ORIGINAL ARTICLE

The lower airway microbiota in early cystic fibrosis lung disease: a longitudinal analysis

Katherine B Frayman,[1,2,3] David S Armstrong,[4,5] Rosemary Carzino,[1,2]
Thomas W Ferkol,[6,7] Keith Grimwood,[8] Gregory A Storch,[6] Shu Mei Teo,[9]
Kristine M Wylie,[6,10] Sarath C Ranganathan[1,2,3]

ABSTRACT

Rationale In infants and young children with cystic fibrosis, lower airway infection and inflammation are associated with adverse respiratory outcomes. However, the role of lower airway microbiota in the pathogenesis of early cystic fibrosis lung disease remains uncertain.

Objectives To assess the development of the lower airway microbiota over time in infants and young children with cystic fibrosis, and to explore its association with airway inflammation and pulmonary function at age 6 years.

Methods Serial, semi-annual bronchoscopies and bronchoalveolar lavage (BAL) procedures were performed in infants newly diagnosed with cystic fibrosis following newborn screening. Quantitative microbiological cultures and inflammatory marker (interleukin 8 and neutrophil elastase) measurements were undertaken contemporaneously. 16S ribosomal RNA gene sequencing was conducted on stored BAL samples. Spirometry results recorded at 6 years of age were extracted from medical records.

Measurements and main results Ninety-five BAL samples provided 16S ribosomal RNA gene data. These were collected from 48 subjects aged 1.2–78.3 months, including longitudinal samples from 27 subjects and 13 before age 6 months. The lower airway microbiota varied, but diversity decreased with advancing age. Detection of recognised cystic fibrosis bacterial pathogens was associated with reduced microbial diversity and greater lower airway inflammation. There was no association between the lower airway microbiota and pulmonary function at age 6 years.

Conclusions In infants with cystic fibrosis, the lower airway microbiota is dynamic. Dominance of the microbiota by recognised cystic fibrosis bacterial pathogens is associated with increased lower airway inflammation, however early microbial diversity is not associated with pulmonary function at 6 years of age.

Key messages

What is the key question?
► How does the lower airway microbiota in infants and young children with cystic fibrosis contribute to disease progression?

What is the bottom line?
► The diversity of the lower airway microbiota in young children with cystic fibrosis decreases with age, and the presence of respiratory bacterial pathogens is associated with reduced microbial diversity and increased airway inflammation.

Why read on?
► This study explored the relationship between the early evolving lower airway microbiota, accompanying airway inflammation, and later pulmonary function in infants and children with cystic fibrosis.

INTRODUCTION

Presymptomatic infants with cystic fibrosis (CF) diagnosed following newborn screening can already have evidence of infection and inflammation in their lower airway samples. These changes are associated with earlier development of bronchiectasis and worse pulmonary function later in childhood.[1][2] Particular pathogens, especially *Staphylococcus aureus* and *Pseudomonas aeruginosa*, are associated with increased airway inflammation.[3][4] Recent molecular techniques have identified the complex, polymicrobial nature of CF lung disease, with obligate and facultative anaerobes recovered increasingly from the airways.[5–7] However, little is known about the role of these communities in the pathogenesis of CF lung disease.

Organisms comprising the human microbiota have complex relationships with one another and their host, where resident bacteria have potentially site-specific roles in immune modulation and response to individual pathogens.[8] Indeed, the upper and lower airways are distinct anatomical compartments with different microbial profiles. In health, the lower airways are not sterile, and while the lower airway microbiota of healthy children is poorly described, it is postulated that in disease states microbial diversity is altered, as either a primary or a secondary phenomenon, by colonisation with proliferating cells and by suppression of innate immunity and airway clearance.[8]

The airway microbiota is linked to progressive CF lung disease. Cross-sectional studies of adults and older children with CF suggest a relatively diverse lower airway microbiota,[9][10] with decreased microbial diversity in older patients correlated with severe

lung disease.[6 7 10–12] Analyses of explanted lung tissue suggest markedly reduced microbial diversity in advanced disease states dominated by recognised CF pathogens, such as *S. aureus*, *P. aeruginosa* and *Burkholderia cepacia* complex.[7 13 14] Evaluating the early lower airway microbiome and its relationship to pulmonary inflammation is essential to understanding the pathogenesis and therefore potentially altering the course of CF lung disease. Here, we describe the early *lower* airway microbiota in children with CF and explore its relationship with lower airway inflammation and pulmonary function at age 6 years. We hypothesised that reduced bacterial diversity during early childhood is associated with lower lung function.

METHODS
Participants
Children identified with CF following newborn screening or as neonates with meconium ileus and sweat chloride concentrations >60 mmol/L were recruited into a single-centre study in Melbourne, Australia from 1992 to 2001. As reported previously, bronchoscopy and bronchoalveolar lavage (BAL) were performed at the time of study entry, annually thereafter, and during hospitalisation for pulmonary exacerbation.[15–17] BAL samples were obtained from subjects' right middle lobe and lingula bronchi, and except for a small subset,[16] these aliquots were pooled at the time of collection. Inflammatory markers, interleukin 8 (IL-8) and neutrophil elastase (NE) were measured and quantitative microbiological culture performed contemporaneously, as described previously.[15 16] NE ≥5 μg/mL was considered elevated, consistent with assay detection limits.[15] Whole BAL samples were stored at −70°C. 16S ribosomal RNA gene (16S rRNA gene) analysis was performed on available samples. Subjects with at least one BAL sample and corresponding 16S rRNA gene data were included in this study. The Royal Children's Hospital Melbourne Human Research Ethics Committee approved the study (DA001-2015-32).

Clinical data
Demographic information was obtained at enrolment. Clinical data, including respiratory symptoms (cough, respiratory hospitalisation) and antibiotic use (oral, inhaled or intravenous) were recorded at 3-monthly clinical assessments and at bronchoscopy. The best FEV$_1$ and FVC at age 6 years were obtained from medical records; per cent predicted values were recalculated using Global Lung Initiative 2012 reference values.[18]

16S rRNA gene sequencing
Total nucleic acid was extracted from 200 μL of whole BAL fluid for 16S rRNA sequence analysis. EasyMAG Lysis Buffer was added to a total volume of 2 mL, and cells were disrupted by bead beating with 0.1 mm MO BIO glass beads (catalogue number 2600-50B2) using the MO BIO vortex adapter at 4°C for 10 min. Beads were pelleted in a microcentrifuge at maximum speed for 1 min, and the supernatant was removed and transferred to the NucliSENS easyMAG instrument (bioMérieux, Inc, Marcy l'Etoile, France). Nucleic acid extraction followed the manufacturer's protocol using off-board lysis. Total nucleic acid was eluted in 50 μL easyMAG Extraction Buffer 3 and 16S rRNA gene sequences were generated on the Roche 454 platform, using the V1-3 variable regions following Human Microbiome Project protocols.[19] In brief, the V1–3 region was amplified using the following primers, adding index sequences to distinguish each sample: 27F–5′ CCTATCCCC TGTGTGCCTTGGCAGTCTCAG_AGAGTTTGATCCTGGCT CAG; 534R–5′ CCATCTCATCCCTGCGTGTCTCCGACTCAG

_INDEX_ATTACCGCGGCTGCTGG. The cycling conditions were: 95°C, 2 min; 30 cycles each at 95°C, 20 s; 56°C, 30 s; 72°C, 5 min; after cycling held at 4°C. PCR products were cleaned using Agencourt AMPure Beads (Beckman Coulter, Indianapolis, Indiana, USA), and DNA concentrations were determined using the Qubit Fluorometer (ThermoFisher, Waltham, Massachusetts, USA). Samples were pooled in equimolar concentrations, except for negative controls and poorly amplified samples, which were added in their entirety. The pools were concentrated using a Qiagen MinElute column following the manufacturer's instructions (Qiagen, Germantown, Maryland, USA) and eluted in 30 μL 1× low TE buffer, pH 8.0. Sequences were generated on the Roche 454 GS FLX Titanium Instrument (Roche/454 Life Sciences, Branford, Connecticut, USA). Sequences were binned, based on the index sequences, allowing one mismatch. Low-quality reads were removed when average quality scores were <Q35 or the read length <200 base pairs. Chimeric sequences were identified using NAST-iEr and Chimera Slayer,[20] and chimeric sequences were removed. Reads passing quality control were classified using Ribosomal Database Project naïve Bayesian classifier,[21] V.2.2 with training set 6. Sequences were classified to the lowest taxonomic level that could be assigned with confidence values >0.5. Negative control samples from PCR reagents and the buffer processed with the extraction protocol in parallel with the samples were included, and these yielded 1–359 sequence reads.

Statistical analysis
Sequence data from patient samples were subsampled to 1000 reads for analysis of community structure,[22–24] and were correlated with culture and inflammatory data from the corresponding lobe. Sequences not classified to phylum level were removed before analysis. The relative abundance of sequences assigned to genus level was calculated from each BAL sample; individual genera that accounted for ≥60% of all sequences in a single sample were defined as 'dominant'. The Shannon Diversity Index (SDI) of individual samples was calculated using 1000 random sequences from each sample. This was repeated 100 times and the mean value reported. Study data were managed using REDCap electronic data capture tools hosted at Murdoch Children's Research Institute.[25] Statistical analysis was performed using 'R' (R Foundation for Statistical Computing, Vienna, Austria). IL-8 and NE values were logarithmically transformed (log$_{10}$) for analysis because of their skewed distributions. When NE activity was below the assay detection threshold, a value of 2.5 μg/mL was assigned. Generalised estimating equations regression with unstructured correlation and robust SEs, linked by subject identifier, were used to account for multiple samples from the same subject. The regression model used Gaussian with identity link and binomial distribution with logit link for continuous and categorical variables respectively. χ^2, paired t tests and linear regression were used if appropriate. β diversity was assessed using Bray-Curtis dissimilarity.

RESULTS
Subjects
The 16S rRNA V1–3 region could be amplified and sequenced for bacterial community analysis in 95 samples from 48 subjects, including serial longitudinal samples from 27 subjects. At least 1000 reads were generated from each of these samples. Amplification failed or was poor in an additional 53 samples, possibly due to low bacterial load and/or sample age (see online supplementary data). Subjects' characteristics are presented in table 1, and except for p.Phe508del homozygosity, these did not differ significantly from the remaining birth cohort at

recruitment (table 1). CF was diagnosed early, at a median age of 1.4 months (IQR 1.0–1.7-months). Samples collected from those with respiratory symptoms were more likely to yield ≥1000 sequences (see online supplementary table S1).

Samples were obtained between 1.2 and 78.3 months of age (median age 20.4 months, IQR 13.0–32.1 months). Two samples were collected from 9 subjects, three samples from 16 subjects, and four samples from 2 subjects (see online supplementary figure S1). Fifty-four (63%) and 45 (51%) samples were obtained from subjects with respiratory symptoms and those receiving antibiotics respectively (table 2). Eleven (13%) samples were collected during respiratory hospitalisation. Thirty samples were obtained from subjects without respiratory symptoms and not taking antibiotics at the time of BAL. Subjects reporting respiratory symptoms were significantly older and more likely to be taking antibiotics (see online supplementary table S2). Seventy-eight (82%) samples were pooled from the right middle lobe and lingula, and 11 (12%)

and six (6%) were separate samples from the right middle lobe and lingula respectively.

Lower airway microbiota in young children with CF

Overall, 1 201 177 individual sequences were identified, 98% of which could be classified to genus level. The most prevalent phyla were Firmicutes (63.7%), Proteobacteria (21.9%), Actinobacteria (8.1%), Bacteroidetes (4.8%) and Fusobacteria (1.3%). *Staphylococcus* was the most prevalent genus (36.6%), followed by *Streptococcus* (13.7%), *Pseudomonas* (6.1%), *Neisseria* (5.4%), *Haemophilus* (4.9%), *Gemella* (4.9%), *Graniculatella* (3.9%), *Prevotella* (2.8%), *Veillonella* (2.6%) and *Streptomyces* (2.1%) (figure 1).

When genera corresponding with recognised CF pathogens were detected on 16S rRNA gene analysis, they were occasionally dominant, accounting for ≥60% of all reads in that sample (figure 2). *Staphylococcus* was detected in 69 (73%) samples and was the dominant genus in 18 (26%). *Pseudomonas* was detected in 62 (65%) samples and was dominant in 5 (8%). *Stenotrophomonas* and *Burkholderia* were detected in 20 (21%) and 18 (19%) samples respectively and each was dominant in one. *Haemophilus* was detected in 59 (62%) samples, but was the dominant genus in just 2 (3%). Both of these latter samples yielded *Haemophilus influenzae* in quantitative culture.

Although considerable variability existed between individuals (see online supplementary figure S2), overall microbial diversity decreased in serial BAL samples in association with increased patient age (β=−0.01, p=0.03), including in the absence of respiratory symptoms and antibiotic use (β=−0.02, p=0.02, figure 3). After adjustment for age and antibiotic use, the presence of respiratory symptoms was associated with reduced microbial diversity (β=−0.35, p=0.02). There was no association between microbial diversity and CF transmembrane

Table 1 Demographic details of subjects with BAL samples included for analysis compared with remainder of original birth cohort at recruitment

|	Study subjects (n=48)	Remainder of birth cohort (n=94)	p Value (χ^2)
Male, n (%)	25 (52%)	38 (40%)	0.19
Mode of diagnosis	n=46	n=87	0.54
Newborn screening, n (%)	33 (72%)	65 (75%)	
Meconium ileus, n (%)	11 (24%)	16 (18%)	
Other clinical presentation, n (%)	2 (4%)	3 (3%)	
Family history, n (%)	0	3 (3%)	
Pancreatic status at recruitment	n=48	n=79	0.93
Pancreatic insufficient, n (%)	41 (85%)	67 (85%)	
Pancreatic sufficient, n (%)	7 (145%)	12 (15%)	
CFTR genotype	n=47	n=85	0.03
P.Phe508del homozygous, n (%)	32 (68%)	38 (45%)	
P.Phe508del heterozygous, n (%)	14 (30%)	42 (49%)	
Other, n (%)	1 (2%)	5 (6%)	

BAL, bronchoalveolar lavage.

Table 2 Bronchoalveolar lavage (BAL) samples (n=95) from which the 16S rRNA V1–3 region could be amplified and sequenced for bacterial community analysis

Age at BAL (months)	Samples with 16S rRNA gene data, n (%)	Respiratory symptoms at time of BAL, n (%)	Antibiotics at time of BAL, n (%)	SDI, median (IQR)
0–12	17 (17.9%)	8 (50.0%)*	7 (43.8%)*	1.30 (0.87–2.01)
12–24	33 (34.7%)	19 (61.3%)*	17 (51.5%)	1.80 (1.39–2.13)
24–36	24 (25.3%)†	12 (57.1%)*	10 (45.5%)*	1.84 (0.23–2.21)
36–48	9 (9.5%)	4 (57.1%)*	2 (33.3%)*	1.37 (0.82–1.87)
48–60	8 (8.4%)	7 (100%)*	6 (85.7%)*	0.7 (0.04–1.41)
>60‡	4 (4.2%)	4 (100%)	3 (75%)	1.08 (0.35–1.77)
Sample total	95	54 (62.8%)*	45 (51.1%)*	1.65 (0.56–2.08)

*Data missing for some subjects within this category.
†One subject contributed two samples to this age category.
‡Age range: 63.6–78.4 months.
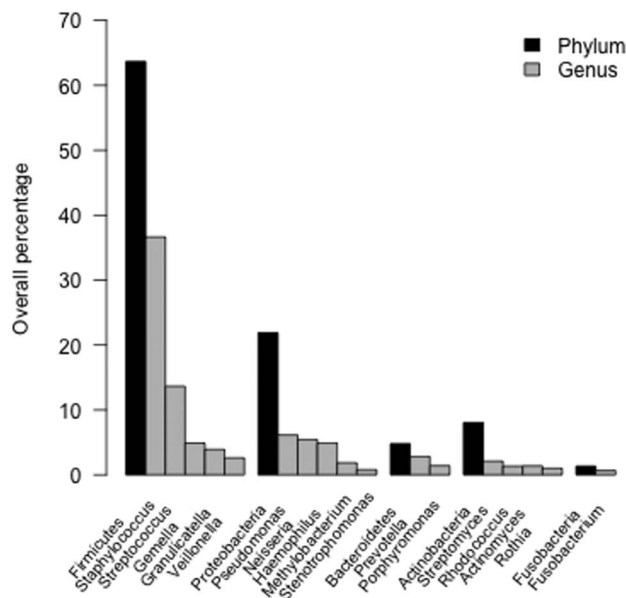
Figure 1 Prevalence of phyla and genera. Sequences were classified to genus level. The overall percentage of reads assigned to individual phyla and genera are displayed. The genera depicted account for 91.6% of sequences detected in the bronchoalveolar lavage samples. The relative abundance values in the respective phylum categories not accounted for by the genera presented are: Firmicutes 1.97%, Proteobacteria 2.74%, Bacteroidetes 0.56%, Actinobacteria 2.23% and Fusobacteria 0.68%.

regulator mutation (β=−0.2; p=0.16), respiratory symptoms (β=−0.27, p=0.20) or antibiotic use (β=−0.002; p=0.97) at bronchoscopy.
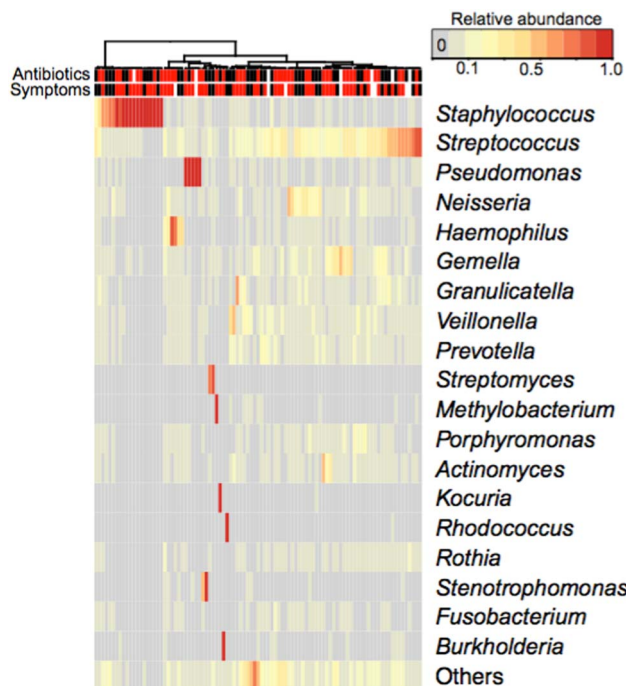

Figure 2 Relative abundance of individual genera at each sampling time. Individual genera with relative abundance >0.5% are depicted in rows. Individual bronchoalveolar lavage (BAL) samples are displayed in columns. Samples are organised according to similarity. The coloured, vertical bars at the top indicate antibiotic treatment and respiratory symptoms at the time of BAL (red—yes, black—no, white—unknown). The relative abundance of each genus in each BAL sample is reflected in the intensity of the heat map colour.

Correlation of the lung microbiota with quantitative culture

16S rRNA gene analysis confirmed the finding of recognised CF pathogens by culture. In all samples where *S. aureus* or *P. aeruginosa* grew in concentrations ≥10^5 colony forming units (cfu)/mL, *Staphylococcus* or *Pseudomonas* respectively were detected in 16S rRNA gene analysis. *S. aureus* was detected by quantitative culture at concentrations ≥10^5 cfu/mL in seven BAL samples. In five of the seven samples, *Staphylococcus* was the dominant genus in 16S rRNA gene analysis. *Staphylococcus* was dominant in 16S rRNA gene analysis of 13 additional samples; in 11 of these, *S. aureus* was also detected by culture. Similarly, *Pseudomonas* was dominant in 16S rRNA gene analysis of five BAL samples; in all five, *P. aeruginosa* cultures yielded ≥10^5 cfu/mL. *P. aeruginosa* was also present in concentrations ≥10^5 cfu/mL in two additional samples; in both, *Pseudomonas* was detected, but not dominant on 16S rRNA gene analysis. *B. cepacia complex* and *Stenotrophomonas maltophilia* each grew in concentrations ≥10^5 cfu/mL in a single BAL sample. In 16S rRNA gene analysis of the respective samples, *Burkholderia* was dominant and *Stenotrophomonas* was detected in a relative abundance of 59.6%. In the only BAL dominated by *Stenotrophomonas*, *S. maltophilia* grew in a concentration of 10^2 cfu/mL. Cultured organisms regarded as upper respiratory commensals (eg. α-haemolytic streptococci and *Neisseria* species) were supported by 16S rRNA gene analysis findings of corresponding genera in variable abundance (table 3). Furthermore, 16S rRNA gene analysis frequently revealed the presence of organisms that went undetected by routine culture.

Lower airway inflammation and the lung microbiota

Dominance of established proinflammatory organisms (*Pseudomonas*, *Staphylococcus*, *Streptococcus*, *Haemophilus*, *Stenotrophomonas* and *Burkholderia*)[3] on 16S rRNA gene analysis was associated with higher inflammation indices in the lower airway (figure 4). In BAL samples with a dominant pathogen genus, the median IL-8 and NE were 1041 (IQR 210–2252) pg/mL and 18.4 (IQR 6.0–37.8) µg/mL respectively,
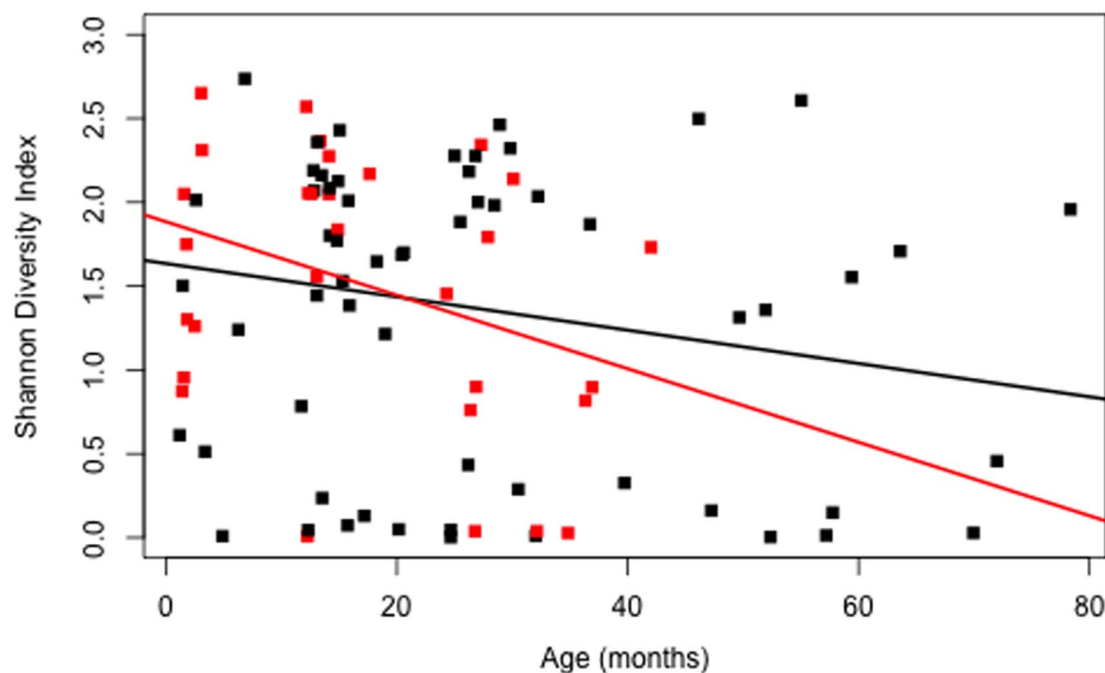
Figure 3 Shannon Diversity Index versus subject age (months) (β=−0.01, p=0.031). Samples obtained during clinical stability, in the absence of respiratory symptoms or antibiotic use, are depicted in red (β=−0.02, p=0.02).

Table 3 Agreement between microbiological culture and 16S rRNA gene analysis

Standard microbiological culture*	16S rRNA gene analysis		
	Not detected	<60% reads	≥60% reads†
Staphylococcus aureus (n‡=31)	*Staphylococcus* (n‡=69)		
Not detected	23	39	2
CFU 10^1–10^4/mL	3	10	11
CFU 10^5–10^7/mL	0	2	5
Pseudomonas aeruginosa (n=18)	*Pseudomonas* (n=62)		
Not detected	30	47	0
CFU 10^1–10^4/mL	3	8	0
CFU 10^5–10^7/mL	0	2	5
Haemophilus influenzae (n=13)	*Haemophilus* (n=59)§		
Not detected	33	49	0
CFU 10^1–10^4/mL	3	3	0
CFU 10^5–10^7/mL	0	5	2
Haemophilus species (n=21)	*Haemophilus* (n=59)§		
Not detected	29	43	2
CFU 10^1–10^4/mL	7	13	0
CFU 10^5–10^7/mL	0	1	0
Streptococcus pneumoniae (n=2)	*Streptococcus* (n=91)§		
Not detected	4	82	7
CFU 10^1–10^4/mL	0	0	0
CFU 10^5–10^7/mL	0	2	0
α haemolytic streptococci (n=46)	*Streptococcus* (n=91)§		
Not detected	2	46	1
CFU 10^1–10^4/mL	2	34	5
CFU 10^5–10^7/mL	0	4	1
β haemolytic streptococci (n=3)	*Streptococcus* (n=91)§		
Not detected	4	82	6
CFU 10^1–10^4/mL	0	1	1
CFU 10^5–10^7/mL	0	1	0
Neisseria species (n=28)	*Neisseria* (n=79)		
Not detected	14	53	0
CFU 10^1–10^4/mL	2	25	0
CFU 10^5–10^7/mL	0	1	0
Moraxella cattarrhalis (n=5)	*Moraxella* (n=25)		
Not detected	69	21	0
CFU 10^1–10^4/mL	0	1	0
CFU 10^5–10^7/mL	1	3	0
Stenotrophomonas maltophilia (n=4)	*Stenotrophomonas* (n=20)		
Not detected	75	16	0
CFU 10^1–10^4/mL	0	2	1
CFU 10^5–10^7/mL	0	1	0
Burkholderia cepacia (n=1)	*Burkholderia* (n=18)		
Not detected	77	17	0
CFU 10^1–10^4/mL	0	0	0
CFU 10^5–10^7/mL	0	0	1

*Twenty-seven BAL samples cultured mixed upper respiratory tract commensals (20 in concentrations of 10^1–10^4 CFU/mL and 7 in concentrations of 10^5–10^7 cfu/mL). These could not be correlated with 16S rRNA gene analysis results.

†'Dominant' genera account for ≥60% of all 16S rRNA gene analysis reads in an individual sample.

‡'n' represents the number of BAL samples in which the specific species or genus was detected by microbiological culture and 16S rRNA analysis respectively.

§All 16S rRNA gene sequences for the genera *Haemophilus* and *Streptococcus* are presented. Data were not split below genus level.

BAL, bronchoalveolar lavage; CFU, colony forming unit.

pathogen genus (IL-8, median 254.5 pg/mL, IQR 136–765 pg/mL; NE, median 3.9 µg/mL, IQR 2.5–6.8 µg/mL) (figure 4).

Reduced microbial diversity was associated with increased BAL NE (β=−0.5, p=0.02, age-adjusted β=−0.43, p=0.03), although not IL-8 (β=−0.23, p=0.05, age-adjusted β=−0.17, p=0.097). Figure 5 illustrates the relationship of microbial diversity, IL-8 levels, NE concentrations, and dominance of pathogen genera, *Staphylococcus* and *Pseudomonas*.

Compared with asymptomatic subjects, children symptomatic at the time of bronchoscopy had increased lower airway inflammatory markers (IL-8: median 864 (IQR 196–1833) vs 195 (IQR 68–547) pg/mL, t-test $\log_{10}$ IL-8, p=0.005; NE: median 13.0 (IQR 3.6–25.3) vs 5.8 (IQR 2.5–12.0) µg/mL, t-test $\log_{10}$ NE, p=0.02). In subjects who had bronchoscopies when both symptomatic and asymptomatic, symptoms were associated with a trend towards increased inflammation (IL-8 (n=14): mean of differences, 1343 pg/mL, 95% CI −888.2 to 3574.2 pg/mL, p=0.2; NE (n=13): mean of differences, 16.3 µg/mL, 95% CI −2.6 to 35.5, p=0.08). A dominant proinflammatory genus was detected in 22/54 (41%) samples from symptomatic subjects and in 8/31 (26%) samples from those without symptoms (p=0.17).

Change in microbiota over time

The composition of individuals' lower airway microbial communities varied considerably over time (see online supplementary figure S2). The β diversity of serial samples obtained from individual subjects was comparable to the β diversity of samples obtained from different subjects (median Bray-Curtis dissimilarity (IQR): 0.91 (0.7–0.99) and 0.9 (0.72–0.99) respectively) with the exception of samples dominated by known proinflammatory genera (figure 6).
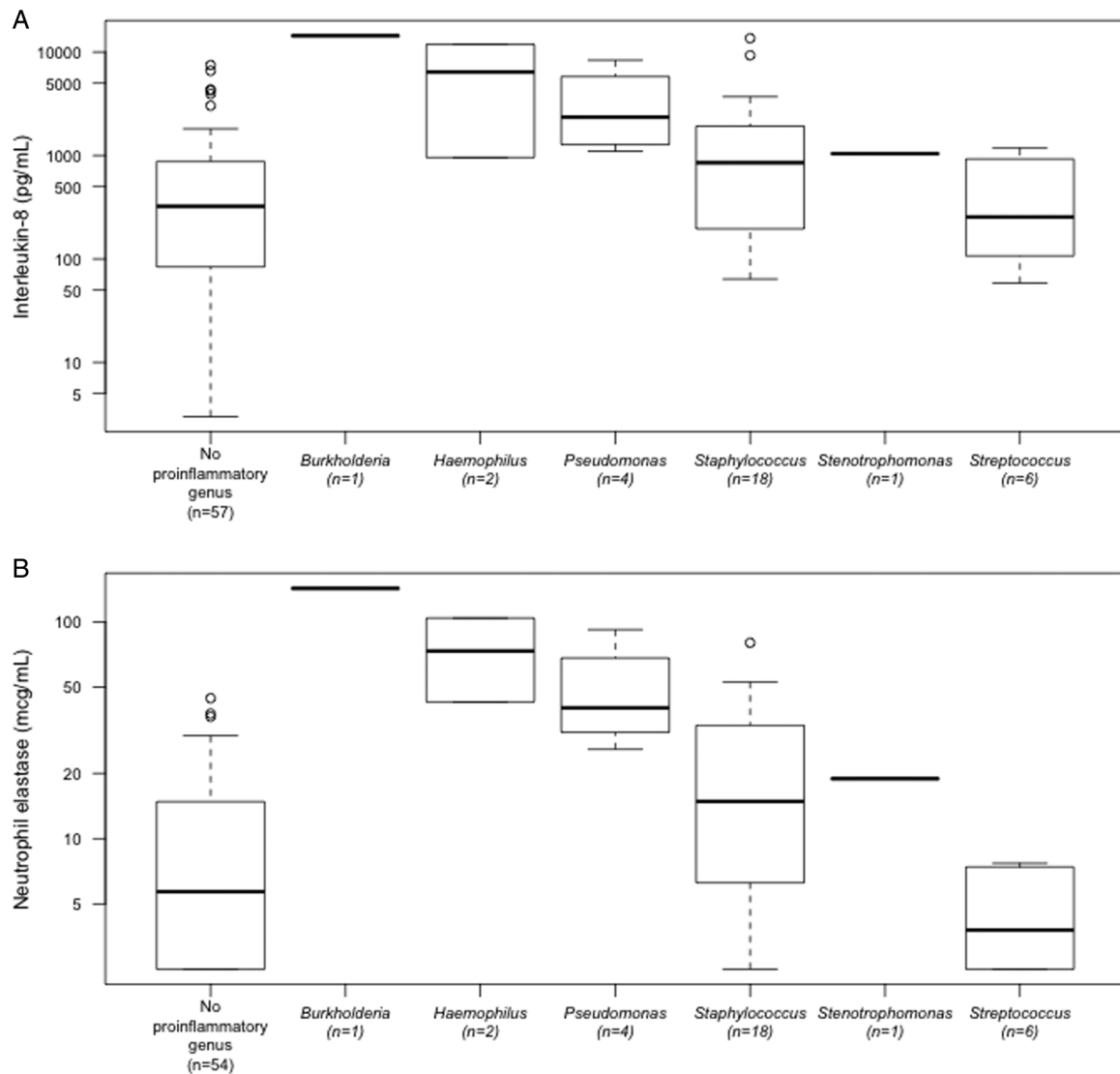
Pulmonary function measurements and microbiota

There was no relationship between subjects' bacterial diversity and their best pulmonary function measures at age 6 years (figure 7). Neither lowest SDI nor reduced early microbial diversity (age <6-months) was associated with pulmonary function measures at age 6 years (r^2=0.03, p=0.26 and r^2=0.001, p=0.9 respectively).

DISCUSSION

We examined the evolution of lower airway microbiota in young children with CF over time using BAL samples that were collected semi-annually up to 24 years ago at the Royal Children's Hospital Melbourne, and its association with airway inflammation and pulmonary function at age 6 years. The lower airway microbiota was diverse, but diversity decreased with subject age. Genera of known CF pathogens often emerged, and their dominance was associated with lower airway inflammation. We were unable to identify any relationship between microbial diversity and pulmonary function at age 6 years.

To our knowledge this is the first longitudinal molecular study of BAL samples from infants with CF and provides unique insights into their lower airway microbiota during early childhood. Recruitment from a single birth cohort not routinely exposed to prophylactic antibiotics and inclusion of long-term outcome data add to the study's ability to provide information about CF lung disease progression and its relationship to the lower airway microbiota. The study does however have important limitations. It involves the secondary analysis of prospectively collected BAL samples from two decades ago. Rigorous storage and handling practices and strict laboratory exclusion criteria were implemented to ensure the quality of the 16S

compared with 323 (IQR 84–875) pg/mL and 5.7 (IQR 2.5–14.4) µg/mL in other samples (IL-8, p=0.002; p=0.0008). IL-8 and NE concentrations in BAL samples dominated by *Streptococcus* were comparable to those without a dominant

Figure 4 Dominant proinflammatory organisms by 16S rRNA gene analysis and (A) interleukin 8 (IL-8) and (B) neutrophil elastase (NE) concentrations. Data for bronchoalveolar lavage samples with corresponding IL-8 (n=89, 94%) and NE (n=86, 91%) are displayed.

rRNA gene data. These data correlated with the participants' own culture results, assessed contemporaneously from the same BAL samples, and previous findings from lower airway samples in older patient cohorts,[26] and indicate that they accurately represent the subjects' lower airway microbiota. In order to address the potential for reagent contamination, negative sequence controls were processed alongside study samples and only the 64% of samples that could be subsampled to 1000 reads were included in analysis of microbial community structure. Further methodological studies are required to explore the molecular analysis of BAL fluid from infants and very young children, who may have both very low sample volumes and total bacterial load. Future prospective, longitudinal studies of BAL fluid are also required to determine the impact of cumulative antibiotic exposure on the airway microbiota. Strict procedural protocols were maintained to minimise potential contamination with upper airway secretions. Nevertheless, it is not possible to fully differentiate between upper airway organisms introduced at the time of bronchoscopy and the resident lung microbiota.

The lower airway microbiota was notably diverse in our cohort of newly diagnosed infants with CF. The most prevalent phylum was Firmicutes, and *Staphylococcus* the most prevalent genus. The facultative anaerobes now frequently detected in molecular analyses of sputum specimens from adults with CF, *Gemella*, *Prevotella* and *Veillonella*, represented 4.9%, 2.8% and 2.6% of overall genera respectively. Interestingly, a recent comparison of BAL samples from 13 older children with CF (mean age 3.95 years, range 1.1–8.1 years) and nine controls (mean age 3.78 years, range 0.9–9.75 years), similarly showed reduced diversity in the CF population.[27]

The association between reduced diversity with increasing subject age is consistent with findings from longitudinal studies of lower airway samples from adults with CF.[10 11] It does, however, contrast with the findings of previous longitudinal analyses of children with CF, which only examined the upper airway microbiota.[10 28–32] During infancy, the nasopharyngeal microbiota is dynamic,[33–35] although differences between that of infants with CF and healthy controls have been reported,[30 31] including increased relative abundance of Staphylococcae at baseline,[31] changes in *Staphylococcus* species,[31] and the emergence of gram-negative bacteria[30] following antibiotic exposure. Two smaller studies examining oropharyngeal swabs from infants[29] and children[10] with CF suggested that upper airway microbial diversity increased with age. Other studies, however,
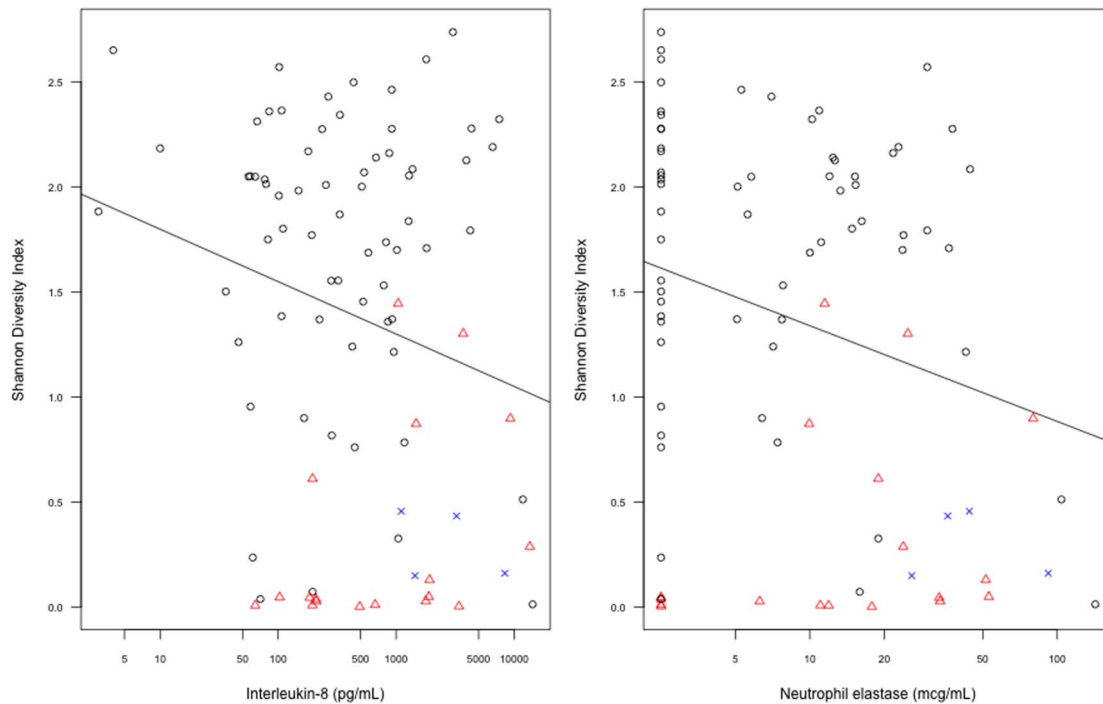
Figure 5 Interleukin-8 and neutrophil elastase concentrations versus the Shannon Diversity Index (age adjusted, p=0.097 and p=0.03 for interleukin 8 and neutrophil elastase respectively). Samples dominated by *Staphylococcus* and *Pseudomonas* are shown in red and blue respectively.
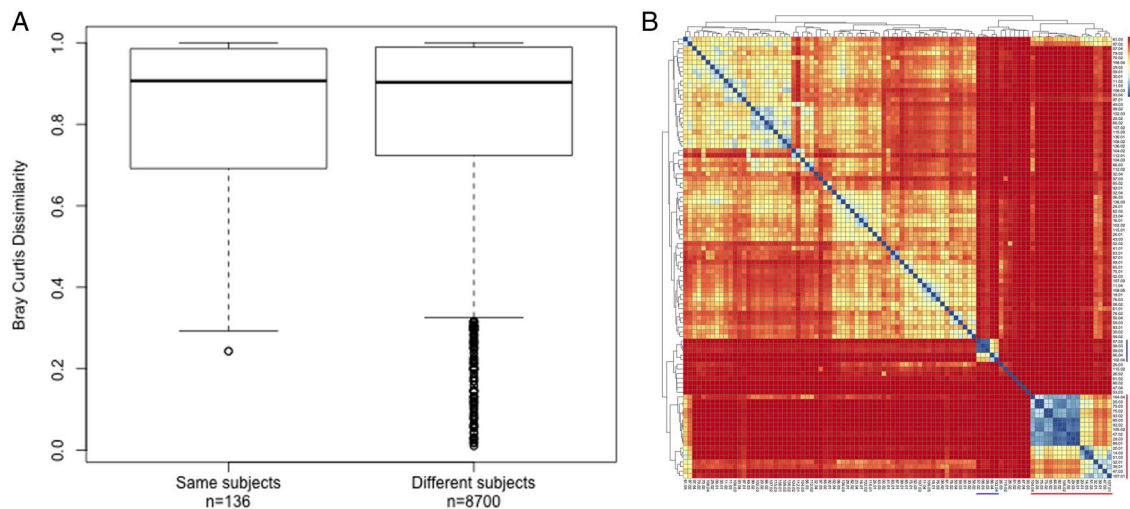

Figure 6 Bray-Curtis dissimilarity of bronchoalveolar lavage (BAL) samples. (A) Box plot summarising the pairwise Bray-Curtis dissimilarity of BAL samples obtained from the same and different subjects, and (B) heatmap of all pairwise distances. Outliers in (A) correlate with pairs of samples dominated by known proinflammatory pathogens.

including those involving this CF infant cohort, have shown that oropharyngeal microbiology does not accurately predict that of the lower airway,[17 26] and our finding of decreasing diversity in the latter environment is therefore not surprising.

Dominance of genera of known CF pathogens was associated with reduced microbial diversity and increased pulmonary inflammation. In this cohort, bacterial density $\geq 10^5$ cfu/mL was associated with greater airway inflammation.[17] A recent study of otherwise healthy children with pneumonia found reduced upper airway microbial diversity,[36] whilst the lower airway microbiota of adults with end-stage CF lung disease is dominated by a relatively homogenous population of typical CF pathogens.[7 13] In cross-sectional studies of adults, increased

microbial diversity is associated with clinical stability and improved pulmonary function,[12 37] although microbial diversity was also maintained in tissue samples obtained at lobectomy from a 3 year old with severe, localised CF lung disease.[38] While it was not possible to exclude the potentially confounding effect of cumulative antibiotic use, our study of young children, whose lifetime antibiotic exposure is likely considerably less than their adult counterparts, suggests that microbial diversity decreases when known pathogens emerge.

The detection of recognised CF pathogens in early childhood using conventional microbiological techniques is associated with greater airway inflammation, earlier onset of structural lung disease, and poorer pulmonary function outcomes.[1–3] Using
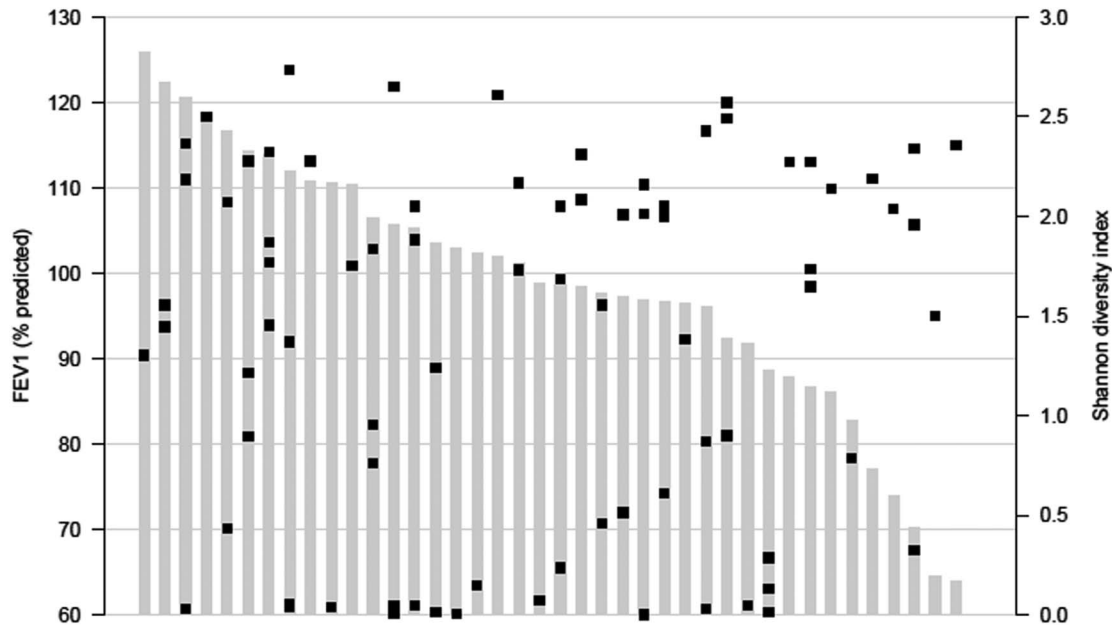
Figure 7 Shannon Diversity Index and best FEV$_1$ at 6 years of age. Individual subjects' with available spirometry data are displayed on the horizontal axis (n=42, 88%); their FEV$_1$ (percentage predicted) and Shannon Diversity Index (•) are displayed on the vertical axes.

molecular techniques to confirm the former, our results clearly demonstrate that dominance of pathogen genera in 16S rRNA gene analysis is associated with greater inflammation in the lung. The impact of this finding on structural lung disease in early childhood could not be tested, as CT of the chest was not routinely performed in this patient population two decades ago. While we did not demonstrate an association between the lower airway microbiota and pulmonary function at age 6 years, normal spirometry results do not exclude bronchiectasis, which occurs in approximately 80% of early school-aged children with CF, including those with normal FEV$_1$ values.[39] Whilst this study was not designed to address causality, it is possible that longer-term follow-up will reveal a relationship between decreased diversity and abnormal pulmonary function later in life.

β diversity was of similar magnitude in longitudinally collected samples from individual subjects as it was in samples from different subjects, highlighting the extent to which the lower airway microbiota changes over time, particularly in the presence of dominant CF pathogens, where exploration of the microbiota may explain increased lower airway inflammation and clinical deterioration in the absence of conventional microbiological findings. In keeping with observations made in adult patients at various stages of an exacerbation cycle, microbial diversity may recover, either returning to baseline or developing a new microbial community structure. The long-term impact of these perturbations is unknown, but finding greater lower airway inflammation in the presence of dominant CF pathogen genera suggests a likely negative impact on structural lung disease and pulmonary function. It is unclear whether regained bacterial diversity alters the trajectory of CF lung disease in young children. Further insights can be achieved by prospective comparisons with BAL samples from age-matched controls and from analysing changes in the lower airway microbiota at frequent, more regular intervals and for a longer time period.

In conclusion, the lower airway microbiota in infants and children with CF is diverse, but bacterial diversity declines with increasing age. When present, CF pathogen genera can become dominant and are associated with greater inflammation, but

individuals can recover diversity or return to their baseline microbiota. Further prospective longitudinal studies are required to understand the factors that change the lower airway microbiota over time, and to assess the impact of these changes on clinical outcome.

Author affiliations
[1]Department of Respiratory and Sleep Medicine, Royal Children's Hospital, Parkville, Victoria, Australia
[2]Respiratory Diseases Group, Murdoch Children's Research Institute, Parkville, Victoria, Australia
[3]Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia
[4]Department of Respiratory Medicine, Monash Children's Hospital, Clayton, Victoria, Australia
[5]Department of Paediatrics, Monash University, Clayton, Victoria, Australia
[6]Department of Pediatrics, Washington University, St Louis, Missouri, USA
[7]Department of Cell Biology and Physiology, Washington University, St Louis, Missouri, USA
[8]Menzies Health Institute Queensland, Griffith University and Gold Coast Health, Gold Coast, Queensland, Australia
[9]Centre for System Genomics, University of Melbourne, Melbourne, Victoria, Australia
[10]McDonnell Genome Institute, Washington University, St Louis, Missouri, USA

Competing interests None declared.

Ethics approval Royal Children's Hospital, Melbourne, Human Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Sly PD, Gangell CL, Chen L, et al. Risk factors for bronchiectasis in children with cystic fibrosis. *N Eng J Med* 2013;368:1963–70.
2. Ramsey KA, Ranganathan S, Park J, et al. Early respiratory infection is associated with reduced spirometry in children with cystic fibrosis. *Am J Respir Crit Care Med* 2014;190:1111–16.
3. Gangell C, Gard S, Douglas T, et al. Inflammatory responses to individual microorganisms in the lungs of children with cystic fibrosis. *Clin Infect Dis* 2011;53:425–32.
4. Pillarisetti N, Williamson E, Linnane B, et al. Infection, inflammation, and lung function decline in infants with cystic fibrosis. *Am J Respir Crit Care Med* 2011;184:75–81.
5. Harris JK, De Groote MA, Sagel SD, et al. Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. *Proc Natl Acad Sci USA* 2007;104:20529–33.
6. Delhaes L, Monchy S, Fréalle E, et al. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community--implications for therapeutic management. *PLoS ONE* 2012;7:e36313.
7. Rudkjøbing VB, Thomsen TR, Alhede M, et al. True microbiota involved in chronic lung infection of cystic fibrosis patients found by culturing and 16S rRNA gene analysis. *J Clin Microbiol* 2011;49:4352–5.
8. Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. *Lancet* 2014;384:691–702.
9. Klepac-Ceraj V, Lemon KP, Martin TR, et al. Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and *Pseudomonas aeruginosa*. *Environ Microbiol* 2010;12:1293–303.
10. Cox MJ, Allgaier M, Taylor B, et al. Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS ONE* 2010;5:e11044.
11. Zhao J, Schloss PD, Kalikin LM, et al. Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc Natl Acad Sci USA* 2012;109:5809–14.
12. Filkins LM, Hampton TH, Gifford AH, et al. Prevalence of streptococci and increased polymicrobial diversity associated with cystic fibrosis patient stability. *J Bacteriol* 2012;194:4709–17.
13. Goddard AF, Staudinger BJ, Dowd SE, et al. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proc Natl Acad Sci USA* 2012;109:13769–74.
14. Rudkjøbing VB, Thomsen TR, Alhede M, et al. The microorganisms in chronically infected end-stage and non-end-stage cystic fibrosis patients. *FEMS Immunol Med Microbiol* 2012;65:236–44.
15. Armstrong DS, Hook SM, Jamsen KM, et al. Lower airway inflammation in infants with cystic fibrosis detected by newborn screening. *Pediatr Pulmonol* 2005;40:500–10.
16. Gutierrez JP, Grimwood K, Armstrong DS, et al. Interlobar differences in bronchoalveolar lavage fluid from children with cystic fibrosis. *Eur Respir J* 2001;17:281–6.
17. Armstrong DS, Grimwood K, Carlin JB, et al. Bronchoalveolar lavage or oropharyngeal cultures to identify lower respiratory pathogens in infants with cystic fibrosis. *Pediatr Pulmonol* 1996;21:267–75.
18. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012;40:1324–43.
19. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–14.
20. Haas BJ, Gevers D, Earl AM, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 2011;21:494–504.
21. Wang Q, Garrity GM, Tiedje JM, et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007;73:5261–7.
22. Huse SM, Ye Y, Zhou Y, et al. A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS ONE* 2012;7:e34242.
23. Zhou Y, Holland MJ, Makalo P, et al. The conjunctival microbiome in health and trachomatous disease: a case control study. *Genome Med* 2014;6:99.
24. Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. *Nature* 2014;509:357–60.
25. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
26. Zemanick ET, Wagner BD, Robertson CE, et al. Assessment of airway microbiota and inflammation in cystic fibrosis using multiple sampling methods. *Ann Am Thorac Soc* 2015;12:221–9.
27. Renwick J, McNally P, John B, et al. The microbial community of the cystic fibrosis airway is disrupted in early life. *PLoS One* 2014;9:e109798.
28. Hoen AG, Li J, Moulton LA, et al. Associations between gut microbial colonization in early life and respiratory outcomes in cystic fibrosis. *J Pediatr* 2015;167:138–47.
29. Madan JC, Koestler DC, Stanton BA, et al. Serial analysis of the gut and respiratory microbiome in cystic fibrosis in infancy: interaction between intestinal and respiratory tracts and impact of nutritional exposures. *MBio* 2012;3:e00251–12.
30. Prevaes SMPJ, de Winter-de Groot KM, Janssens HM, et al. Development of the nasopharyngeal microbiota in infants with cystic fibrosis. *Am J Respir Crit Care Med* 2016;193:504–15.
31. Mika M, Korten I, Qi W, et al. The nasal microbiota in infants with cystic fibrosis in the first year of life: a prospective cohort study. *Lancet Respir Med* 2016;4:627–35.
32. Keravec M, Mounier J, Prestat E, et al. Insights into the respiratory tract microbiota of patients with cystic fibrosis during early *Pseudomonas aeruginosa* colonization. *SpringerPlus* 2015;4:405.
33. Mika M, Mack I, Korten I, et al. Dynamics of the nasal microbiota in infancy: a prospective cohort study. *J Allergy Clin Immunol* 2015;135:905–12.
34. Teo SM, Mok D, Pham K, et al. The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development. *Cell Host Microbe* 2015;17:704–15.
35. Biesbroek G, Tsivtsivadze E, Sanders EA, et al. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am J Respir Crit Care Med* 2014;190:1283–92.
36. Sakwinska O, Bastic Schmid V, Berger B, et al. Nasopharyngeal microbiota in healthy children and pneumonia patients. *J Clin Microbiol* 2014;52:1590–4.
37. Blainey PC, Milla CE, Cornfield DN, et al. Quantitative analysis of the human airway microbial ecology reveals a pervasive signature for cystic fibrosis. *Sci Transl Med* 2012;4:153ra130.
38. Brown PS, Pope CE, Marsh RL, et al. Directly sampling the lung of a young child with cystic fibrosis reveals diverse microbiota. *Ann Am Thorac Soc* 2014;11:1049–55.
39. Owens CM, Aurora P, Stanojevic S, et al. Lung Clearance Index and HRCT are complementary markers of lung abnormalities in young children with CF. *Thorax* 2011;66:481–8.