



# Changes in the Composition of the Gut Microbiota and the Blood Transcriptome in Preterm Infants at Less than 29 Weeks Gestation Diagnosed with Bronchopulmonary Dysplasia

✉ Feargal J. Ryan,<sup>a</sup> Damian P. Drew,<sup>a,b</sup> ✉ Chloe Douglas,<sup>a,c,d</sup> Lex E. X. Leong,<sup>a,b</sup> Max Moldovan,<sup>a</sup> Miriam Lynn,<sup>a</sup> Naomi Fink,<sup>a,c,d</sup> Anastasia Sribnaia,<sup>a</sup> Irmeli Penttila,<sup>c,d</sup> Andrew J. McPhee,<sup>c,f</sup> Carmel T. Collins,<sup>c,d</sup> Maria Makrides,<sup>c,d</sup> Robert A. Gibson,<sup>c,e</sup> Geraint B. Rogers,<sup>a,b</sup> David J. Lynn<sup>a,b</sup>

<sup>a</sup>Precision Medicine Theme, South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia

<sup>b</sup>College of Medicine and Public Health, Flinders University, Bedford Park, South Australia, Australia

<sup>c</sup>SAHMRI Women and Kids, South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia

<sup>d</sup>Adelaide Medical School, The University of Adelaide, Adelaide, South Australia, Australia

<sup>e</sup>School of Agriculture, Food, and Wine, The University of Adelaide, Adelaide, South Australia, Australia

<sup>f</sup>Neonatal Medicine, Women's and Children's Hospital, North Adelaide, South Australia, Australia

**ABSTRACT** Bronchopulmonary dysplasia (BPD) is a common chronic lung condition in preterm infants that results in abnormal lung development and leads to considerable morbidity and mortality, making BPD one of the most common complications of preterm birth. We employed RNA sequencing and 16S rRNA gene sequencing to profile gene expression in blood and the composition of the fecal microbiota in infants born at <29 weeks gestational age and diagnosed with BPD in comparison to those of preterm infants that were not diagnosed with BPD. 16S rRNA gene sequencing, performed longitudinally on 255 fecal samples collected from 50 infants in the first months of life, identified significant differences in the relative levels of abundance of *Klebsiella*, *Salmonella*, *Escherichia/Shigella*, and *Bifidobacterium* in the BPD infants in a manner that was birth mode dependent. Transcriptome sequencing (RNA-Seq) analysis revealed that more than 400 genes were upregulated in infants with BPD. Genes upregulated in BPD infants were significantly enriched for functions related to red blood cell development and oxygen transport, while several immune-related pathways were downregulated. We also identified a gene expression signature consistent with an enrichment of immunosuppressive CD71<sup>+</sup> early erythroid cells in infants with BPD. Intriguingly, genes that were correlated in their expression with the relative abundances of specific taxa in the microbiota were significantly enriched for roles in the immune system, suggesting that changes in the microbiota might influence immune gene expression systemically.

**IMPORTANCE** Bronchopulmonary dysplasia (BPD) is a serious inflammatory condition of the lung and is the most common complication associated with preterm birth. A large body of evidence now suggests that the gut microbiota can influence immunity and inflammation systemically; however, the role of the gut microbiota in BPD has not been evaluated to date. Here, we report that there are significant differences in the gut microbiota of infants born at <29 weeks gestation and subsequently diagnosed with BPD, which are particularly pronounced when infants are stratified by birth mode. We also show that erythroid and immune gene expression levels are significantly altered in BPD infants. Interestingly, we identified an association between the composition of the microbiota and immune gene expression in blood in early life. Together, these findings suggest that the composition of the microbiota may influence the risk of developing BPD and, more generally, may shape systemic immune gene expression.

**Citation** Ryan FJ, Drew DP, Douglas C, Leong LEX, Moldovan M, Lynn M, Fink N, Sribnaia A, Penttila I, McPhee AJ, Collins CT, Makrides M, Gibson RA, Rogers GB, Lynn DJ. 2019. Changes in the composition of the gut microbiota and the blood transcriptome in preterm infants at less than 29 weeks gestation diagnosed with bronchopulmonary dysplasia. *mSystems* 4:e00484-19. <https://doi.org/10.1128/mSystems.00484-19>.

**Editor** Thomas J. Sharpton, Oregon State University

**Copyright** © 2019 Ryan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to David J. Lynn, [David.Lynn@sahmri.com](mailto:David.Lynn@sahmri.com).

**Received** 8 August 2019

**Accepted** 9 October 2019

**Published** 29 October 2019

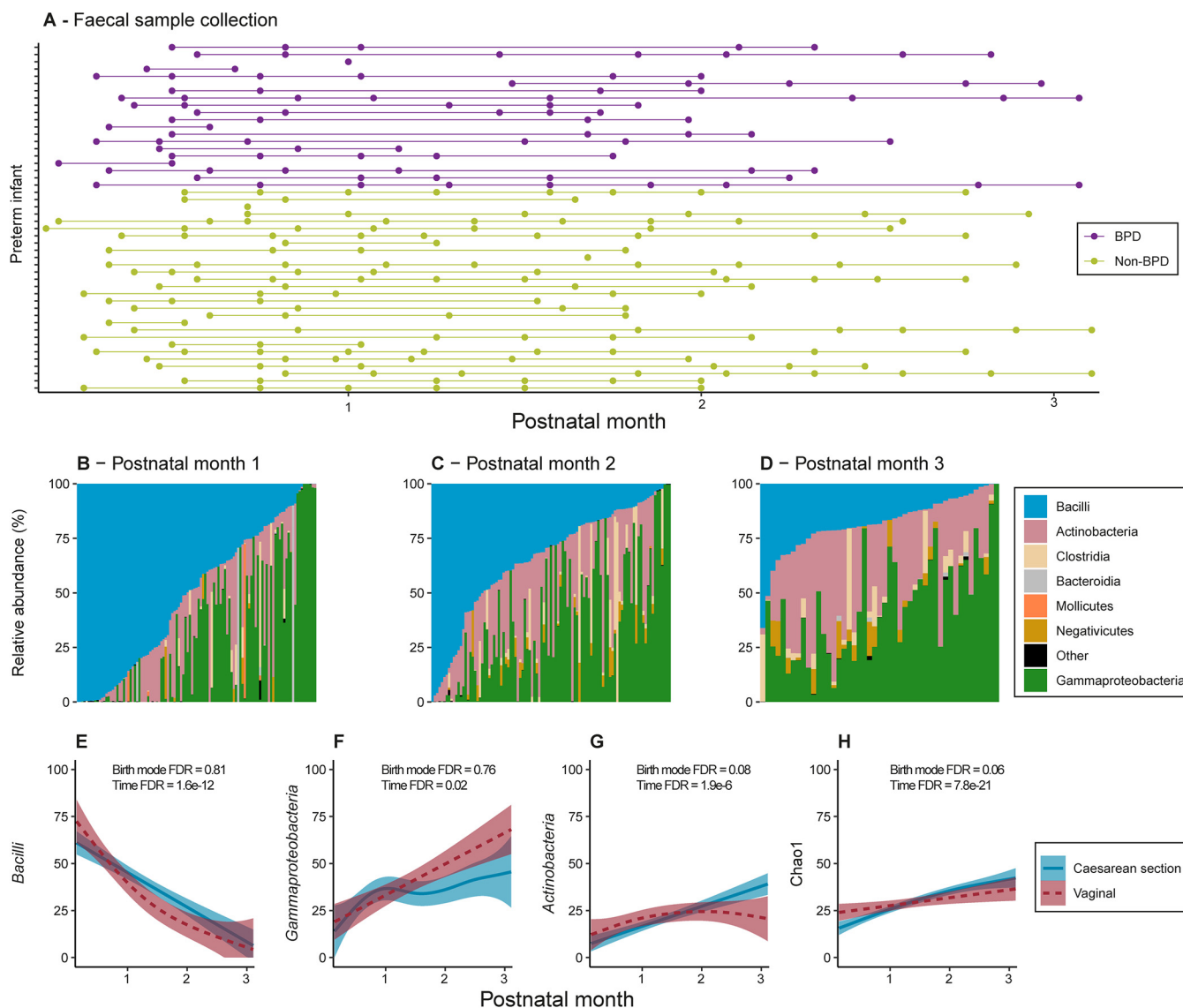
**KEYWORDS** BPD, RNA-Seq, VLBW, fecal organisms, microbiota, neonates

**B**ronchopulmonary dysplasia (BPD) is a chronic lung condition affecting approximately two-thirds of extremely preterm (born at <28 weeks gestational age) infants (1–3). BPD results in abnormal lung development and considerable morbidity, with complications continuing into adulthood (1). BPD was first described in 1967, but its clinical definition has since undergone several refinements. BPD as described now reflects an interruption to lung development (1, 4). BPD diagnosis and assessment of severity are based on the requirement for supplemental oxygen at 36 weeks postmenstrual age (gestational age plus postnatal age) (3).

The pathogenesis of BPD is recognized to be multifactorial, with potential impacts of both pre- and postnatal conditions on its development and severity (3, 5). Currently, gestational age and birth weight remain the best described risk factors for BPD (3, 6), and therefore, there is significant clinical interest in identifying other factors that can more accurately predict which infants are at risk of developing BPD. Several studies have identified a large contribution of genetic factors to BPD susceptibility, with 53 to 82% of the variance explained by genetics in twin cohort studies in North America (7, 8). Genome-wide association studies (GWAS) of BPD, however, have produced inconclusive results, with one study finding an association with the SPOCK2 gene and another with the C-reactive protein (CRP) gene but others finding no significant associations (9–12). The association of rare variants with BPD has been evaluated through multiple exome sequencing studies, which have suggested a role for kinase A- and mitogen-activated protein (MAP) kinase-related pathways in BPD (13–15). To date, microarray-based gene expression studies of BPD have been performed on both blood and lung tissue samples collected from infants and from animal models and have identified several thousand genes as potentially being differentially expressed in BPD (16–19). These studies reported that pathways involved in the inflammatory response are downregulated in BPD infants, while pathways related to the cell cycle are upregulated (18). These signatures were, however, dependent on the timing of sample collection for gene expression analysis.

The gut microbiota in very low birth-weight (VLBW; born weighing <1,500 g) preterm infants has been shown to follow a markedly different pattern of colonization than that of healthy term-born infants (20–22), leading us to hypothesize that the gut microbiota could influence BPD susceptibility and/or severity. Consistent with this hypothesis, prolonged antibiotic use in VLBW preterm infants is associated with an increased risk of developing BPD (23, 24). Recent studies have identified differences in the lung microbiota in infants who develop BPD (25); however, to date, only one small ( $n = 13$ ) PCR-based study has investigated whether the composition of the gut microbiota was altered in infants with BPD (26). Increasing evidence from preclinical and clinical studies strongly suggests that the gut microbiota plays a key role in healthy immune development, particularly in early life (27, 28). Importantly, several studies have demonstrated the potential of the gut microbiota to influence immune responses in the lung (28, 29), suggesting that the gut microbiota could influence the severity of BPD by modulating inflammatory responses systemically and in the lung.

To investigate whether changes in the gut microbiota are associated with BPD, we used 16S rRNA gene sequencing to perform longitudinal profiling of the microbiota in >250 fecal samples collected from a cohort of 50 infants born at <29 weeks gestation. We identified significant differences in the relative abundances (RAs) of several taxa in the fecal microbiota of vaginally born infants that were subsequently diagnosed with BPD. Interestingly, these differences were not evident in infants born by cesarean section. Blood samples were also collected from a subset of the infants at baseline and at approximately the time that BPD diagnosis was made. Gene expression in venous blood collected from BPD infants was not significantly different at baseline compared to the gene expression in venous blood from non-BPD infants, but >400 genes were significantly differentially expressed at the time of BPD diagnosis. We identified a gene



**FIG 1** 16S rRNA gene sequencing was used to longitudinally profile the composition of the fecal microbiota in 50 preterm infants subsequently diagnosed with or without BPD. (A) Filled circles represent time points at which fecal samples were collected from each infant (mean of 5 per infant, total of 255). (B to D) The relative abundances of the major classes of bacteria identified in the preterm infants in this study. (E to G) The mean relative abundances (line graph) of the top three most abundant bacterial classes over time in preterm infants that were born vaginally or by cesarean section. Shaded areas represent the 95% confidence intervals. (H) Chao1 alpha diversity. A generalized linear mixed-effects model was used to assess changes in relative abundance or alpha diversity over time (Time FDR) or by birth mode (Birth mode FDR).

expression signature in BPD infants that suggests an enrichment of immunosuppressive CD71<sup>+</sup> early erythroid cells. Finally, we also uncovered a potential association between the composition of the microbiota and peripheral blood immune gene expression.

**RESULTS**

Two hundred fifty-five fecal samples (mean of 5 per infant) were collected from 50 preterm infants (mean gestational age, 26.4 ± 1.7 weeks [mean ± standard deviation]; mean birth weight, 869.9 ± 240.1 g), recruited from the Women’s and Children’s Hospital, Adelaide, Australia. Samples were collected longitudinally from recruitment (shortly after birth) to discharge home or 40 weeks postmenstrual age, whichever occurred first (Fig. 1A). These infants were recruited as a subset of infants enrolled in the larger, multisite n-3 Fatty Acids for Improvement in Respiratory Outcomes (N3RO)

**TABLE 1** Infant clinical data according to BPD diagnosis

Characteristic	BPD ( <i>n</i> = 20)	Non-BPD ( <i>n</i> = 28)	Unadjusted <i>P</i> value
Mean gestational age ± SD (wks)	25.7 ± 1.8	26.8 ± 1.4	0.018 <sup>a</sup>
Mean birth wt ± SD (g)	746.8 ± 187.4	967.9 ± 233.5	0.00095 <sup>a</sup>
Female (%)	45	46.4	1 <sup>b</sup>
Vaginally born (%)	40	28.5	0.537 <sup>b</sup>
Sepsis (% yes)	35	17.9	0.19 <sup>b</sup>
Necrotizing enterocolitis (% yes)	10	3.6	0.57 <sup>b</sup>
Mean no. of fecal samples per subject ± SD	5.0 ± 2.1	5.9 ± 2.4	0.29 <sup>a</sup>
Postnatal steroid treatment (% yes)	60	10.7	0.0004 <sup>b</sup>

<sup>a</sup>*P* value was generated by Wilcoxon test for continuous variables.

<sup>b</sup>*P* value was generated by Fisher's exact test for categorical variables.

clinical trial, which investigated the impact of enteral docosahexaenoic acid (DHA) supplementation on the risk of developing BPD (30). BPD was defined as the requirement for supplemental oxygen and/or respiratory support at 36 weeks' postmenstrual age or discharge home, whichever occurred first, using modified criteria (see Materials and Methods) from Walsh et al. (31). A subset of infants in this study also had venous blood collected for gene expression profiling via RNA-Seq at recruitment/baseline (*n* = 10; mean day of life, day 4.6) and again at the time of BPD diagnosis (*n* = 21; mean day of life, day 64.4) (Fig. 1A; see Table S1 in the supplemental material).

Of the 50 infants recruited, hereinafter referred to as BPD/non-BPD infants, 20 were subsequently diagnosed with BPD at 36 weeks postmenstrual age. One infant died prior to 36 weeks postmenstrual age and did not receive a BPD diagnosis. Another infant could not be diagnosed with/without BPD as the physiological challenge was not performed (although the infant was receiving respiratory support at 36 weeks postmenstrual age). Fecal microbiota data from these two infants were excluded when testing for associations with BPD diagnosis. Ten of the 20 BPD infants and 15/28 non-BPD infants received DHA supplementation. As is common for this population, all but two received antibiotics over the course of the study, both of whom were non-BPD cesarean section-born infants. Similarly, all but two infants (one BPD and one non-BPD, both cesarean section born) received the probiotic Infloran (Berna, Switzerland), a mixture of *Bifidobacterium* and *Lactobacillus* species. Probiotics are commonly administered to preterm infants in Australia to reduce necrotizing enterocolitis (32, 33). There was no significant difference in the number of days of probiotic supplementation between the BPD diagnosis groups (data not shown). Lower gestational age is a well-documented risk factor for BPD. While all infants in this cohort were born very premature, BPD infants were on average 1.2 weeks more preterm than non-BPD infants (mean of 25.6 versus 26.8 weeks, *P* = 0.017). Further cohort demographics are shown in Table 1.

**Composition of the fecal microbiota in VLBW preterm infants.** The fecal microbiota of the preterm infants in this study were assessed by 16S rRNA gene sequencing. Samples were sequenced to an average of 31,464 (2 × 300 bp) reads on an Illumina MiSeq (interquartile range [IQR], 11,779). The DADA2 algorithm (34), as implemented in QIIME2 (35), was used to denoise sequences, resulting in a total of 676 unique ribosomal sequence variants that were detected in at least 1 sample (hereinafter referred to as exact sequence variants [ESVs]). Six samples that had <1,000 denoised reads were excluded from further analysis. Multiple fecal samples were collected per infant, allowing a longitudinal assessment of changes in the composition of the microbiota during the first months of life (Fig. 1B to D; Fig. S1).

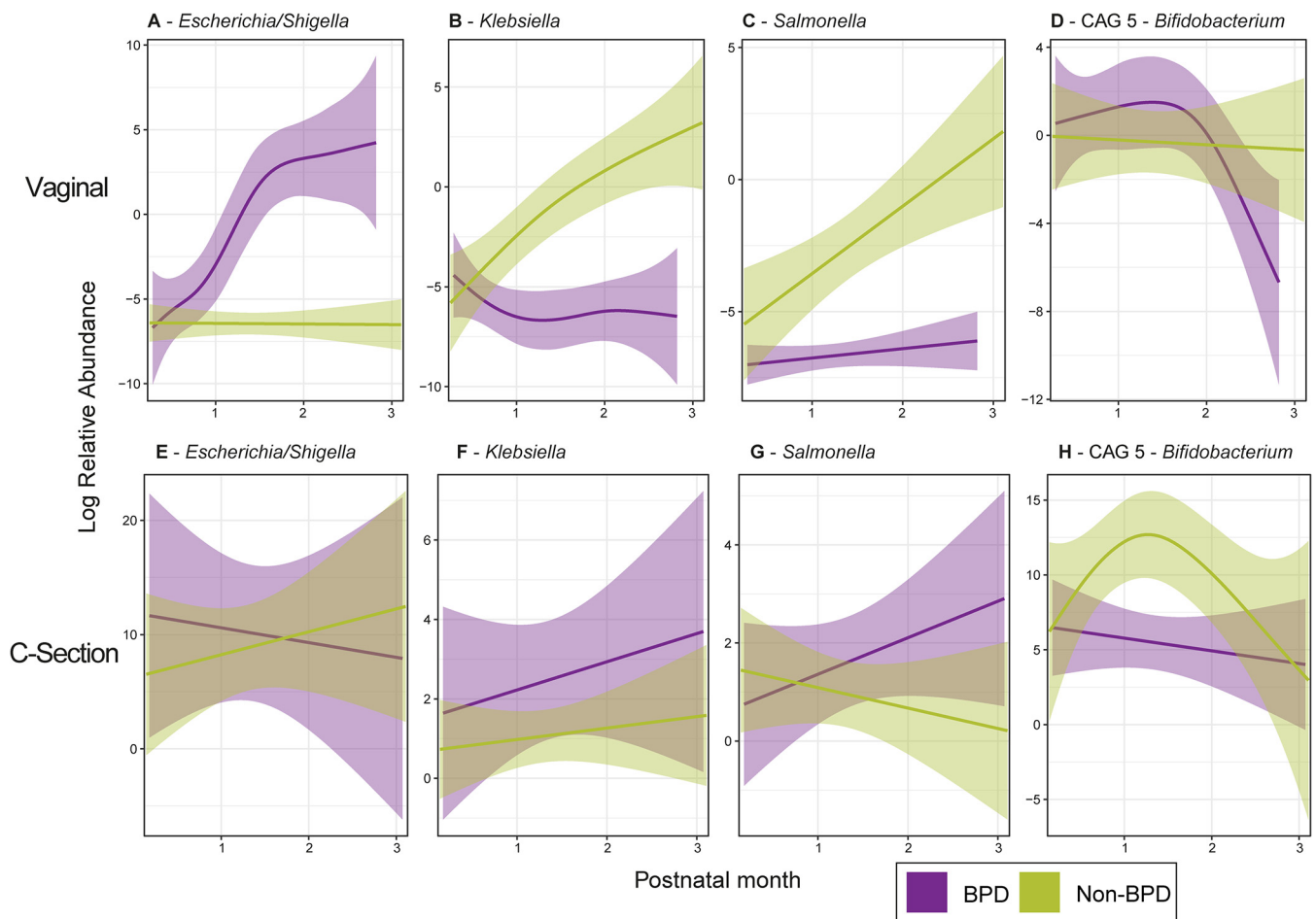
In the first month of life, three bacterial classes dominated the fecal microbiota: *Bacilli* (mean relative abundance [RA], 38.14%; IQR, 43.94%), *Gammaproteobacteria* (mean RA, 35.19%; IQR, 57.42%), and *Actinobacteria* (mean RA, 20.23%; IQR, 32.95%). After the first month of life, the *Bacilli* RA decreased sharply, while the RAs of both the *Gammaproteobacteria* and *Actinobacteria* increased (Fig. 1E to G). *Actinobacteria* were observed to be highly abundant, which may be a result of the Infloran probiotic that

was administered to all but 2 of the infants. The increased RA of *Actinobacteria* observed in this study was consistent with previous reports demonstrating significant increases in the RAs of *Bifidobacterium* in the microbiota of infants receiving *Bifidobacterium*-based probiotics (36). *Clostridia* have previously been reported to be highly abundant in the VLBW preterm infant fecal microbiota, at levels comparable to the levels of *Gammaproteobacteria* and *Bacilli* (20, 37). In our study, only a small subset of infants had a high RA of *Clostridia*, while in most infants, they were either not detected or were present at a low RA (mean RA, 3.55%; IQR, 1.35%) (Fig. 1B to D; Fig. S1). Alpha diversity (Chao1 index) increased with age (Fig. 1H), as has been previously reported (37).

Due to the limitations of species-level taxonomic classification with the V4 region of the 16S rRNA gene (38), we examined subgenus-level variation by defining coabundant groups (CAGs) of ESVs based on their Spearman correlations (39). Correlation analysis, limited to ESVs present in at least 20 samples, indicated that ESV-level abundance could largely be attributed to 16 highly correlated CAGs of ESVs (Fig. S2A). While certain genera were spread over multiple CAGs (*Bifidobacterium*, *Streptococcus*, and *Lactobacillus*), the majority of CAGs were made up of ESVs which could be classified as either a single bacterial genus or a family. This is in contrast to the complex coabundance relationships commonly observed in the adult microbiota (40). The intra-CAG taxonomic homogeneity and high intra-CAG correlation observed may indicate that CAGs represent individual species (or groups of species not discernible with the taxonomic resolution of 16S rRNA gene sequencing). Sequences classified as part of the family *Enterobacteriaceae* were also spread across multiple CAGs that showed little to no correlation with each other (Fig. S2A). Examination of CAG RAs over time indicated that CAGs largely capture the same variation as shown at the genus level of classification (Fig. S2B), with the exception of those taxa mentioned above.

**Differences in the compositions of the fecal microbiota in preterm infants diagnosed with BPD.** To identify taxa in the fecal microbiota that were significantly altered in BPD infants over time, we implemented a generalized linear mixed-effects model (GLMM) using the lme4 R package (41). The longitudinal RAs of taxa at the class and genus levels, as well as CAG RAs, were analyzed for an association with BPD diagnosis while controlling for day of life, birth mode, DHA treatment, and subject. Three genera (*Escherichia/Shigella*, *Klebsiella*, and *Salmonella*) from the *Enterobacteriaceae* (class *Gammaproteobacteria*) were significantly associated (false discovery rate [FDR]  $\leq 0.01$ ) with BPD diagnosis (Fig. 2A, C, and E to G). Interestingly, these associations were only evident in vaginally born BPD infants. *Escherichia/Shigella* were significantly increased in vaginally born BPD infants but were not significantly altered in cesarean section-born infants, while *Klebsiella* and *Salmonella* were significantly less abundant in vaginally born (but not cesarean section-born) BPD infants. Additionally, the RA of one of the two *Bifidobacterium* CAGs (CAG 5) was significantly associated with BPD diagnosis in cesarean section-born infants (FDR = 0.0028) (Fig. 2D and H; Table S2). The GLMM approach did not identify any significant association between BPD diagnosis and any taxon at the higher taxonomic levels or with alpha diversity (Table S2).

As antibiotics are among the most frequently prescribed medications in neonatal intensive care units (42), we sought to establish whether a relationship existed between antibiotic treatment and the relative abundances of BPD-associated taxa. All but two infants in this cohort received antibiotics; however, vaginally born infants in this cohort had a significantly higher number of days of antibiotic exposure than those born by cesarean section, both in the first week of life and over the entire study period (Fig. S3A and C). There were no significant differences in numbers of days of antibiotic exposure by BPD diagnosis, either in the first week of life or over the entire study period (Fig. S3B and D). Importantly, there was no significant difference in numbers of days of antibiotic exposure in vaginally born BPD infants compared to the exposure of vaginally born non-BPD infants, suggesting that antibiotic exposure cannot explain the association between the composition of the fecal microbiota and BPD diagnosis observed in these infants. Furthermore, we examined the correlation between the numbers of days of



**FIG 2** Bacterial genera and CAGs that were identified as being significantly associated ( $FDR \leq 0.01$ ) with BPD diagnosis using a generalized linear mixed-effects model. The model was implemented using the lme4 package in R and incorporated day of life, birth mode, BPD diagnosis, DHA treatment, and subject (as a random effect). Three genera were found to be significantly associated with BPD, *Escherichia/Shigella* ( $FDR = 5.27e-08$ ), *Klebsiella* ( $FDR = 0.0059$ ), and *Salmonella* ( $FDR = 0.014$ ), as well as a single *Bifidobacterium* CAG ( $FDR = 0.0039$ ). (A to D) Log relative abundances of genera associated with BPD in fecal samples collected from BPD and non-BPD vaginally born infants. (E to H) Log relative abundances of genera associated with BPD in fecal samples collected from BPD and non-BPD cesarean section-born infants. Smoothed mean values (line graph) and 95% confidence intervals (shaded areas) are shown.

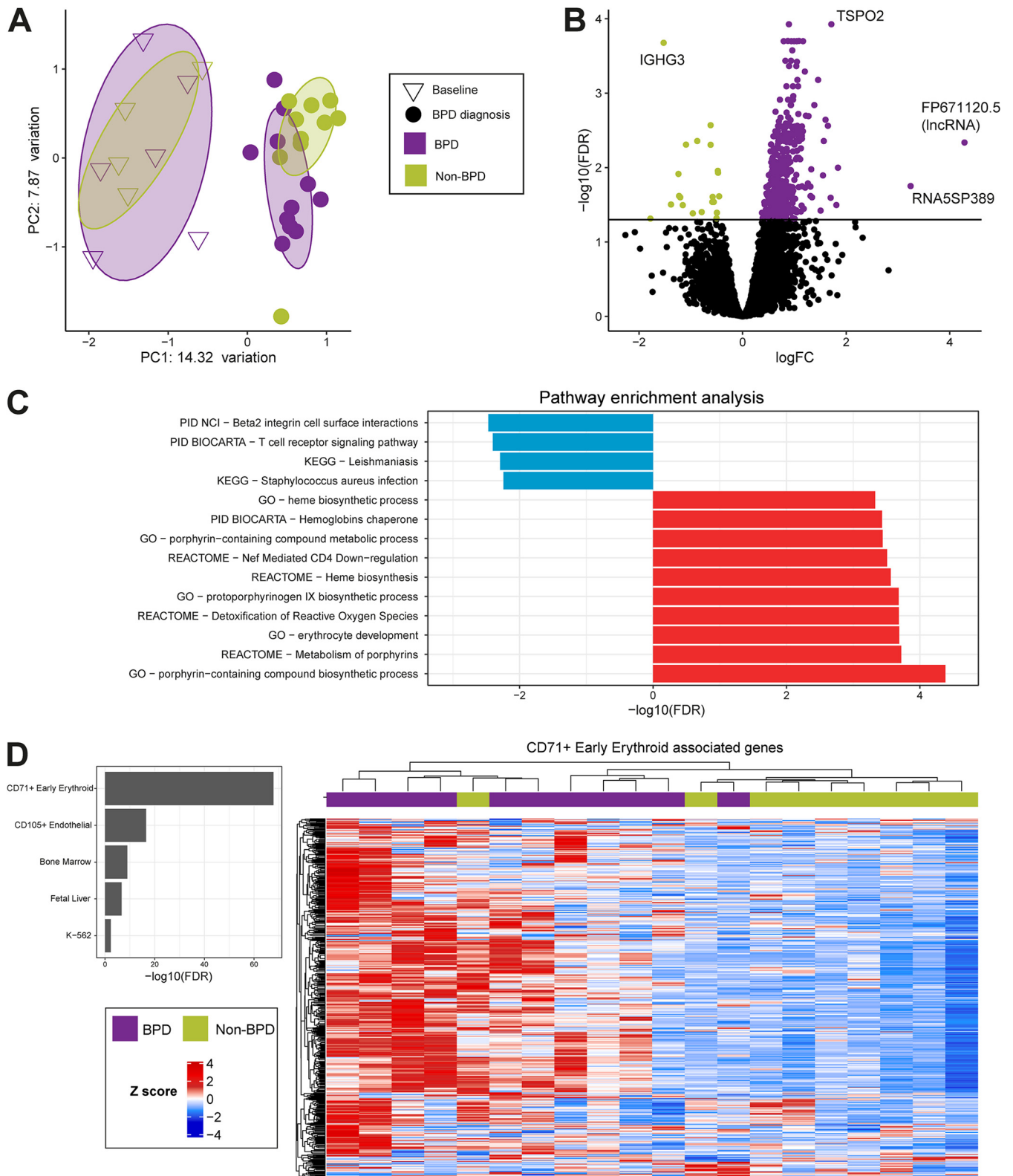
antibiotic exposure and the RAs over time (summarized by the area under the curve [AUC] for each taxon) of each genus and CAG that we found to be associated with BPD in the vaginally born infants. No significant correlation between the number of days of antibiotic exposure and the RA of any of the taxa associated with BPD was found (Fig. S3E to H).

**Impact of DHA supplementation on the fecal microbiota.** Infants in the N3RO study were randomized to receive a daily enteral soy-based placebo or supplementation with 60 mg of DHA per kg of body weight within 3 days of commencing enteral feeding, finishing at 36 weeks postmenstrual age (mean of 60 days of DHA supplementation). In the larger N3RO trial, supplementation with DHA did not decrease and may have increased the risk of developing BPD (30). Interestingly, using the GLMM approach described above, we found a significant association between DHA treatment and the RA of *Escherichia/Shigella* that was independent of BPD diagnosis ( $FDR = 4.32e-8$ ) (Fig. S4). These data suggest that the RA of *Escherichia/Shigella* is associated with both BPD diagnosis and DHA treatment. We detected no significant association between DHA treatment and alpha or beta diversity (Fig. S4C and D).

**Differential gene expression in preterm infants diagnosed with BPD.** Gene expression in peripheral venous blood samples collected from a subset of infants at baseline ( $n = 10$ ; mean day of life, day 4.6) and again at BPD diagnosis ( $n = 21$ ; mean

day of life, day 64.4), was profiled using RNA-Seq. More than 1.2 billion reads were sequenced (mean, 56 million single-end, 100-bp reads per sample; IQR, 11.3 million reads). Reads were aligned to the human genome (GRCh38) using HISAT2, with a mean alignment rate of 95.6% (Table S1). Infant gene expression profiles clustered primarily by age at time of blood collection, indicating that large-scale differences in blood gene expression occur between the first week of life and time of BPD diagnosis (Fig. 3A). One non-BPD infant was a notable outlier in gene expression compared to the other non-BPD infants (Fig. 3A). While this infant did not meet either the physiological or the clinical definition of BPD used in the N3RO clinical trial (30), this infant received 61 days of supplemental oxygen, which put this infant in the 92nd percentile of supplemental oxygen among all non-BPD infants in the study. We excluded this infant from further analysis to avoid confounding results. More than 2,000 genes were identified as differentially expressed at 36 weeks postmenstrual age compared to baseline (Table S3). No genes differentially expressed between BPD and non-BPD infants were identified in samples collected at baseline ( $FDR < 0.05$ ). In contrast, 431 genes were identified as being significantly differentially expressed in BPD infants compared to their expression levels in non-BPD infants at time of diagnosis ( $FDR < 0.05$ ) (Fig. 3B; Table S3). Pathway and Gene Ontology (GO) analyses revealed that BPD infants had increased expression of genes involved in red blood cell development and oxygen transport, whereas immune-related pathways were downregulated (Fig. 3C; Table S3). Postnatal steroids may be administered to preterm infants as a treatment for chronic lung disease (43), and as expected, BPD infants received significantly more steroids than non-BPD infants (Table 1). While increased steroid administration could contribute to the observed downregulation of immune-related genes in BPD infants (44), we observed a similar pattern of expression in BPD infants regardless of postnatal steroid treatment (Fig. S5A). Changes in whole blood gene expression could reflect changes in the frequencies of immune or erythroid cell populations. The small volume of the blood samples collected from the VLBW preterm infants precluded analysis by flow cytometry. To computationally assess whether specific cell populations were enriched among genes upregulated in BPD infants, cell type enrichment analysis was performed using the CTen (Cell Type ENrichment) platform (45). This analysis revealed that genes upregulated in BPD were significantly enriched for genes expressed in CD71<sup>+</sup> early erythroid cells (Fig. 3D; Fig. S5B). Transcription factor binding site analysis identified a single motif enriched among the promoters of genes upregulated in BPD ( $FDR < 0.05$ ). The motif GATA:SCL was found in 4.53% of those upregulated genes but only 0.89% of background sequences. GATA motifs have been found to regulate hematopoietic development (46), which is consistent with the upregulation of genes involved in red blood cell development.

**Specific taxa in the preterm infant gut microbiota are correlated with gene expression in blood.** Given the increasingly well-recognized influence of the gut microbiota on immune responses in the periphery (29), particularly in early life (28), we speculated that the composition of the gut microbiota in preterm infants might be associated with changes in (immune) gene expression in blood at the time of BPD diagnosis. A partial Spearman correlation analysis, which controlled for birth mode, sex, and gestational age, was used to assess the association between CAG relative abundance and normalized blood gene expression. This analysis was not limited to only immune gene expression but considered all genes that were at least modestly expressed in blood (count per million [cpm] of  $>10$  in 15 samples). The approach used was similar to that used in a publication from the Human Microbiome Project (47). Due to the relatively small sample size, all infants (i.e., both BPD and non-BPD) with 16S rRNA gene sequencing and RNA sequencing data were included in the analysis. Our analysis was not sufficiently powered to identify statistically significant individual gene-CAG correlations after correction for multiple testing. Instead, we assessed whether genes identified as being correlated with CAG relative abundance at a  $P$  value of  $<0.05$  were enriched for specific biological pathways or processes. This approach is analogous to methods that have been successfully adopted to detect statistical signa-



**FIG 3** RNA-Seq was used to profile gene expression in peripheral blood samples collected from BPD and non-BPD infants at recruitment (baseline) and at the time of BPD diagnosis. (A) Multidimensional scaling (MDS) analysis of RNA-Seq data in BPD and non-BPD infants at baseline (mean day of life, day 4.6) and at the time of BPD diagnosis (mean day of life, day 64.4). PC1 and -2, principal coordinates 1 and 2. (B) Volcano plot showing genes that were differentially expressed at the time of BPD diagnosis in BPD versus non-BPD infants. Horizontal line corresponds to an FDR value of 0.05. Positive log fold change (logFC) values correspond to genes with increased expression in BPD relative to non-BPD infants. (C) Top pathways and GO terms enriched among differentially expressed genes. Red bars represent pathways/terms upregulated in BPD. Blue bars represent downregulation. (D) Right, heatmap showing the expression levels of CD71<sup>+</sup> early erythroid cell-associated genes at time of BPD diagnosis; bottom left, heat map key; top left, CTen enrichment analysis results.



tures in genome-wide association (GWA) data when individual associations do not reach genome-wide significance (48–51). Additionally, the statistical significance of enriched terms was further assessed using a permutation-based approach, whereby we randomized the gene labels and repeated the enrichment analysis 1,000 times. Any pathways/GO terms found to be significantly enriched ( $FDR < 0.01$ ) in any permutation with randomized gene labels were not considered to be significantly associated with CAG relative abundance.

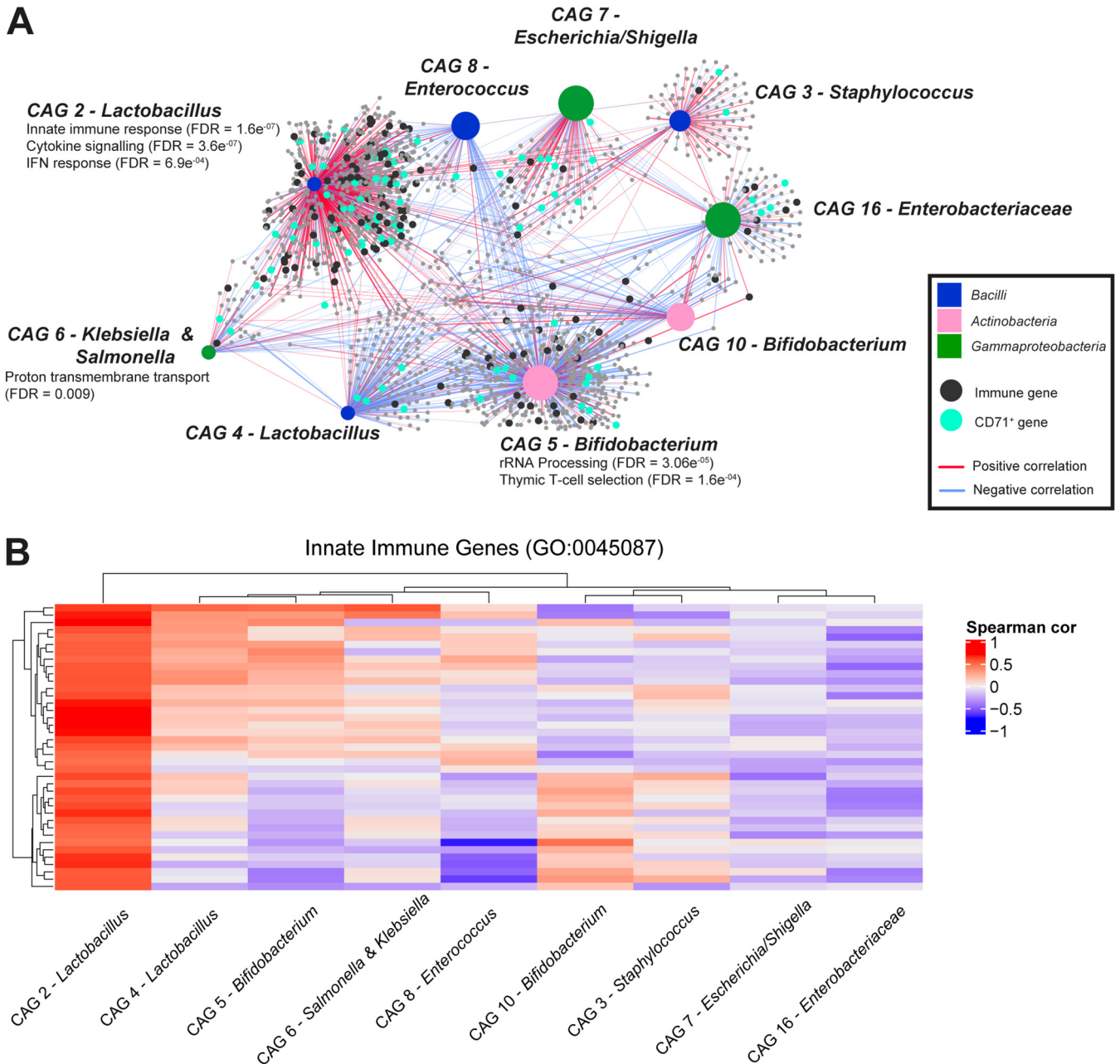
Given the multiple samples collected longitudinally per infant, the highly variable nature of the composition of the microbiota in early life, and reports of time-dependent microbiota-immune interactions (52), we performed these analyses independently for each month of life (selecting the last sample collected per month for each infant). First, we investigated which pathways or GO terms were statistically enriched among genes correlated with the RA of at least one CAG (Table S4). We found that genes correlated with CAG RAs were significantly enriched for multiple immune response pathways and GO terms (Table S4). Next, we assessed genes correlated with the relative abundances of specific CAGs and found that genes correlated with 7 of the 16 CAGs were enriched for a range of immune processes (Fig. 4; Fig. S6A and B; Table S4). For example, genes negatively correlated with the RA of *Bifidobacterium* in the first month of life were enriched for inflammatory response genes ( $FDR = 5.1e-09$ ) (Table S4), suggesting that a higher relative abundance of *Bifidobacterium* might be associated with lower inflammatory gene expression in blood. Consistent with these data, the potential anti-inflammatory effects of *Bifidobacterium* have been demonstrated in a number of studies (53).

Genes correlated with the RA of *Lactobacillus* in the second month of life of life also showed a strong enrichment for immune-related genes, including genes involved in interferon signaling (Fig. 4A) and the innate immune response (Fig. 4B). Different *Lactobacillus* strains have been shown to induce type I and II interferons in multiple previous human and mouse studies (54–57). The strong correlation between type I interferon-inducible gene expression and the RA of *Lactobacillus* was also evident at the per-gene level (Fig. S6C). We also identified an association between the RA of *Staphylococcus* in the third month of life and other immune-related processes implicated in the pathogenesis of BPD, such as the inflammatory response and MAP kinase activation (Fig. S6B; Table S4). Finally, we evaluated whether any of the correlated gene sets were also enriched for CD71<sup>+</sup> early erythroid cell-associated genes. This cell type has been shown to regulate inflammation induced by the microbiota in neonates (58). Genes correlated with the RAs of 5 of the CAGs were found to be significantly enriched for CD71<sup>+</sup>-associated genes. Multiple members of the *Bacilli* showed particularly strong enrichments for CD71<sup>+</sup>-associated genes across multiple months (Table S4), which is consistent with previous reports that members of this bacterial class can induce the expression of CD71<sup>+</sup>-associated genes (59).

While this analysis does not prove a causative link between the composition of the microbiota and gene expression in blood, our approach has identified a relatively small set of host-microbe associations, some of which are supported by the literature, that may help in understanding the role the microbiota plays in shaping the immune system in preterm infants. Further work is now needed to validate these associations in animal models and better-powered clinical studies.

## DISCUSSION

Bronchopulmonary dysplasia (BPD) is a chronic inflammatory condition of the lung that is one of the most common complications associated with early preterm birth. A large body of evidence now suggests that the gut microbiota can influence immunity and inflammation systemically, including in the lung (28, 29). The composition of the gut microbiota in early life has also been associated with the risk of necrotizing enterocolitis and sepsis (60–62). However, whether the gut microbiota influences susceptibility to BPD is currently unknown. To investigate whether changes in the gut microbiota are associated with BPD, we used 16S rRNA gene sequencing to longitudinally



**FIG 4** Correlation analysis between microbiota relative abundances and gene expression levels at time of BPD diagnosis. (A) Network showing genes expressed in blood at 36 weeks postmenstrual age that were correlated ( $P < 0.05$ ) in their expression with the relative abundance of at least one fecal microbiota coabundant group (CAG) in postnatal month 2 (larger colored nodes). Gene nodes are colored according to whether they are immune associated (i.e., in GO terms GO:0002376, GO:0045087, or GO:0006955) or associated with CD71<sup>+</sup> early erythroid cells. The size of each CAG node is proportional to its relative abundance. Pathways and GO terms that were significantly enriched (FDR < 0.01) among genes correlated with the relative abundances of specific CAG nodes are shown. See Table S4 in the supplemental material for further details. IFN, interferon. (B) Heatmap of the Spearman correlation coefficients between innate immune genes (GO:0045087) and CAG relative abundances in postnatal month 2. Only genes that were correlated ( $P < 0.05$ ) with at least one CAG are shown.

nally profile the composition of the microbiota in >250 fecal samples collected from a cohort of 50 preterm infants born at <29 weeks gestation. Adjusting for several potentially confounding factors, we identified three genera (*Escherichia/Shigella*, *Klebsiella*, and *Salmonella*) from the *Enterobacteriaceae* (class *Gammaproteobacteria*) that were significantly associated with BPD diagnosis, but interestingly, these differences were only evident in vaginally born BPD infants. Members of the *Gammaproteobacteria* have previously been identified to be associated with vaginal birth in VLBW infants (37). The relative abundance of one of the *Bifidobacterium* CAGs was also significantly

associated with BPD diagnosis. All four of the genera that showed significant association with BPD diagnosis have previously been identified as colonizers of the VLBW infant gut microbiota (20–22, 37), with *Escherichia coli* in particular being implicated in both necrotizing enterocolitis (62) and late-onset sepsis (63) in preterm infants. Differences in antibiotic exposures, which were higher in vaginally born infants, could also potentially explain why different associations were detected in vaginally and cesarean section-born infants. Importantly, however, there was no significant difference in antibiotic or probiotic exposure between vaginally born BPD and non-BPD infants, indicating that these factors cannot fully explain associations between taxa in the fecal microbiota and BPD diagnosis. Previous studies of the VLBW infant gut microbiota have primarily focused on investigating changes in the gut microbiota at the phylum or class level (20, 21, 37), but these data show that important intrafamily and intragenus variation may be present, suggesting that methods allowing higher taxonomic resolution should be implemented where possible.

Aside from the differences between BPD- and non-BPD-diagnosed infants, the preterm infants in our study displayed a pattern of colonization over time broadly similar to that described previously, with *Gammaproteobacteria* increasing with age as *Bacilli* sharply decreased (20–22, 37). Interestingly, *Clostridia* were absent from the majority of infants in our cohort, whereas *Clostridia* have been reported to be a major constituent of the VLBW preterm infant fecal microbiota in other studies (20–22, 37). We also observed much higher levels of *Actinobacteria* than have previously been reported (20–22, 37). These differences in the relative abundances of *Actinobacteria* and *Clostridia* may in part be due to the administration of the probiotic Infloran, a mixture of *Bifidobacterium bifidum*, *Bifidobacterium infantis*, and *Lactobacillus acidophilus* (36). Probiotic supplementation has routinely been given to very preterm infants in Sweden, Japan, and other nations for over a decade (33) and, as of 2015, has been widely adopted in Australia for this population (33). Importantly, Infloran was administered to almost all infants in our study and, therefore, does not confound comparisons of the different infant groups.

In addition to profiling the composition of the microbiota, we also profiled gene expression in peripheral blood samples collected from a subset of infants at baseline and again at the time of BPD diagnosis. Gene expression profiles differed significantly between time points, in agreement with previous reports of changes in cell populations in neonatal blood during this developmental period (64). While there was no significant difference in gene expression between BPD and non-BPD infants at baseline, more than 400 genes were identified as differentially expressed at the time of BPD diagnosis. Using microarray analysis, Pietrzyk et al. reported that there were between 324 and 3,498 genes differentially expressed in BPD infant blood samples, depending on day of life (18). We identified considerably fewer alterations in gene expression; however, this may be in part due to different technologies used to profile gene expression (microarray versus RNA-Seq) or differences in the time points. Pathway overrepresentation analysis found a significant enrichment for processes associated with oxygen transport and red blood cell development among genes upregulated in BPD infants in our study. Consistent with these results, we also found an enrichment of a GATA motif, which has been associated with hematopoietic development, in the promoters of genes upregulated in BPD (46). Cell type enrichment analysis also revealed enrichment for CD71<sup>+</sup> early erythroid cells among genes that were upregulated in BPD infants. CD71<sup>+</sup> cells have been shown to prevent excessive inflammation induced by commensal microbes colonizing following birth (58). Future studies should assess changes in circulating cell populations in BPD infants using flow cytometry. Intriguingly, we also uncovered associations between the relative abundances of certain taxa in the microbiota and host gene expression levels in blood, including an association between *Lactobacillus* and interferon signaling, which is supported by multiple human and mouse studies showing that a number of different *Lactobacillus* strains can regulate interferon gene expression (54–57). This analysis also pointed to a number of associations between the microbiota and several pathways relevant to BPD pathogenesis and the expression of

CD71<sup>+</sup>-associated genes. These data suggest that changes in the gut microbiota might influence immune gene expression systemically; however, these associations need to be replicated in further studies or in animal models before a causative link can be supported.

Our study is not without its limitations. As the gut microbiota at this stage of life is volatile, we focused on a modest number of infants ( $n = 50$ ) with many repeated samples to assess whether changes in the gut microbiota are associated with BPD diagnosis. The sample size limits our statistical power to detect weaker associations, and further studies with larger sample sizes are now warranted to confirm our findings. BPD was diagnosed based on the requirement for supplemental oxygen at 36 weeks postmenstrual age based on an assessment of oxygen saturation in infants meeting prespecified criteria (65). Due to our relatively limited sample size, we did not assess the relationship between the microbiota and BPD subtypes. Furthermore, all infants were recruited from the neonatal intensive care unit (NICU) in a single hospital, which may have been the source of some commonly detected microbes. Replication of our findings in larger multicenter studies is a key next step. Preterm infants are routinely exposed to medication other than antibiotics, including corticosteroids, which may impact the composition of the microbiota (66). Assessing what role, if any, current treatment regimens play in shaping the BPD gut microbiota will be a crucial, albeit challenging task. Finally, 16S rRNA sequencing has limited species- and strain-level resolution (67), multiple copies of the 16S rRNA gene per genome may inflate diversity estimates (68), and the choice of primer set may impact taxon detection. Future studies should investigate the use of shotgun metagenomics approaches to profile the composition of the microbiota.

## MATERIALS AND METHODS

**Study design and sample collection.** The infants in this study were recruited as part of the n-3 Fatty Acids for Improvement in Respiratory Outcomes (N3RO) trial (30). In N3RO, infants born before 29 weeks gestation who had commenced enteral feeding in the previous 3 days were eligible to participate. Infants were randomized to receive an enteral emulsion of DHA (60 mg/kg of body weight/day) or a control emulsion (soy) from randomization to 36 weeks postmenstrual age. The primary outcome in N3RO was BPD, defined as the requirement for supplemental oxygen and/or respiratory support at 36 weeks postmenstrual age or discharge home, whichever occurred first, using modified criteria from Walsh et al. (31) involving a physiological challenge. For a complete description of how this was performed, see the supplementary appendix in Collins et al. (30). A subset of 50 infants enrolled in the N3RO trial at the Women's and Children's Hospital, Adelaide, Australia, consented to participate in this study. Ethics approval for the additional sample collection was obtained from the Human Research Ethics Committee of the Women's and Children's Health Network (HREC 2434/12/16). Fecal samples for microbiota profiling were collected from the nappies of each infant between baseline and time of BPD diagnosis (Table 1; Fig. 1). Fecal samples were aseptically transferred to cryotubes and frozen immediately at  $-80^{\circ}\text{C}$  for later DNA extraction. A capillary blood sample (0.5 ml) was obtained from infants via heel prick at baseline (induction in the trial) and at 36 weeks postmenstrual age. Blood was collected in anticoagulant tubes containing potassium and sodium EDTA; 0.1 ml was separated into RNase-free microcentrifuge tubes, mixed with 0.3 ml RNAlater (Ambion, Inc.), and stored at  $-80^{\circ}\text{C}$  until processing.

**Fecal DNA extraction and 16S rRNA library preparation.** Approximately 0.2 g of stool sample was extracted using the PowerLyzer PowerSoil DNA isolation kit (Mo Bio Laboratories, CA, USA) according to the manufacturer's instructions, with minor modifications. Samples were eluted in 50  $\mu\text{l}$  of distilled water rather than solution C6. A FastPrep-24 instrument (MP Biomedicals, Santa Ana, USA) was used for sample homogenization, and samples were homogenized with two pulses at 6.5 m/s for 60 s. Total DNA concentrations of all samples were calculated on a Qubit 2.0 fluorometer (Thermo Fisher Scientific, MA, USA) with a high-sensitivity double-stranded DNA (dsDNA) assay kit (Life Technologies Corp., Carlsbad, CA) using 2  $\mu\text{l}$  of extract. Sequence libraries were prepared as previously described by Choo et al. (69). Briefly, the V4 hypervariable region of the 16S rRNA gene was amplified from DNA using the universal bacterial primer pair 515F (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGCCAGCAGCGG TAA-3') and 806R (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACHVGGGTWTCTAAT-3'). Amplicons were then generated, cleaned, indexed, and sequenced according to the Illumina MiSeq 16S metagenomic sequencing library preparation protocol (Illumina, Inc., San Diego, CA, USA). The resulting libraries were then sequenced ( $2 \times 300$  bp) on an Illumina MiSeq instrument.

**16S rRNA gene sequence data analysis.** 16S rRNA gene sequences were demultiplexed and imported into QIIME2 (release 2018.8) for processing (70). Sequences were error corrected, and counts of error-corrected reads per sample, which we refer to herein as exact sequence variants (ESVs), were generated with DADA2 version 1.8 (34). A phylogenetic tree of error-corrected sequences was constructed with FastTree (71). Taxonomy was assigned to sequences with the RDP Naive Bayesian Classifier algorithm (72) as implemented in the assignTaxonomy function in QIIME2 (version 1.8.0). All statistical

analysis was carried out in R version 3.6.0, with graphing performed using ggplot2 (73). Alpha diversity and Bray-Curtis distances were generated using PhyloSeq version 1.24.2 (74). Principle-coordinate analysis was conducted using the R package Ape version 5.1 (75). Differences in the relative abundances of microbial taxa were assessed using a generalized linear mixed-effects model (GLMM) implemented in the lme4 R package (41). The subject was modelled as a random effect, whereas BPD status, day of life, and DHA treatment were fixed effects. The area under the curve (AUC) was calculated using the AUC function in the DescTools package (76). Coabundant groups (CAGs) of microbes were defined by hierarchical clustering of the Spearman's rank correlation coefficients of sequence relative abundances. Only those CAGs that were present in at least 20 samples were included in the correlation analysis. R code for all analysis is provided as described in the data availability statement below.

**RNA extraction from blood and library preparation.** RNA extraction and genomic DNA elimination were carried out using the RiboPure kit (Ambion, Inc.) according to the manufacturer's instructions. Final elution into 10  $\mu$ l RNase-free water yielded 5 to 20  $\mu$ g total RNA as determined by analysis of samples using a Bioanalyzer 2100 (Agilent). RNA was transcribed into cDNA in a strand-dependent manner using the Ovation human blood RNA-Seq library systems kit (NuGen Technologies) with 500  $\mu$ g RNA as the input material. cDNA was sheared into fragments of 200 to 300 bp using a Covaris S220 focused ultrasonicator with empirically determined settings, and the samples were selectively enriched for non-ribosomal RNA and nonglobin sequences via targeted depletion of selected sequences using Insert Dependent Adaptor Cleavage technology. Additional oligonucleotides were also designed and incorporated to deplete the infant-predominant  $\gamma$ -globin, and unique adaptors were incorporated for RNA-Seq multiplexing. The resulting cDNA libraries were visualized on a Bioanalyzer 2100 to confirm the correct size distribution and to determine the cDNA concentration. Libraries were pooled and sequenced using an Illumina HiSeq 2500 machine (1  $\times$  100-bp single-end reads).

**RNA-Seq data analysis.** The quality and number of the reads for each sample were assessed using FastQC version 0.11.4 (77). Read trimming was carried out using Trimmomatic version 0.38 (78) with a window size of 2 and an average quality score of 20. Following this, reads which were <50 nucleotides after trimming were discarded. Reads that passed all quality control steps were then aligned to the human genome (GRCh38 assembly) using HISAT2 version 2.1.0 (79). The gene count matrix was generated with FeatureCounts version 1.5.0-p2 (80) using the union model with Ensembl version 93 annotation. This was then imported into R version 3.5.0 for further analysis. Counts were normalized using the trimmed mean of M values (TMM) method in EdgeR version 3.22.3 (81) prior to multidimensional scaling analysis and differential gene expression analysis (as performed with the glmLRT function). Gene sets were filtered to remove genes with <1 count per million (cpm) in 50% of samples prior to differential expression analysis. Pathway and Gene Ontology (GO) overrepresentation analysis were carried out with InnateDB (82), and cell type expression enrichment was undertaken using CTen (45). HOMER (Hypergeometric Optimization of Motif EnRichment) was used to identify transcription factor binding sites enriched among differentially expressed genes (83).

**Correlating the relative abundances of taxa in the microbiota with gene expression levels in peripheral blood.** Spearman correlation analysis was used to identify associations between the relative abundances of CAGs of ESVs in the microbiota at 1, 2, and 3 months of life and normalized blood gene expression levels at BPD diagnosis in an approach similar to that previously described by the Human Microbiome Project in reference 47. Only genes that were expressed with a cpm of >10 in at least 15 samples were included in the analysis. Samples from both BPD and non-BPD infants from whom both fecal and blood samples were collected were included in the analysis. Prior to correlation analysis, CAG counts were adjusted by fitting each to a mixed-effects model in lme4 where the subject was modelled as a random effect and adjusting for gestational age, sex, and birth mode. The residuals from each model were then used for the correlation analysis in place of the CAG counts. Genes that were correlated with at least one CAG ( $P < 0.05$ ; Spearman's  $\rho > 0.3$ ) were identified and analyzed to identify significantly enriched (FDR < 0.01) pathways or Gene Ontology terms. This approach is analogous to methods that have been successfully implemented to detect signatures in genome-wide association data when individual associations do not reach genome-wide significance (48–51). Pathway, GO, or cell type (45) (using the highly expressed, cell-specific HECS gene database) enrichment analysis was performed using a hypergeometric test implemented in R version 3.5.0. The statistical significance of enriched terms was further assessed using a permutation-based approach whereby we randomized the gene labels and repeated the enrichment analysis 1,000 times. Any pathways/GO terms found to be significantly enriched (FDR < 0.01) in any permutation with randomized gene labels were not considered to be significantly associated with CAG relative abundance. The R code for all analysis is provided as described in the data availability statement below.

**Data availability.** 16S rRNA gene sequence data have been deposited in the NCBI Sequence Read Archive under BioProject accession number [PRJNA517768](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA517768). RNA-Seq data have been deposited in the Gene Expression Omnibus (GEO) under accession number [GSE125873](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE125873). Count tables, metadata, and R code are available via the N3RO analysis repository in the Lynn Laboratory BitBucket ([https://bitbucket.org/lynnlab/n3ro\\_data\\_analysis](https://bitbucket.org/lynnlab/n3ro_data_analysis)).

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00484-19>.

**FIG S1**, PDF file, 0.4 MB.

**FIG S2**, PDF file, 0.4 MB.

**FIG S3**, PDF file, 0.1 MB.

**FIG S4**, PDF file, 0.1 MB.

**FIG S5**, PDF file, 0.3 MB.

**FIG S6**, PDF file, 2.6 MB.

**TABLE S1**, XLSX file, 0.02 MB.

**TABLE S2**, XLSX file, 0.01 MB.

**TABLE S3**, XLSX file, 0.3 MB.

**TABLE S4**, XLSX file, 0.02 MB.

## ACKNOWLEDGMENTS

We gratefully acknowledge the infants and their families who participated in this study. We also gratefully acknowledge the clinical team at Women's and Children's Hospital, without whom this study would not have been possible. We also thank Mark van der Hoek for assistance with RNA-Seq library preparation and sequencing. We thank Thomas Sullivan and Kerry Curtain for their considerable work on collating subject metadata.

This study was supported by NHMRC fellowships to Maria Makrides (grants number APP1061704 and APP1154912) and Robert A. Gibson (grant number APP1046207), CRE Foods for Future Australians funding (grant number APP1035530), and an EMBL Australia Group Leader award to David J. Lynn.

Maria Makrides reports grants from National Health & Medical Research Council during the conduct of the study, personal fees from Nestle Scientific Advisory Board membership, and personal fees from Fonterra Scientific Advisory Board membership, outside the submitted work.

Feargal J. Ryan, Damian P. Drew, and Max Moldovan performed bioinformatic analysis under the supervision of David J. Lynn. Chloe Douglas and Lex E. X. Leong performed laboratory work to generate 16S rRNA gene sequencing data under the supervision of Geraint B. Rogers. Naomi Fink collected samples and managed the clinical metadata under the supervision of Irmeli Penttila, Carmel T. Collins, Andrew J. McPhee, Maria Makrides, and Robert A. Gibson. Damian P. Drew, Miriam Lynn, and Anastasia Sribnaia performed laboratory work to generate RNA-Seq data. Feargal J. Ryan and David J. Lynn wrote the manuscript. All authors read and approved the final manuscript.

## REFERENCES

- Day CL, Ryan RM. 2017. Bronchopulmonary dysplasia: old becomes new again! *Pediatr Res* 81:210–213. <https://doi.org/10.1038/pr.2016.201>.
- Australian & New Zealand Neonatal Network, Chow SS, Creighton P, Kander V, Haslam R, Lui K. 2018. Report of the Australian and New Zealand Neonatal Network, 2016. ANZNN, Sydney, Australia.
- Higgins RD, Jobe AH, Koso-Thomas M, Bancalari E, Viscardi RM, Hartert TV, Ryan RM, Kallapur SG, Steinhorn RH, Konduri GG, Davis SD, Thebaud B, Clyman RI, Collaco JM, Martin CR, Woods JC, Finer NN, Raju TNK. 2018. Bronchopulmonary dysplasia: executive summary of a workshop. *J Pediatr* 197:300–308. <https://doi.org/10.1016/j.jpeds.2018.01.043>.
- Northway WH, Jr., Rosan RC, Porter DY. 1967. Pulmonary disease following respirator therapy of hyaline-membrane disease: bronchopulmonary dysplasia. *N Engl J Med* 276:357–368. <https://doi.org/10.1056/NEJM196702162760701>.
- Shahzad T, Radajewski S, Chao C-M, Bellusci S, Ehrhardt H. 2016. Pathogenesis of bronchopulmonary dysplasia: when inflammation meets organ development. *Mol Cell Pediatr* 3:23. <https://doi.org/10.1186/s40348-016-0051-9>.
- Trembath A, Laughon M. 2012. Predictors of bronchopulmonary dysplasia. *Clin Perinatol* 39:585–601. <https://doi.org/10.1016/j.clp.2012.06.014>.
- Bhandari V, Bizzarro MJ, Shetty A, Zhong X, Page GP, Zhang H, Ment LR, Gruen JR. 2006. Familial and genetic susceptibility to major neonatal morbidities in preterm twins. *Pediatrics* 117:1901–1906. <https://doi.org/10.1542/peds.2005-1414>.
- Lavoie PM, Pham C, Jang KL. 2008. Heritability of bronchopulmonary dysplasia, defined according to the consensus statement of the National Institutes of Health. *Pediatrics* 122:479–485. <https://doi.org/10.1542/peds.2007-2313>.
- Cuna A, Halloran B, Faye-Petersen O, Kelly D, Crossman DK, Cui X, Pandit K, Kaminski N, Bhattacharya S, Ahmad A, Mariani TJ, Ambalavanan N. 2015. Alterations in gene expression and DNA methylation during murine and human lung alveolar septation. *Am J Respir Cell Mol Biol* 53:60–73. <https://doi.org/10.1165/rcmb.2014-0160OC>.
- Lal CV, Ambalavanan N. 2015. Genetic predisposition to bronchopulmonary dysplasia. *Semin Perinatol* 39:584–591. <https://doi.org/10.1053/j.semperi.2015.09.004>.
- Mahlman M, Karjalainen MK, Huusko JM, Andersson S, Kari MA, Tammela OKT, Sankilampi U, Lehtonen L, Marttila RH, Bassler D, Poets CF, Lacaze-Masmonteil T, Danan C, Delacourt C, Palotie A, Muglia LJ, Lavoie PM, Hadchouel A, Rämetsä M, Hallman M. 2017. Genome-wide association study of bronchopulmonary dysplasia: a potential role for variants near the CRP gene. *Sci Rep* 7:9271. <https://doi.org/10.1038/s41598-017-08977-w>.
- Wang H, St Julien KR, Stevenson DK, Hoffmann TJ, Witte JS, Lazzeroni LC, Krasnow MA, Quaintance CC, Oehlert JW, Jelliffe-Pawłowski LL, Gould JB, Shaw GM, O'Brodovich HM. 2013. A genome-wide association study (GWAS) for bronchopulmonary dysplasia. *Pediatrics* 132:290–297. <https://doi.org/10.1542/peds.2013-0533>.
- Carrera P, Di Resta C, Volonteri C, Castiglioni E, Bonfiglioli S, Lazarevic D, Cittaro D, Stupka E, Ferrari M, Somaschini M. 2015. Exome sequencing and pathway analysis for identification of genetic variability relevant for bronchopulmonary dysplasia (BPD) in preterm newborns: a pilot study. *Clin Chim Acta* 451:39–45. <https://doi.org/10.1016/j.cca.2015.01.001>.

14. Hamvas A, Feng R, Bi Y, Wang F, Bhattacharya S, Mereness J, Kaushal M, Cotten CM, Ballard PL, Mariani TJ. 2018. Exome sequencing identifies gene variants and networks associated with extreme respiratory outcomes following preterm birth. *BMC Genet* 19:94. <https://doi.org/10.1186/s12863-018-0679-7>.
15. Li J, Yu K-H, Oehlert J, Jelliffe-Pawlowski LL, Gould JB, Stevenson DK, Snyder M, Shaw GM, O'Brodovich HM. 2015. Exome sequencing of neonatal blood spots and the identification of genes implicated in bronchopulmonary dysplasia. *Am J Respir Crit Care Med* 192:589–596. <https://doi.org/10.1164/rccm.201501-0168OC>.
16. Bao TP, Wu R, Cheng HP, Cui XW, Tian ZF. 2016. Differential expression of long non-coding RNAs in hyperoxia-induced bronchopulmonary dysplasia. *Cell Biochem Funct* 34:299–309. <https://doi.org/10.1002/cbf.3190>.
17. Bhattacharya S, Go D, Krenitsky DL, Huyck HL, Solleti SK, Lunger VA, Metlay L, Srisuma S, Wert SE, Mariani TJ, Pryhuber GS. 2012. Genome-wide transcriptional profiling reveals connective tissue mast cell accumulation in bronchopulmonary dysplasia. *Am J Respir Crit Care Med* 186:349–358. <https://doi.org/10.1164/rccm.201203-0406OC>.
18. Pietrzyk JJ, Kwinta P, Wollen EJ, Bik-Multanowski M, Madetko-Talowska A, Günther C-C, Jagła M, Tomasik T, Saugstad OD. 2013. Gene expression profiling in preterm infants: new aspects of bronchopulmonary dysplasia development. *PLoS One* 8:e78585. <https://doi.org/10.1371/journal.pone.0078585>.
19. Yang M, Chen B-L, Huang J-B, Meng Y-N, Duan X-J, Chen L, Li L-R, Chen Y-P. 2017. Angiogenesis-related genes may be a more important factor than matrix metalloproteinases in bronchopulmonary dysplasia development. *Oncotarget* 8:18670–18679. <https://doi.org/10.18632/oncotarget.14722>.
20. La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM, Stevens HJ, Bennett WE, Shaikh N, Linneman LA, Hoffmann JA, Hamvas A, Deych E, Shands BA, Shannon WD, Tarr PI. 2014. Patterned progression of bacterial populations in the premature infant gut. *Proc Natl Acad Sci U S A* 111:12522–12527. <https://doi.org/10.1073/pnas.1409497111>.
21. Warner BB, Deych E, Zhou Y, Hall-Moore C, Weinstock GM, Sodergren E, Shaikh N, Hoffmann JA, Linneman LA, Hamvas A, Khanna G, Rouggy-Nickless LC, Ndao IM, Shands BA, Escobedo M, Sullivan JE, Radmacher PG, Shannon WD, Tarr PI. 2016. Gut bacteria dysbiosis and necrotizing enterocolitis in very low birthweight infants: a prospective case-control study. *Lancet* 387:1928–1936. [https://doi.org/10.1016/S0140-6736\(16\)00081-7](https://doi.org/10.1016/S0140-6736(16)00081-7).
22. Patel AL, Mutlu EA, Sun Y, Koenig L, Green S, Jakubowicz A, Mryan J, Engen P, Fogg L, Chen AL, Pombar X, Meier PP, Keshavarzian A. 2016. Longitudinal survey of microbiota in hospitalized preterm very low birth weight infants. *J Pediatr Gastroenterol Nutr* 62:292. <https://doi.org/10.1097/MPG.0000000000000913>.
23. Cantey JB, Huffman LW, Subramanian A, Marshall AS, Ballard AR, Lefevre C, Sagar M, Pruszyński JE, Mallett LH. 2017. Antibiotic exposure and risk for death or bronchopulmonary dysplasia in very low birth weight infants. *J Pediatr* 181:289–293.e1. <https://doi.org/10.1016/j.jpeds.2016.11.002>.
24. Novitsky A, Tuttle D, Locke RG, Saiman L, Mackley A, Paul DA. 2015. Prolonged early antibiotic use and bronchopulmonary dysplasia in very low birth weight infants. *Am J Perinatol* 32:43–48. <https://doi.org/10.1055/s-0034-1373844>.
25. Lal CV, Travers C, Aghai ZH, Eipers P, Jilling T, Halloran B, Carlo WA, Keeley J, Rezonzew G, Kumar R, Morrow C, Bhandari V, Ambalavanan N. 2016. The airway microbiome at birth. *Sci Rep* 6:31023–31023. <https://doi.org/10.1038/srep31023>.
26. Berkhout DJC, Niemarkt HJ, Benninga MA, Budding AE, van Kaam AH, Kramer BW, Pantophlet CM, van Weissenbruch MM, de Boer NKH, de Meij T. 2018. Development of severe bronchopulmonary dysplasia is associated with alterations in fecal volatile organic compounds. *Pediatr Res* 83:412. <https://doi.org/10.1038/pr.2017.268>.
27. Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, Belzer C, Delgado Palacio S, Arboleya Montes S, Mancabelli L, Lugli GA, Rodriguez JM, Bode L, de Vos W, Gueimonde M, Margolles A, van Sinderen D, Ventura M. 2017. The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiol Mol Biol Rev* 81:e00036-17. <https://doi.org/10.1128/MMBR.00036-17>.
28. Gray J, Oehrle K, Worthen G, Alenghat T, Whitsett J, Deshmukh H. 2017. Intestinal commensal bacteria mediate lung mucosal immunity and promote resistance of newborn mice to infection. *Sci Transl Med* 9:eaa9412. <https://doi.org/10.1126/scitranslmed.aaa9412>.
29. Budden KF, Gellatly SL, Wood DLA, Cooper MA, Morrison M, Hugenholtz P, Hansbro PM. 2017. Emerging pathogenic links between microbiota and the gut–lung axis. *Nat Rev Microbiol* 15:55. <https://doi.org/10.1038/nrmicro.2016.142>.
30. Collins CT, Makrides M, McPhee AJ, Sullivan TR, Davis PG, Thio M, Simmer K, Rajadurai VS, Travadi J, Berry MJ, Liley HG, Opie GF, Tan K, Lui K, Morris SA, Stack J, Stark MJ, Chua M-C, Jayagobi PA, Holberton J, Bolisetty S, Callander IR, Harris DL, Gibson RA. 2017. Docosaehexaenoic acid and bronchopulmonary dysplasia in preterm infants. *N Engl J Med* 376:1245–1255. <https://doi.org/10.1056/NEJMoa1611942>.
31. Walsh MC, Yao Q, Gettner P, Hale E, Collins M, Hensman A, Everette R, Peters N, Miller N, Muran G, Auten K, Newman N, Rowan G, Grisby C, Arnell K, Miller L, Ball B, McDavid G. 2004. Impact of a physiologic definition on bronchopulmonary dysplasia rates. *Pediatrics* 114:1305–1311. <https://doi.org/10.1542/peds.2004-0204>.
32. AlFaleh K, Anabrees J. 2014. Probiotics for prevention of necrotizing enterocolitis in preterm infants. *Evid Based Child Health* 9:584–671. <https://doi.org/10.1002/ebch.1976>.
33. Deshpande G, Rao S, Patole S. 2015. Probiotics in neonatal intensive care—back to the future. *Aust N Z J Obstet Gynaecol* 55:210–217. <https://doi.org/10.1111/ajo.12328>.
34. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581. <https://doi.org/10.1038/nmeth.3869>.
35. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852–857. <https://doi.org/10.1038/s41587-019-0209-9>.
36. Plummer EL, Bulach DM, Murray GL, Jacobs SE, Tabrizi SN, Garland SM, Garland SM, Jacobs SE, Tobin JM, Tabrizi SN, Pirota M, Donath S, Opie GF, Tang MLK, Morley CJ, Hickey L, Tan K, Lewis A, Veldman A, Travadi J, Wright IMR, Osborn DA, Sinn J, Levison J, Stack JA, DePaoli AG, Austin NC, Darlow BA, Alsweiler JM, Buksh MJ, ProPrens Study Group. 2018. Gut microbiota of preterm infants supplemented with probiotics: sub-study of the ProPrens trial. *BMC Microbiol* 18:184. <https://doi.org/10.1186/s12866-018-1326-1>.
37. Ho TTB, Groer MW, Kane B, Yee AL, Torres BA, Gilbert JA, Maheshwari A. 2018. Dichotomous development of the gut microbiome in pre-term infants. *Microbiome* 6:157. <https://doi.org/10.1186/s40168-018-0547-8>.
38. Edgar RC. 2018. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *PeerJ PrePrints* 6:e4652. <https://doi.org/10.7717/peerj.4652>.
39. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L, Alm EJ, Birmingham A, Cram JA, Fuhrman JA, Raes J, Sun F, Zhou J, Knight R. 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J* 10:1669. <https://doi.org/10.1038/ismej.2015.235>.
40. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. 2012. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* 8:e1002606. <https://doi.org/10.1371/journal.pcbi.1002606>.
41. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J Stat Soft* 67:1–48. <https://doi.org/10.18637/jss.v067.i01>.
42. Schulman J, Dimand RJ, Lee HC, Duenas GV, Bennett MV, Gould JB. 2015. Neonatal intensive care unit antibiotic use. *Pediatrics* 135:826–833. <https://doi.org/10.1542/peds.2014-3409>.
43. Halliday HL. 2017. Update on postnatal steroids. *Neonatology* 111:415–422. <https://doi.org/10.1159/000458460>.
44. Coutinho AE, Chapman KE. 2011. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. *Mol Cell Endocrinol* 335:2–13. <https://doi.org/10.1016/j.mce.2010.04.005>.

45. Shoemaker JE, Lopes TJ, Ghosh S, Matsuoka Y, Kawaoka Y, Kitano H. 2012. CTen: a web-based platform for identifying enriched cell types from heterogeneous microarray data. *BMC Genomics* 13:460. <https://doi.org/10.1186/1471-2164-13-460>.
46. Kobayashi-Osaki M, Ohneda O, Suzuki N, Minegishi N, Yokomizo T, Takahashi S, Lim K-C, Engel JD, Yamamoto M. 2005. GATA motifs regulate early hematopoietic lineage-specific expression of the Gata2 gene. *Mol Cell Biol* 25:7005–7020. <https://doi.org/10.1128/MCB.25.16.7005-7020.2005>.
47. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, Casero D, Courtney H, Gonzalez A, Graeber TG, Hall AB, Lake K, Landers CJ, Mallick H, Plichta DR, Prasad M, Rahnavard G, Sauk J, Shungin D, Vázquez-Baeza Y, White RA, Bishai J, Bullock K, Deik A, Dennis C, Kaplan JL, Khalili H, McIver LJ, Moran CJ, Nguyen L, Pierce KA, Schwager R, Sirota-Madi A, Stevens BW, Tan W, ten Hove JJ, Weingart G, Wilson RG, Yajnik V, Braun J, Denson LA, Jansson JK, Knight R, Kugathasan S, McGovern DPB, et al. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569:655–662. <https://doi.org/10.1038/s41586-019-1237-9>.
48. Carbonetto P, Stephens M. 2013. Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for IL-2 signaling genes in type 1 diabetes, and cytokine signaling genes in Crohn's disease. *PLoS Genet* 9:e1003770. <https://doi.org/10.1371/journal.pgen.1003770>.
49. Zhu X, Stephens M. 2018. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nat Commun* 9:4361. <https://doi.org/10.1038/s41467-018-06805-x>.
50. Zhang Z, Ma P, Li Q, Xiao Q, Sun H, Olasege BS, Wang Q, Pan Y. 2018. Exploring the genetic correlation between growth and immunity based on summary statistics of genome-wide association studies. *Front Genet* 9:393. <https://doi.org/10.3389/fgene.2018.00393>.
51. White MJ, Yaspan BL, Veatch OJ, Goddard P, Risse-Adams OS, Contreras MG. 2019. Strategies for pathway analysis using GWAS and WGS data. *Curr Protoc Hum Genet* 100:e79. <https://doi.org/10.1002/cphg.79>.
52. Al Nabhani Z, Dulauroy S, Marques R, Cousu C, Al Bounny S, Déjardin F, Sparwasser T, Bérard M, Cerf-Bensussan N, Eberl G. 2019. A weaning reaction to microbiota is required for resistance to immunopathologies in the adult. *Immunity* 50:1276–1288. <https://doi.org/10.1016/j.immuni.2019.02.014>.
53. Ruiz L, Delgado S, Ruas-Madiedo P, Sánchez B, Margolles A. 2017. Bifidobacteria and their molecular communication with the immune system. *Front Microbiol* 8:2345. <https://doi.org/10.3389/fmicb.2017.02345>.
54. Weiss G, Maaetoft-Udsen K, Stifter SA, Hertzog P, Goriely S, Thomsen AR, Paludan SR, Frøkiær H. 2012. MyD88 drives the IFN- $\beta$  response to *Lactobacillus acidophilus* in dendritic cells through a mechanism involving IRF1, IRF3, and IRF7. *J Immunol* 189:2860–2868. <https://doi.org/10.4049/jimmunol.1103491>.
55. Pereyra BS, Falcoff R, Falcoff E, Lemonnier D. 1991. Interferon induction by *Lactobacillus bulgaricus* and *Streptococcus thermophilus* in mice. *Eur Cytokine Netw* 2:299–303.
56. Weiss G, Christensen HR, Zeuthen LH, Vogensen FK, Jakobsen M, Frøkiær H. 2011. *Lactobacilli* and *bifidobacteria* induce differential interferon- $\beta$  profiles in dendritic cells. *Cytokine* 56:520–530. <https://doi.org/10.1016/j.cyto.2011.07.024>.
57. Kishi A, Uno K, Matsubara Y, Okuda C, Kishida T. 1996. Effect of the oral administration of *Lactobacillus brevis* subsp. *coagulans* on interferon-alpha producing capacity in humans. *J Am Coll Nutr* 15:408–412. <https://doi.org/10.1080/07315724.1996.10718617>.
58. Elahi S, Ertelt JM, Kinder JM, Jiang TT, Zhang X, Xin L, Chaturvedi V, Strong BS, Qualls JE, Steinbrecher KA, Kalfa TA, Shaaban AF, Way SS. 2013. Immunosuppressive CD71+ erythroid cells compromise neonatal host defence against infection. *Nature* 504:158–162. <https://doi.org/10.1038/nature12675>.
59. Banchereau R, Jordan-Villegas A, Ardura M, Mejias A, Baldwin N, Xu H, Saye E, Rossello-Urgell J, Nguyen P, Blankenship D, Creech CB, Pascual V, Banchereau J, Chaussabel D, Ramilo O. 2012. Host immune transcriptional profiles reflect the variability in clinical disease manifestations in patients with *Staphylococcus aureus* infections. *PLoS One* 7:e34390. <https://doi.org/10.1371/journal.pone.0034390>.
60. Madan JC, Salari RC, Saxena D, Davidson L, O'Toole GA, Moore JH, Sogin ML, Foster JA, Edwards WH, Palumbo P, Hibberd PL. 2012. Gut microbial colonisation in premature neonates predicts neonatal sepsis. *Arch Dis Child Fetal Neonatal Ed* 97:F456–F462. <https://doi.org/10.1136/fetalneonatal-2011-301373>.
61. Stewart CJ, Marrs EC, Nelson A, Lanyon C, Perry JD, Embleton ND, Cummings SP, Berrington JE. 2013. Development of the preterm gut microbiome in twins at risk of necrotising enterocolitis and sepsis. *PLoS One* 8:e73465. <https://doi.org/10.1371/journal.pone.0073465>.
62. Ward DV, Scholz M, Zolfo M, Taft DH, Schibler KR, Tett A, Segata N, Morrow AL. 2016. Metagenomic sequencing with strain-level resolution implicates uropathogenic *E. coli* in necrotizing enterocolitis and mortality in preterm infants. *Cell Rep* 14:2912–2924. <https://doi.org/10.1016/j.celrep.2016.03.015>.
63. Stewart CJ, Embleton ND, Marrs ECL, Smith DP, Fofanova T, Nelson A, Skeath T, Perry JD, Petrosino JF, Berrington JE, Cummings SP. 2017. Longitudinal development of the gut microbiome and metabolome in preterm neonates with late onset sepsis and healthy controls. *Microbiome* 5:75. <https://doi.org/10.1186/s40168-017-0295-1>.
64. Olin A, Henckel E, Chen Y, Lakshminanth T, Pou C, Mikes J, Gustafsson A, Bernhardsson AK, Zhang C, Bohlin K, Brodin P. 2018. Stereotypic immune system development in newborn children. *Cell* 174:1277–1292. <https://doi.org/10.1016/j.cell.2018.06.045>.
65. Jobe AH, Bancalari E. 2001. Bronchopulmonary dysplasia. *Am J Respir Crit Care Med* 163:1723–1729. <https://doi.org/10.1164/ajrccm.163.7.2011060>.
66. Huang EY, Inoue T, Leone VA, Dalal S, Touw K, Wang Y, Musch MW, Theriault B, Higuchi K, Donovan S, Gilbert J, Chang EB. 2015. Using corticosteroids to reshape the gut microbiome: implications for inflammatory bowel diseases. *Inflamm Bowel Dis* 21:963–972. <https://doi.org/10.1097/MIB.0000000000000332>.
67. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12:635. <https://doi.org/10.1038/nrmicro3330>.
68. Kembel SW, Wu M, Eisen JA, Green JL. 2012. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput Biol* 8:e1002743. <https://doi.org/10.1371/journal.pcbi.1002743>.
69. Choo JM, Leong LEX, Rogers GB. 2015. Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep* 5:16350. <https://doi.org/10.1038/srep16350>.
70. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.
71. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26:1641–1650. <https://doi.org/10.1093/molbev/msp077>.
72. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. <https://doi.org/10.1128/AEM.00062-07>.
73. R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
74. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>.
75. Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290. <https://doi.org/10.1093/bioinformatics/btg412>.
76. Signorell A. 2016. DescTools: tools for descriptive statistics. R package version 99.18.
77. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.
78. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
79. Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12:357–360. <https://doi.org/10.1038/nmeth.3317>.
80. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose



- program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
81. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>.
82. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS, Lynn DJ. 2013. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res* 41:D1228–D1233. <https://doi.org/10.1093/nar/gks1147>.
83. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.