion of never-smokers argues for potential relevance to COPD pathophysiology. BDR measures can be variable over time, but in our subjects were not significantly different between their baseline visit and closest annual study visit before bronchoscopy. We observed similar associations with BDR when we used the baseline measures (not shown). Prior studies have reported relationships between airway dysbiosis and the presence of wheeze or airway hyper-reactivity in asthmatic subjects<sup>28,29</sup>, including involvement of *Streptococcus* and *Staphylococcus* species<sup>29</sup>. A recent analysis from COPDGene<sup>30</sup>, a more severe disease cohort, found that the presence of BDR was associated with radiographic evidence of small airway disease and more exacerbations.

Overall, the relationships observed lead us to speculate that airflow obstruction results in or could also reflect a mutually reinforcing relationship within microenvironmental conditions that exert selective pressure on the compositional make-up and functional behavior of microbiota in the niche. We previously reported that total mucin concentrations in sputum inversely associated with FEF<sub>25–75</sub> in SPIROMICS subjects<sup>31</sup>, which may reflect one aspect of changes in the lung environment that could shape the microbiome. COPD is also associated with loss of several host defense factors that may facilitate bacterial invasion of airway epithelium<sup>32–34</sup>. Over time, evolving interactions between host and microbiota may provoke further immune responses and contribute to perpetuation of airway inflammation. Synergistic interactions between microbiota, as has been shown for *Streptococcus* and *Veillonella* species<sup>35–38</sup>, may also contribute. This ecological perspective of host-microbial interactions aligns with the “vicious cycle hypothesis” of inflammation and ‘infection’ in COPD<sup>39</sup> and is important to maintain in considering how such interactions may affect airway disease pathogenesis in COPD.



**Fig. 3** Post-bronchodilator FVC response and 6-min walk distance correlate with variation in lung bacterial composition along principal component 1, which explained 23.5% of the total variation in lung bacterial community structure among ever-smokers. Each dot represents an individual BAL sample and is colored according to its PC score. OTUs contributing significantly ( $q$ -value  $< 0.10$ ) to PC1 are shown in the accompanying table, ranked by Pearson's correlation coefficient. The relative abundances of *Prevotella*, *Pseudomonas*, and *Streptococcus* are primary drivers of PC1 variation. **a** Post-bronchodilator FVC response measured by change in percent and volume in mL. **b** Six-minute walk distance (meters).

Findings from this study suggest additional avenues for further translational and mechanistic investigations into lung microbiota-host interactions in early stage COPD.

## METHODS

### Subject characterization and research bronchoscopy

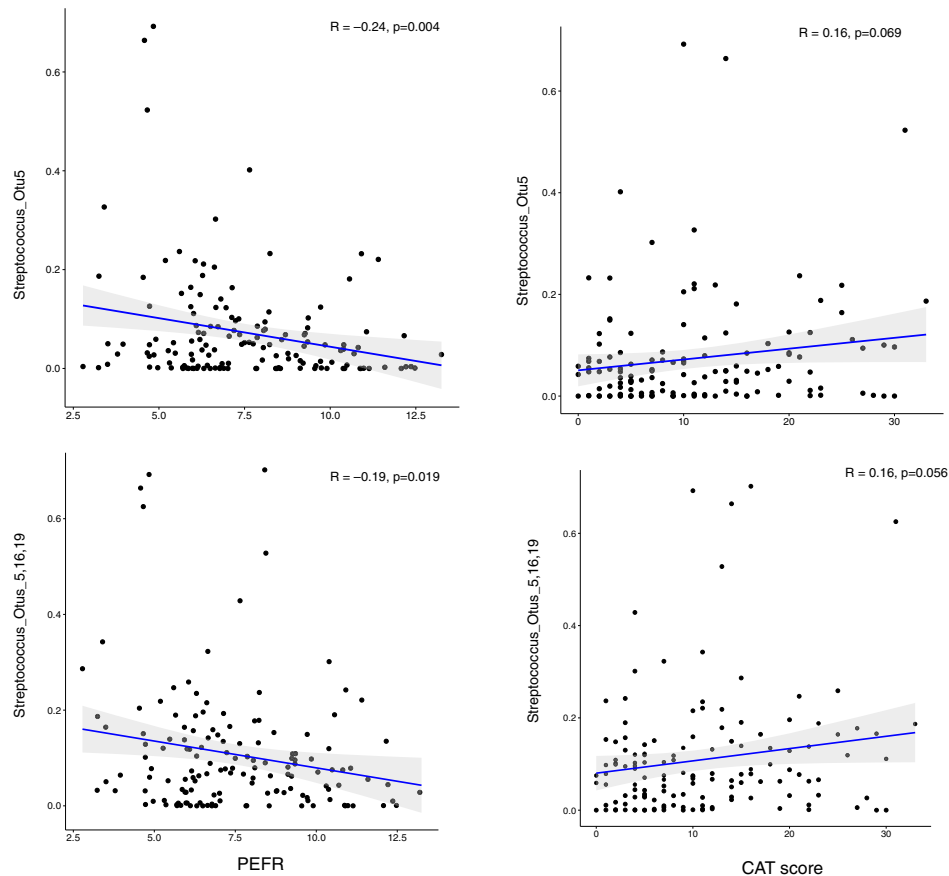
SPIROMICS is a multicenter observational study (NCT01969344; clinicaltrials.gov) of never-smokers, smokers without airflow obstruction or with mild, moderate or severe COPD. All SPIROMICS participants ( $n = 2981$ ) underwent detailed clinical characterization and collection of biological and radiographic data, as previously described<sup>6,40</sup>. The institutional review boards of all SPIROMICS sites approved the main study protocol, and all participants provided written informed consent. Pulmonary function testing was performed according to the 2005 ATS/ERS guidelines. Ever-smokers were defined as current or former smokers with a smoking history of  $\geq 20$  pack-years; former smokers were those who had been free from use of any tobacco products for 6 months preceding the bronchoscopy. A subset of subjects ( $n = 215$ ; inclusive of never-smokers, smokers with normal lung function, mild-moderate COPD, and severe COPD) additionally participated in a bronchoscopy sub-study<sup>14</sup>, which was approved by the institutional review boards of the participating SPIROMICS sites and for which subjects provided separate informed consent. The institutions that enrolled subjects for the research bronchoscopy included Columbia University/Weill Cornell Medical College, National Jewish Health (Denver, Colorado), University of Alabama at Birmingham, University of California at Los Angeles, University of California at San Francisco, University of Michigan, University of Utah, and Wake Forest University School of Medicine. Bronchoscopy was performed a median of 62 days after the

preceding annual study visit (Supplementary Fig. 1), from which most of the clinical measures used in our analyses were obtained and did not differ significantly from earlier measurements (Supplementary Table 1). Samples of BAL allocated for microbiome analysis (10 mL into RNALater) were obtained from 181 subjects, along with biological control specimens (e.g. oral wash/tongue scraping, instrument samples such as bronchoscope suction channel flushes with sterile saline, negative controls of sterile saline or lidocaine). Remaining BAL was saved for other purposes in SPIROMICS including a 5-mL aliquot for potential microbial cultivation experiments. Samples were stored long-term at  $-80^{\circ}\text{C}$  until processing for this study.

### Sample processing, microbiota sequencing, and raw data processing

Five mL of BAL, oral rinse/tongue scraping, and bronchoscope flushes were centrifuged at high-speed (13,000 rpm) to pellet cellular material. Total genomic DNA was extracted from cell pellets using a commercial kit (DNeasy Blood and Tissue kit, Qiagen) with the following modifications: after re-suspension in ATL buffer the samples were transferred to Lysing Matrix E tubes (MP Biomedicals) and subjected to bead beating for 30 s before proceeding. The manufacturer's protocol was modified to use 40  $\mu\text{L}$  proteinase K instead of the recommended 20  $\mu\text{L}$ , and samples were eluted with 100  $\mu\text{L}$  of buffer AE instead of the suggested 200  $\mu\text{L}$ . Negative extraction controls were generated by performing the extraction protocol without addition of sample but with sterile PBS.

The Microbial Systems Molecular Biology Laboratory (Microbiome) Core at the University of Michigan performed 16S rRNA gene (V4 region) sequencing using barcoded dual-index primers<sup>41</sup> on an Illumina MiSeq (Illumina, San Diego, CA). Given the low biomass of samples, we employed a touchdown PCR amplification strategy<sup>42</sup>, including a mock reference



**Fig. 4** *Streptococcus*, PEFR and CAT score in ever-smokers with or without COPD. The relative abundance of *Streptococcus* Otu5, either alone (top panels) or together with other *Streptococcus* OTUs (bottom panels), correlate negatively with peak expiratory flow rate (PEFR) and positively with COPD Assessment Test (CAT) score.

community (ZymoBIOMICS; Zymo Research; Irvine, CA). Library preparation/sequencing controls (including elution buffer, sterile water, PBS, empty wells) also were sequenced. Several of these were performed in replicate. A mock community (Zymo) was included for quality control. Normalized pooled libraries were sequenced on the Illumina MiSeq platform using the 500 cycle MiSeq V2 Reagent kit. For internal testing purposes, BAL lavage samples were sequenced in triplicate (batch 2 only), while all other samples were sequenced once. The distribution of subjects by SPIROMICS group assignment did not differ between the two runs (Fisher exact test  $p > 0.05$ ).

16S rRNA gene sequence reads were processed using mothur (v1.40.5)<sup>43</sup>, aligned to the SILVA reference alignment (release 102)<sup>44</sup> and classified using the RDP 16S rRNA reference training set (version 16)<sup>45</sup>. The range of read counts in samples was 577 to 53,155, with the average being 21,444. Operational taxonomic units (OTUs) were defined using 97% similarity threshold.

Results of negative control samples (raw data prior to decontamination analysis using R package *decontam*<sup>46</sup>) are provided as Supplementary Table 5 and Supplementary Fig. 2. The prevalence method in *decontam* was used. The two batches of sample runs were decontaminated separately based on the negative controls collected/generated in parallel for the samples in a given batch. Initially, there were a total of 5172 OTUs and 986 OTUs were removed after *decontam* analysis. Eight of these OTUs were in common between the two decontamination runs (Otu0056, Otu0300, Otu0078, Otu1812, Otu0111, Otu0154, Otu0047, and Otu0204) and OTU 0011 was also removed because it appeared to be a contaminant. This brought the total of OTUs to 4193 prior to filtering by relative abundance. To compare further the two batches of samples, it was necessary to choose a single replicate sample from each subject in batch 2. For each patient, the geometric mean of normalized OTU counts was calculated to create a representative community to compare samples against. Then the Bray–Curtis distance between each individual sample

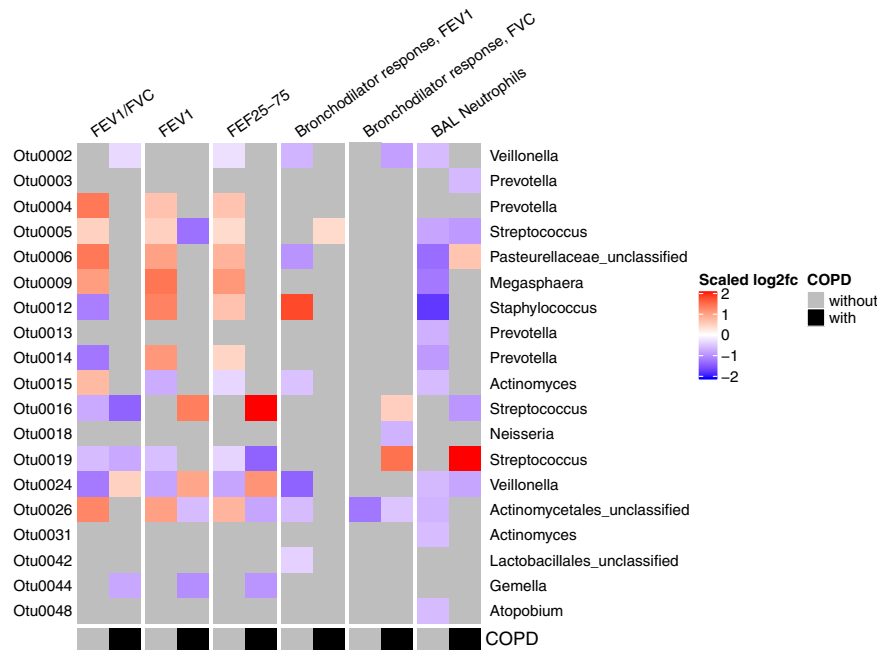
and the mean sample was calculated and the sample with the lowest distance to the representative sample kept.

The count table, taxonomy table and phylogenetic tree file (*picante*)<sup>47</sup> were imported into R for analysis using *phyloseq* (v1.26.0)<sup>48</sup>. For alpha-diversity calculations, reads were subsampled to the lowest count (577). All other analyses were performed with non-subsampled data, using relative abundance. The OTUs were filtered by average relative abundance  $>0.1\%$  (final total of 125 OTUs analyzed). Principal component analysis of Hellinger-transformed abundance data from 84 paired BAL and oral rinse/tongue scrape samples demonstrated distinct segregation (clustering) of the two sample types (Supplementary Fig. 3; PERMANOVA  $p < 0.05$ ).

### Statistical/data analysis

Clinical data from the most recent annual SPIROMICS visit preceding bronchoscopy were used wherever possible (SPIROMICS Core dataset CORE5\_20180910). For exploratory analyses, we also included information available only from the baseline SPIROMICS visit. Non-parametric tests and Fisher's exact tests were used where appropriate. Spearman correlation ( $r_s$ ) was used to compare continuous measures including alpha-diversity indices, with Benjamini–Hochberg correction for multiple comparisons<sup>49</sup>. To identify variables associated with bacterial community differences between subjects (i.e. beta-diversity), univariate distance-based PERMANOVA<sup>50,51</sup> based on Bray–Curtis or weighted Unifrac distances was performed (*vegan* v2.5-4<sup>27</sup>), with 1000 permutations of the input matrices for determination of significance. The Unifrac distance incorporates phylogenetic information to capture the relatedness of bacterial communities between samples. Taxon-level differential abundance was determined by DESeq2 which utilizes a negative binomial generalized linear model with Wald's test<sup>52</sup>.





**Fig. 5** Heatmap summarizing results of DESeq analysis showing lung bacterial taxa significantly associated ( $q$ -value  $< 0.10$ ) with measures of lung function, symptoms (CAT score), and percentage of BAL neutrophils. Further information is provided in Supplementary Table 3.

### Assessment of bacterial burden and culture-based studies

Copy numbers of the 16S rRNA gene were assessed by quantitative (real-time) PCR using a standard curve comprising a 10-fold dilution series from  $10^8$  to  $10^2$  copies per microliter. TaqMan reagents (ABI) were used to amplify 3  $\mu$ L of DNA; nuclease-free water was used in place of DNA for a no-template control. All reactions were performed in triplicate. Primers and probe were as follows: forward primer (TGG AGC ATG TGG TTT AAT TCG A), reverse primer (TGC GGG ACT TAA CCC AAC), probe (FAM-CAC GAG CTG ACG ACA CCC ATG CA-BHQ). Amplification was performed on the ABI Step One Plus thermocycler and data analyzed using Step One software version 2.1. To further identify primary species of interest (e.g. *Streptococcus*, *Staphylococcus*), aliquots of raw frozen BAL were cultured onto blood agar and in aerobic broth (SBHI medium) at 37 °C overnight, followed by sub-culture of individual colonies which by Gram stain revealed Gram-positive cocci in chains and clusters. Fresh liquid cultures were grown from single colonies and DNA isolated. Sequencing of the 16S ribosomal RNA gene and alignment in BLAST indicated two isolates in the genus *Streptococcus* and one isolate of *Staphylococcus hominis*. CLUSTAL (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) was used to align the latter sequence with the representative sequences obtained from the BAL community and identify *Staphylococcus hominis* (OTU0012). Due to high sequence similarity in the 16S rRNA gene for the genus *Streptococcus*, sequencing of the *mpB* locus was performed<sup>16</sup> to determine species identity.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

All data supporting the findings of this study are available within the paper and in the Supplementary Information. The 16S rRNA gene sequence data have been submitted to the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA673153. These data also are available from the SPIROMICS Genomics and Informatics Coordinating Center (<https://www2.csc.unc.edu/spiromics/contact-gic>).

### CODE AVAILABILITY

Information regarding software, data inputs and code used to support the findings of this study are in the Reporting Summary. No custom code was developed for this

study. The specific code implemented, based on published analysis packages, can be obtained from the authors upon request.

Received: 11 August 2020; Accepted: 7 January 2021;

Published online: 05 February 2021

### REFERENCES

- Erb-Downward, J. R. et al. Analysis of the lung microbiome in the 'healthy' smoker and in COPD. *PLoS ONE* **6**, e16384 (2011).
- Pragman, A. A., Kim, H. B., Reilly, C. S., Wendt, C. & Isaacson, R. E. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PLoS ONE* **7**, e47305 (2012).
- García-Núñez, M. et al. Severity-related changes of bronchial microbiome in chronic obstructive pulmonary disease. *J. Clin. Microbiol.* **52**, 4217–4223 (2014).
- Einarsson, G. G. et al. Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. *Thorax* **71**, 795–803 (2016).
- Sze, M. A. et al. Host response to the lung microbiome in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **192**, 438–445 (2015).
- Couper, D. et al. Design of the subpopulations and intermediate outcomes in COPD study (SPIROMICS). *Thorax* **69**, 491–494 (2014).
- Woodruff, P. G. et al. Clinical significance of symptoms in smokers with preserved pulmonary function. *N. Engl. J. Med.* **374**, 1811–1821 (2016).
- Cabrera-Rubio, R. et al. Microbiome diversity in the bronchial tracts of patients with chronic obstructive pulmonary disease. *J. Clin. Microbiol.* **50**, 3562–3568 (2012).
- Segal, L. N. et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome* **1**, 19 (2013).
- Zakharkina, T. et al. Analysis of the airway microbiota of healthy individuals and patients with chronic obstructive pulmonary disease by T-RFLP and clone sequencing. *PLoS ONE* **8**, e68302 (2013).
- Grønseth, R. et al. The Bergen COPD microbiome study (MicroCOPD): rationale, design, and initial experiences. *Eur. Clin. Respir. J.* **1**, <https://doi.org/10.3402/ecrj.v1.26196>. (2014). eCollection 2014.
- Leiten, E. O. et al. Complications and discomfort after research bronchoscopy in the MicroCOPD study. *BMJ Open Respir. Res.* **7**, e000449, <https://doi.org/10.1136/bmjresp-2019-000449>. (2020).
- Mika, M. et al. Microbial and host immune factors as drivers of COPD. *ERJ Open Res.* **4**, 00015-2018 (2018).

14. Wells, J. M. et al. Safety and tolerability of comprehensive research bronchoscopy in chronic obstructive pulmonary disease. Results from the SPIROMICS Bronchoscopy Substudy. *Ann. Am. Thorac. Soc.* **16**, 439–446 (2019).
15. Freeman, C. M. et al. Design of a multi-center immunophenotyping analysis of peripheral blood, sputum and bronchoalveolar lavage fluid in the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS). *J. Transl. Med.* **13**, 19 (2015).
16. Tapp, J., Tholleson, M. & Herrmann, B. Phylogenetic relationships and genotyping of the genus *Streptococcus* by sequence determination of the RNase P RNA gene, *mpbB*. *Int. J. Syst. Evol. Microbiol.* **53**, 1861–1871 (2003).
17. Scott, J. E. & O'Toole, G. A. The yin and yang of streptococcus lung infections in cystic fibrosis: a model for studying polymicrobial interactions. *J. Bacteriol.* **201**, e00115-19 (2019).
18. Jacobs, D. M., Ochs-Balcom, H. M., Zhao, J., Murphy, T. F. & Sethi, S. Lower airway bacterial colonization patterns and species-specific interactions in chronic obstructive pulmonary disease. *J. Clin. Microbiol.* **56**, e00330-18 (2018).
19. Mayhew, D. et al. Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. *Thorax* **73**, 422–430 (2018).
20. Wang, Z. et al. Lung microbiome dynamics in COPD exacerbations. *Eur. Respir. J.* **47**, 1082–1092 (2016).
21. Millares, L. et al. Relationship between the respiratory microbiome and the severity of airflow limitation, history of exacerbations and circulating eosinophils in COPD patients. *BMC Pulm. Med.* **19**, 112 (2019).
22. Wang, Z. et al. Airway host-microbiome interactions in chronic obstructive pulmonary disease. *Respir. Res.* **20**, 113 (2019).
23. Dickson, R. P. et al. Cell-associated bacteria in the human lung microbiome. *Microbiome* **2**, 28 (2014).
24. Morris, A. et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am. J. Respir. Crit. Care Med.* **187**, 1067–1075 (2013).
25. Vargas Buonfiglio, L. G. et al. Airway surface liquid from smokers promotes bacterial growth and biofilm formation via iron-lactoferrin imbalance. *Respir. Res.* **19**, 42 (2018).
26. Molyneaux, P. L. et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **190**, 906–913 (2014).
27. Takahashi, Y. et al. Impaired diversity of the lung microbiome predicts progression of idiopathic pulmonary fibrosis. *Respir. Res.* **19**, 34 (2018).
28. Huang, Y. J. et al. Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *J. Allergy Clin. Immunol.* **127**, 372–381.e1–3 (2011).
29. Kim, B.-S. et al. Different functional genes of upper airway microbiome associated with natural course of childhood asthma. *Allergy* **73**, 644–652 (2018).
30. Fortis, S. et al. Combined forced expiratory volume in 1 second and forced vital capacity bronchodilator response, exacerbations, and mortality in chronic obstructive pulmonary disease. *Ann. Am. Thorac. Soc.* **16**, 826–835 (2019).
31. Kesimer, M. et al. Mucin concentrations and peripheral airway obstruction in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **198**, 1453–1456 (2018).
32. Polosukhin, V. V. et al. Secretory IgA deficiency in individual small airways is associated with persistent inflammation and remodeling. *Am. J. Respir. Crit. Care Med.* **195**, 1010–1021 (2017).
33. Martinez, F. J. et al. At the root: defining and halting progression of early chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **197**, 1540–1551 (2018).
34. Lacho-Contreras, M. E. et al. Protective role for club cell secretory protein-16 (CC16) in the development of COPD. *Eur. Respir. J.* **45**, 1544–1556 (2015).
35. Mashima, I. & Nakazawa, F. The interaction between *Streptococcus* spp. and *Veillonella tobetsuensis* in the early stages of oral biofilm formation. *J. Bacteriol.* **197**, 2104–2111 (2015).
36. van den Bogert, B., Meijerink, M., Zoetendal, E. G., Wells, J. M. & Kleerebezem, M. Immunomodulatory properties of *Streptococcus* and *Veillonella* isolates from the human small intestine microbiota. *PLoS ONE* **9**, e114277 (2014).
37. Ng, S. K. C. & Hamilton, I. R. Lactate metabolism by *Veillonella parvula*. *J. Bacteriol.* **105**, 999–1005 (1971).
38. Scheiman, J. et al. Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism. *Nat. Med.* **25**, 1104–1109 (2019).
39. Mammen, M. J. & Sethi, S. COPD and the microbiome. *Respirology* **21**, 590–599 (2016).
40. O'Neal, W. K. et al. Comparison of serum, EDTA plasma and P100 plasma for luminex-based biomarker multiplex assays in patients with chronic obstructive pulmonary disease in the SPIROMICS study. *J. Transl. Med.* **12**, 9 (2014).
41. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
42. Korbie, D. J. & Mattick, J. S. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat. Protoc.* **3**, 1452–1456 (2008).
43. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
44. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
45. Cole, J. R. et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**, D633–D642 (2014).
46. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).
47. Kembel, S. W. et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
48. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
49. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).
50. Anderson, M. J. A new method for non-parametric multivariate analysis of variance: NON-PARAMETRIC MANOVA FOR ECOLOGY. *Austral Ecol.* **26**, 32–46 (2001).
51. Anderson M. J. Permutational Multivariate Analysis of Variance (PERMANOVA) (Wiley Online Library, 2017).
52. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

## ACKNOWLEDGEMENTS

The authors thank the SPIROMICS participants and all participating physicians, investigators, and staff for making this research possible. More information about the study and how to access SPIROMICS data is at [www.spiromics.org](http://www.spiromics.org). We would like to acknowledge the following current and former investigators of the SPIROMICS sites and reading centers: Neil E Alexis, MD; Wayne H Anderson, PhD; Mehrdad Arjomandi, MD; Igor Barjaktarevic, MD, PhD; R Graham Barr, MD, DrPH; Lori A Bateman, MSc; Surya P Bhatt, MD; Eugene R Bleeker, MD; Richard C Boucher, MD; Russell P Bowler, MD, PhD; Stephanie A Christenson, MD; Alejandro P Comellas, MD; Christopher B Cooper, MD, PhD; David J Couper, PhD; Gerard J Criner, MD; Ronald G Crystal, MD; Jeffrey L Curtis, MD; Claire M Doerschuk, MD; Mark T Dransfield, MD; Brad Drummond, MD; Christine M Freeman, PhD; Craig Galban, PhD; MeiLan K Han, MD, MS; Nadia N Hansel, MD, MPH; Annette T Hastie, PhD; Eric A Hoffman, PhD; Yvonne Huang, MD; Robert J Kaner, MD; Richard E Kanner, MD; Eric C Kleerup, MD; Jerry A Krishnan, MD, PhD; Lisa M LaVange, PhD; Stephen C Lazarus, MD; Fernando J Martinez, MD, MS; Deborah A Meyers, PhD; Wendy C Moore, MD; John D Newell Jr, MD; Robert Paine, III, MD; Laura Paulin, MD, MHS; Stephen P Peters, MD, PhD; Cheryl Pirozzi, MD; Nirupama Putcha, MD, MHS; Elizabeth C Oelsner, MD, MPH; Wanda K O'Neal, PhD; Victor E Ortega, MD, PhD; Sanjeev Raman, MBBS, MD; Stephen I. Rennard, MD; Donald P Tashkin, MD; J Michael Wells, MD; Robert A Wise, MD; and Prescott G Woodruff, MD, MPH. The project officers from the Lung Division of the National Heart, Lung, and Blood Institute were Lisa Postow, PhD, and Lisa Viviano, BSN; SPIROMICS was supported by contracts from the NIH/NHLBI (HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C, HHSN268200900019C, HHSN268200900020C), grants from the NIH/NHLBI (U01 HL137880 and U24 HL141762), and supplemented by contributions made through the Foundation for the NIH and the COPD Foundation from AstraZeneca/MedImmune; Bayer; Bellerophon Therapeutics; Boehringer-Ingelheim Pharmaceuticals, Inc.; Chiesi Farmaceutici S.p.A.; Forest Research Institute, Inc.; GlaxoSmithKline; Grifols Therapeutics, Inc.; Icaria, Inc.; Novartis Pharmaceuticals Corporation; Nycomed GmbH; ProterixBio; Regeneron Pharmaceuticals, Inc.; Sanofi; Sunovion; Takeda Pharmaceutical Company; and Theravance Biopharma and Mylan. Additional support for this work came from R01HL121774-S1 (to Y.J.H. and G.B.H.), R01AI129958 (Y.J.H.), R01HL121774 (to G.B.H. and P.G.W.), R03HL138310 (Y.J.H.).

## AUTHOR CONTRIBUTIONS

Y.J.H., K.O., L.A.B., S.M., G.B.H., and J.R.E.D.: designed and executed sample processing and data analyses. Y.J.H., L.A.B., and K.O.: drafted first version of the manuscript. K.O. and L.A.B. contributed equally as co-first authors. P.G.W.: chaired the SPIROMICS Bronchoscopy committee that developed the original protocol, together with committee members. D.J.C.: oversaw the SPIROMICS data coordinating center. J.L.C.: oversaw BAL immune cell analyses. All other authors led efforts to enroll subjects, obtain samples, and data used in this study. All authors reviewed and contributed to revisions of the manuscript for critically important intellectual content and approved the version of the manuscript to be published.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41522-021-00185-9>.

**Correspondence** and requests for materials should be addressed to Y.J.H.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021