

OPEN

# Gut Microbiota in Children With Cystic Fibrosis: A Taxonomic and Functional Dysbiosis

Michael J. Coffey<sup>1</sup>, Shaun Nielsen<sup>2</sup>, Bernd Wemheuer<sup>2</sup>, Nadeem O. Kaakoush<sup>3</sup>, Millie Garg<sup>1</sup>, Bronwen Needham<sup>1,2</sup>, Russell Pickford<sup>4</sup>, Adam Jaffe<sup>1,5,6</sup>, Torsten Thomas<sup>1,2</sup> & Chee Y. Ooi<sup>1,5,7\*</sup>

Intestinal dysbiosis has been observed in children with cystic fibrosis (CF), yet the functional consequences are poorly understood. We investigated the functional capacity of intestinal microbiota and inflammation in children with CF. Stool samples were collected from 27 children with CF and 27 age and gender matched healthy controls (HC) (aged 0.8–18 years). Microbial communities were investigated by iTag sequencing of 16S rRNA genes and functional profiles predicted using Tax4Fun. Inflammation was measured by faecal calprotectin and M2-pyruvate kinase. Paediatric CF gastrointestinal microbiota demonstrated lower richness and diversity compared to HC. CF samples exhibited a marked taxonomic and inferred functional dysbiosis when compared to HC. In children with CF, we predicted an enrichment of genes involved in short-chain fatty acid (SCFA), antioxidant and nutrient metabolism (relevant for growth and nutrition) in CF. The notion of pro-inflammatory GI microbiota in children with CF is supported by positive correlations between intestinal inflammatory markers and both genera and functional pathways. We also observed an association between intestinal genera and both growth z-scores and FEV1%. These taxonomic and functional changes provide insights into gastrointestinal disease in children with CF and future gastrointestinal therapeutics for CF should explore the aforementioned pathways and microbial changes.

Gastrointestinal (GI) disease in cystic fibrosis (CF) begins *in utero*, continues throughout childhood and into adulthood<sup>1,2</sup>. Dysfunction of the cystic fibrosis transmembrane conductance regulator (CFTR) results in an altered intestinal milieu with proposed disease factors including: (i) reduced bicarbonate secretion and low intestinal pH, (ii) thick and inspissated mucus, (iii) a lack of endogenous pancreatic enzymes, (iv) delayed intestinal transit and (v) possibly an impaired innate immunity<sup>3,4</sup>. These mechanisms in combination with an energy- and fat-dense diet<sup>5</sup> and frequent antibiotic usage likely contribute to a dysbiosis of the intestinal bacterial community, which has been observed in CF<sup>6–9</sup>. Individuals with CF suffer from maldigestion and malabsorption, which contribute to malnutrition. Nutrition is of paramount importance in CF, as it has prognostic value for growth, lung function, overall well-being and mortality<sup>10</sup>.

Intestinal inflammation has been reported in several studies indirectly using various biomarkers, including calprotectin, rectal nitric oxide and M2-pyruvate kinase (M2-PK), as well as directly by capsule endoscopy<sup>11–16</sup>. The pathogenesis of intestinal inflammation in CF is poorly understood, however the altered intestinal environment and microbiota likely play a role. Intestinal inflammation in CF may have significant clinical relevance due to its link with growth and nutrition<sup>11,17,18</sup>, and to the increased risk of GI cancers in the adult patient population<sup>19–21</sup>.

Recently, Manor *et al.*<sup>8</sup> utilised metagenomic shotgun sequencing of faecal samples and suggested that in children with CF, enteric fat abundance selects for a pro-inflammatory GI microbiota. Debyser *et al.*<sup>7</sup> utilised

<sup>1</sup>Discipline of Paediatrics, School of Women's and Children's Health, University of New South Wales, Sydney, NSW, Australia. <sup>2</sup>Centre for Marine Science and Innovation, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW, Australia. <sup>3</sup>School of Medical Sciences, University of New South Wales, Sydney, NSW, Australia. <sup>4</sup>Bioanalytical Mass Spectrometry Facility, Mark Wainwright Analytical Centre (MWAC), University of New South Wales, Sydney, NSW, Australia. <sup>5</sup>Molecular and Integrative Cystic Fibrosis (miCF) Research Centre, High Street, Randwick, NSW, Australia. <sup>6</sup>Department of Respiratory, Sydney Children's Hospital, High Street, Randwick, NSW, Australia. <sup>7</sup>Department of Gastroenterology, Sydney Children's Hospital, High Street, Randwick, NSW, Australia. \*email: [keith.ooi@unsw.edu.au](mailto:keith.ooi@unsw.edu.au)

liquid chromatography-mass spectrometry (LC-MS) on faecal protein extracts from children with CF and their unaffected siblings to highlight the presence of an intestinal dysbiosis and inflammation in CF. Vernocchi *et al.*<sup>9</sup> utilised 16S rRNA sequencing and metabolomics (LC-MS) in children with CF and age-matched healthy controls (HC) to suggest that the gut microbiota enterotypes of CF, together with endogenous and bacterial CF biomarkers, reflect the expression of functional alteration at the intestinal level.

These studies all demonstrate an altered functional environment and with established and recent tools for microbial modulation (e.g. prebiotics, probiotics, synbiotics, antibiotics, diet and faecal microbiota transplantation), understanding these changes in the CF gut is essential for developing potential therapeutics. It is therefore essential to explore and define the functional differences between children with CF and well-matched HC. Furthermore, the links between intestinal bacteria and markers of inflammation have yet to be characterised in CF.

In this study we investigate composition and function of bacterial communities inhabiting the intestines of children with CF. We hypothesise that the composition and functional capacity of bacterial communities in the intestine of children with CF are different when compared to HC. We also hypothesise that intestinal bacteria and their functional profiles will be associated with biomarkers of intestinal inflammation. Furthermore, we hypothesise that these changes will be associated with growth and lung function in children with CF.

## Results

**Population characteristics.** Initially, 39 children with CF and 39 paired HC (matched for age and gender) were included, whose microbiota was assessed using 16S rRNA gene sequencing. After removal of low-coverage samples (sequencing depth <22,578, see below), the final study population comprised of 27 children with CF and 27 paired. The mean (SD) age of CF and HC subjects was 8.2 (5.0) and 7.7 (5.3) years respectively ( $p = 0.2$ ). Each group consisted of 14 males (51.9%). Of the CF children, 24 were pancreatic insufficient (PI) and 3 pancreatic sufficient (PS). Twenty-two of the CF children were homozygous F508del (81%), whilst the remaining 5 were heterozygotes. Clinical characteristics are presented in Table 1.

**Bacterial community characteristics.** A total of 3,179,543 bacterial 16S rRNA sequences were retrieved covering more than 72% of the bacterial diversity per sample. This is further supported by the rarefaction curves, which shows that the majority of bacterial communities in our samples are being covered by the sequencing effort undertaken here (Supp. Fig. 1).

**Alpha diversity.** Children with CF compared with HC had a significantly lower bacterial richness as assessed by the number of zOTUs (mean difference (95% CI) of  $-168.8$  ( $-219.5$  to  $-118.1$ ),  $p = 2.9 \times 10^{-07}$ ) and a lower diversity as assessed by the Shannon index (mean difference (95% CI) of  $-0.74$  ( $-1.00$  to  $-0.49$ ),  $p = 2.2 \times 10^{-06}$ ) (Fig. 1A,B). Controlling for age, children with CF had consistently lower richnesses (estimate (SE)  $-175.1$  (21.6),  $p = 1.0 \times 10^{-10}$ ) and Shannon diversity indices (estimate (SE)  $-0.78$  (0.15),  $p = 3.2 \times 10^{-06}$ ) (Fig. 1C,D).

**Phylogeny-based beta diversity.** Visualisation of phylogeny-based beta diversity (weighted and unweighted UniFrac distances) is presented in Fig. 2. PERMANOVA showed a significant difference in bacterial communities between the CF and HC cohorts (weighted UniFrac distance  $R^2 = 0.139$ ,  $p = 0.001$ , unweighted UniFrac distance  $R^2 = 0.121$ ,  $p = 0.001$ ). PERMANOVA showed no significant difference in bacterial communities between males and females (weighted UniFrac distance  $R^2 = 0.029$ ,  $p = 0.2$ , unweighted UniFrac distance  $R^2 = 0.018$ ,  $p = 0.5$ ), or with age (weighted UniFrac distance  $R^2 = 0.013$ ,  $p = 0.7$ , unweighted UniFrac distance  $R^2 = 0.026$ ,  $p = 0.09$ ).

**Differences in microbial communities between CF and HC populations.** The relative abundance of all bacterial phyla in CF and HC subjects are presented in Supplementary Fig. 2. Relative abundances of the top 10 most abundant bacterial genera in CF and HC are presented in Fig. 3. Bacterial taxa with a significant different abundance (ANCOM with  $q < 0.05$ ) between CF and HC populations are presented in Table 2 and Supplementary Fig. 3.

**Predicted functional analysis using Tax4Fun.** To predict the functional profiles of intestinal bacteria, we used Tax4Fun. On average per sample, 84.6% of zOTUs could be used by Tax4Fun to predict a functional profile of the community based on KEGG Orthology (KO) terms, demonstrating high predictive power in our analysis. In total, abundances of 275 unique KO pathways (based on 6,286 unique functions) were predicted in CF and HC populations. The abundance of 56 KO pathways was significantly different between CF and HC pairs ( $p < 0.00019$ ) (Table 3). Exploring the effect of age, linear models of pathways of interest were constructed and presented in Fig. 4.

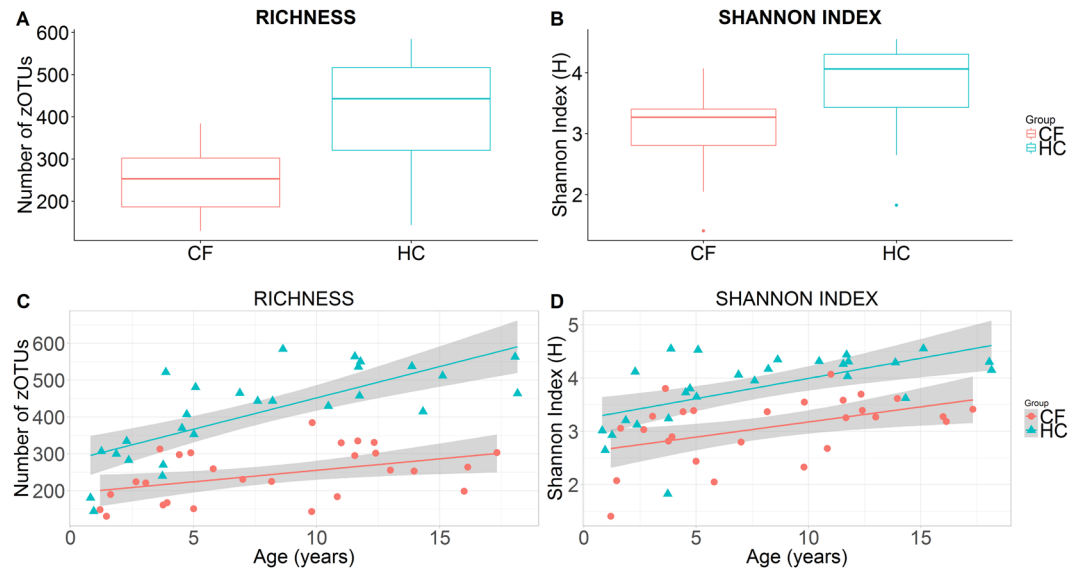
**Intestinal inflammation and its correlation with the bacterial community.** Faecal calprotectin levels were measured in 19/27 pairs (70.4%). Calprotectin levels were significantly elevated in CF compared with paired HC subjects (median (IQR) 92.0 mg/kg (58.7–142.3) vs. 19.5 mg/kg (19.5–19.8), respectively,  $p = 3.8 \times 10^{-6}$ ) (Fig. 5A). Controlling for age, children with CF had consistently higher faecal calprotectin levels than HC (estimate (SE) 110.6 mg/kg (32.9),  $p = 0.0019$ ) (Fig. 5C).

Faecal M2-PK levels were measured in 18/27 pairs (66.7%). M2-PK levels were significantly elevated in CF compared with paired HC subjects (median (IQR) 10.5 U/mL (4.7–46.0) vs. 1.0 U/mL (1.0–1.0) respectively,  $p = 0.00032$ ) (Fig. 5B). Controlling for age, children with CF had consistently higher faecal M2-PK levels than HC (estimate (SE) 36.8 U/mL (13.4),  $p = 0.0096$ ) (Fig. 5D).

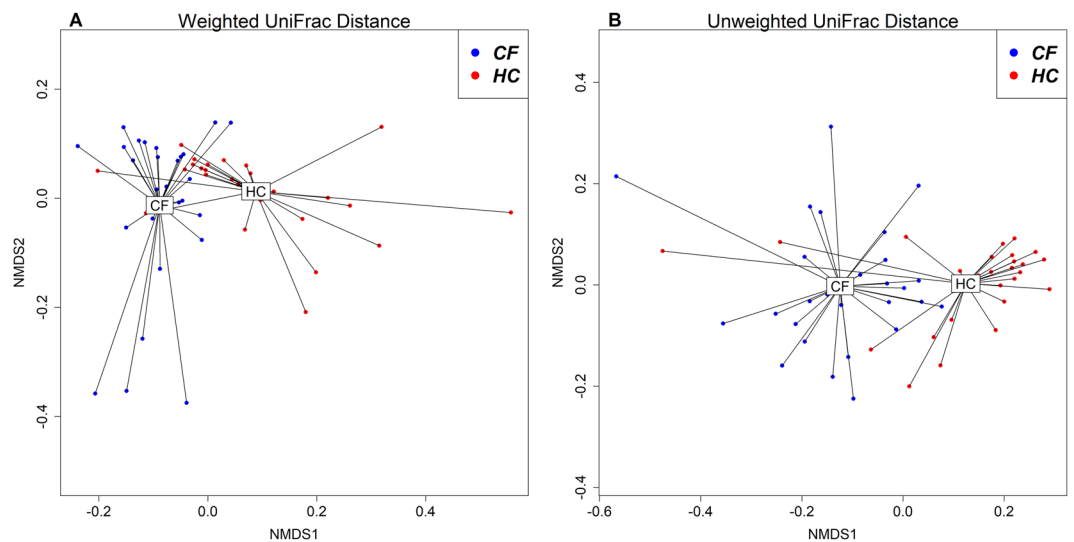
Correlations between bacterial taxa (at each taxonomic rank) and inflammatory markers (calprotectin ( $n = 27$ ) and M2-PK ( $n = 26$ )) within CF subjects are presented in Supplementary Fig. 4. Ten genera (*Acidaminococcus*

ID	Gender	Group	Pancreas	Age (yr)	Genotype	Calprotectin (mg/kg)	M2-PK (U/ml)
CF1	M	CF	PS	1.2	F508del/D1152H	101.3925	1
HC1	M	HC	—	0.8	—/—	<19.5325	—
CF3	F	CF	PI	1.5	F508del/F508del	27.915	8.029
HC3	F	HC	—	0.9	—/—	<19.5325	1
CF4	F	CF	PI	1.6	F508del/F508del	66.19	13.984
HC4	F	HC	—	1.3	—/—	<19.5325	1
CF5	M	CF	PI	2.7	F508del/F508del	60.545	3.814
HC5	M	HC	—	1.9	—/—	<19.5325	1
CF7	M	CF	PI	3.1	F508del/F508del	56.7975	35.035
HC7	M	HC	—	3.8	—/—	<19.5325	1
CF11	F	CF	PI	3.6	F508del/F508del	124.7	101.03
HC11	F	HC	—	2.3	—/—	20.0775	1
CF12	M	CF	PI	3.8	F508del/F508del	203.7725	203.95
HC12	M	HC	—	3.9	—/—	32.5975	1
CF13	M	CF	PI	3.9	F508del/F508del	<19.5325	7.147
HC13	M	HC	—	4.5	—/—	—	—
CF14	M	CF	PI	4.4	F508del/F508del	157.485	48.4625
HC14	M	HC	—	4.7	—/—	<19.5325	1
CF15	F	CF	PI	4.9	F508del/W1282X	61.3075	5.875
HC15	F	HC	—	2.4	—/—	24.9825	1
CF17	M	CF	PI	5.0	F508del/F508del	40.185	38.6475
HC17	M	HC	—	5.1	—/—	<19.5325	1
CF20	F	CF	PI	5.8	F508del/F508del	171.4025	6.133
HC20	F	HC	—	3.7	—/—	24.23	—
CF22	F	CF	PI	7.0	F508del/F508del	55.865	9.128
HC22	F	HC	—	5.0	—/—	—	—
CF25	M	CF	PI	8.2	F508del/F508del	<19.5325	50
HC25	M	HC	—	8.6	—/—	—	—
CF26	F	CF	PI	9.8	F508del/F508del	<19.5325	13.553
HC26	F	HC	—	6.9	—/—	—	—
CF27	F	CF	PI	9.8	F508del/F508del	106.29	3.962
HC27	F	HC	—	7.6	—/—	—	—
CF28	M	CF	PI	10.8	F508del/R334W	48.1175	1.397
HC28	M	HC	—	11.6	—/—	—	—
CF29	F	CF	PI	11.0	F508del/F508del	47.4175	3.227
HC29	F	HC	—	8.2	—/—	<19.5325	1
CF31	M	CF	PI	11.6	F508del/F508del	92	12.938
HC31	M	HC	—	11.7	—/—	<19.5325	1
CF32	M	CF	PI	11.7	F508del/F508del	50.775	52.43
HC32	M	HC	—	13.9	—/—	—	—
CF33	F	CF	PS	12.3	F508del/unknown	33.0875	1
HC33	F	HC	—	10.5	—/—	<19.5325	1
CF34	F	CF	PI	12.4	F508del/F508del	49.145	14.923
HC34	F	HC	—	11.7	—/—	—	—
CF35	M	CF	PI	13.0	F508del/F508del	360.3575	3.181
HC35	M	HC	—	15.1	—/—	<19.5325	1
CF36	M	CF	PS	14.0	F508del/V232D	127.1925	5.493
HC36	M	HC	—	18.1	—/—	<19.5325	3.264
CF37	F	CF	PI	16.0	F508del/F508del	645.5825	4.458
HC37	F	HC	—	11.8	—/—	<19.5325	1.816
CF38	M	CF	PI	16.1	F508del/F508del	95.7875	50
HC38	M	HC	—	18.2	—/—	27.675	1
CF39	F	CF	PI	17.3	F508del/F508del	80.0575	138.93
HC39	F	HC	—	14.3	—/—	<19.5325	1

**Table 1.** Clinical characteristics of included participants.



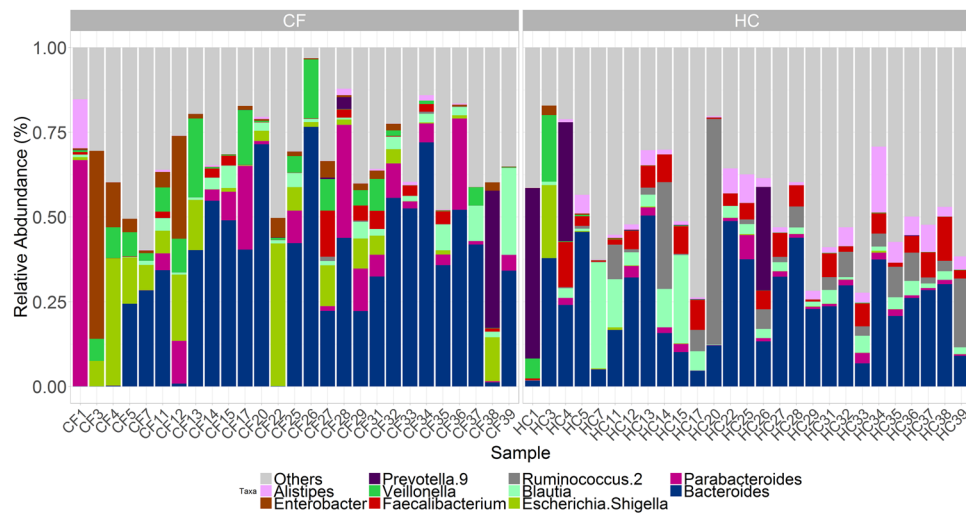
**Figure 1.** Boxplots of sample richness (number of zOTUs) (A) and Shannon index (B) in CF and HC cohorts. Scatterplots of sample richness (number of zOTUs) (C) and Shannon index (D) against age in CF and HC cohorts. Cohort mean and 95% confidence intervals are constructed from generalised linear models and presented as lines and shaded regions, respectively (C,D).



**Figure 2.** NMDS plots based on weighted (A) and unweighted (B) UniFrac distances between CF and HC cohorts.

( $r = 0.8$ ,  $q = 3.9 \times 10^{-7}$ ), *Allisonella* ( $r = 0.8$ ,  $q = 3.5 \times 10^{-7}$ ), *Eubacterium coprostanoligenes* group ( $r = 0.7$ ,  $q = 0.0002$ ), *Howardella* ( $r = 0.8$ ,  $q = 1.7 \times 10^{-7}$ ), *Lachnospiraceae* UCG-010 ( $r = 0.8$ ,  $q = 6.9 \times 10^{-7}$ ), *Mogibacterium* ( $r = 0.8$ ,  $q = 4.0 \times 10^{-7}$ ), *Olsenella* ( $r = 0.8$ ,  $q = 1.8 \times 10^{-7}$ ), *Sutterella* ( $r = 0.8$ ,  $q = 2.1 \times 10^{-5}$ ), uncultured *Lachnospiraceae* ( $r = 0.5$ ,  $q = 0.03$ ) and uncultured *Porphyromonadaceae* ( $r = 0.8$ ,  $q = 6.5 \times 10^{-6}$ ) were positively correlated with faecal calprotectin. *Lachnospiraceae* AC2044 group positively correlated with faecal M2-PK ( $r = 0.5$ ,  $q = 0.04$ ).

Correlations between predicted KEGG pathways and inflammatory markers (calprotectin ( $n = 27$ ) and M2-PK ( $n = 26$ )) within CF subjects are presented in Supplementary Fig. 5. Six predicted pathways (benzoate degradation (KO362) ( $r = 0.6$ ,  $q = 0.005$ ), carbon fixation pathways in prokaryotes (KO720) ( $r = 0.5$ ,  $q = 0.02$ ), DDT degradation (KO351) ( $r = 0.6$ ,  $q = 0.01$ ), phenylalanine metabolism (KO360) ( $r = 0.6$ ,  $q = 0.009$ ), styrene degradation (KO643) ( $r = 0.5$ ,  $q = 0.048$ ) and tropine, piperidine and pyridine alkaloid biosynthesis (KO960) ( $r = 0.6$ ,  $q = 0.007$ )) were positively correlated with faecal calprotectin. Indole alkaloid biosynthesis (KO901) ( $r = 0.5$ ,  $q = 0.01$ ) and Photosynthesis – antenna proteins (KO196) ( $r = 0.5$ ,  $q = 0.04$ ) positively correlated with faecal M2-PK.



**Figure 3.** Relative abundance of top 10 most abundant bacterial genera for CF and HC subjects. Samples ordered in increasing age (from left to right).

**Associations with clinical measures in children with cystic fibrosis.** The mean (SD) weight, height and BMI z-scores in children with CF were  $-0.0004$  (1.1),  $0.07$  (1.2), and  $0.1$  (1.0) respectively. No significant correlations between anthropometric z-scores and alpha diversity indices (richness or Shannon index) were identified. Correlations between bacterial genera and anthropometric z-scores identified positive correlations between: (i) *Ruminococcaceae UCG 014* and BMI ( $r = 0.6$ ,  $q = 0.04$ ), (ii) uncultured *VadinBE97* and weight ( $r = 0.5$ ,  $q = 0.049$ ), and (iii) uncultured *VadinBE97* and BMI ( $r = 0.6$ ,  $q = 0.02$ ). No significant correlations between anthropometric z-scores and predicted KEGG pathways were identified.

Lung function was recorded in 19 (70%) children, with a mean (SD) FEV1% of 98.9% (15.4). No significant correlations between FEV1% and alpha diversity indices (richness or Shannon index) were identified. FEV1% positively correlated with five genera: (i) *Adlercreutzia* ( $r = 0.6$ ,  $q = 0.03$ ), (ii) *Ruminococcaceae NK4A214* group ( $r = 0.6$ ,  $q = 0.04$ ), (iii) *Lachnospiraceae NC2004* group ( $r = 0.6$ ,  $q = 0.04$ ), (iv) *Tyzzereella 3* ( $r = 0.6$ ,  $q = 0.04$ ), and (v) *Candidatus Soleaferrea* ( $r = 0.5$ ,  $q = 0.047$ ). No significant correlations between FEV1% and predicted KEGG pathways were identified.

**Metabolomics.** Untargeted metabolomics data was available for 12 CF and 4 HC children as part of a previous study<sup>22</sup>. A total of 24,814 and 13,341 hits were detected in positive and negative ion modes, respectively, from all stool samples. Metabolites were identified using the human metabolome database (HMDB)<sup>23</sup> and the mean (SD) number of identified metabolites per sample in positive and negative ion modes were 6,409 (690) and 1,845 (150), respectively. These lists of identified metabolites were used for all subsequent analyses. The metabolomes of CF and HC cohorts revealed distinct clustering (Supp. Figs. 7, 8). Principal component analysis (PCA) revealed a strong separation between CF and HC cohorts (Supp. Fig. 8A,C). A sensitivity analysis with 4 CF and 4 age and gender matched HC children also revealed a strong separation between CF and HC cohorts (Supp. Fig. 8B,D). Metabolites with a significantly different abundance between CF and HC cohorts ( $q < 0.05$ ) are presented in a volcano plot (Supp. Fig. 9) and hierarchical clustering presented in Supplementary Fig. 10. Correlations between metabolites and bacterial genera (significant genera identified in Table 2) in children with CF are presented in Supplementary Fig. 10 and Supplementary Dataset 6. Correlations between metabolites and predicted KEGG pathways (significant pathways identified in Table 3) in children with CF are presented in Supplementary Fig. 12 and Supplementary Dataset 6.

A post hoc analysis of metabolites related to short-chain fatty acids (SCFA) was performed. A total of 25 and 2 hits were putatively identified as metabolites related to propanoate and/or butanoate metabolism<sup>23</sup> in positive and negative ion modes, respectively. Two metabolites identified as butyric acid (positive ion mode): (i)  $m/z$  106, RT 10.0 min and (ii)  $m/z$  106, RT 9.8 min were constantly lower in children with CF compared to HC across all ages (normalised abundance estimate (SE)  $-10,134.4$  (2,631.8),  $p = 0.002$  and  $-6,841.5$  (3,011.4),  $p = 0.04$ , respectively) (Supp. Fig. 13). One metabolite identified as pantetheine (negative ion mode;  $m/z$  277, RT 11.5 min), an amide of pantothenic acid (vitamin B5), was significantly lower in children with CF than HC (normalised abundance median (IQR) 2,795 (1,343–16,628) vs. 35,418 (24,798–43,056),  $p = 0.045$ ) (Supp. Fig. 14). Pantetheine is involved in the metabolism of propanoate to propionyl coenzyme A<sup>23</sup>.

## Discussion

We have demonstrated that the intestines of children with CF exhibit a marked taxonomic and inferred functional dysbiosis when compared to well-matched healthy controls. To our knowledge this is the first study to infer the function of intestinal bacterial communities in a paediatric CF population. We predicted an enrichment of genes involved in SCFA, antioxidant and nutrient metabolism, all of which are relevant to growth and nutrition in CF (Table 3; Fig. 4). We identified in a subset of samples, that children with CF have distinct metabolic profiles compared to HC. Reduced levels of butyric acid and pantetheine in the stool of children with CF provides

Phylum	Class	Order	Family	Genus	
↑ Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	↑ <i>Fusobacterium</i>	
↑ Proteobacteria	↑ Gammaproteobacteria	↑ Enterobacteriales	↑ Enterobacteriaceae	↑ <i>Enterobacter</i>	
				↑ <i>Escherichia</i> , <i>Shigella</i>	
				↑ <i>Morganella</i>	
↓ Verrucomicrobia	↓ Verrucomicrobiae	↓ Verrucomicrobiales	↓ Verrucomicrobiaceae		
Firmicutes	↑ Negativicutes	↑ Selenomonadales	↑ Veillonellaceae	↑ <i>Veillonella</i>	
	Mollicutes	↓ Mollicutes RF9	↓ uncultured Mollicutes RF9		
				↓ Christensenellaceae	↓ <i>Christensenellaceae</i> R7 group
				↓ Family XIII	
				↓ Ruminococcaceae	↓ <i>Ruminiclostridium</i> 6
					↓ <i>Ruminococcaceae</i> UCG 002
					↓ <i>Ruminococcaceae</i> UCG 005
					↓ <i>Ruminococcaceae</i> UCG 013
					↓ <i>Ruminococcaceae</i> UCG 014
					↓ <i>Ruminococcus</i> 1
					↓ <i>Ruminococcus</i> 2
					↓ <i>Eubacterium coprostanoligenes</i> group
				Eubacteriaceae	↓ <i>Eubacterium eligens</i> group
				Lachnospiraceae	↓ <i>Lachnospira</i>
	↓ <i>Lachnospiraceae</i> NK4A136 group				
↑ <i>Ruminococcus gnavus</i> group					
↓ <i>Eubacterium ventriosum</i> group					
↓ <i>Eubacterium xylanophilum</i> group					
	↑ <i>Tyzzereella</i> 4				
Bacilli	Lactobacillales	↑ Enterococcaceae	↑ <i>Enterococcus</i>		
Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	↓ <i>Erysipelotrichaceae</i> UCG 003		
			↑ <i>Clostridium innocuum</i> group		
Bacteroidetes	Bacteroidia	Bacteroidales	↓ Rikenellaceae	↓ <i>Alistipes</i>	

**Table 2.** Bacterial taxa with a significant different abundance (ANCOM with  $q < 0.05$ ) between CF and HC populations at each taxonomic rank (bold). Arrows indicate if the relative abundance of each taxa is higher (↑) or lower (↓) in CF compared with HC populations. Taxa without an arrow describe phylogeny and were not significantly different in abundance. ANCOM analysis presented in Supplementary Fig. 3.

support for the notion of increased SCFA (butanoate and propanoate) metabolism. Furthermore, the notion of pro-inflammatory GI microbiota in children with CF is supported by positive correlations between intestinal inflammatory markers and both genera and functional pathways. We also demonstrated an association between intestinal genera and both growth z-scores and FEV1%. Thus, this study provides evidence and insights into the links between alterations in the composition of the intestinal bacterial community with (i) predicted functional changes, (ii) intestinal inflammation, and (iii) clinical outcomes (e.g. growth and lung function). However, considering the nature of 16S rRNA-derived gene data, the precise pathogenic mechanisms in CF GI disease need further exploration and validation.

The paediatric CF GI microbiota demonstrated significantly lower  $\alpha$ -diversity indices (richness and Shannon index) across all ages (Fig. 1). Increased GI diversity has been repeatedly associated with health, whilst decreased diversity has been associated with several inflammatory, metabolic, immune-mediated and systemic diseases<sup>24</sup>. These changes are clinically relevant in CF as the GI microbiota of young children has been proposed to be a determinant of respiratory and systemic disease progression<sup>25</sup>. Intestinal dysbiosis is further supported by the significant difference in bacterial communities between CF and HC cohorts using both weight and unweighted UniFrac distances (Fig. 2). Similarly Vernocchi *et al.*<sup>9</sup> demonstrated that children (aged 1 to 6 years) with CF have lower alpha diversity (Chao1) and a distinct beta diversity (unweighted UniFrac distance) compared to HC.

Similar to previous reports, we identified a marked increase in the relative abundance of *Proteobacteria* (genera *Escherichia*, *Shigella*, *Enterobacter* and *Morganella*), particularly in the first few years of life (Fig. 3; Table 2 and Supp. Figure 3A,E)<sup>6,8,26</sup>. In our samples, *Escherichia* was predominantly comprised of *Escherichia coli* species, which have been shown to positively correlate with faecal measures of nutrient malabsorption and inflammation in CF<sup>26</sup>. The relative abundances of functions related to two intestinal pathogens were increased in CF, shigellosis ( $p = 3.0 \times 10^{-8}$ ) and *Vibrio cholerae* (pathogenic cycle) ( $p = 0.0001$ ) (Table 3). This likely reflects the enrichment of *Escherichia* and *Enterobacter*, which share several common genes, rather than the presence of the pathogens themselves. The clinical implications of this however are unclear.

The increased relative abundance of *Veillonella* in CF samples (Fig. 3; Table 2 and Supp. Figure 3E) has also previously been observed<sup>8,25,27</sup> and *Veillonella* comprises part of the core respiratory microbiota in CF<sup>25,27</sup>.

PATHWAY	KO	CF*	Association	CF Rel Abund Median % (IQR)	HC Rel Abund Median % (IQR)	Mean Difference** Est % (95% CI)	P-value
Pyrimidine metabolism	240	↓	Glucose, Fatty acids	2.78 (2.6–2.905)	3.3 (3.09–3.566)	−0.733 (−1.025–0.44)	8.32E-06
Alanine aspartate and glutamate metabolism	250	↓	Amino acids	1.41 (1.32–1.5)	1.78 (1.61–1.945)	−0.459 (−0.687–0.232)	1.88E-05
Thiamine metabolism	730	↓	Vitamins	0.792 (0.736–0.914)	1.14 (1.002–1.33)	−0.43 (−0.632–0.228)	1.31E-06
Lipopolysaccharide biosynthesis	540	↑	Fatty acids	0.913 (0.819–1.075)	0.54 (0.347–0.691)	0.365 (0.227–0.503)	3.52E-05
Pertussis	5133	↑	Bacteria	0.376 (0.226–0.669)	0.134 (0.093–0.164)	0.353 (0.217–0.489)	1.64E-06
Bacterial chemotaxis	2030	↓	Bacteria	0.615 (0.455–0.756)	0.922 (0.758–1.05)	−0.301 (−0.417–0.185)	1.35E-05
Cell cycle Caulobacter	4112	↓	Bacteria	1.31 (1.245–1.365)	1.56 (1.49–1.635)	−0.271 (−0.338–0.204)	5.93E-06
Pantothenate and CoA biosynthesis	770	↓	Vitamins	0.753 (0.74–0.831)	0.94 (0.857–1.015)	−0.209 (−0.307–0.111)	3.07E-05
Homologous recombination	3440	↓	DNA repair	1.61 (1.535–1.635)	1.72 (1.67–1.885)	−0.208 (−0.293–0.123)	3.97E-05
Glutathione metabolism	480	↑	Antioxidants	0.608 (0.556–0.651)	0.444 (0.401–0.477)	0.182 (0.13–0.233)	0.00002
Glyoxylate and dicarboxylate metabolism	630	↑	Glucose, Fatty acids	0.767 (0.723–0.912)	0.677 (0.617–0.696)	0.153 (0.101–0.205)	5.53E-06
Pyruvate metabolism	620	↑	Glucose, Fatty acids	1.19 (1.14–1.255)	1.08 (1.045–1.105)	0.143 (0.102–0.184)	9.29E-06
Ubiquinone and other terpenoid quinone biosynthesis	130	↑	Antioxidants	0.541 (0.488–0.631)	0.423 (0.379–0.486)	0.123 (0.074–0.171)	3.52E-05
Phenylalanine tyrosine and tryptophan biosynthesis	400	↓	Amino acids	1.15 (1.125–1.21)	1.24 (1.22–1.29)	−0.11 (−0.159–0.061)	9.64E-05
Citrate cycle TCA cycle	20	↑	Fatty acids, Glucose	0.553 (0.517–0.595)	0.444 (0.386–0.482)	0.109 (0.072–0.146)	6.66E-06
Vibrio cholerae pathogenic cycle	5111	↑	Bacteria	0.189 (0.163–0.33)	0.149 (0.112–0.164)	0.105 (0.061–0.149)	0.00014
Lysine degradation	310	↑	Amino acids	0.357 (0.331–0.387)	0.27 (0.237–0.293)	0.1 (0.073–0.128)	1.04E-07
Propanoate metabolism	640	↑	Fatty acids, SCFA	0.433 (0.415–0.471)	0.362 (0.339–0.383)	0.099 (0.073–0.125)	6.64E-06
Nicotinate and nicotinamide metabolism	760	↓	Vitamins	0.743 (0.718–0.774)	0.824 (0.797–0.903)	−0.093 (−0.135–0.05)	0.00013
Butanoate metabolism	650	↑	Fatty acids, SCFA	0.657 (0.624–0.692)	0.591 (0.57–0.608)	0.092 (0.058–0.125)	1.98E-05
Photosynthesis	195	↓	Bacteria	0.284 (0.26–0.304)	0.356 (0.318–0.375)	−0.084 (−0.116–0.052)	3.59E-05
Phenylalanine metabolism	360	↑	Amino acids	0.37 (0.322–0.452)	0.307 (0.295–0.337)	0.082 (0.044–0.119)	0.00018
Sulfur metabolism	920	↑	Amino acids	0.432 (0.399–0.45)	0.344 (0.305–0.379)	0.081 (0.055–0.108)	2.76E-05
Polycyclic aromatic hydrocarbon degradation	624	↓	Benzenes	0.23 (0.197–0.254)	0.292 (0.27–0.318)	−0.081 (−0.118–0.045)	1.35E-05
Valine leucine and isoleucine degradation	280	↑	Amino acids	0.383 (0.346–0.414)	0.313 (0.268–0.351)	0.077 (0.041–0.113)	0.00014
Protein processing in endoplasmic reticulum	4141	↓	Proteins	0.081 (0.066–0.099)	0.142 (0.116–0.166)	−0.075 (−0.111–0.04)	1.14E-05
D-Alanine metabolism	473	↓	Glucose	0.278 (0.256–0.316)	0.361 (0.33–0.379)	−0.071 (−0.096–0.045)	8.14E-05
Biosynthesis of siderophore group nonribosomal peptides	1053	↑	Iron	0.098 (0.082–0.138)	0.064 (0.055–0.07)	0.07 (0.041–0.099)	1.04E-06
Nitrotoluene degradation	633	↑	Benzenes	0.138 (0.104–0.184)	0.095 (0.085–0.115)	0.058 (0.029–0.087)	0.00013
Fatty acid metabolism	71	↑	Fatty acids	0.197 (0.183–0.252)	0.169 (0.144–0.174)	0.052 (0.034–0.07)	1.64E-06
Phosphonate and phosphinate metabolism	440	↑	Bacteria	0.168 (0.144–0.184)	0.113 (0.093–0.134)	0.048 (0.031–0.066)	8.14E-05
Peroxisome	4146	↑	Lipids	0.19 (0.176–0.207)	0.148 (0.127–0.164)	0.047 (0.028–0.065)	9.45E-05
Plant pathogen interaction	4626	↑	Bacteria	0.266 (0.253–0.278)	0.23 (0.215–0.243)	0.045 (0.024–0.067)	3.52E-05
Biosynthesis of unsaturated fatty acids	1040	↑	Fatty acids	0.102 (0.098–0.137)	0.085 (0.074–0.088)	0.036 (0.026–0.046)	1.49E-08
Tryptophan metabolism	380	↑	Amino acids	0.06 (0.054–0.117)	0.051 (0.044–0.054)	0.036 (0.022–0.05)	0.00002
Inositol phosphate metabolism	562	↑	Cellular function	0.204 (0.19–0.211)	0.174 (0.156–0.182)	0.034 (0.022–0.046)	3.78E-05
Carbon fixation in photosynthetic organisms	710	↓	Bacteria	0.429 (0.421–0.441)	0.456 (0.448–0.471)	−0.033 (−0.049–0.016)	0.00012
Taurine and hypotaurine metabolism	430	↑	Antioxidants	0.114 (0.104–0.132)	0.088 (0.082–0.092)	0.031 (0.024–0.038)	1.49E-08
Shigellosis	5131	↑	Bacteria	0.009 (0.003–0.024)	0.001 (0.0004–0.002)	0.026 (0.007–0.046)	2.98E-08
Toluene degradation	623	↑	Benzenes	0.075 (0.068–0.098)	0.059 (0.052–0.062)	0.023 (0.016–0.031)	6.41E-07
Bisphenol degradation	363	↑	Benzoates	0.196 (0.188–0.208)	0.176 (0.167–0.185)	0.023 (0.015–0.031)	5.15E-05
Glycosphingolipid biosynthesis lacto and neolacto series	601	↓	Cell membrane	0.018 (0.014–0.025)	0.036 (0.027–0.043)	−0.022 (−0.033–0.011)	4.75E-05
Chlorocyclohexane and chlorobenzene degradation	361	↑	Benzoates	0.051 (0.044–0.07)	0.041 (0.037–0.044)	0.018 (0.01–0.025)	1.88E-05
Penicillin and cephalosporin biosynthesis	311	↑	Antibiotics	0.076 (0.067–0.082)	0.06 (0.055–0.068)	0.017 (0.011–0.024)	4.57E-06
Phosphatidylinositol signaling system	4070	↑	Lipids	0.066 (0.058–0.07)	0.053 (0.046–0.058)	0.015 (0.008–0.021)	4.57E-06
Ethylbenzene degradation	642	↓	Benzenes	0.094 (0.086–0.098)	0.106 (0.102–0.111)	−0.014 (−0.02–0.008)	2.21E-05
Basal transcription factors	3022	↓	Transcription	0.013 (0.007–0.019)	0.024 (0.017–0.03)	−0.011 (−0.015–0.006)	1.59E-05
Fluorobenzoate degradation	364	↑	Benzoates	0.012 (0.01–0.024)	0.008 (0.005–0.009)	0.01 (0.006–0.014)	3.77E-06
Huntingtons disease	5016	↑	Genetic	0.031 (0.03–0.034)	0.024 (0.019–0.026)	0.009 (0.006–0.013)	3.41E-05
Type I diabetes mellitus	4940	↓	Autoimmune	0.04 (0.038–0.042)	0.047 (0.043–0.049)	−0.008 (−0.012–0.005)	1.25E-05
Amyotrophic lateral sclerosis ALS	5014	↑	Autoimmune	0.012 (0.009–0.017)	0.005 (0.004–0.007)	0.008 (0.006–0.011)	1.49E-08
alpha Linolenic acid metabolism	592	↑	Fatty acids	0.006 (0.004–0.009)	0.003 (0.003–0.004)	0.005 (0.003–0.007)	4.47E-08
Continued							

PATHWAY	KO	CF*	Association	CF Rel Abund Median % (IQR)	HC Rel Abund Median % (IQR)	Mean Difference** Est % (95% CI)	P-value
Chagas disease American trypanosomiasis	5142	↑	Parasites	0.013 (0.011–0.014)	0.009 (0.006–0.011)	0.005 (0.002–0.008)	8.37E-05
VEGF signaling pathway	4370	↓	Blood vessel formation	0.0005 (0.0002–0.002)	0.005 (0.003–0.006)	–0.004 (–0.007–0.002)	1.59E-05
Calcium signaling pathway	4020	↓	Cell signalling	0.001 (0.001–0.003)	0.005 (0.003–0.007)	–0.004 (–0.007–0.002)	0.00013
Influenza A	5164	↓	Viruses	0.002 (0.002–0.003)	0.005 (0.004–0.006)	–0.003 (–0.005–0.002)	6.41E-07

**Table 3.** Predicted KEGG pathways with a significant different abundance between CF and HC pairs ( $p < 0.00019$ ). Samples are ordered by absolute mean difference (decreasing). \*Arrows indicate if the relative abundance of each pathway is higher (↑) or lower (↓) in CF compared with HC. \*\*Estimate of mean difference between pairs based on Wilcoxon signed-rank testing. Est, estimate. KO, KEGG Orthology number. Rel Abund, relative abundance.

*Veillonella* is known to produce propionate (i.e. propanoate) from lactate fermentation<sup>28</sup> and the relative abundance of airway *Veillonella* has been reported to inversely correlate with airway inflammation<sup>29</sup>.

*Fusobacterium* was significantly enriched in children with CF compared to HC, and this genus has been extensively linked to colorectal cancer<sup>30</sup>. This finding is relevant given the 5–10 times greater risk of colorectal cancer in adults with CF compared to the general population<sup>31</sup>.

Several members of the *Ruminococcaceae* family were reduced in CF samples (Fig. 3; Table 2 and Supp. Figure 3D,E) and they are known producers of SCFAs (from fermentation of carbohydrates, including resistant starch)<sup>32</sup>. Interestingly in CF children, the relative abundance of *Ruminococcaceae* UCG 014 positively correlated with BMI z-scores ( $r = 0.6$ ,  $q = 0.04$ ) and *Ruminococcaceae* NK4A214 group positively correlated with FEV1% ( $r = 0.6$ ,  $q = 0.04$ ). Dietary studies examining resistant starch intake and potential interventional studies may be worth exploring in children with CF.

The relative abundance of *Alistipes* was decreased in CF samples (Fig. 3; Table 2 and Supp. Figure 3E) and species of this genus are known to produce succinic acid<sup>33</sup>. In mice, dietary succinate was identified as a substrate for intestinal gluconeogenesis and improved glucose homeostasis<sup>34</sup>. This may be relevant to the exploration of factors associated with glucose abnormalities in children with CF, an increasingly recognised problem<sup>35</sup>.

Intestinal inflammation as measured by faecal calprotectin is clinically relevant given its negative correlation with height and weight z-scores in CF children<sup>11</sup>. Both faecal calprotectin and M2-PK were significantly elevated in our CF cohort compared with HC (Fig. 5). Interestingly, *Acidaminococcus* showed a strong positive correlation with calprotectin ( $r = 0.83$ ,  $q = 3.9 \times 10^{-7}$ ) in CF (Supp. Figure 4D). Increased *Acidaminococcus* relative abundance has been associated with lower future height z scores in twin-cohorts of children from Malawi and Bangladesh<sup>36</sup>. *Acidaminococcus* sp. consume glutamate (as their sole source of carbon and energy) which is important to gut amino acid metabolism, nitrogen balance, barrier function and epithelial restitution<sup>36</sup>.

Children with CF appeared to have distinct metabolic profiles compared to HC (Supp. Figs. 7–10). Interestingly, those children with CF and PS (samples CF33 and CF36) appeared similar to HC rather than PI CF children (Supp. Figs. 7, 8, 10A), providing support for a gradation effect with the degree of CFTR dysfunction. The taxonomic and functional dysbiosis which we have identified, along with the known CFTR dysfunction are plausible factors contributing to the altered metabolic profiles in children with CF (compared to HC). In children with CF, we identified several significant correlations between metabolites and (i) bacterial genera (Supp. Fig. 11), and (ii) predicted KEGG pathways (Supp. Fig. 12), suggesting that intestinal dysbiosis has an influence on metabolic profiles.

In our CF cohort there was a predicted enrichment of genes involved in the metabolism of SCFAs propanoate, butyrate and a precursor, pyruvate (Table 3 and Fig. 4A,B). Short-chain fatty acids have several beneficial effects on GI tract health including, improved intestinal motility, reduced intestinal inflammation, promotion of differentiation of regulatory T cells and regulation of both fatty acid and glucose metabolism<sup>37,38</sup>. Recently, Vernocchi *et al.*<sup>9</sup> reported butyrate and propionate levels to be lower in children with CF compared to HC. Similarly, in a subset of samples, we identified lower levels of butyric acid and pantotheine in children with CF compared to HC. It is likely that in children with CF, intestinal microbiota have an increased propensity for the metabolism of beneficial SCFAs. The variable trend of propanoate metabolism with advancing age in CF subjects (Fig. 4A) highlights the need to perform a quantitative analysis of dietary-fibre intake and faecal excretion of SCFAs.

In children with CF, we predicted an enrichment of bacterial genes associated with glutathione, taurine and hypotaurine metabolism and ubiquinone (coenzyme Q10) biosynthesis (Table 3 and Fig. 4G–I). Both glutathione and taurine are antioxidants which are protective against oxidative stress and exhibit anti-inflammatory properties<sup>39</sup>. This is notable in light of a recent randomised controlled trial on oral reduced glutathione demonstrating improved growth z-scores and decreased faecal calprotectin in CF children<sup>18</sup>. Antioxidants have been explored as a therapeutic intervention in CF lung disease, however evidence is lacking to support their use<sup>40</sup>. Little is known about the role of oxidative stress in the intestinal metabolism of CF and further exploration into the role of antioxidants for CF GI disease is warranted.

The relative abundance of predicted functions associated with phenylalanine metabolism were increased in CF ( $p = 0.0002$ ) and also positively correlated with calprotectin ( $r = 0.6$ ,  $q < 0.01$ ) (Table 3, Fig. 4J and Supp. Fig. 14). Phenylalanine is an essential and aromatic amino acid which may play several key roles including: (i) attenuating intestinal inflammation through activating calcium-sensing receptors (in piglets)<sup>41</sup>; (ii) inhibiting TNF- $\alpha$  production<sup>42</sup>; and (iii) enhancing immune responses<sup>42</sup>. Quantification of phenylalanine intake may be a useful measure in future CF studies.



In our CF cohort predicted functions related to the metabolism of thiamine (vitamin B<sub>1</sub>) ( $p = 1.3 \times 10^{-6}$ ) and biosynthesis of pantothenate (vitamin B<sub>5</sub>) and coenzyme A (CoA) ( $p = 3.1 \times 10^{-5}$ ) were decreased (Table 3 and Fig. 4K,L). These two water-soluble vitamins help metabolise carbohydrates, fats and proteins and their depletion may provide insights into the pathophysiology of malnutrition in CF (despite pancreatic enzyme replacement therapy, high-fat high-calorie diets and vitamin supplementation<sup>43</sup>).

Some limitations of this current study need to be considered. Firstly, functional pathways were inferred from 16S rRNA gene data and not measured directly. However, given that 84.6% of our 16S rRNA genes could be assigned a functional profile, we believe our analysis still has strong predictive power. Secondly, although our sample size is small, it is still the largest study to date analysing intestinal microbiota function in CF children and HC (age and gender matched) and exceeded the minimum number required (calculation below). Given only three CF children were PS, we were unable to make any meaningful comparisons with PI children. Furthermore, for children with CF, FEV1% was not recorded at the time of stool sampling but within the preceding 12 months. And finally, confounding factors, including antibiotic usage and altered dietary regimes between CF and HC, were not controlled for with this study. A quantitative analysis of dietary intake and direct measures of faecal proteins and metabolites (i.e. targeted metabolomics for measurement of SCFA levels) would be required to validate the findings presented in this study. Although our comparison of metabolomes between CF and HC cohorts was unmatched, we performed a sensitivity analysis with 4 CF and 4 age and gender matched HC which yielded similar results (Supp. Fig. 8B,D).

In conclusion, there exists both a taxonomic and inferred functional dysbiosis, which provides insights into paediatric CF GI disease. Future observational or interventional studies should simultaneously evaluate dietary intake, abdominal symptoms and faeces. Further exploration of potential CF GI therapeutics including antioxidants (e.g. glutathione), SCFAs (e.g. propanoate and butyrate), amino acids (e.g. phenylalanine) and gut microbiota modulators (e.g. prebiotics including resistant starch, probiotics and synbiotics) is warranted.

## Methods

**Study population.** We performed a prospective, cross-sectional, observational study in children with CF and HC. The subjects and samples for this analysis were collected as part of three prior studies evaluating the progression of intestinal microbiota<sup>6</sup> and inflammation in children with CF<sup>12,14</sup>. Children with CF and HC (aged 0 to 18 years) were prospectively recruited from the outpatient clinics (CF and Orthopaedic/Plastics clinics respectively) at Sydney Children's Hospital Randwick, Australia. In addition, inclusion and exclusion criteria were as follows:

Inclusion criteria:

- Children with CF diagnosed according to the United States Cystic Fibrosis Foundation consensus criteria<sup>44</sup>.
- Exocrine pancreatic insufficiency (PI) or pancreatic sufficiency (PS) defined based on the 72-hour faecal fat and/or faecal elastase-1<sup>45,46</sup>.
- Children with CF on oral or inhaled antibiotic therapy were not excluded.
- Healthy controls included children without CF or any gastrointestinal disease.

Exclusion criteria:

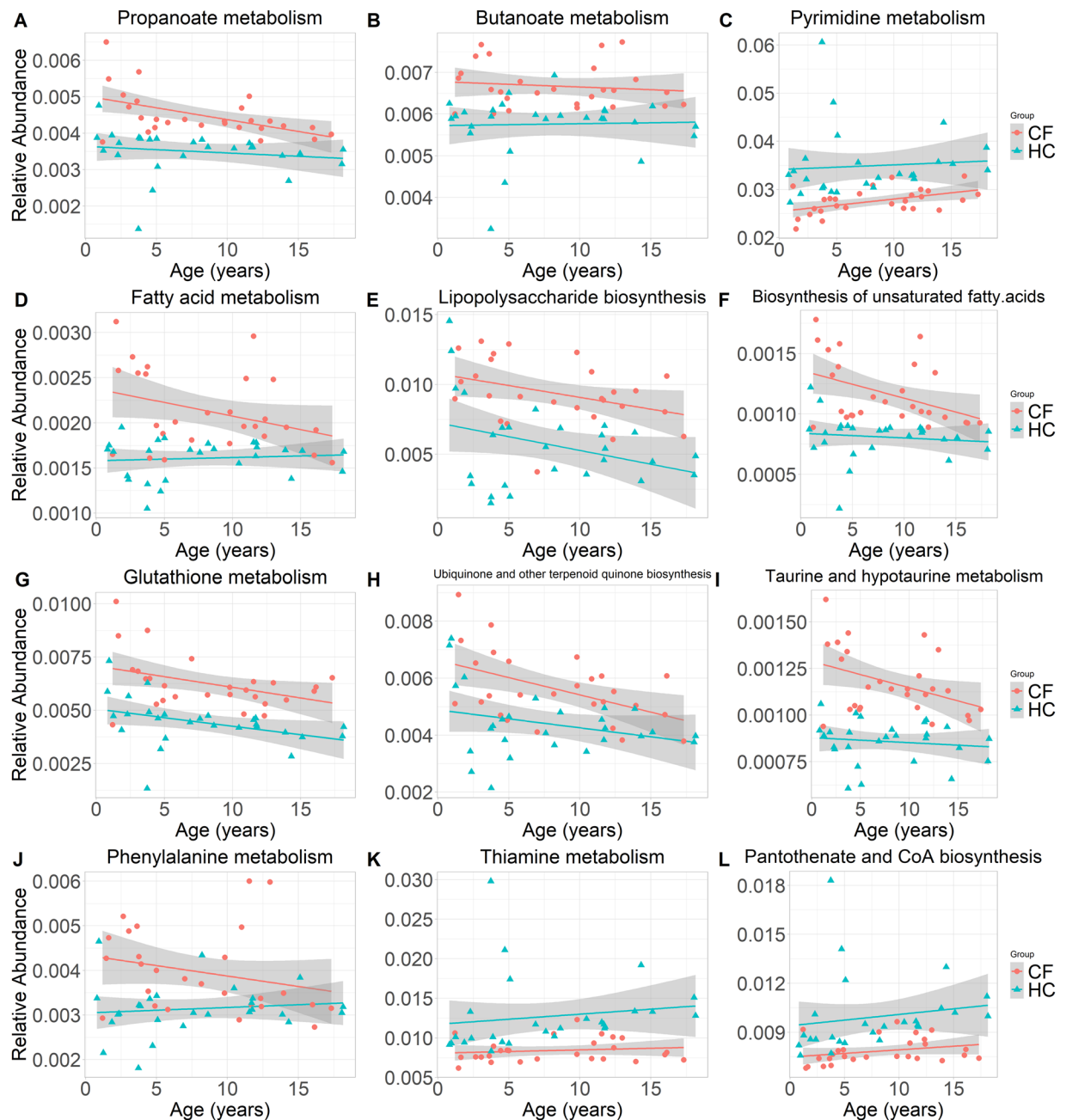
- Patients requiring antibiotic therapy for a pulmonary exacerbation<sup>47</sup> or intravenous antibiotics in the preceding four weeks before sampling.
- Any child (CF or HC) with gastroenteritis, on oral corticosteroids, probiotics and/or non-steroidal anti-inflammatory drugs in the preceding two weeks before sampling.

Children with CF were matched to healthy controls for gender and age (as closely as possible).

**Sample collection and processing.** A single stool sample from each subject was collected and stored immediately at  $-80^{\circ}\text{C}$ , or stored at  $-20^{\circ}\text{C}$  (home freezer) until transport to the laboratory for storage at  $-80^{\circ}\text{C}$ . Thawing of the sample during transport did not occur<sup>6</sup>. At the time of sample collection, demographic and anthropometric (height, weight and body mass index (BMI) z-scores) data was recorded. In CF children older than four years, the forced expiratory volume in one second, percent predicted (FEV1%) from their most recent lung function test was recorded (within the preceding 12 months).

DNA was extracted from stool samples and sequenced as previously described<sup>6</sup>. The 16S rRNA genes of the gut microbiota were amplified with primers 27F (AGAGTTTGATCMTGGCTCAG) and 519R (GWATTACCGCGKCGCTG) spanning the V1–V3 regions and sequenced using the Illumina MiSeq platform (v3,  $2 \times 300$  bp).

**Processing of 16S rRNA data.** Paired-end sequences were quality filtered using Trimmomatic v0.36<sup>48</sup>: low quality reads were truncated if the quality dropped below 15 in a sliding window of 4 bp. Reads shorter than 36 bp were removed. Remaining paired-end reads were merged and then filtered using USEARCH v9.2.64<sup>49</sup>. Filtering included the removal of reads shorter than 400 bp or longer than 500 bp as well as the removal of low-quality reads (expected error  $> 1$ ) and reads with more than one ambiguous base. Processed sequences of all samples were concatenated into one file and subsequently dereplicated into unique sequences. These sequences were clustered in unique sequences (zero-distance operational taxonomic unit; zOTU) with the unoise<sup>50</sup> algorithm implemented in USEARCH. A denovo chimera removal step was included in the unoise step. Subsequently, the uchime algorithm was utilized in reference mode with the SILVA 123 database<sup>51</sup> to remove remaining chimeric sequences.



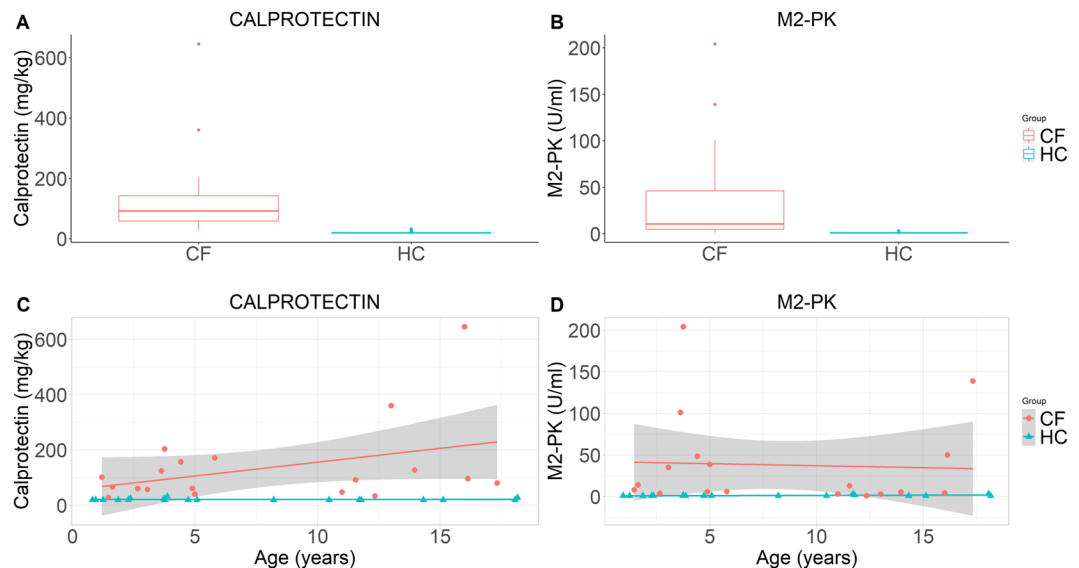
**Figure 4.** Scatterplots of the relative abundance of predicted KEGG pathways against age in CF and HC cohorts. Predicted functions associated with: (i) short-chain fatty acids (A,B), (ii) fatty acid metabolism (C–F), (iii) antioxidants (G–I), (iv) essential amino acid (J), and (v) vitamins (K–L). Cohort mean and 95% confidence intervals are constructed from generalised linear models and presented as lines and shaded regions, respectively.

Chimera-free sequences were classified by BLASTn alignment against the SILVA database. Concatenated sequences of all sequences were mapped on the final set of zOTUs to calculate the abundance of each zOTU in any given sample. All non-bacterial sequences were removed based on their taxonomic classification. A phylogenetic tree was constructed using MUSCLE v3.8.31<sup>52</sup>.

Samples with <22,578 sequences per sample were removed, with this threshold being based on a trade-off between sequencing depth and the number of samples for paired analysis. Coverage per sample was estimated with a Michaelis-Menten Fit.

Functional profiles were inferred from 16S rRNA data using Tax4Fun<sup>53</sup>. Tax4Fun predicts the functional capabilities of the microbial communities by mapping the 16S rRNA gene sequences with the functional annotation of sequenced prokaryotic genomes in the Kyoto Encyclopedia of Genes and Genomes (KEGG).

**Inflammatory biomarkers.** Calprotectin was extracted and measured as described in a previous study<sup>12</sup> using the PhiCal kit (Calpro, San Diego, CA, US). The lower limit of detection for the assay was 19.5 mg/kg.



**Figure 5.** Boxplots of calprotectin ( $n = 19$  pairs) and M2-PK ( $n = 18$  pairs) in CF and HC cohorts (A,B). Scatterplots of calprotectin and M2-PK against age in CF and HC cohorts (C,D). Cohort mean and 95% confidence intervals are constructed from generalised linear models and presented as lines and shaded regions, respectively (C,D).

Calprotectin  $> 50$  mg/kg is considered elevated. Faecal M2-PK was extracted and measured as described in a previous study<sup>14</sup> using the ScheBo Tumour M2-PK kit (ScheBo Biotech, Giessen, Germany). The lower limit of detection of the assay was 1 U/mL.

**Stool metabolomics.** A subset of samples underwent untargeted ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) as previously reported, with the primary aim of exploring inflammatory pathways<sup>22</sup>. Metabolites were identified using the human metabolome database (HMDB)<sup>23</sup> and the normalised abundances of these identified metabolites were uploaded into Perseus v1.6.6.0. The normalised abundances were then  $\log(2)$  transformed. Metabolite rows were filtered for at least 70% valid values in at least one group. Missing values were replaced from a normal distribution (total matrix). Proteins with a significantly different abundance between CF and HC cohorts were determined using a two sample Student's T-test (permutation-based FDR  $< 0.05$ ). The normalised abundances were then Z-score normalised which were used for downstream analysis. A post hoc analysis of metabolites related to propanoate and butanoate metabolism (Supp. Dataset 7)<sup>23</sup> was also performed.

**Sample size.** The minimum required sample size for this study was calculated using GLIMPSE 2.0.0 (<https://glimpse.samplesizeshop.org/#/>) and based on the age-adjusted mean (95% CI) Shannon index reported in our initial gut microbiota study for CF and HC cohorts (2.75 (2.51–2.98) and 3.90 (3.72–4.08), respectively)<sup>6</sup>. A minimum sample size of 10 subjects was required to reject the null hypothesis that the population means of the CF and HC groups are equal with probability (power) 0.8 and Type I error of 0.05.

**Statistical analysis.** Statistical analysis was performed in R v3.4.4. Alpha and phylogeny-based beta diversity indices were calculated with a dataset random subsampled to 22,578 sequences per sample. Pairwise weighted and unweighted UniFrac distances between samples<sup>54</sup> were calculated using the GUniFrac package<sup>55</sup> and used to generate non-metric multidimensional scaling (NMDS) plots. Continuous variables were compared using a paired t-test or Wilcoxon signed-rank test for parametric and non-parametric data, respectively ( $p < 0.05$  considered statistically significant). Generalised linear models (glm function; using a gaussian distribution) were constructed to control for age when comparing continuous variables between pairs. Permutational multivariate analysis of variance (PERMANOVA) tests (permutations = 1000) were utilised to test if beta diversity significantly differed between groups ('CF vs HC' and 'male vs female') and for age using the vegan function *adonis*<sup>56</sup>. Graphs were generated using ggplot2 in R<sup>57</sup>. A significant difference in abundance of taxa or zOTU between CF and HC groups was assessed using the ANCOM package v1.1-3 (Benjamini & Hochberg correction for multiple comparisons,  $q < 0.05$ )<sup>58</sup>. A significant difference in abundance of functional pathways was assessed with multiple Wilcoxon signed-rank tests corrected for multiple testing using a Dunn-Sidak correction ( $n = 275$ ,  $p < 0.00019$ ). Correlations between continuous variables were assessed using Pearson or Spearman correlations (Benjamini & Hochberg correction,  $q < 0.05$ )<sup>59</sup>.

**Patient consent.** Written informed consent was obtained from each subject or caregiver(s) and the study was carried out in accordance with the approved guidelines.

**Ethics approval.** This study was approved by the South Eastern Sydney Area Health Service, Human Research Ethics Committee, Sydney Australia (HREC ref no: 10/240).

**Data sharing statement.** The datasets generated and analysed during the current study are available upon reasonable request to the corresponding author.

Received: 23 November 2018; Accepted: 22 November 2019;

Published online: 09 December 2019

## References

- Imrie, J. R., Fagan, D. G. & Sturgess, J. M. Quantitative evaluation of the development of the exocrine pancreas in cystic fibrosis and control infants. *The American Journal of Pathology* **95**, 697–707 (1979).
- Couper, R. T., Corey, M., Durie, P. R., Forstner, G. G. & Moore, D. J. Longitudinal evaluation of serum trypsinogen measurement in pancreatic-insufficient and pancreatic-sufficient patients with cystic fibrosis. *The Journal of pediatrics* **127**, 408–413 (1995).
- Ooi, C. Y. & Durie, P. R. Cystic fibrosis from the gastroenterologist's perspective. *Nature reviews. Gastroenterology & hepatology* **13**, 175–185, <https://doi.org/10.1038/nrgastro.2015.226> (2016).
- Ooi, C. Y. *et al.* Fecal Human beta-Defensin 2 in Children with Cystic Fibrosis: Is There a Diminished Intestinal Innate Immune Response? *Digestive diseases and sciences* **60**, 2946–2952, <https://doi.org/10.1007/s10620-015-3842-2> (2015).
- Sutherland, R. *et al.* Dietary intake of energy-dense, nutrient-poor and nutrient-dense food sources in children with cystic fibrosis. *Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society*. <https://doi.org/10.1016/j.jcf.2018.03.011> (2018).
- Nielsen, S. *et al.* Disrupted progression of the intestinal microbiota with age in children with cystic fibrosis. *Scientific reports* **6**, 24857, <https://doi.org/10.1038/srep24857> (2016).
- Debyser, G. *et al.* Faecal proteomics: A tool to investigate dysbiosis and inflammation in patients with cystic fibrosis. *Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society* **15**, 242–250, <https://doi.org/10.1016/j.jcf.2015.08.003> (2016).
- Manor, O. *et al.* Metagenomic evidence for taxonomic dysbiosis and functional imbalance in the gastrointestinal tracts of children with cystic fibrosis. *Scientific reports* **6**, 22493, <https://doi.org/10.1038/srep22493> (2016).
- Vernocchi, P. *et al.* Gut microbiota signatures in cystic fibrosis: Loss of host CFTR function drives the microbiota enterophenotype. *PLoS one* **13**, e0208171, <https://doi.org/10.1371/journal.pone.0208171> (2018).
- Corey, M., McLaughlin, F. J., Williams, M. & Levison, H. A comparison of survival, growth, and pulmonary function in patients with cystic fibrosis in Boston and Toronto. *Journal of clinical epidemiology* **41**, 583–591 (1988).
- Dhaliwal, J. *et al.* Intestinal inflammation and impact on growth in children with cystic fibrosis. *Journal of pediatric gastroenterology and nutrition* **60**, 521–526, <https://doi.org/10.1097/MPG.0b013e3181d1b013> (2015).
- Garg, M. *et al.* Age-dependent variation of fecal calprotectin in cystic fibrosis and healthy children. *Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society* **16**, 631–636, <https://doi.org/10.1016/j.jcf.2017.03.010> (2017).
- Werlin, S. L. *et al.* Evidence of intestinal inflammation in patients with cystic fibrosis. *Journal of pediatric gastroenterology and nutrition* **51**, 304–308, <https://doi.org/10.1097/MPG.0b013e3181d1b013> (2010).
- Garg, M. *et al.* Age-related levels of fecal M2-pyruvate kinase in children with cystic fibrosis and healthy children 0 to 10 years old. *Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society* **17**, 109–113, <https://doi.org/10.1016/j.jcf.2017.07.011> (2018).
- Bruzzese, E. *et al.* Intestinal inflammation is a frequent feature of cystic fibrosis and is reduced by probiotic administration. *Alimentary pharmacology & therapeutics* **20**, 813–819, <https://doi.org/10.1111/j.1365-2036.2004.02174.x> (2004).
- Pang, T. *et al.* Elevated fecal M2-pyruvate kinase in children with cystic fibrosis: a clue to the increased risk of intestinal malignancy in adulthood? *Journal of gastroenterology and hepatology* **30**, 866–871, <https://doi.org/10.1111/jgh.12842> (2015).
- De Lisle, R. C. Decreased Expression of Enterocyte Nutrient Assimilation Genes and Proteins in the Small Intestine of Cystic Fibrosis Mouse. *Journal of pediatric gastroenterology and nutrition* **62**, 627–634, <https://doi.org/10.1097/MPG.0000000000001030> (2016).
- Visca, A., Bishop, C. T., Hilton, S. & Hudson, V. M. Oral reduced L-glutathione improves growth in pediatric cystic fibrosis patients. *Journal of pediatric gastroenterology and nutrition* **60**, 802–810, <https://doi.org/10.1097/MPG.0000000000000738> (2015).
- Yamada, A. *et al.* Risk of gastrointestinal cancers in patients with cystic fibrosis: a systematic review and meta-analysis. *The Lancet. Oncology* **19**, 758–767, [https://doi.org/10.1016/s1470-2045\(18\)30188-8](https://doi.org/10.1016/s1470-2045(18)30188-8) (2018).
- Garg, M. & Ooi, C. Y. The Enigmatic Gut in Cystic Fibrosis: Linking Inflammation, Dysbiosis, and the Increased Risk of Malignancy. *Curr Gastroenterol Rep* **19**, 6, <https://doi.org/10.1007/s11894-017-0546-0> (2017).
- Slæ, M. & Wilschanski, M. Cystic fibrosis: a gastrointestinal cancer syndrome. *The Lancet. Oncology* **19**, 719–720, [https://doi.org/10.1016/s1470-2045\(18\)30250-x](https://doi.org/10.1016/s1470-2045(18)30250-x) (2018).
- Kaakoush, N. O., Pickford, R., Jaffe, A. & Ooi, C. Y. Is there a role for stool metabolomics in cystic fibrosis? *Pediatrics international: official journal of the Japan Pediatric Society* **58**, 808–811, <https://doi.org/10.1111/ped.13063> (2016).
- Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic acids research* **46**, D608–d617, <https://doi.org/10.1093/nar/gkx1089> (2018).
- Albenberg, L. & Kelsen, J. Advances in Gut Microbiome Research and Relevance to Pediatric Diseases. *The Journal of pediatrics* **178**, 16–23, <https://doi.org/10.1016/j.jpeds.2016.08.044> (2016).
- Hoen, A. G. *et al.* Associations between Gut Microbial Colonization in Early Life and Respiratory Outcomes in Cystic Fibrosis. *Journal of Pediatrics* **167**(138–147), e131–133 (2015).
- Hoffman, L. R. *et al.* Escherichia coli dysbiosis correlates with gastrointestinal dysfunction in children with cystic fibrosis. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **58**, 396–399, <https://doi.org/10.1093/cid/cit715> (2014).
- Madan, J. C. *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* **3** (2012).
- Johns, A. T. The mechanism of propionic acid formation by Veillonella gazogenes. *Journal of general microbiology* **5**, 326–336, <https://doi.org/10.1099/00221287-5-2-326> (1951).
- Zemanick, E. T. *et al.* Airway microbiota across age and disease spectrum in cystic fibrosis. *The European respiratory journal* **50**, <https://doi.org/10.1183/13993003.00832-2017> (2017).
- Hussan, H., Clinton, S. K., Roberts, K. & Bailey, M. T. Fusobacterium's link to colorectal neoplasia sequenced: A systematic review and future insights. *World journal of gastroenterology* **23**, 8626–8650, <https://doi.org/10.3748/wjg.v23.i48.8626> (2017).
- Hadjiiladis, D. *et al.* Cystic Fibrosis Colorectal Cancer Screening Consensus Recommendations. *Gastroenterology* **154**, 736–745, <https://doi.org/10.1053/j.gastro.2017.12.012> (2018).
- Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J* **6**, 1535–1543, <https://doi.org/10.1038/ismej.2012.4> (2012).
- Song, Y. *et al.* Alistipes onderdonkii sp. nov. and Alistipes shahii sp. nov., of human origin. *International journal of systematic and evolutionary microbiology* **56**, 1985–1990, <https://doi.org/10.1099/ijs.0.64318-0> (2006).

34. De Vadder, F. *et al.* Microbiota-Produced Succinate Improves Glucose Homeostasis via Intestinal Gluconeogenesis. *Cell metabolism* **24**, 151–157, <https://doi.org/10.1016/j.cmet.2016.06.013> (2016).
35. Prentice, B. *et al.* Diagnosing cystic fibrosis-related diabetes: current methods and challenges. *Expert review of respiratory medicine* **10**, 799–811, <https://doi.org/10.1080/17476348.2016.1190646> (2016).
36. Gough, E. K. *et al.* Linear growth faltering in infants is associated with *Acidaminococcus* sp. and community-level changes in the gut microbiota. *Microbiome* **3**, 24, <https://doi.org/10.1186/s40168-015-0089-2> (2015).
37. den Besten, G. *et al.* The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of lipid research* **54**, 2325–2340, <https://doi.org/10.1194/jlr.R036012> (2013).
38. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446–450, <https://doi.org/10.1038/nature12721> (2013).
39. Bao, X., Feng, Z., Yao, J., Li, T. & Yin, Y. Roles of Dietary Amino Acids and Their Metabolites in Pathogenesis of Inflammatory Bowel Disease. *Mediators of inflammation* **2017**, 6869259, <https://doi.org/10.1155/2017/6869259> (2017).
40. Ciofu, O. & Lykkesfeldt, J. Antioxidant supplementation for lung disease in cystic fibrosis. The Cochrane database of systematic reviews, Cd007020, <https://doi.org/10.1002/14651858.CD007020.pub3> (2014).
41. Tan, B. *et al.* 408 Aromatic amino acids alleviate intestinal inflammation in piglets through calcium-sensing receptor activation. *Journal of Animal Science* **95**, 201–202, <https://doi.org/10.2527/asasann.2017.408> (2017).
42. Li, P., Yin, Y. L., Li, D., Kim, S. W. & Wu, G. Amino acids and immune function. *The British journal of nutrition* **98**, 237–252, <https://doi.org/10.1017/S000711450769936x> (2007).
43. Saxby, N. *et al.* Nutrition Guidelines for Cystic Fibrosis in Australia and New Zealand, ed. Scott C. Bell, Thoracic Society of Australia and New Zealand. (2017).
44. Farrell, P. M. *et al.* Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report. *The Journal of pediatrics* **153**, S4–S14, <https://doi.org/10.1016/j.jpeds.2008.05.005> (2008).
45. Jeejeebhoy, K. N., Ahmad, S. & Kozak, G. Determination of fecal fats containing both medium and long chain triglycerides and fatty acids. *Clinical biochemistry* **3**, 157–163 (1970).
46. Loser, C., Mollgaard, A. & Folsch, U. R. Faecal elastase 1: a novel, highly sensitive, and specific tubeless pancreatic function test. *Gut* **39**, 580–586 (1996).
47. Fuchs, H. J. *et al.* Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. *The New England journal of medicine* **331**, 637–642, <https://doi.org/10.1056/nejm199409083311003> (1994).
48. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
49. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)* **26**, 2460–2461, <https://doi.org/10.1093/bioinformatics/btq461> (2010).
50. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv*, <https://doi.org/10.1101/081257> (2016).
51. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590–D596, <https://doi.org/10.1093/nar/gks1219> (2013).
52. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797, <https://doi.org/10.1093/nar/gkh340> (2004).
53. Aßhauer, K. P., Wemheuer, B., Daniel, R. & Meinicke, P. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics (Oxford, England)* **31**, 2882–2884, <https://doi.org/10.1093/bioinformatics/btv287> (2015).
54. Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and environmental microbiology* **73**, 1576–1585, <https://doi.org/10.1128/aem.01996-06> (2007).
55. Chen, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics (Oxford, England)* **28**, 2106–2113, <https://doi.org/10.1093/bioinformatics/bts342> (2012).
56. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**, 32–46, <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x> (2001).
57. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer Publishing Company, Incorporated, 2009).
58. Mandal, S. *et al.* Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease* **26**, 27663, <https://doi.org/10.3402/mehd.v26.27663> (2015).
59. Torondel, B. *et al.* Assessment of the influence of intrinsic environmental and geographical factors on the bacterial ecology of pit latrines. *Microbial biotechnology* **9**, 209–223, <https://doi.org/10.1111/1751-7915.12334> (2016).

## Acknowledgements

The authors thank the parents and families of our children with CF and our healthy participants. We thank the CF Clinic at Sydney Children's Hospital Randwick (Dr Yvonne Belessis, Dr Penny Field, Dr John Widger, the late Dr John Morton, Rhonda Bell, Rebecca McDonald and Amanda Thomsen). We also thank Lily Nahidi (Sydney Children's Hospital Clinical Research Centre), Roxanne Strachan, Jung M. Lee, Tamara Pang and Joanna Michalowski for their assistance in recruitment and sample collection. This work was supported by the Sydney Children's Hospital Foundation, the Australian Government Department of Education and Training (Australian Postgraduate Award), the Cystic Fibrosis Australia Research Trust Grant and the Royal Australasian College of Physicians Research and Education Foundation Award (Servier Staff Research Fellowship).

## Author contributions

Conception and design of the study: M.J.C., B.W., S.N., T.T. and C.Y.O.; data and sample collection: B.N. and M.G.; sample processing: S.N., N.O.K., M.G. and R.P.; sequencing processing: M.J.C. and B.W.; data analysis: M.J.C. and B.W.; statistical analysis and interpretation: M.J.C., B.W., N.O.K., T.T., A.J. and C.Y.O.; writing and critical revision of the manuscript: M.J.C., S.N., B.W., N.O.K., M.G., B.N., R.P., A.J., T.T. and C.Y.O. All authors approved the final version for publication.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-55028-7>.

**Correspondence** and requests for materials should be addressed to C.Y.O.

**Reprints and permissions information** is available at [www.nature.com/reprints](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) 2019