Airway microbiome in adult survivors of extremely preterm birth: the EPICure study

Sylvia A.D. Rofael[1,2], Timothy D. McHugh[1], Rachael Troughton[1], Joanne Beckmann[3], David Spratt[4], Neil Marlow[3] and John R. Hurst[5]

Affiliations: [1]Centre for Clinical Microbiology, Division of Infection and Immunity, UCL, London, UK. [2]Faculty of Pharmacy, University of Alexandria, Alexandria, Egypt. [3]Academic Neonatology, UCL Elizabeth Garret Anderson Institute for Women's Health, UCL, London, UK. [4]Dept of Microbial Diseases, UCL Eastman Dental Institute, UCL, London, UK. [5]UCL Respiratory, University College London, London, UK.

Correspondence: John R. Hurst, UCL Respiratory, University College London, 1st floor Royal Free Campus, Pond Street, London, NW4 2PF, UK. E-mail: j.hurst@ucl.ac.uk

ABSTRACT Bronchopulmonary dysplasia (BPD) is a major complication of preterm birth that leads to lifelong respiratory morbidity. The EPICure study has investigated the longitudinal health outcomes of infants born extremely preterm (EP; <26 weeks gestation). Our aim was to characterise the airway microbiome in young adults born extremely preterm, with and without neonatal BPD, in comparison to matched term-born controls.

Induced sputum was collected from 92 young adults aged 19 years (51 EP and 41 controls). Typical respiratory pathogens were detected using quantitative PCR. 16S rRNA gene sequencing was completed on 74 samples (29 EP with BPD; 9 EP without BPD; and 36 controls).

The preterm group with BPD had the least diverse bacterial communities. The relative abundance of Bacteriodetes, particularly *Prevotella melaninogenica* was significantly lower in the preterm group compared to controls. This decline was balanced by a nonsignificant increase in Firmicutes. Total *Prevotella* relative abundance correlated with forced expiratory volume in 1 s z-score (ρ=0.272; p<0.05). Typical respiratory pathogen loads and prevalence were similar between groups.

In conclusion, extremely preterm birth is associated with a significant dysbiosis in airway microbiome in young adulthood regardless of neonatal BPD status. This is characterised by a shift in the community composition away from Bacteriodetes as manifested in a significant drop in *Prevotella* relative abundance.

Introduction

Global and national survival rates of extremely preterm birth have steadily increased over the past decades with advances in perinatal and neonatal care; however, the prevalence of bronchopulmonary dysplasia (BPD) remains high [1, 2]. BPD is a multifactorial lung disease that develops primarily in preterm infants and may lead to lifelong respiratory morbidity. Initially, the aetiology was thought to be mechanical lung injury, but more recently in more immature infants a "new" form of BPD has been defined, characterised by disrupted distal lung development, arrest of alveolarisation and interference with normal vascularisation [3]. The long-term sequelae of BPD include obstructive lung disease and reduced aerobic capacity in adulthood, which can be confused with, and may meet spirometric criteria for asthma and chronic obstructive pulmonary disease (COPD) [4].

The EPICure study is a national cohort study investigating the health outcomes of babies born at <26 weeks of gestation in the United Kingdom and Ireland between March and December 1995. This unique cohort has been followed-up and assessed at 2.5, 6, 9, 11 and now 19 years of age [2, 5, 6]. Aged 11 years, children who were born extremely preterm had significantly more chest deformities, respiratory symptoms, lung function abnormalities with evidence of airway obstruction, ventilation inhomogeneity, gas trapping and airway hyperresponsiveness and twice the prevalence of asthma, compared to their classmates who were born full-term [5, 7].

Studies have provided evidence that bacteria may play a role in the development of BPD in preterm infants [8–11]. One longitudinal study described a characteristic pattern of airway microbial dysbiosis prior to the development of BPD. This was characterised by a remarkable decrease in the richness and α-diversity with time; in addition to a shift in the bacterial community composition, in contrast to a relatively diverse and stable community in the preterm infants who did not develop BPD [12]. However, the extent to which microbial succession sustains the dysbiosis of the airway microbiome described in infancy into later life stages has not been investigated. In addition, how the airway microbiome of BPD survivors differs from the healthy microbiome in adulthood is not known. The aim of this study is to characterise the airway microbiome in the EPICure cohort at the age of 19 years and compare it with the healthy microbiome of their counterparts who were born full term.

Materials and methods

The EPICure study was approved by the Southampton and South West Hampshire research ethics committee [5]. All participants gave informed consent for tests to be performed on their biological samples in relation to the EPICure@19 study.

Details on sputum sample collection and processing are described in the supplementary material. Briefly, induced sputum was collected from 92 participants in the EPICure@19 study during their follow-up visit aged 19 years. The average percentage of squamous epithelial cells in one aliquot of each sputum sample was measured by microscopy for quality control.

Samples were allocated to three groups based on the medical history of participants; those who were born extremely preterm without BPD history (EP no BPD); those who were born extremely preterm with neonatal BPD (EP + BPD); and full term-born participants of the same age, who had been evaluated as classmates in earlier studies (EPICure control group). Microbiome investigators were blinded to the group assignment until analysis was complete.

Metagenomic DNA was extracted from 500 μL of Sputasol-treated sputum samples using Qiagen DNeasy Blood and Tissue kit (Qiagen, Manchester, UK) as per the manufacturer's protocol. Extraction negative controls of the saline used for sputum induction and diluted Sputasol were also performed. A mock community composed of equal proportions of *Streptococcus pneumoniae* ATCC 49619, *Haemophilus influenzae* ATCC 8468 and *Moraxella catarrhalis* ATCC 25240 was run as a positive control.

The bacterial loads of *H. influenzae*, *M. catarrhalis* and *S. pneumoniae* were quantified using multiplex quantitative (q)PCR, as previously described [13]. The bacterial load of each tested organism was calculated in colony forming units per mL (CFU·mL^{-1}) of sputum and a mean of technical triplicates was taken for each sample. Additional details are provided in the supplementary material.

A sequence library was created by amplification of V5–V7 regions of the bacterial 16S *rrna* gene from metagenomic DNA using 785 forward primer and 1175 reverse primer. Each sample was assigned a unique pair combination of standard Illumina dual indexed primers (Sigma-Aldrich, Gillingham, UK), 74 samples produced an amplicon at the expected size (504 bp). The PCR products were cleaned up using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK) to remove amplicons <200 bp, DNA was quantified using Qubit dsDNA HS kit (Invitrogen, Life Technologies, Paisley, UK). The samples were pooled in equimolar ratio at 10 nM. Sequencing was performed using Illumina MiSeq Platform using

costume sequencing primers, MiSeq Reagent Kit v2 (500 cycles) and PhiX control V3 KIT as internal control for sequencing (Illumina Inc., Eindhoven, Netherlands). The extraction negative controls and a no-template PCR control (water) were run throughout the amplification and sequencing process to reveal any potential contamination. Additional details on the PCR and primers sequences are provided in the supplementary material. Sequence data are deposited in the European Nucleotide Archive; study accession number is PRJEB27216.

In bioinformatics analysis, we adopted the workflow established by Microbiome Helper for stitching paired reads, quality filtering reads with Q<30 over 10% of bases and length <350 bp, and chimera screening [14]. The subsequent steps were done through QIIME pipeline (v.1.9.1) [15], the sequences were clustered based on 97% similarity into operational taxonomic units (OTUs) and taxonomic classification was assigned to OTUs using open-reference OTU picking against Greengenes database (v.13_8). The OTU table was then rarefied per sample to 4000 reads removing four samples with <1000 reads (1 EP + BPD and three controls). α- and β-diversity indices were calculated on the rarefied OTU table. The appropriate statistical significance tests were calculated using SPSS (v.23) or QIIME wrapper scripts. STAMP (v2.1.3) [16] was used to visualise the data. Whenever applicable the p-values were corrected using Benjamini–Hochberg false discovery rate method for multiple comparisons.

Results

General characteristics of the participants

Induced sputum samples were collected from 92 young-adult participants; 51 were born extremely preterm at <26 weeks' gestation and 41 were born full-term (control group). Mean±SD age was 19±0.5 years. Microbiome analysis was completed on 74 samples with an amplified 16S rRNA gene: 36 controls and 38 extremely preterm; 29 of whom had neonatal BPD, defined as receiving supplemental oxygen or respiratory support at 36 weeks postmenstrual age. The demographic clinical and medical data of the whole cohort and the 74 participants whose samples were sequenced recorded at the 19-year follow-up visit are reported in table 1.

Within the sequenced cohort, forced expiratory volume in 1 s (FEV1) was significantly lower in the EP + BPD group compared to controls (mean difference −0.91 L, 95% CI −1.24−−0.59 L) and the EP no BPD group (mean difference −0.81 L, 95% CI −1.31−−0.32 L). In addition, after adjustment for age, sex and body size using z-scores, the mean FEV1 z-score of the EP + BPD group was significantly lower compared to the control group and the EP no BPD group; mean differences (95% CI) were −1.55 (−2.05−−1.05) and −1.13 (−1.88−−0.37), respectively.

The prevalence of self-reported asthma was relatively higher in the EP + BPD group (59%) and controls (38%), compared to the EP no BPD group (22%), although this did not reach statistical significance. As discussed further later, it is likely that there is significant overdiagnosis of "asthma" here, but we did not consider it ethical to stop asthma treatment for the purposes of research. At the time of sample collection, the mean±SD exhaled nitric oxide fraction (F_{eNO}) concentration was 16.59±14.10 ppb and 25.57±27.71 ppb in the preterm and control groups, respectively (p>0.05). The mean±SD eosinophil counts in blood were 190±136 cells·µL^{-1} and 232±156 cells·µL^{-1} in the preterm group and controls, respectively (p>0.05). 47% of the EP group and 50% of the control group were prescribed inhalers. No statistically significant differences were found across the three groups with respect to the number of patients who were prescribed bronchodilator inhalers, inhaled corticosteroids or those who had been treated with antibiotics for respiratory problems in the year prior to sample collection. The numbers of males and females, smokers and those who were exposed to passive smoking (>30 min per week) were similar across the three groups.

Microbial community composition

The bacterial communities were significantly less diverse and less rich in the sputum samples from the whole EP group compared to controls; the mean±SEM difference in Chao 1 and Fisher α indices between the whole EP group and the controls were −39±13 (p<0.05) and −4.8±2.3 (p<0.05), respectively (supplementary figure E1). Other richness and α-diversity indices including the number of observed OTUs and PD whole tree also showed significantly less diverse and less rich microbiota in the EP group (supplementary figure E2). Within the EP group, the trend observed in all previously mentioned α-diversity indices suggests that those with neonatal BPD had the least diverse and rich microbial communities, while controls had the highest values (figure 1a and b and supplementary figure E2). This trend was statistically significant only in Chao 1 when tested by ANOVA (p<0.05). Nevertheless, the *post hoc* comparisons of differences between the EP group with neonatal BPD and controls were significant (figure 1).

In principal coordinate analysis of weighted Unifrac β-diversity index, the samples from EP participants clustered significantly, regardless of neonatal BPD status; whereas the samples from controls were scattered (p<0.01 by ANOSIM and p<0.05 by PERMANOVA) (figure 1c and supplementary figure E1).

TABLE 1 Demographic and clinical characteristics of all EPICure participants and those with sequenced samples

| | Whole cohort[#] | | | | Sequenced samples[#] | | | |
| | EP[#] | | Controls | p-value | EP[#] | | Controls | p-value |
	BPD	No BPD			BPD	No BPD		
Subjects n	37	14	41		29	9	36	
Age years	19.02±0.54	19.02±0.41	19.09±0.52	0.803[§]	18.93±0.66	18.82±0.48	19.02±0.55	0.474[f]
Males	35	43	39	0.866[##]	34	56	42	0.521[##]
Females	65	57	61		66	44	58	
Asthma diagnosis	57	21*	37	0.047[##]	59	22	38	0.094[##]
Current smoker	22	29	27	0.821[##]	28	33	33	0.873[##]
Passive smoke exposure >30 min·week^{-1}	22	29	35	0.471[##]	32	22	60	0.348[##]
Squamous epithelial cell %	17.4±17.7	17.9±11.9	13.8±11.1	0.573[f]	20.9±19.2	19.5±14.4	14.6±11.3	0.566[f]
Prescribed inhalers[¶]	60	38	64	0.557[¶¶]	62	17	50	0.239[¶¶]
Prescribed bronchodilator inhalers[+]	53	25	55	0.417[¶¶]	54	17	50	0.298[¶¶]
Prescribed ICS	27	13	36	0.476[¶¶]	31	17	33	0.883[¶¶]
Antibiotic treatment in past year	7	13	30	0.397[¶¶]	8	0	27	0.395[¶¶]
Treated for respiratory problem in the past year	20	21	25	0.393[¶¶]	22	11	28	0.556[##]
FEV1 L	2.66**±0.54	3.22±0.76	3.57±0.65	0.000[§]	2.63**±0.66	3.45±0.76	3.55±0.61	0.000[§]
FEV1 z-score	−1.66**±1.09	−0.911±0.04	−0.37±0.87	0.000[§]	−1.87**±1.17	−0.75±0.69	−0.32±0.89	0.000[§]
Percentage change in FEV1 with bronchodilator	7.93±6.25	7.75±7.81	5.26±5.60	0.077[§]	9.52±7.77	7.19±4.91	5.50±5.85	0.058[§]
F_{eNO} ppb	16.33±12.63	18.00±14.72	26.47±26.30	0.683[f]	16.50±13.98	16.89±15.34	25.57±27.71	0.830[f]
Eosinophils cells·µL^{-1}	181±127	164±104	229±148	0.184[f]	194±141	179±128	232±156	0.399[f]

Data are presented as mean±SD or %, unless otherwise stated. EP: extremely preterm birth; BPD: bronchopulmonary dysplasia; ICS: inhaled corticosteroids; FEV1: forced expiratory volume in 1 s; F_{eNO}: exhaled nitric oxide fraction. [#]: whole cohort n=92, of which EP n=51, sequenced samples n=74, of which EP n=38; [¶]: β_2-adrenoreceptor agonists: salbutamol (Ventolin), terbutaline (Bricanyl), salmeterol (Serevent, Seretide); muscarinic receptor antagonist: ipratropium (Atrovent); leukotriene receptor antagonist: montelukast (Singulair); ICS: beclomethasone (Becotide), budesonide (Pulmicort), fluticasone (Flixotide, Seretide); [+]: β_2-agonists and muscarinic receptor antagonists; [§]: calculated by ANOVA; [f]: p-value calculated by Kruskal–Wallis test; [##]: p-value calculated by Chi-squared test; [¶¶]: p-value calculated by Fisher's exact test. *: p<0.05; **: p<0.01.

The bacterial community at phylum level was dominated by Firmicutes, followed by Bacteriodetes, then Proteobacteria and Actinobacteria. The samples from both EP groups, with and without BPD, had a significantly lower relative abundance (RA) of the phylum Bacteriodetes compared to the control group (p<0.05 by Kruskal–Wallis test). Differences were compensated by a nonsignificant increase in the relative abundance of Firmicutes (figure 2a).

Looking more closely at the composition of the microbial communities at genus level, although natural interindividual differences were obvious in microbiome profiles of sputum from different participants, the relative abundance of *Prevotella* was significantly lower in both EP groups, with and without BPD, in comparison to the control group (p<0.05 by Kruskal–Wallis test) (figures 2b and 3a). This was compensated for by a nonsignificant and inconsistent increase in relative abundance of other genera such as *Streptococcus*, *Veillonella*, *Rothia* and *Neisseria*, which are normal microbiota in airways (figure 2b).

Prevotella was completely absent in two negative controls and present at 0.4% RA in the extraction negative control of the sputum induction matrix, which had 5198 reads classified into three main taxa: *Propionibacterium* (RA 62%), *Staphylococcus* (RA 14%) and *Streptococcus* (RA 6%) (supplementary figure E6). The Sputasol extraction negative control had 17 reads and the no-template PCR negative control had two reads; in contrast, the mean±SD from the study samples was 24089±5751 reads (figure 2b). This gives confidence that the impact of environmental contamination was minimal. The relative abundance of *Prevotella* did not correlate with the percentage of squamous epithelial cells in the sputum samples (Spearman ρ=0.07, p=0.61). Nevertheless, *Prevotella* did correlate significantly with the FEV1 z-score (Spearman ρ=0.272, p=0.02) (figure 3b), but not with a self-reported diagnosis of asthma or use of asthma inhalers (supplementary material).

Mining the sequencing data, out of 121 operational taxonomic units (OTUs) belonging to genus *Prevotella*, the relative abundance of OTU 4458304 accounted for most of the observed difference in the
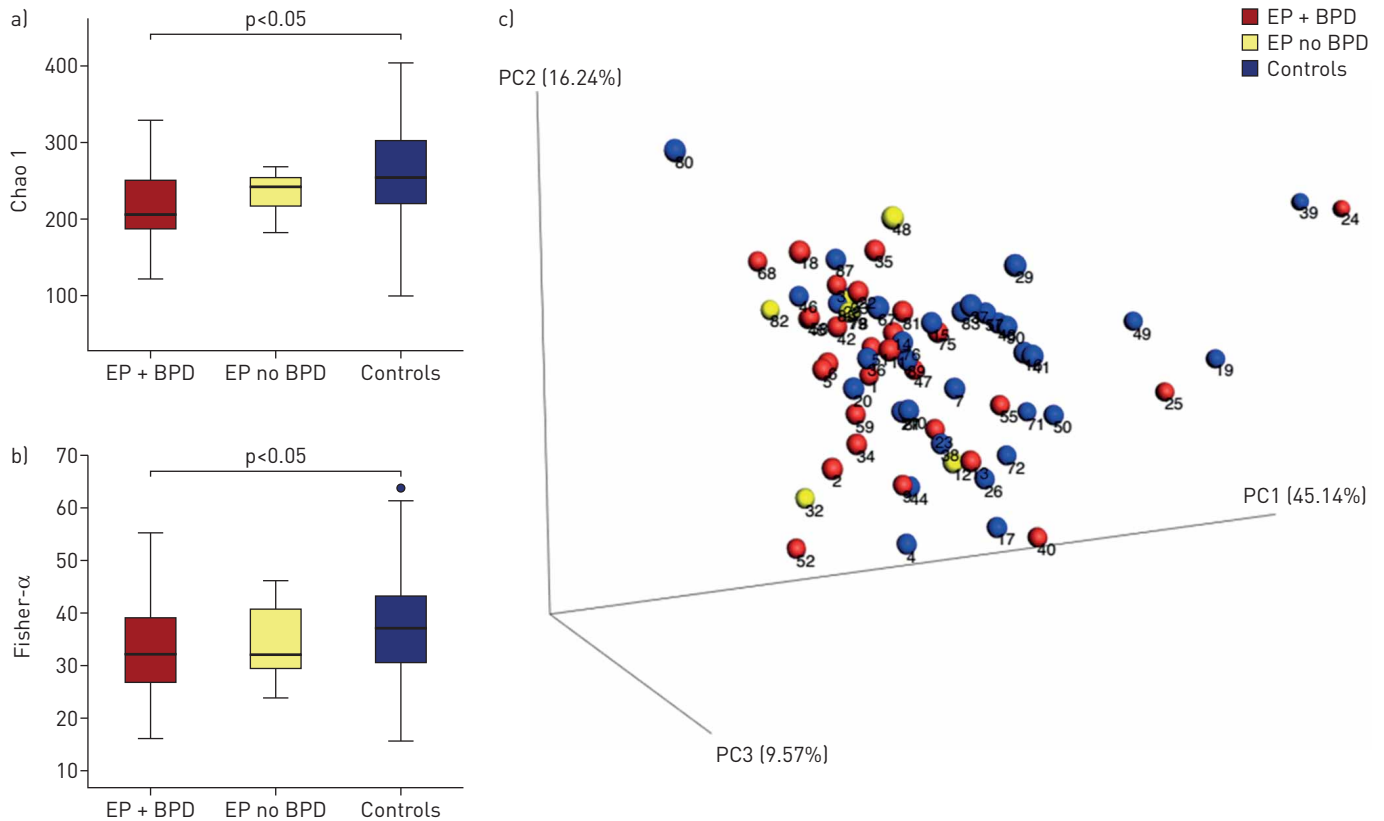
FIGURE 1 Comparison of the richness and α-diversity of microbial communities in sputum between the extreme preterm birth survivors (EP), with and without neonatal bronchopulmonary dysplasia (BPD), and controls. Richness and α-diversity measured by a) Chao 1 (p<0.05 by ANOVA); b) Fisher α-diversity index (p=0.07 by ANOVA); nevertheless, Fisher-α was significantly lower in the whole EP group compared to controls (p<0.05 by t-test). c) Principal coordinate analysis of weighted UniFrac β-diversity index (p<0.01 by ANOSIM and p<0.05 by PERMANOVA comparing the whole EP group (n=37) and controls (n=33); p>0.05 by ANOSIM and PERMANOVA comparing the three groups: EP + BPD (n=28), EP no BPD (n=9), controls (n=33)).

total *Prevotella* relative abundance between the study groups (figure 3c). By extracting and basic local alignment search tool (BLAST)-searching the 350-bp representative sequence of this OTU against the National Center for Biotechnology Information (NCBI) 16S ribosomal RNA database using the nucleotide BLAST tool [17], the sequence was identified as 100% identical to *Prevotella melaninogenica* strains (figure 3d).

Load and prevalence of airway bacteria using multiplex qPCR

The loads and prevalence of the three bacteria *S. pneumoniae*, *H. influenzae* and *M. catarrhalis* were similar within the three study groups (figure 4). In the EP + BPD group, 28% had *S. pneumoniae*, 14% had *H. influenzae* and 11% had both organisms together; a similar pattern was observed in controls. The mean±SD bacterial load of *S. pneumoniae* and *H. influenzae* were similar in these two groups; 4.44 ± 0.1 $\log_{10}$ CFU·mL^{-1}. In the EP no BPD group history, *H. influenzae* was the most prevalent organism, detected in 57% of the sputum samples from this group with a mean±SD load of 4.51 ± 0.88 $\log_{10}$ CFU·mL^{-1} of sputum. It was found in 36% of the samples alone and in 21% of the samples with *S. pneumoniae*. *S. pneumoniae* was the most populous, with a mean±SD bacterial load of 4.96 ± 1.08 $\log_{10}$ CFU·mL^{-1} (figure 4).

M. catarrhalis was the least prevalent and the least populous organism. The mean load of *M. catarrhalis* was 0.7 $\log_{10}$ higher in the EP + BPD group compared to the other two groups (p>0.05, ANOVA). None of the differences in prevalence and load of the three organisms across the study groups were statistically significant.

There was a significant correlation between the sequencing relative abundance of genus *Moraxella* and the corresponding qPCR bacterial loads (ρ=0.472, p<0.01; figure 5c). However, the correlation was not statistically significant for *Haemophilus* and *Streptococcus* (figure 5a and d). This can be attributed to the presence of commensal members of these two genera in sputum as normal respiratory microbiota. Within
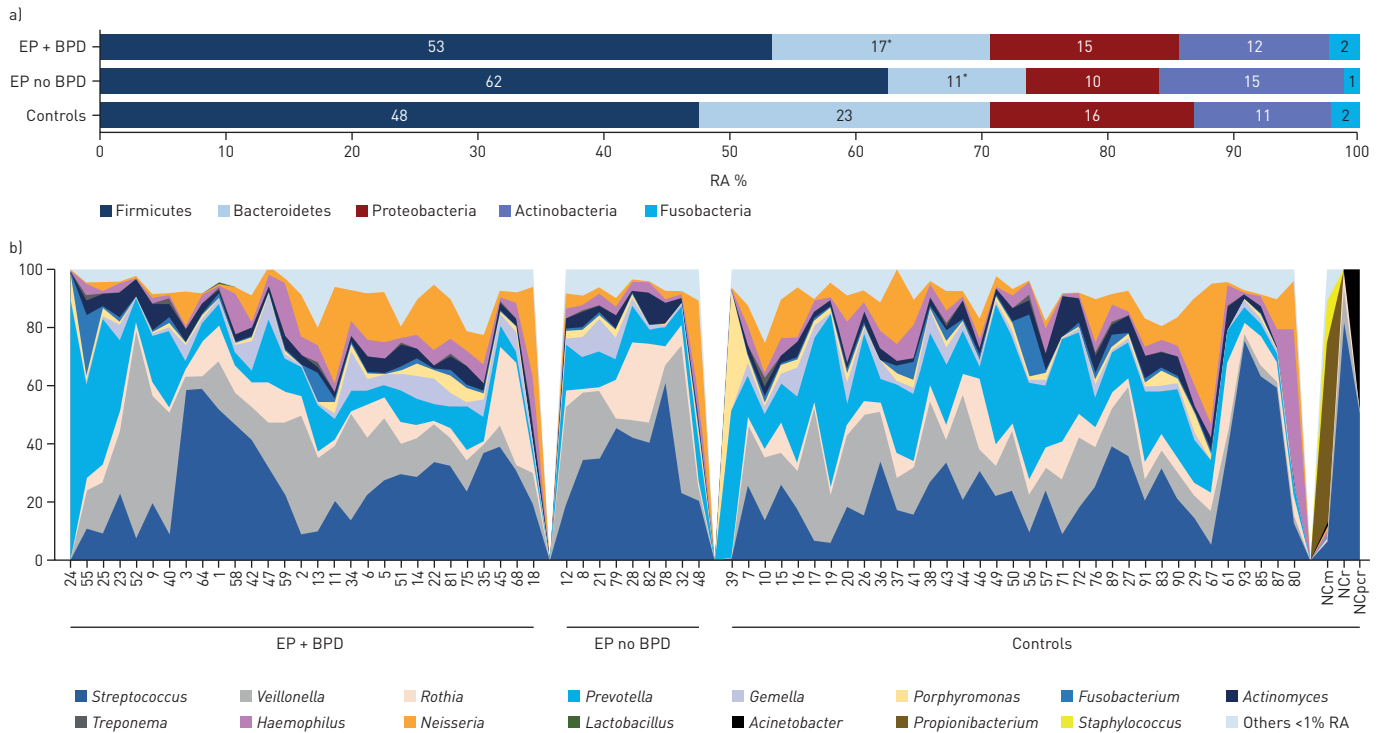
FIGURE 2 Comparison of the airway microbiome profile in the EPICure groups at a) phylum level and b) genus level. a) At phylum level, the Bacteriodetes relative abundance (RA) was significantly lower in both extremely preterm-born (EP) groups compared to controls (p<0.05 by Kruskal–Wallis test); b) at the genus level, the *Prevotella* relative abundance was significantly lower in EP groups compared to controls (p<0.05 by Kruskal–Wallis test). NCm: extraction negative control of the saline matrix used for sputum induction; NCr: extraction negative control of the diluted Sputasol and reagents; NCpcr: PCR negative control. Sample size: EP + BPD n=29, EP no BPD n=9 and controls n=36. *: p<0.05 by Mann–Whitney test.

the genus *Haemophilus*, five OTUs were classified as *H. influenzae*, and the correlation between the relative abundance of *H. influenzae* OTUs in sequencing and *H. influenzae* load in qPCR was statistically significant ($\rho=0.43$, $p<0.01$; figure 5b). Comparing the relative abundance by sequencing to the qPCR loads and bacterial count by quantitative cultures in the mock community, *Streptococcus* was overrepresented by 15% and *Haemophilus* was underrepresented by 21% (figure 5d). Further data on the sensitivity and specificity of the methods are provided in the supplementary material.

Discussion

To the best of our knowledge, this is the first study to investigate the airway microbiome in adult survivors of preterm birth in comparison to matched full term born controls. Thus far, previous studies investigating the association between bacteria and BPD involved preterm-born infants [10–12, 18, 19]. In the current study, we have compared the airway microbiome in 38 young adults who were born at <26 weeks' gestation (29 of them had neonatal BPD and nine did not) and 36 term-born controls, recruited from classmates at ages 6 and 11 years.

The airway microbiome profile in the three study groups was consistent with previous studies [20, 21]. It was dominated by Firmicutes, Bacteriodetes, Proteobacteria and Actinobacteria. The extremely preterm group had significantly less diverse and less rich microbial communities in comparison with the control group. We did not find significant differences between the airway microbiome profiles of the extremely preterm groups with and without neonatal BPD. The samples from both groups clustered together in principal coordinate analysis of weighted Unifrac β-diversity index regardless of neonatal BPD status (figure 1). No significant differences were observed in the richness and α-diversities between the two groups; however, a trend was observed in which the α-diversity and richness of the microbial communities in the BPD group was slightly, but not significantly lower than the group without BPD, and significantly lower than the control group. This trend may be important, but our study did not have sufficient statistical power to be able to confirm a difference in all microbial diversity indices. We demonstrated a significant shift away from Bacteriodetes, driven particularly by reduction in the relative abundance of genus *Prevotella* in both extremely preterm groups, with and without BPD, relative to the control group. Moreover, *Prevotella* relative abundance correlated significantly with FEV1 z-score, but had no association
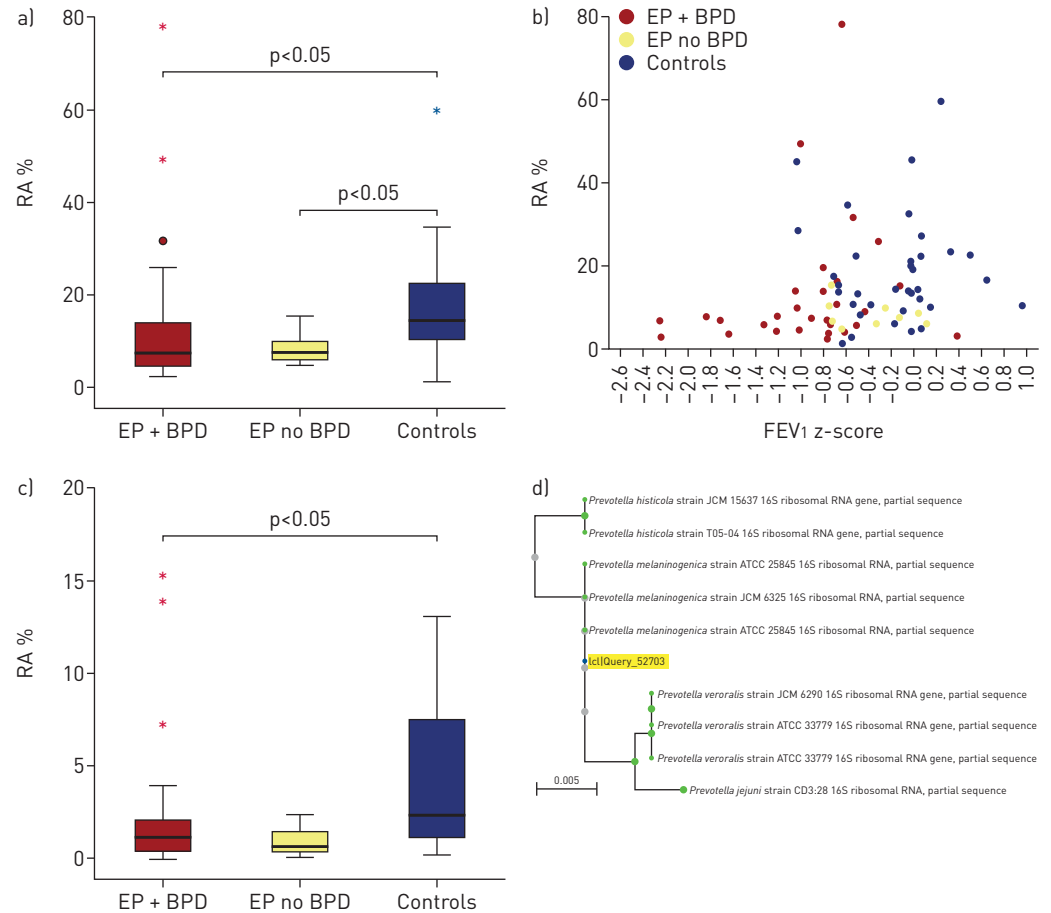
FIGURE 3 Comparison of *Prevotella* relative abundance (RA) across study groups. a) Genus *Prevotella* RA was significantly lower in both extremely preterm (EP) groups regardless of neonatal bronchopulmonary dysplasia (BPD) status compared to controls (p<0.05 by Kruskal–Wallis test); b) genus *Prevotella* RA correlated with the forced expiratory volume in 1 s (FEV1) z-score (Spearman ρ=0.272, p=0.02); c) operational taxonomic unit (OTU) 4458304 contributed most to the observed difference in genus *Prevotella* RA across the study groups (p<0.05 by Kruskal–Wallis test); d) phylogenetic tree of OTU 4458304 was 100% identical to *Prevotella melaninogenica* strains as obtained by basic local alignment search tool (BLAST) analysis of the representative sequence against National Center for Biotechnology Information (NCBI) 16S ribosomal RNA database. Sample size: EP + BPD n=29, EP no BPD n=9 and controls n=36.

with other clinical parameters such as smoking status, exposure to passive smoking, self-reported diagnosis of asthma, FeNO, blood eosinophil count or use of inhalers.

Observational studies that have investigated the lung microbiome have commonly detected *Prevotella* in lung tissues and bronchoalveolar lavage of healthy subjects [20, 22, 23]. It is suggested that Bacteriodetes relative abundance is linked with healthy lung microbiome [24]. In addition, an association between reduced relative abundance of Bacteriodetes and chronic lung disease has been frequently reported [23–26]. A shift in community composition away from Bacteriodetes towards Proteobacteria has been observed in people with COPD [27, 28] and those with asthma [28]. A similar trend in which there is a shift from Bacteriodetes to Firmicutes, mainly streptococci, has been reported by ZHANG *et al.* [29] in severe asthma. HILTY *et al.* [28] reported a shift in the community membership away from Bacteriodetes (mainly *Prevotella*), towards Proteobacteria (mainly *Haemophilus*), as well as Firmicutes in children with asthma. In the study by LAL *et al.* [30], which compared the airway microbiome at birth in both preterm- and full term born infants, *Prevotella* was detected in tracheal aspirates at day one of life. In their results, *Prevotella* RA was lower in extremely low birthweight preterm infants compared to controls. Among the preterm infants, those who were predisposed to BPD had the least *Prevotella* abundance; however, this trend was not statistically significant.

We analysed the sequencing data to go beyond the genus level and found that *P. melaninogenica* was the species that contributed most to the observed reduction in total *Prevotella* abundance in the preterm groups in comparison to the control group. *Prevotella* species are obligate anaerobes that have been
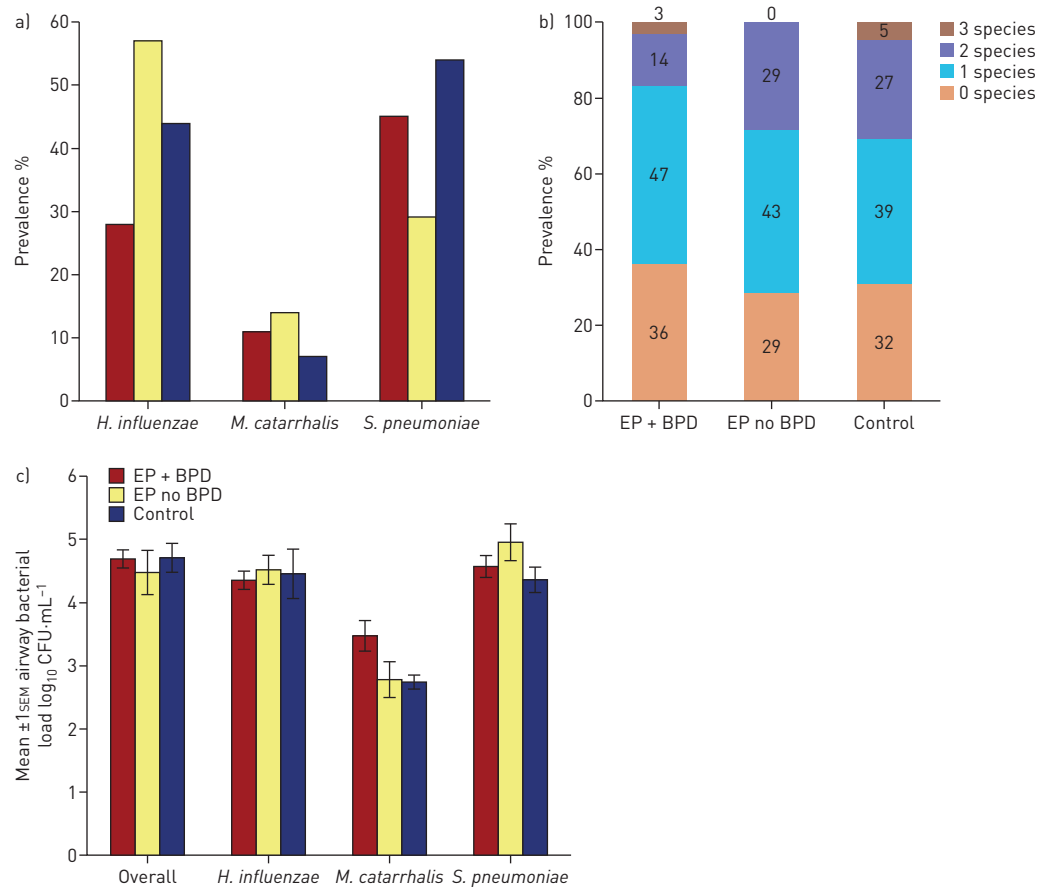
FIGURE 4 Prevalence and loads of pathogenic airway bacteria: *Haemophilus influenzae*, *Moraxella catarrhalis* and *Streptococcus pneumoniae*. a) Prevalence of the three airway bacteria within the three study groups ($p > 0.05$, Chi-squared test); b) co-existence of the three bacteria in the sputum samples in each group ($p > 0.05$, Fisher's exact test); c) mean bacterial load of each of the three bacteria as determined by the multiplex quantitative PCR ($p > 0.05$, MANOVA). Loads ($CFU \cdot mL^{-1}$) of original sputum sample for each bacterium and the sum of the three (overall) were calculated for each sample, then means were calculated for each study group. Samples which gave negative results for a given bacterium were excluded from the analysis. Sample size: extreme preterm birth (EP) + bronchopulmonary dysplasia (BPD) n=37, EP no BPD n=14, controls n=41.

regarded as opportunistic members of the oral microbiota as well as other body sites and have been isolated with other bacteria in mixed anaerobic infections and lower respiratory tract infections [31]. Increased abundance of *Prevotella* has been associated with some diseases such as periodontitis, rheumatoid arthritis, bacterial vaginosis and inflammatory bowel disease [32]. Nevertheless, other *Prevotella* species, *e.g. P. intermedia* and *P. nigrescens* are usually associated with disease [33]. Very little has been published on *P. melaninogenica*, but this species was the most frequently isolated from the sputum in cystic fibrosis in previous studies and it was noticed that the presence of anaerobes was associated with clinical stability [34, 35].

Currently, it is not clear what role *Prevotella* plays in lower respiratory microbial homeostasis. Many studies that compared the lung microbiome in health and disease suggest that *Prevotella* is associated with health, and is quickly replaced by members of Proteobacteria or Firmicutes in various chronic lung conditions [24, 26–29]. Conversely, some studies have proposed that *Prevotella* might be contributing to the pathogenesis of lung diseases [32, 36–39]. However, caution must be exercised in extrapolating observations to the whole genus. The role of *Prevotella* species in pathogenesis has received little attention, possibly because *Prevotella* is difficult to culture and is not usually isolated from specimens in routine microbiology laboratories. Consequently, further research is required to understand the interactions of *Prevotella* species with the host immune system and with other microbes within the lung microbial communities, and to characterise the strains that may be beneficial to respiratory health.

Conventionally, *S. pneumoniae*, *H. influenzae* and *M. catarrhalis* have been classified as typical airway pathogens. These organisms normally reside harmlessly within the human nasopharynx [40]. Numerous
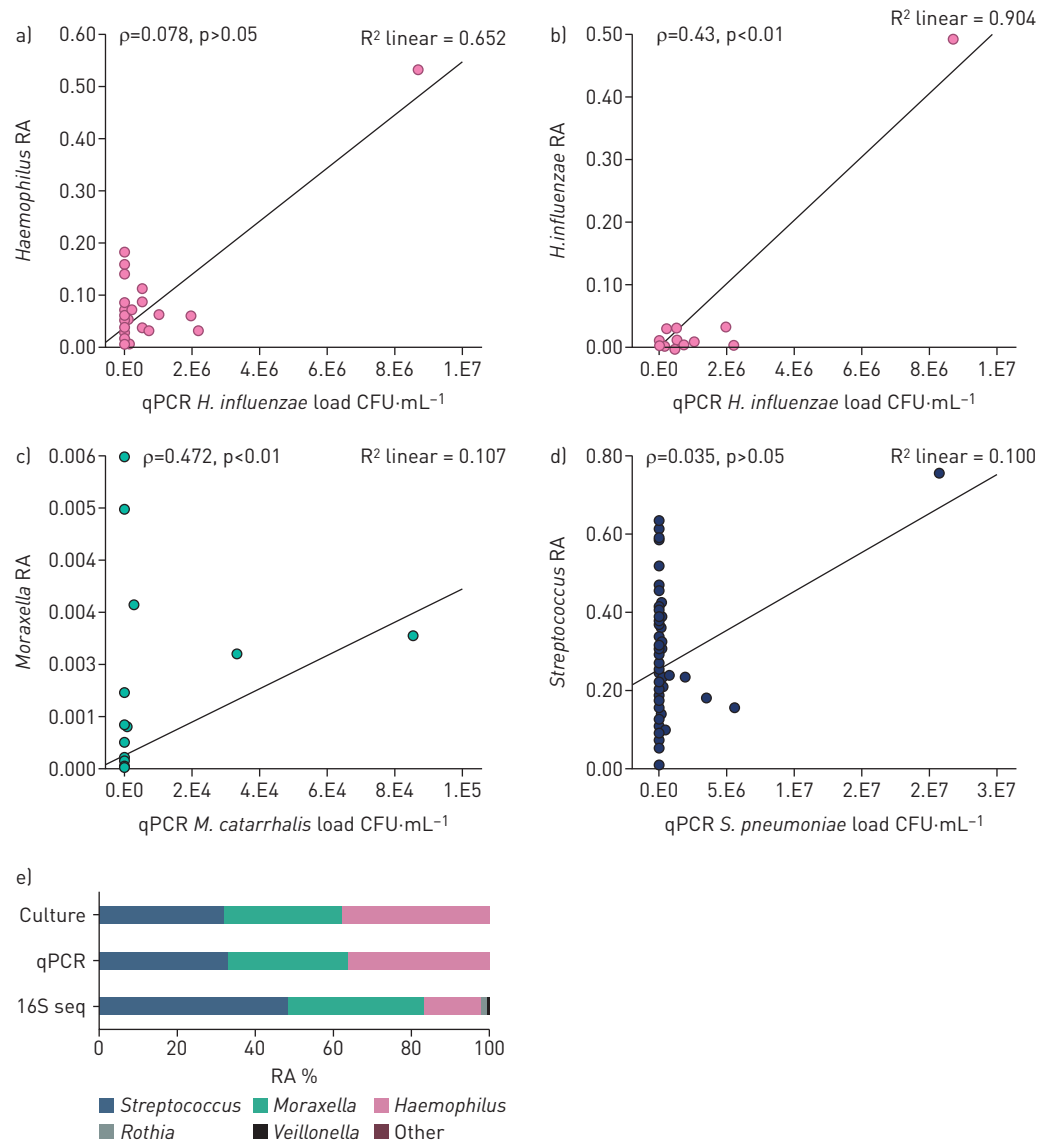
FIGURE 5 Correlation between the 16S sequencing relative abundance (RA) results and quantitative (q)PCR results (CFU·mL^{-1}) in sputum samples and mock community. a) Correlation between relative abundance of genus *Haemophilus* and load of *H. influenzae* in sputum samples; b) correlation between relative abundance of *H. influenzae* operational taxonomic units and load of *H. influenzae* in sputum samples; c) correlation between relative abundance of genus *Moraxella* and load of *M. catarrhalis* in sputum samples; d) correlation between relative abundance of genus *Streptococcus* and load of *S. pneumoniae* in sputum samples; e) comparison of the relative abundance of the three bacteria in a lab-prepared mock community by 16S sequencing, qPCR and quantitative culture. ρ: Spearman's correlation. n=74.

studies have demonstrated the potential consequences of these organisms in COPD [13] and asthma [28] exacerbations. Therefore, it was important to investigate the prevalence and loads of these three bacteria as pathogens in our cohort to investigate the possibility that reduced abundance of normal flora may open niches for opportunistic pathogens. Unfortunately, one limitation of 16S sequencing methods is that 16S rRNA hypervariable regions exhibit different degrees of sequence diversity, and no single hypervariable region can discriminate between all bacterial groups. In many instances, the bioinformatic pipelines provide reliable identification down to the genus level for most of organisms [41]. For this reason, despite the high sensitivity of the 16S sequencing and good correlation with the qPCR loads, specificity was relatively low especially for *S. pneumoniae* and *H. influenza*, probably due to the presence of commensal members of these genera as normal respiratory microbiota. Therefore, qPCR may be a more appropriate method to determine the prevalence and load of specific pathogenic bacteria. We did not find significant differences in the bacterial load nor the prevalence of these pathogens between the three study groups.

In addition, we compared the airway microbiome profiles of our young-adult cohort with published microbiological data for preterm infants during infancy. Numerous studies have identified an important role for respiratory colonisation with *Ureaplasma* in the development of BPD [10, 11, 18]. In our results, *Ureaplasma* was rarely identified in the EP group (RA <0.01% in six sputum samples) (data not shown). In the study by LOHMANN *et al.* [12], notable changes were reported in airway microbiome of preterm infants who subsequently developed BPD. Immediately after birth, the preterm airway microbiome was mainly dominated by Proteobacteria, particularly *Acinetobacter* species, and over time the relative abundance of Firmicutes increased, driven mainly by *Staphylococcus* in those infants who developed BPD, in contrast to the relatively diverse and stable community in the non-BPD group. *Staphylococcus* was also associated with BPD development in another study [19]. In our results, *Acinetobacter* and *Staphylococcus* were rarely identified in our cohort (RA for both was <0.1%). Interestingly, when detected, *Staphylococcus* RA was slightly higher in the BPD group compared to the other two groups; however, this observation was not statistically significant (supplementary figure E5). *Corynebacterium* has also been associated with the development of severe BPD in premature infants [42]. In our results, *Corynebacterium* was present at similar RA in the three groups at ~1%.

The main limitations of this study relate to the small sample size and an inability to induce sputum in all subjects, which may not be random, as it is more challenging to induce sputum in healthy subjects compared to those with respiratory pathology. For clinical safety reasons, lower hypertonic saline strength was used for sputum induction in those who were labelled as asthmatics which might have accounted for some of the variations between the groups. Although the microbial signature in the hypertonic saline matrix negative control was quite different from that in the sputum samples, it would have been useful to investigate the effect of the different saline strengths on the quality of sample, DNA extraction and PCR inhibition. In addition, asthma may be regarded as a potential confounder. The asthma prevalence was self-reported; we do not have objective testing to confirm or refute this; and many subjects were prescribed asthma inhalers. EPICure and other preterm cohort studies have previously reported that BPD survivors often have airflow obstruction later in life which can be mislabelled as asthma [5, 43, 44]. Nevertheless, in covariate analysis, asthma status was not a significant covariate ($p > 0.05$ by both adonis and ANOSIM). Moreover, none of the microbiome describing parameters (α- and β-indices) nor the *Prevotella* RA were different between subjects labelled as asthmatics and nonasthmatics, either in the whole cohort or the control group (supplementary figure E3). In all sputum-based studies, upper airway contamination is a concern. Induced sputum was the best available sample in a group of young adults with lung disease many of whom have additional developmental problems preventing research bronchoscopy. It would have been useful to directly compare the microbial signature in parallel upper and lower respiratory samples.

In conclusion, young adults born extremely preterm exhibit significant dysbiosis at 19 years of age. This is characterised by a shift in the microbial community structure away from Bacteriodetes and specifically manifest by a significant reduction in the relative abundance of genus *Prevotella*, as frequently described in other chronic lung diseases. *P. melaninogenica* was the species showing most variation within this genus. The prevalence and loads of typical respiratory opportunistic pathogens were not affected by such dysbiosis, and we did not identify persistence of dysbiosis patterns related to the development of neonatal BPD reported in studies of BPD in infancy.

References

1   Santhakumaran S, Statnikov Y, Gray D, *et al.* Survival of very preterm infants admitted to neonatal care in England 2008–2014: time trends and regional variation. *Arch Dis Child Fetal Neonatal Ed* 2018; 103: F208–F215.

2   Costeloe KL, Hennessy EM, Haider S, *et al.* Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *BMJ* 2012; 345: e7976.

3   Gien J, Kinsella JP. Pathogenesis and treatment of bronchopulmonary dysplasia. *Curr Opin Pediatr* 2011; 23: 305–313.

4   Lovering AT, Elliott JE, Laurie SS, *et al.* Ventilatory and sensory responses in adult survivors of preterm birth and bronchopulmonary dysplasia with reduced exercise capacity. *Ann Am Thorac Soc* 2014; 11: 1528–1537.

5   Fawke J, Lum S, Kirkby J, *et al.* Lung function and respiratory symptoms at 11 years in children born extremely preterm: the EPICure study. *Am J Respir Crit Care Med* 2010; 182: 237–245.

6   Hennessy EM, Bracewell MA, Wood N, *et al.* Respiratory health in pre-school and school age children following extremely preterm birth. *Arch Dis Child* 2008; 93: 1037–1043.

7   Lum S, Kirkby J, Welsh L, *et al.* Nature and severity of lung function abnormalities in extremely pre-term children at 11 years of age. *Eur Respir J* 2011; 37: 1199–1207.

8   Stressmann FA, Connett GJ, Goss K, *et al.* The use of culture-independent tools to characterize bacteria in endo-tracheal aspirates from pre-term infants at risk of bronchopulmonary dysplasia. *J Perinat Med* 2010; 38: 333–337.

9   Mourani PM, Harris JK, Sontag MK, *et al.* Molecular identification of bacteria in tracheal aspirate fluid from mechanically ventilated preterm infants. *PLoS One* 2011; 6: e25959.

10   Kotecha S, Hodge R, Schaber JA, *et al.* Pulmonary *Ureaplasma urealyticum* is associated with the development of acute lung inflammation and chronic lung disease in preterm infants. *Pediatr Res* 2004; 55: 61–68.

11   Beeton ML, Maxwell NC, Davies PL, *et al.* Role of pulmonary infection in the development of chronic lung disease of prematurity. *Eur Respir J* 2011; 37: 1424–1430.

12   Lohmann P, Luna RA, Hollister EB, *et al.* The airway microbiome of intubated premature infants: characteristics and changes that predict the development of bronchopulmonary dysplasia. *Pediatr Res* 2014; 76: 294–301.

13   Garcha DS, Thurston SJ, Patel AR, *et al.* Changes in prevalence and load of airway bacteria using quantitative PCR in stable and exacerbated COPD. *Thorax* 2012; 67: 1075–1080.

14   Comeau AM, Douglas GM, Langille MG. Microbiome Helper: a custom and streamlined workflow for microbiome research. *mSystems* 2017; 2: e00127-16.

15   Caporaso JG, Kuczynski J, Stombaugh J, *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010; 7: 335–336.

16   Parks DH, Tyson GW, Hugenholtz P, *et al.* STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 2014; 30: 3123–3124.

17   National Center for Biotechnology Information (NCBI). National Library of Medicine (US), National Center for Biotechnology Information. 1988. https://www.ncbi.nlm.nih.gov/ Date last accessed: February 16, 2018.

18   Van Marter LJ, Dammann O, Allred EN, *et al.* Chorioamnionitis, mechanical ventilation, and postnatal sepsis as modulators of chronic lung disease in preterm infants. *J Pediatr* 2002; 140: 171–176.

19   Wagner BD, Sontag MK, Harris JK, *et al.* Airway microbial community turnover differs by BPD severity in ventilated preterm infants. *PLoS One* 2017; 12: e0170120.

20   Charlson ES, Bittinger K, Haas AR, *et al.* Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011; 184: 957–963.

21   Morris A, Beck JM, Schloss PD, *et al.* Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med* 2013; 187: 1067–1075.

22   Pragman AA, Lyu T, Baller JA, *et al.* The lung tissue microbiota of mild and moderate chronic obstructive pulmonary disease. *Microbiome* 2018; 6: 7.

23   Erb-Downward JR, Thompson DL, Han MK, *et al.* Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PLoS One* 2011; 6: e16384.

24   Dickson RP, Erb-Downward JR, Martinez FJ, *et al.* The microbiome and the respiratory tract. *Annu Rev Physiol* 2016; 78: 481–504.

25   Garcia-Nuñez M, Millares L, Pomares X, *et al.* Severity-related changes of bronchial microbiome in chronic obstructive pulmonary disease. *J Clin Microbiol* 2014; 52: 4217–4223.

26   Wu J, Liu W, He L, *et al.* Sputum microbiota associated with new, recurrent and treatment failure tuberculosis. *PLoS One* 2013; 8: e83445.

27   Einarsson GG, Comer DM, McIlreavey L, *et al.* Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. *Thorax* 2016; 71: 795–803.

28   Hilty M, Burke C, Pedro H, *et al.* Disordered microbial communities in asthmatic airways. *PLoS One* 2010; 5: e8578.

29   Zhang Q, Cox M, Liang Z, *et al.* Airway microbiota in severe asthma and relationship to asthma severity and phenotypes. *PLoS One* 2016; 11: e0152724.

30   Lal CV, Travers C, Aghai ZH, *et al.* The airway microbiome at birth. *Sci Rep* 2016; 6: 31023.

31   Kedzia A, Kwapisz E, Wierzbowska M. Wystepowanie bakterii beztlenowych w zakazeniach układu oddechowego [Incidence of anaerobic bacteria in respiratory tract infections]. *Pneumonol Alergol Pol* 2003; 71: 68–73.

32   Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. *Immunology* 2017; 151: 363–374.

33   Stingu CS, Schaumann R, Jentsch H, *et al.* Association of periodontitis with increased colonization by *Prevotella nigrescens*. *J Investig Clin Dent* 2013; 4: 20–25.

34   Field TR, Sibley CD, Parkins MD, *et al.* The genus *Prevotella* in cystic fibrosis airways. *Anaerobe* 2010; 16: 337–344.

35   Tunney MM, Field TR, Moriarty TF, *et al.* Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *Am J Respir Crit Care Med* 2008; 177: 995–1001.

36   Sherrard LJ, McGrath SJ, McIlreavey L, *et al.* Production of extended-spectrum β-lactamases and the potential indirect pathogenic role of *Prevotella* isolates from the cystic fibrosis respiratory microbiota. *Int J Antimicrob Agents* 2016; 47: 140–145.

37   Flynn JM, Niccum D, Dunitz JM, *et al.* Evidence and role for bacterial mucin degradation in cystic fibrosis airway disease. *PLoS Pathog* 2016; 12: e1005846.

38   Segal LN, Clemente JC, Tsay JC, *et al.* Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat Microbiol* 2016; 1: 16031.

39   Twigg HL 3rd, Knox KS, Zhou J, *et al.* Effect of advanced HIV infection on the respiratory microbiome. *Am J Respir Crit Care Med* 2016; 194: 226–235.

40   Sulikowska A, Grzesiowski P, Sadowy E, *et al.* Characteristics of *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis* isolated from the nasopharynges of asymptomatic children and molecular analysis of *S. pneumoniae* and *H. influenzae* strain replacement in the nasopharynx. *J Clin Microbiol* 2004; 42: 3942–3949.

41   Chakravorty S, Helb D, Burday M, *et al.* A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 2007; 69: 330–339.

42   Imamura T, Sato M, Go H, *et al.* The microbiome of the lower respiratory tract in premature infants with and without severe bronchopulmonary dysplasia. *Am J Perinatol* 2017; 34: 80–87.

43   Northway WH Jr, Moss RB, Carlisle KB, *et al.* Late pulmonary sequelae of bronchopulmonary dysplasia. *N Engl J Med* 1990; 323: 1793–1799.

44   Vrijlandt EJ, Gerritsen J, Boezen HM, *et al.* Gender differences in respiratory symptoms in 19-year-old adults born preterm. *Respir Res* 2005; 6: 1465–9921.