Bacteroidota and *Lachnospiraceae* Integration Into the Gut Microbiome at Key Time Points in Early Life are Critical for Neurodevelopment

Kaitlyn Oliphant

University of Chicago Division of the Biological Sciences <https://orcid.org/0000-0002-6921-7094>

Mehneez Ali

University of Chicago Division of the Biological Sciences

Mark D'Souza

University of Chicago Division of the Biological Sciences

Patrick D. Hughes

University of Chicago Comer Children's Hospital

Dinanath Sulakhe

University of Chicago Division of the Biological Sciences

Annie Z. Wang

University of Chicago Comer Children's Hospital

Bingqing Xie

University of Chicago Division of the Biological Sciences

Rummanu Yeasin

University of Chicago Division of the Biological Sciences

Michael M. Msall

University of Chicago Division of the Biological Sciences

Bree Andrews

University of Chicago Division of the Biological Sciences

Erika C. Claud (✉ EClaud@peds.bsd.uchicago.edu)

Department of Pediatrics, Biological Sciences Division, University of Chicago, Chicago, IL, USA

<https://orcid.org/0000-0002-2408-2114>

Research

Keywords: Human gut microbiome, Infant microbiome succession, Infant neurodevelopment

Posted Date: July 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-668952/v1>

Abstract

Background: The early life microbiome plays critical roles in host development, shaping long-term outcomes including brain functioning. It is not known which initial infant colonizers elicit optimal neurodevelopment; thus, this study investigated the association between gut microbiome succession from the first week of life and head circumference growth (HCG), the earliest validated marker for neurodevelopment. Faecal samples were collected weekly from a preterm infant cohort during their neonatal intensive care unit stay and subjected to 16S rRNA gene sequencing for evaluating gut microbiome composition, in conjunction with clinical data and head circumference measurements.

Results: Preterm infants with suboptimal HCG trajectories had a depletion in the abundance/prevalence of *Bacteroidota* and *Lachnospiraceae*, independent of morbidity and caloric restriction. The severity of gut microbiome depletion matched the timing of significant HCG pattern separation between study groups at 30 weeks postmenstrual age. Consideration of the clinical variables indicated that optimal infant microbiome succession is primarily driven by dispersal limitation (i.e., delivery mode) and secondarily by habitat filtering (i.e., antibiotics and enteral feeding).

Conclusions: *Bacteroidota* and *Lachnospiraceae* are known core taxa of the adult microbiome, with roles in dietary glycan foraging, beneficial metabolite production and immunity, and our work provides strong evidence that their integration into the gut microbiome needs to occur early for optimal neurodevelopment.

Background

The first five years of life are widely recognized to be critical for an individual's educational and vocational success[1], as well as lifelong health and wellbeing[2]. On a cellular neurodevelopmental level, synaptogenesis and myelination occur at a net positive during this timeframe increasing brain volume until 90% of adult size is reached by age five[3,4]. These processes are known to be both programmed by genetics and responsive to environmental cues[5], and the consequence of their impediment is developmental disability, a group of conditions resultant from physical, learning, language, or behavioural impairment that often lasts throughout an individual's lifetime[6]. Developmental disability was estimated by the Global Burden Disease Study 2016 to affect 8.4% of children under five years of age worldwide[7], and although global initiatives specifically targeting this timeframe for improving overall human health and economics have achieved success in reducing child mortality[8], the rate of developmental disability remains minimally changed[7]. Thus, there is a vital need to identify modifiable environmental factors to reduce the incidence of developmental disabilities. Infancy represents a potentially crucial stage for intervention, as diagnosis of these developmental disabilities can already occur by 2-3 years of age[9,10]. An early marker of neurodevelopment is necessary, and head circumference growth (HCG) has been

found to sufficiently proxy increase in brain volume, thus correlating well with later neurodevelopmental testing results[11–15].

Studies to date have focused on nutritional interventions for improving HCG in infants[16–23]. However, associations between infant diet and neurodevelopment are not always clear; for example, a recent meta-analysis on human milk feeding for improving neurodevelopmental outcomes demonstrated only weak or inconclusive evidence for its beneficial effects[23]. One overlooked factor that is gaining attention is the ecosystem of microorganisms that populate the gastrointestinal tract, the gut microbiome. The gut microbiome diversifies in parallel with infant development[24–27], and antibiotic use in the first year of life has been correlated with later developmental outcomes, including lower cognitive capabilities, attention deficit hyperactivity disorder (ADHD), anxiety and depression[28,29]. The gut microbiome also intersects with many environmental factors already associated with neurodevelopment[30–33], importantly including nutrition. Critical roles of the gut microbiome in digestion include breaking-down macronutrients, producing neuroactive metabolites after fermentation, and synthesizing micronutrients[5,34]. The gut microbiome additionally indirectly affects neurodevelopment via its interactions with the immune system and intestinal barrier, impacting systemic inflammation and circulation of metabolites[5,34]. Alterations in the composition and function of the gut microbiome have been associated with developmental disabilities in children including autism spectrum disorder (ASD) [35] and ADHD[36]. However, it is not known which early life microbes occurring prior to developmental disability onset are vital to optimal neurodevelopment, and no study to date has examined the relationship between the earliest neurodevelopmental marker, HCG, and the gut microbiome in infants over time.

To address this knowledge gap, prospectively collected longitudinal faecal samples from 58 infants born <34 weeks gestational age at the University of Chicago Comer Children's Hospital were subjected to 16S rRNA gene sequencing[37,38] to determine the microbiome composition and predicted functional profile[39]. Preterm infants were selected as they represent a substantive cohort (1 in every 10 infants worldwide is born preterm[40]) at an increased risk for both developmental disability[41] and altered gut microbiome configurations from clinical exposures such as antibiotics[24–27]. Further, preterm infants can be strictly monitored within the neonatal intensive care unit (NICU) environment, allowing statistical and machine learning models to be built that incorporate HCG trajectories, time, and clinical variables. We hypothesized that features of the early gut microbiome are associated with HCG from birth to term-equivalent age in the NICU and found that infants with suboptimal HCG trajectories (SHCGT) experienced a concurrent loss in HCG and the abundances of *Bacteroidota* and *Lachnospiraceae* specifically from 31–36 weeks postmenstrual age (PMA) independent of clinical morbidity and caloric restriction. Thorough examination of the clinical variables revealed preferential effects of infant microbiome successional drivers, with dispersal limitation (i.e., delivery mode) superseding habitat filtering (i.e., antibiotics and enteral feeding) in shaping optimum microbiome and thus host development.

Methods

Study participants:

The Microbiome In Neonatal Development (MIND) study received approval from an institutional review board (IRB16-1431) in accordance with the human subjects' research policies of the U.S. Food & Drug Administration and the Declaration of Helsinki. Study participant enrolment took place in the NICU at the University of Chicago Comer Children's Hospital between January 2010 and December 2018. Infants born prior to 37 weeks gestational age were eligible for the study and were enrolled after receiving written informed consent from the parent. Infants with a genetic syndrome or severe congenital anomalies, including major congenital heart disease or major kidney, lung, or brain malformation, were excluded, as were infants judged not to be viable by the attending physician. There was no exclusion based on race, sex, or socioeconomic status.

Clinical data and sample collection:

The primary outcome was infant head circumference growth (HCG). HCG trajectory was evaluated by the difference between head circumference z-scores (HCZs) calculated from the Fenton 2013 growth curve[42] by the completed weeks method at birth and 36 weeks PMA. If the infant was discharged prior to 36 weeks PMA, the measurement at NICU discharge was taken instead (no infant was discharged prior to 34 weeks PMA). Infants that expired in the NICU were excluded. Head circumferences were measured weekly by nursing staff.

For assessment of the infant faecal microbiome, infant diapers were collected weekly by the nursing staff and stored immediately at -20 °C. Research staff aliquoted the faecal samples into sterile microcentrifuge tubes under an anaerobic atmosphere (5/5/90 H₂/CO₂/N₂). Samples were immediately cryopreserved at -80 °C after aliquoting.

Clinical variables identified as possible confounders were gathered by abstracting information from patient charts using the electronic medical record system at the University of Chicago Comer Children's Hospital. Data collected included delivery mode, gestational age at birth, sex, birthweight, head circumference at birth, enteral feeding regimens, antibiotics administration, and clinical morbidities.

Enteral feeding data was extracted as the total volume in mL of formula, human milk and total enteral feeds (formula + human milk) for each day the infant was in the NICU up until the head circumference measurement end point. Human milk comprised of mother's own milk except for two infants that were fed donor human milk. Infants were additionally weighed daily by nurses, and the feeding volumes in mL/kg were computed using the bodyweight recorded. The daily feeding amounts were then averaged over the time periods of interest, and the percent of human milk was calculated by dividing the amount by the total enteral feeding amount. The number of days of total parenteral nutrition (TPN - 0 mL/kg total enteral feeds) over the time periods of interest was also calculated, in addition to the first day of life that full enteral feeds (120 mL/kg) was achieved.

Antibiotics data was extracted up until the head circumference measurement end point as the total number of days and longest number of consecutive days administered over the time periods of interest. The administered antibiotics (via IV or injection unless otherwise noted) included antifungals (amphotericin B except for one infant that received oral nystatin), cephalosporins (3rd generation cefotaxime, ceftazidime or ceftriaxone except for two infants that received 1st generation cefazolin, four that received 4th generation cefepime and one that received 1st generation oral cephalixin), clindamycin, oral erythromycin, gentamicin, metronidazole, penicillins (ampicillin except for one infant that received penicillin, two that received piperacillin, two that received oxacillin and two that received oral amoxicillin) and vancomycin.

In addition to the total length of stay in the NICU and the PMA at discharge, the morbidities reported include bronchopulmonary dysplasia (BPD), severe brain injury (SBI), severe retinopathy of prematurity (SROP), necrotizing enterocolitis (NEC), seizures and sepsis. BPD was defined as the need for supplemental oxygen at 36 weeks PMA[43]. SBI was defined as the presence of periventricular leukomalacia and/or grade 3 or 4 intraventricular haemorrhage on a cranial ultrasonogram[44]. SROP was defined as unilateral or bilateral retinopathy of prematurity (ROP) at stage 4 or 5 and/or ROP requiring treatment by laser or anti-vascular endothelial growth factor drugs[44]. NEC was defined by the modified Bell's criteria, with at least stage 2 being requisite for disease classification[45]. Presence of seizures required both clinical and EEG-activity confirmation. Sepsis encompassed both blood culture-positive early- (<72 h) and late- (>72 h) onset.

Illumina 16S rRNA gene sequencing and processing:

Patient faecal samples were submitted to the Environmental Sample Preparation and Sequencing Facility at Argonne National Laboratory (Lemont, IL, USA) for genomic DNA extraction and Illumina 16S rRNA gene sequencing[37,38]. FASTQ files received from the facility were imported into the QIIME2 version 2019.7[46] pipeline per batch and demultiplexed using the demux tool. The demultiplexed samples were then denoised by the sample inference tool DADA2[47] from amplicon data within QIIME2. Batches were then exported and merged prior to being imported into R statistical software version 3.6.2 for subsequent classification and post-processing. Amplicon sequence variants (ASVs) were classified to the genus level by the IDTAXA method[48] via R package DECIPHER version 2.14.0 using the Genome Taxonomy Database[49] version 89. ASVs were additionally classified into species-like groups by the online NCBI Nucleotide Basic Local Alignment Search Tool[50] (BLAST - <https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Top hits were selected by first the highest percentage identity, and second the lowest e-value. ASVs were amalgamated by identical lists of top hits, taking the sequence with the highest count in the dataset as the representative. Species were called if the top hit was of $\geq 97\%$ identity and matched the genus identified by IDTAXA; multiple species calls were allowed. After classification, low quality samples with <1000 total sequence counts were removed, and then species-like groups that represented <0.1% mean abundance were culled.

The α -diversity metrics of richness and Shannon diversity were computed by R package iNEXT[51] version 2.0.20. For β -diversity analysis, the taxonomic levels of phylum, family, genus, and species were individually considered. These data were centre-log ratio (CLR) transformed using R package ALDEx2[52] version 1.18.0 and taking the median of the Monte-Carlo instances as the value. This CLR transformation allowed standard statistical analysis to be conducted on the inherently compositional data. Predicted functional profiles were obtained via the R package Tax4Fun2[39] version 1.1.5 using their downloaded version 2 reference dataset from NCBI BLAST (Ref99NR) and KEGG[53]. As the output of Tax4Fun2 yields not raw counts but percent abundances, this data was CLR transformed manually, with zeroes imputed by the non-parametric multiplicative simple method using a detection threshold of 1×10^{-12} through R package zCompositions version 1.3.4.

Statistical analysis:

All statistical analysis was done in R statistical software version 3.6.2 with plots generated by R package ggplot2 version 3.3.0. Original study groups were defined by the loss in HCZ from birth to 36 weeks PMA: appropriate head circumference growth trajectory (AHCGT - ≥ 0.5 ; n=28 patients, n=118 faecal samples), mildly SHCGT ($< 0.5-1$; n=16 patients, n=67 faecal samples), moderately SHCGT ($< 1-1.5$; n=8 patients, n=32 faecal samples) and severely SHCGT (< 1.5 ; n=6 patients, n=23 faecal samples). Later analysis combined all SHCGT groups in comparison to the AHCGT group, and the moderately to severely SHCGT groups in comparison to the AHCGT to mildly SHCGT groups, for both the complete and limited morbidity (LM) datasets. The LM subset comprised of infants without SBI, SROP, NEC, seizures, or sepsis; BPD was not excluded as it had an equivalent risk ratio between study groups (**Table 2**): AHCGT (n=23 patients, n=101 faecal samples), mildly SHCGT (n=13 patients, n=50 faecal samples) and moderately to severely SHCGT (n=8 patients, n=32 faecal samples).

Differences in the clinical characteristics of delivery mode, sex, and incidence of morbidities between study groups were assessed by the Fisher's exact test and presented as the percentages (number) per group. Differences in the remaining clinical variables between study groups were assessed by Welch's ANOVA (multiple groups) or Welch's t-test (two groups) and presented as the mean $\pm$ standard deviation per group. Post-hoc analysis was conducted by the Games-Howell method using R package userfriendlyscience version 0.7.2, and pairwise Cohen's Ds were computed by R package rstatix version 0.6.0 without assuming equal variances. Two-sample tests with two-sided p values were completed unless the mean and standard deviation for a study group was zero, which occurred for the clinical variables of clindamycin, erythromycin, and metronidazole administration. In these cases, a series of Welch's t-tests were conducted with the lowest p value reported, and one-sample tests with one-sided (greater) p values were completed *in lieu* of the all zero study groups. A standard p value < 0.05 was considered statistically significant. Detailed statistics are provided for the complete dataset [see **Additional file 1**] and for the LM subset [see **Additional file 2**].

For multidimensional β -diversity analysis, redundancy analysis (RDA) was performed by R package vegan version 2.5.6 on the genus level data, with HCG trajectory and PMA as environmental variables.

Both overall significance of the model and term significance was determined by an ANOVA like permutation test, with constraints on the permutations by patient, through this R package. A standard p value <0.05 was considered statistically significant. Vector or factor averages of the environmental variables were similarly computed by this R package, for both reporting the R^2 values and locating the study group centroids on the RDA plot. Details of the RDA are provided [see **Additional file 3**].

For determining statistically significant differences in α -diversity metrics, individual microbial taxon abundances or predicted function (KEGG orthologies - KOs) abundances between study groups, mixed-effect linear models were constructed via R package lme4 version 1.1.23 using HCG trajectory and PMA as fixed effects, and patient as a random effect. Additionally, an equivalent approach was utilized to test which microbial taxon abundances were resultant from vertical transmission, by replacing the fixed effect HCG trajectory with delivery mode. Variables were standardized using R package standardize version 0.2.1 prior to fitting the model, and p values were determined by Satterthwaite's method using R package lmerTest version 3.1.2. For the α -diversity metrics, a standard p value <0.05 was considered significant. To correct for multiple testing of the individual microbial taxa and KOs, the Benjamini-Hochberg method was utilized, and a false-positive rate of $<1\%$ was the set threshold for significance. The marginal coefficient of determination for generalized mixed-effect models (variance explained by fixed effects) calculated by R package MuMIn version 1.43.15 was also reported as the R^2 . Change point analysis (see below) uncovered the key time windows of 24-30 weeks PMA and 31-36 weeks PMA. The differences in individual microbial taxon abundances and KO abundances between study groups were thus again evaluated during these time windows using the same method as above, except PMA was removed as a fixed effect since it was already inherently controlled. The standardization step was also omitted as there would be no issues of scaling due to the model having only one fixed effect.

Prevalence of individual microbial taxa or KOs amongst patients within study groups was also computed overall and separately for the key time windows of 24-30 weeks PMA and 31-36 weeks PMA. If a given microbial taxa or KO had >0 sequence counts in at least one faecal sample during the time window of interest, it was considered to be present in the patient. Statistical significance was then evaluated between study groups by the Fisher's exact test, with the Benjamini-Hochberg correction for multiple testing applied. A false-positive rate of $<1\%$ was the set threshold for significance. The ratios of study group percent prevalence are also reported.

Discussed key microbial taxa for HCG were both significantly differentially abundant or prevalent between study groups by the set threshold and were present in at least three patients of one study group. For KOs, statistically significant features were tallied by their respective KEGG[53] pathways using R package KEGGREST version 1.26.1. The importance of KEGG pathways for discussion was then determined through ranking them by their total number of significant features. Detailed statistics from all studied datasets and time windows are provided for both the abundance and prevalence of significant microbial taxa and KOs [see **Additional file 4**]. This same procedure was applied to discussed vertically transmitted microbial taxa that were significantly differentially abundant by delivery mode; detailed statistics are provided [see **Additional file 5**].

Change point analysis:

Individual microbial taxa were selected for change point analysis if they significantly varied in abundance by PMA as determined from the above mixed-effect linear models. Change point analysis was conducted on the mean percent abundances of the microbial taxon over PMA, and separately for each study group, in case the microbial taxon had a distinct pattern over time by the clinical variable of interest. Only the patients for which the microbial taxon was present in at least one faecal sample (>0 sequence counts) were used for determining change points, as membership of individual microbial taxa was expectedly variant between patients, and thus the number of all zeroes would obscure the patterns of mean microbial taxon abundances over time. A data point for each PMA week from 24-36 was required for this analysis, and missing points were imputed by R package imputeTS version 3.0. Change points were evaluated using the pruned exact linear time (PELT) algorithm with a non-parametric cost function based on the empirical distribution of the data using R package changepoint.np[54] version 1.01, with the number of quartiles set to four times the log length of the time series. Change points were evaluated for a range of penalty values between a minimum of the log length of the time series and a maximum of ten times that value. The ideal penalty value (and thus number of change points) was selected from the diagnostic plot as outlined by Lavielle[55], i.e., the value at which the difference in test statistic (cost of penalty) has the largest improvement, or the elbow of the plot. This elbow was automatically detected through determination of the value that had the largest distance from a straight line between the minimum and maximum plot points. The change points were then tallied by PMA week for each examined taxonomic level (phylum, family, genus, and species), which revealed the time point at which the most changes in microbial taxon abundances were occurring.

Random forest classification:

To determine the relative importance of clinical factors versus faecal microbiome features in predicting infant HCG trajectories, the random forest classification approach was applied. The R package caret version 6.0.86 was utilized to build random forest classifiers with 500 trees, and the number of variables randomly sampled as candidates at each split (argument *mtry*) tuned from a random search using 10-fold adaptive cross-validation and 3 repeats. For evaluating feature importance, statistical significance was estimated for the decrease in Gini coefficient from permuting the response variable by R package pRF version 1.2. Features were then ranked by their permutation importance[56], or the number of permutations yielding a lower importance than observed out of 1001. The key time windows of 24-30 weeks PMA and 31-36 weeks PMA were each separately considered for building random forest classifiers to predict the binary outcome of AHCGT versus any SHCGT.

Classifiers were first built using the faecal microbial taxonomic abundance data at the studied individual taxonomic levels on a per faecal sample basis with indexing to control against samples from the same patient being included in both the training and testing subsets. Microbial taxa that had a statistically significant ($p < 0.05$) feature importance, were present in at least three patients' faecal samples and were classified to the respective taxonomic level examined were subsequently selected to be included in the

final multi-taxonomic level random forest classifiers. In order to be combined with the clinical data, it was necessary to reduce the microbial taxonomic abundance data to one value per patient, and thus the median abundance across each of the patient's faecal samples for the respective time windows was taken as the value. Clinical data included mode of delivery, gestational age at birth, sex, birthweight, birth head circumference, morbidities (BPD, SBI, SROP, NEC, seizures, sepsis, presence of any morbidity, presence of two or more morbidities and number of morbidities), enteral feeding (total in mL/kg, formula in mL/kg, human milk in mL/kg, percentage of human milk, and days of TPN over time window of interest), and antibiotics (all, penicillins, gentamicin, vancomycin, cephalosporins, antifungals, clindamycin and metronidazole in number of days and longest number of consecutive days over time windows of interest). Faecal microbial taxonomic prevalence data were also added to the multi-taxonomic level random forest classifiers in the same fashion as the faecal microbial taxonomic abundance data, except it was not necessary to reduce these data to one value as the values were already in the format of microbial taxon presence versus absence per patient over the respective time windows.

After the multi-taxonomic level random forest classifiers were built, a further feature reduction step took place to remove redundant variables. Features were ranked by importance as previously described, and this information was used to select if morbidities were better indicated individually or by one of the summary variables, if enteral feeding-type was better indicated in mL/kg or by percentage, if enteral feeding withdrawal was better indicated by days of TPN or total enteral feeds in mL/kg, and if antibiotics were better indicated as a total or by separate classes and by total days or by longest number of consecutive days. Further, faecal microbial features were reduced by determining if a given microbial taxon was better indicated by prevalence or median abundance, and if a given higher taxonomic level was a better indicator than its lower taxonomies. The final random forest classifiers accuracies are described by the subtraction of the out of bag error from 100, and features are displayed by their permutation importance. Details of the random forest classifiers are provided [see **Additional file 6**].

Moderation analysis:

To assess the impact of faecal microbial taxon abundances and clinical factors on the association between delivery mode and infant HCG trajectories, moderation analysis was conducted. Cumulative link mixed regression models were built using R package ordinal version 2019.12.10, with infant HCG trajectory (AHCGT -> Mildly SHCGT -> Moderately SHCGT -> Severely SHCGT) as the outcome, PMA, delivery mode, the microbial taxon abundance/clinical factor and the interaction between delivery mode and the microbial taxon abundance/clinical factor as fixed effects, and patient as a random effect. Data were standardized using R package standardize version 0.2.2 prior to fitting the models, and models were fitted with the adaptive Gauss-Hermite quadrature approximation with 10 quadrature points. Significance of the delivery mode and microbial taxon abundance/clinical factor interaction was assessed by the Wald statistic implemented in the ordinal package. McFadden's R^2 was calculated manually through dividing the log likelihood of the model by the log likelihood of the null model and

subtracting this value from one. The null model was built through the same method as above except with removal of all fixed effects.

The Benjamini-Hochberg method was used to correct for multiple testing, and a false-positive rate of <1% was the set threshold for significance. Additionally, only the microbial taxa that met the prevalence threshold for assessing the effects of delivery mode on microbial taxon abundances as described in the statistical analysis section were considered. As it was of specific interest to determine which microbial taxa moderated the positive impact of vaginal delivery on infant HCG trajectories, Cohen's D effect sizes between vaginally delivered infants with AHCGT versus vaginally delivered infants with each of the three SHCGT groupings were calculated using R package rstatix version 0.7.0 without assuming equal variances, and microbial taxa that were consistently augmented or diminished in abundance across the three groupings were reported. Detailed statistics are provided for the significant microbial taxon abundances [see **Additional file 7**].

The clinical factors examined included gestational age at birth, birthweight, birth head circumference, sex, total days of all antibiotics, longest number of consecutive days of all antibiotics, total amount of enteral feeds, total days of TPN, day of life full enteral feeds achieved, total amount of human milk, total amount of formula and number of morbidities. Individual antibiotics or morbidities were not examined for this analysis, as some were too rare in incidence to be properly assessed. A standard $p < 0.05$ was the set threshold for significance. Cohen's D effect sizes between infants with AHCGT versus any SHCGT were calculated separately for each delivery mode using R package rstatix without assuming equal variances. Detailed statistics are provided for the significant clinical factors [see **Additional file 8**].

Results

The gut microbiome differentiates suboptimal head circumference growth trajectories from appropriate head circumference growth trajectories in infants specifically after 30 weeks PMA, the time point of head circumference growth trajectory divergence:

It was not previously known if gut microbiome succession could be linked to HCG, the earliest marker of human neurodevelopment. This study not only determined that β -diversity of the gut microbiome significantly differentiated infants with SHCGT from AHCGT, but also that the key time point when most faecal microbial taxa exhibited a shift in mean abundance of 30 weeks PMA matched the onset of significant SHCGT. Preterm infant HCG trajectory was assessed as the difference in HCZ from birth to 36 weeks PMA by the Fenton growth curve[14,42], and study groups were stratified by 0.5 interval losses in HCZ: AHCGT (≥ 0.5), mildly SHCGT ($< 0.5-1$), moderately SHCGT ($< 1-1.5$) and severely SHCGT (< 1.5). Alterations in α -diversity between study groups were first considered using ANOVA on multivariate regression with PMA as a covariate and patient as a random effect, but no significant differences were found [see **Additional file 9**]. Next, β -diversity (genus level) was examined by RDA with study groups and PMA as terms and permutations blocked by patient (**Figure 1a**). A significant RDA model was built ($p = 0.001$), and all terms had a statistically significant effect on gut microbiome composition ($p = 0.001$),

listed from most to least impact (R^2): PMA (0.6391), AHCGT (0.3795), moderately SHCGT (0.2160), severely SHCGT (0.1447) and mildly SHCGT (0.1060).

Defining key time points of gut microbiome succession is critical, as previous research has shown that the inability of the microbiome to achieve a particular state on time (i.e., “mature”) is associated with poor developmental outcomes such as weight gain failure in children[57]. To identify successional time points within the infant dataset, a novel exploitation of change point analysis was employed. Change point analysis presents a distinct advantage over previous time series methods applied to microbiome succession[24,27,57], as it is embedded in a statistical framework, does not require *a priori* knowledge of important microbial taxa or time ranges, and unlike machine learning strategies does not need a separate dataset for model training. Change point analysis has been successfully utilized in a human microbiome study to link the abundance of individual microbial taxa to a clinical state over time[58], but has been more widely used on the macroecological scale for detecting whole community changes[59,60]. Therefore, successional time points of the gut microbiome can be identified as the maxima of the number of change points in microbial taxon abundances similar to the macroecological approach, given relevant parametrization to account for the complexity and multi-community nature of the microbiome as described in the methods section. After application of this approach, 30 weeks PMA was determined to be the key time point of gut microbiome succession within our studied time range, regardless of taxonomic level (**Figure 1b**).

To observe how this key time point correlated with HCG patterns, the loss in HCZ from birth for each PMA week was plotted for each study group (**Figure 1c**). ANOVA with post-hoc analysis revealed that the severely SHCGT group significantly separated from the AHCGT group starting at 28 weeks PMA, but significantly differentiating all three SHCGT groups from AHCGT could not be done until 31 weeks PMA [see **Additional file 10**]. Additionally, all three SHCGT groups could not be significantly distinguished from each other until the measurement end point of 36 weeks PMA. Together, these results indicate that SHCGT could generally be treated as a loss in HCZ of greater than 0.5 that occurred from 31-36 weeks PMA. Therefore, further analysis of the infant gut microbiome was conducted on the time windows of 24-30 weeks PMA (before SHCGT) and 31-36 weeks PMA (during SHCGT), which is also concordant with the gut microbiome successional pattern.

Infants with suboptimal head circumference growth trajectories have a depleted abundance of faecal Bacteroidota and Lachnospiraceae from 31-36 weeks PMA:

Having identified the critical time point of microbiome change relevant to head circumference growth, we next investigated which microbial taxa are putatively important for infant neurodevelopment.

Bacteroidota and *Lachnospiraceae* were specifically identified as biomarkers of AHCGT. Further, it was determined that their significant reduction in abundance occurred during the time period of HCZ loss (31-36 weeks PMA), solidifying a key time point for interventional strategies. Finally, through predictive metagenomic profiling[39,53], the depletion of these taxa was found to result in a reduced carbohydrate utilization capacity of the gut microbiome, suggesting a mechanistic link in regards to energy resource

utilization and short-chain fatty acid (SCFA) production[5,34]. The differences in microbial taxon and KO abundances between study groups were assessed by ANOVA on multivariate regression with patient as a random effect during the time windows of 24-30 weeks PMA (before SHCGT) and 31-36 weeks PMA (during SHCGT). The number of patients with microbial taxon or KO presence versus absence for each study group was statistically compared by the Fisher's exact test. Study group comparisons also included infants with any SHCGT versus AHCGT, and infants with moderately SHCGT to severely SHCGT versus AHCGT to mildly SHCGT. However, the latter yielded no statistically significant differences in abundance or prevalence of microbial taxa, and thus will not be discussed further.

From 24-30 weeks PMA (before SHCGT), few statistically significant differences in faecal microbial taxa were observed between study groups, and these taxa were relatively more abundant in the gut microbiome of infants with SHCGT. The relative abundance of *Firmicutes* was significantly higher ($p=0.009$, $R^2=0.10$) in infants with any SHCGT versus AHCGT (**Figure 2a**). No specific sub-taxon from within *Firmicutes* could account for the observed significant difference. *Actinobacteriota* was also significantly differentially abundant between study groups ($p=0.02$, $R^2=0.15$) from 24-30 weeks PMA, mainly attributed to its relatively high abundance in the mildly SHCGT group (**Figure 2b**), specifically the abundances of *Mycobacteriaceae* ($p=0.0009$, $R^2=0.20$) and its genus *Corynebacterium* ($p=0.001$, $R^2=0.19$).

From 31-36 weeks PMA (during SHCGT), several faecal microbial taxa were significantly less abundant or prevalent in infants with any SHCGT compared to AHCGT. The bacterial phylum *Bacteroidota* was found to be significantly differentially abundant ($p=0.0009$, $R^2=0.11$) between infants with any SHCGT and AHCGT (**Figure 2c**), in addition to its family *Bacteroidaceae* ($p=0.002$, $R^2=0.087$) and its genus *Bacteroides_B* ($p=0.009$, $R^2=0.087$). The bacterial family *Lachnospiraceae* was also found to be both significantly differentially abundant ($p=0.004$, $R^2=0.095$) and prevalent ($p=0.009$) between infants with any SHCGT and AHCGT (**Figure 2d**). No singular genera from within *Lachnospiraceae* could account for the observed significant differences. The prevalence of the bacterial family *Ruminococcaceae* was significantly different between study groups ($p=0.007$) during this timeframe, which could be attributed to the species *Faecalibacterium prausnitzii* that was significantly less prevalent ($p=0.004$) in infants with any SHCGT versus AHCGT (8% vs. 48%). Finally, two families from the bacterial phylum *Firmicutes_C* were additionally significantly less prevalent in infants with any SHCGT versus AHCGT: *Megasphaeraceae* ($p=0.004$; 0% vs. 25%) and *Negativicoccaceae* ($p=0.009$; 0% vs. 21%). Main genera present in the dataset from these bacterial families included *Megasphaera* and *Anaeroglobus*, and *Negativicoccus*, respectively.

To examine differences in function associated with these taxonomic differences, predictive metagenomic profiling was performed by Tax4Fun2[39]. Significant KOs ($p<0.05$; FP<1%) were tallied by their KEGG pathway classifications to identify where the most changes in function were occurring. Out of the major categories, the greatest number of significant differences in both KO abundance and prevalence occurred in "Metabolism", and the majority of significant changes in metabolism were related to "Carbohydrate

metabolism" (**Table 1**). The highest number of these KOs significantly depleted in the gut microbiome of infants with any SHCGT compared to AHCGT belonged to the pathway "Starch and sucrose metabolism", and the second highest number belonged to the pathway "Amino sugar and nucleotide sugar metabolism" (**Table 1**). All KOs that were significantly less abundant or prevalent were only noted from 31-36 weeks PMA. The changes in starch and sucrose metabolism could be attributed to the loss of representatives from the phylum *Bacteroidota*, whereas most changes in amino sugar and nucleotide sugar metabolism could be related to the loss of *Lachnospiraceae* (with some overlap of *Bacteroidota*).

Faecal microbiome features are better predictors of infant head circumference growth trajectories than clinical factors:

Clinical factors could potentially confound the relationships between the infant faecal microbiome and HCG trajectories, and thus required investigation to probe the robustness of this study's findings. Both statistical testing and machine learning modelling were used to determine the independent effect of a given clinical factor on infant HCG trajectories, and the relative importance of clinical factors compared to faecal microbiome features in predicting infant HCG trajectories, respectively. Separate random forest classifiers were built for the key time windows of 24-30 weeks PMA (**Figure 3a**) and 31-36 weeks PMA (**Figure 3b**), and in each case faecal microbiome features out ranked the majority of clinical factors in importance for binary classification of infants into AHCGT versus any SHCGT with 77.5% and 84% accuracy, respectively. From 24-30 weeks PMA, the top patient demographic factor was gestational age at birth (ranked #13), antibiotic factor was total days of all antibiotics (ranked #16), and enteral feeding factor was percentage of enteral feeds as human milk (ranked #17), which were all in the 2/3rds ranked section of important features. From 31-36 weeks PMA, the top patient demographic factor was again gestational age at birth (ranked #29), antibiotic factors were total days of metronidazole (ranked #20), total days of erythromycin (ranked #24) and consecutive days of clindamycin (ranked #26), and enteral feeding factor was total days of TPN (ranked #30), which were all in the 2/3rds ranked section of important features. Further, no statistically significant differences in patient demographics (**Table 2**), or clinical care including antibiotics and enteral feeding were found between study groups [see **Additional file 1**]. Critically, these results indicate that changes in infant HCG trajectories were not due to caloric restriction.

The random forest classifiers identified different sub-taxa from within the found key microbial taxa for infant HCG as important predictors across the two time windows. From 24-30 weeks PMA, the top microbial taxa were the abundances of proteolytic *Bacteroidota* (*Alistipes onderdonkii*, *Alistipes putredinis* and family *Marinifilaceae* comprised of *Odoribacter* spp. and *Butyricimonas* spp.), with remaining important sub-taxa including the abundance of *Bacteroides_B* (phylum *Bacteroidota*), the presence of an unclassified *Coproccoccus* species and abundance of *Blautia wexlerae* (family *Lachnospiraceae*), the presence of *Actinobacteriota* and the abundance of *Firmicutes*. Other unique microbial taxa identified amongst the top third of important features were the abundances of *Veillonella dispar* and an unclassified *Klebsiella* species, and the presence of the bacterial family *Peptoniphilaceae*. From 31-36 weeks PMA, the top microbial taxon was the presence of the genus *Faecalibacterium*, with

remaining important sub-taxa including the abundance of *Bacteroides vulgatus* and presence of family *Tannerellaceae* (phylum *Bacteroidota*), the presence of families *Lachnospiraceae* and *Megasphaeraceae*, and the presence of *Negativicoccus succinicivorans* (family *Negativicoccaceae*). Other unique microbial taxa identified amongst the top third of important features were the abundances of bacterial family *Enterococcaceae* and *Peptostreptococcus anaerobius*, and the presences of bacterial family *Actinobacteraceae* and *Gemmiger formicilis*.

However, it was found that infants with any SHCGT experienced a significantly greater incidence of specific morbidities in the NICU (**Table 2**). These results contrasted the random forest classifiers, as the model built from 24-30 weeks PMA had ranked the incidence of any morbidity amongst the bottom three predictors, and the model built from 31-36 weeks PMA had its highest ranked morbidities amongst the 2/3rds ranked features including incidence of seizures (ranked #16) and SROP (ranked #25). Part of the discrepancy may lie with the incidence of morbidities being correlated with the structure of the infant microbiome, as prior studies have associated specific morbidities, such as NEC and sepsis, with the infant faecal microbiota[61]. Evaluating differences between the infant faecal microbiome amongst patients with and without specific morbidities is beyond the scope of this study; it was instead important to ensure that morbidity did not significantly alter our found relationships between infant faecal microbial taxa and HCG trajectories. A robust permutation approach was utilized to select and rank features for the random forest classifiers[56], which more frequently identified microbiome characteristics over morbidities as important for predicting infant HCG trajectories. As an additional precaution, a limited morbidity (LM) subset was created to determine the dependence of the faecal microbial taxon abundance associations with HCG trajectories on this confounder. This LM subset comprised of infants without SBI, SROP, NEC, seizures, or sepsis. Not only were the previously established gut microbiome and infant HCG trajectory relationships retained after consideration of the clinical variables, but *Actinobacteriota* abundance newly emerged as a biomarker distinguishing moderately to severely SHCGT from mildly SHCGT and AHCGT.

RDA analysis confirmed the significant difference ($p=0.001$) in gut microbiome β -diversity between study groups of the LM subset for both the model and terms, including PMA ($R^2=0.6829$) and any SHCGT ($R^2=0.3706$). For infants with any SHCGT versus AHCGT, significant reductions of *Bacteroidota* (**Figure 2c**) and *Lachnospiraceae* (**Figure 2d**) abundance were also confirmed in the LM subset overall ($p=0.008$, $R^2=0.086$ and $p=0.005$, $R^2=0.085$ respectively). No sub-taxon from within *Bacteroidota* and *Lachnospiraceae* was significantly differentially abundant between study groups [see **Additional file 11**]. Furthermore, predicted metagenomic profiling again revealed that the greatest number of changes in function occurred in “Metabolism”, specifically “Carbohydrate metabolism” and its subcategories “Starch and sucrose metabolism” and “Amino sugar and nucleotide sugar” metabolism [see **Additional file 12**].

A finding unique to the LM subset as compared to the complete dataset was that infants with moderately to severely SHCGT had a significant depletion in overall *Actinobacteriota* abundance ($p=0.008$, $R^2=0.086$) compared to infants with AHCGT or mildly SHCGT. Again, no specific sub-taxon could account for this significant difference [see **Additional file 11**].

Clinical characteristics were statistically analogous between the complete dataset and the LM subset [see **Additional file 2**]. However, a lower proportion of infants with any SHCGT in the complete dataset were vaginally delivered (**Table 2**), and this difference became statistically significant ($p=0.02$) in the LM subset [see **Additional file 2**]. Interestingly, vaginal delivery was also identified as a highly important predictor for the random forest classifiers across both time windows, ranking #5 from 24-30 weeks PMA and as the top feature from 31-36 weeks PMA. We hypothesized that the positive influence of vaginal delivery on infant HCG trajectories resulted from the vertical transmission of beneficial microbes[62], and thus investigated the interaction of this clinical factor with the infant microbiome in further detail.

The impact of delivery mode on infant head circumference growth trajectories reveals dispersal limitation and supersedes habitat filtering as a driver of optimal infant microbiome succession:

Vaginal delivery was significantly positively associated with infant HCG trajectories, and we hypothesized that this resulted from the vertical transmission[62] of microbes that elicit AHCGT. To address this hypothesis, it was first evaluated if the effect of delivery mode on infant HCG trajectories was significantly moderated by the abundances of faecal microbial taxa. Such a result would indicate dependence of vaginal delivery's benefit to infant HCG trajectories on the structure of the microbiome, and indeed significant moderation by the abundances of faecal microbial taxa previously significantly associated with AHCGT was found, including *Lachnospiraceae*, *Bacteroides_B*, *Faecalibacterium* and *Megasphaera*. Next, it was investigated if the abundances of faecal microbial taxa were significantly associated with delivery mode to determine which microbes were more likely to be vertical transmitted. It was discovered that the abundance of *Bacteroidota* was significantly enhanced by vaginal delivery, which was one of the found key microbial taxa for infant HCG. Taken together, these results affirm our hypothesis that delivery mode can be considered as a dispersal limitation factor influencing optimal infant microbiome succession, which impacts infant health and development. Understanding the relative importance of successional drivers is vital for identifying potential indirect routes of modifying the gut microbiome-brain axis, or variables that could alter the effectiveness of direct interventions. Despite clinical variables typically attributed as habitat filtering factors impacting microbiome succession, such as enteral feeding and antibiotics, not having a significant direct effect on infant HCG trajectories, it was interrogated if these clinical factors could significantly moderate the influence of vaginal delivery. That was indeed found to be the case, therefore delineating that optimal infant microbiome succession is primarily driven by dispersal limitation and secondarily by habitat filtering, at least in the context of host neurodevelopmental markers.

To test the influence of the microbiome on the relationship between delivery mode and infant HCG trajectories, moderation analysis was conducted by building cumulative link mixed regression models of the four categories of infant HCG trajectories (AHCGT -> Mildly SHCGT -> Moderately SHCGT -> Severely SHCGT) for evaluating the significance of the interaction between delivery mode and the abundance of a given faecal microbial taxon through the Wald statistic, with PMA, and delivery mode and the given faecal microbial taxon individually as fixed effects, plus patient as a random effect. Significant moderation ($p<0.05$; FP<1%) by faecal microbial taxa that were consistently augmented or depleted in

abundance (as determined by Cohen's D effect sizes) for vaginally delivered infants with AHCGT versus all three SHCGT categories are displayed in **Table 3**. Vaginally delivered infants with any SHCGT compared to vaginally infants with AHCGT had reductions in faecal microbial taxa previously found to be directly significantly associated with infant HCG trajectories from both standard statistics and machine learning modelling, including bacterial family *Bacteroidaceae* and its genus *Bacteroides_B* (also its species *Bacteroides dorei* and *Bacteroides vulgatus*), bacterial family *Lachnospiraceae* (also its genus *Eubacterium_E* and *Hungatella effluvii*), *Faecalibacterium prausnitzii* and bacterial genus *Megasphaera* (from family *Megasphaeraceae*). Additional diminished faecal microbial taxa that were selected as predictors of infant HCG trajectories for the random forest classifiers only included the bacterial family *Acutalibacteriaceae* and *Alistipes putredinis*. Faecal microbial taxa that were decreased in abundance in vaginally delivered infants specifically across all SHCGT categories compared to AHCGT included *Dialister invisus*, the bacterial genera *Erysipelatoclostridium* and *Cloacibacterium*, and the bacterial family *Mycobacteriaceae*, whereas the bacterial genus *Coprobacillus* was the only faecal microbial taxon found to be increased specifically in vaginally delivered infants across all SHCGT categories versus AHCGT. The large number of significant faecal bacterial modifiers, with many of these having a significant direct impact on infant HCG trajectories, indicate that the positive effect of vaginal delivery on infant HCG trajectories is dependent on the microbiome (**Figure 3c**).

To statistically test the hypothesis of vertical transmission, the differences in faecal microbial taxon abundances between vaginally delivered versus Caesarean-section delivered infants were assessed by ANOVA on multivariate regression with PMA as a fixed effect and patient as a random effect. *Bacteroidota* ($p=0.01$, $R^2=0.08$), a microbial taxon directly significantly associated with infant HCG trajectories by both standard statistics and machine learning modelling, was found to be significantly increased in abundance in vaginally delivered infants compared to Caesarean-section delivered infants. This result also included its genus *Bacteroides_B* ($p=0.009$, $R^2=0.08$), as well as its family *Tannerellaceae* ($p=0.008$, $R^2=0.04$), genus *Parabacteroides* ($p=0.007$, $R^2=0.05$) and species *Parabacteroides distasonis* ($p=0.009$, $R^2=0.06$). An additional faecal microbial taxon significantly elevated in abundance in vaginally versus Caesarean-section delivered infants but not found to be directly related to infant HCG trajectories was the bacterial family *Methanobacteraceae* ($p=0.005$, $R^2=0.03$) and its genus *Methanobrevibacter_A* ($p=0.009$, $R^2=0.03$). Therefore, *Bacteroidota* is likely vertically transmitted in vaginally delivered infants, which can be considered as a mediator of the positive influence of vaginal delivery on infant HCG trajectories due to both its significant direct effects on infant HCG trajectories and its significance as a moderator of the relationship between delivery mode and infant HCG trajectories (**Figure 3c**). Other faecal microbial taxa that colonize the infant faecal microbiome less specifically via vertical transmission additionally served as moderators, as previously described.

Finally, to test how clinical factors change infant HCG trajectory outcomes of vaginally delivered versus Caesarean-section delivered infants, moderation analysis equivalent to the approach used for the faecal microbial taxon abundances as above was performed. Time was found to be an important moderator of delivery mode and infant HCG trajectory relationships as demonstrated by the significance of gestational

age at birth ($p < 2 \times 10^{-16}$, $R^2 = 0.11$), and its proxies' birthweight ($p < 2 \times 10^{-16}$, $R^2 = 0.11$) and birth head circumference ($p = 0.002$, $R^2 = 0.08$). The number of morbidities was expectedly also a significant modifier ($p = 3 \times 10^{-11}$, $R^2 = 0.13$), and sex was not ($p = 0.7$, $R^2 = 0.02$). Both time and morbidities were found to have a significant direct effect on infant HCG trajectories, and thus these results further evidenced our previous interrogation of critical time points (change point analysis) and morbidity as a confounder (LM subset). Intriguingly, several clinical factors known to affect the infant microbiome were found to be significant modifiers of the effect of delivery mode on infant HCG trajectories, and these had at least a five-fold greater impact on vaginally delivered infants, whose Cohen's D effect sizes (AHCGT/any SHCGT) were all large, compared to Caesarean-section delivered infants, whose Cohen's D effect sizes were all small to negligible (**Figure 3d**). These clinical factors included both the total days ($p = 2 \times 10^{-5}$, $R^2 = 0.18$) and longest number of consecutive days ($p = 4 \times 10^{-8}$, $R^2 = 0.13$) of all antibiotics, and the total amount of enteral feeds ($p < 2 \times 10^{-16}$, $R^2 = 0.14$), along with the additional proxies of total days of TPN ($p = 2 \times 10^{-6}$, $R^2 = 0.14$) and the day of life full enteral feeding was achieved ($p = 0.002$, $R^2 = 0.12$). As for enteral feeding-type, the total amount of human milk was a significant moderator ($p < 2 \times 10^{-16}$, $R^2 = 0.07$), whereas the total amount of formula was not ($p = 0.1$, $R^2 = 0.04$). Notably, none of these clinical factors had a significant direct effect on infant HCG trajectories. As diet and antibiotics are known to be primary influencers of the intestinal environment, these results demonstrate that delivery mode, i.e., a dispersal limitation factor from vertical transmission, supersedes habitat filtering as a driver of infant gut microbiome succession.

Discussion

There is a vital need to identify modifiable environmental factors for reducing the incidence of developmental impairments at a time point early enough for successful intervention. Thus, it was the aim of this study to determine if infant gut microbiome composition was associated with HCG, the earliest marker of neurodevelopment[13,14]. This study is the first to show that β -diversity of the gut microbiome was significantly distinct between infants with AHCGT versus any SHCGT, and that reduced abundances of *Bacteroidota* and *Lachnospiraceae* were specifically associated with SHCGT independent of concurrent morbidities and caloric restriction. Notably, this study's novel application of change point analysis further linked microbiome and HCG by revealing that the timing of peak gut microbiome composition alteration exactly matched the timing of significant HCG separation between study groups at 30 weeks PMA. Clinical factors were additionally thoroughly examined to determine their influence on the association between infant HCG trajectory and the gut microbiome, of which the significant effect of delivery mode provided further innovative insights into the primary drivers of optimal infant microbiome succession.

There exists evidence for the potential mechanistic impact of *Bacteroidota* and *Lachnospiraceae* on host neurodevelopment. *Bacteroidota* may affect neurodevelopment through altering intestinal barrier integrity and systemic metabolite availability. Hsiao and colleagues[63] have demonstrated that oral treatment of the maternal immune activation mouse model of ASD with human-derived *Bacteroides fragilis* both altered intestinal tight junction protein expression and the serum metabolite profile, and ameliorated

behavioural defects. *Lachnospiraceae* are key producers of SCFAs after fermentation, particularly butyrate[64,65], that are utilized by intestinal epithelial cells as a fuel source[66], have potent anti-inflammatory effects[67], and promote immune system maturation by increasing the colonic population of T regulatory cells[68–71]. Stolp and colleagues[72] have shown that neonatal systemic inflammation in rats can lead to altered blood-brain barrier permeability and behaviour, and therefore, *Lachnospiraceae* could affect neurodevelopment through influencing energy resources and immunity.

Vaginal delivery was the clinical factor significantly associated with improved HCG but was dependent upon the successful vertical transmission[62] of key microbial taxa [see **Additional file 5**], which could be impeded by certain clinical factors, including prolonged antibiotics, delayed enteral feeding, formula feeding and younger gestational age at birth leading to SHCGT even with vaginal delivery [see **Additional file 8**]. The significance of gestational age connects to the change point analysis demonstrating 30 weeks PMA to be a key transition point for the preterm infant gut microbiome, which complements other studies[24,27]. This transition is thought to occur due to intestinal maturation[73], at which point the preterm infant intestinal environment is like the term infant at birth. Studies in term infants using different neurodevelopmental metrics and timelines have also indicated *Bacteroidota* and *Lachnospiraceae* as potentially important taxa; Kelsey et al.[74] found functional brain network connectivity was associated with *Bacteroides* and *Lachnospiraceae* abundance, and Loughman and colleagues[75] associated behavioural problems at 24 months of age with a low *Prevotella* abundance at 12 months. Therefore, the results from this study taken in the context of the current literature interestingly suggest that *Bacteroidota* and *Lachnospiraceae* should broadly colonize infants born >30 weeks gestational age at birth through vaginal delivery.

Diet and antibiotics are known to have a large impact on the infant microbiome, exceeding stochasticity[33,62,76]. Further, a seminal study by Feng and colleagues has demonstrated that the influence of relative microbial fitness on infant microbiome succession appears to supersede historical contingency, i.e., the order of introduction of microbes[77]. These results have together indicated that habitat filtering, i.e., variables that shape the intestinal environment, is a dominant driver of infant microbiome succession. Other studies have additionally shown that delivery mode shapes the initial colonizers of the infant gastrointestinal tract[62,76], which includes faecal microbes for vaginally delivered infants, such as *Bacteroidota*[78,79] complementing our study, and skin microbes for Caesarean-section infants. However, infants delivered at term possess an advantage as they can access the home environment usually relatively quickly, whereas preterm infants usually remain in the hospital environment at length with sanitation protocols preventing normal microbial dissemination[80]. This difference allows the effect of dispersal limitation on infant microbiome succession to be more closely examined in preterm infants, and a major finding of this study is that its influence exceeds habitat filtering. That could explain why studies on enteral feeding[23,76] and antibiotics[76,81,82] for their health benefits and effects on microbiome composition have yielded somewhat inconsistent results for preterm infants, despite their known effects on the term infant microbiome.

However, this study's findings do not preclude the importance of habitat filtering as a driver, since after considering dispersal limitation, significance for antibiotics, caloric restriction and human milk feeding was found. Human milk provides microbially fermentative human milk oligosaccharides[83], which are known to be degraded by *Bacteroidota*, *Lachnospiraceae* and *Bifidobacterium*[83,84] and additionally supported by our predicted functional profiling analysis, as well as notably acting as a source of microbes and thus a secondary dispersal limitation factor[85]. Further, the maternal microbiome can also exhibit dysbiosis, as the faecal and vaginal microbiome composition of mothers that deliver preterm has exhibited significant differences compared to mothers that deliver at term in other studies[86,87]. The effects of delivery mode on preterm infant health outcomes and microbiome succession can thus also be masked depending on a given study's cohort distribution[76]. Therefore, it is recommend that clinical variables should be considered sequentially for infant microbiome studies, first by dispersal limitation factors (e.g., delivery mode, environment such as hospital versus home or urban versus rural, probiotics) then by habitat filtering factors (e.g., diet, prebiotics, antibiotics).

Conclusions

Bacteroidota and *Lachnospiraceae* are both known core taxa of the adult gut microbiome[88], and our work provides strong evidence that *Bacteroidota* and *Lachnospiraceae* need to integrate into the gut microbiota early in infancy for optimal developmental outcomes. Ultimately, the promising findings from this study encourage future research of microbiome modification to improve infant developmental trajectories, either from clinical investigations to verify microbial functions or metabolites, or animal model experiments for direct probing of the impact of *Bacteroidota* and *Lachnospiraceae* on neurodevelopment. Optimizing the gut microbiome in infancy could also reduce the incidence of developmental disability, as evidenced by the associations between early antibiotics and later cognitive outcomes[28,29] or the altered gut microbiome of children with ASD[35] and ADHD[36], and future work could continue to follow the faecal microbiome of infants with AHCGT and SHCGT over time to determine the potential life-long effectiveness of an infant microbiome based intervention.

List Of Abbreviations

HCG Head circumference growth

ADHD Attention deficit hyperactivity disorder

ASD Autism spectrum disorder

NICU Neonatal intensive care unit

SHCGT Suboptimal head circumference growth trajectory

PMA Postmenstrual age

MIND Microbiome In Neonatal Development study

HCZ Head circumference z-score

TPN Total parenteral nutrition

BPD Bronchopulmonary dysplasia

SBI Severe brain injury

SROP Severe retinopathy of prematurity

NEC Necrotizing enterocolitis

ROP Retinopathy of prematurity

ASV Amplicon sequence variant

BLAST Basic local alignment search tool

CLR Centre-log ratio

AHCGT Appropriate head circumference growth trajectory

LM Limited morbidity

RDA Redundancy analysis

KO KEGG orthology

SCFA Short-chain fatty acid

Declarations

Ethics approval and consent to participate

The MIND study received approval from an institutional review board (IRB16-1431) in accordance with the human subjects' research policies of the U.S. Food & Drug Administration and the Declaration of Helsinki. Patients were enrolled after receiving written informed consent from the parent.

Consent for publication

Not applicable.

Availability of data and material

The datasets generated and/or analysed during the current study are available in the NCBI SRA repository, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA739139/>.

Competing interests

Bree Andrews, MD/MPH, is an equity partner in the B-Corporation PreeMe+You (PMY), a technology start-up whose goal is to provide bedside technology to parents of NICU patients that ameliorates health disparities. There is no current income from the company or royalties and her research on the microbiome and neurodevelopment is not related to her work with PMY. There were no overlapping patients that were studied in the current manuscript who also interfaced with the technology from PMY. Dinanath Sulakhe is a co-founder and equity partner in Navipoint Genomics, LLC (NG), a genomics start-up that was incubated at the University of Chicago, whose goal is to provide a cloud-based genomic data analysis platform. NG currently has an active SBIR Phase-1 grant and Mr. Sulakhe is currently providing consulting services for this grant. Mr. Sulakhe is also a co-founder and equity partner in Navipoint Health, Inc (NH), a genetic testing start-up. There is no current income or royalties from NH. The research presented in the paper is not relevant to NG and NH, or his work at NG and NH. It will not have any impact in the form of financial gain or loss to NG and NH. The remaining authors declare no competing interests.

Funding

We would like to acknowledge the National Institutes of Health UG3 OD023281 to EC and BA and R01 HD083481 to EC for providing funding. The funders did not contribute to the design of the study, writing, and the collection, analysis, or interpretation of data.

Authors' contributions

KO conducted the classification, post-processing, and all analyses of the 16S rRNA gene sequencing data, generated all figures and tables, and primarily wrote the manuscript. MA and RY consented patients, and collected, prepared, and sent infant faecal samples for 16S rRNA gene sequencing; MA additionally contributed to the Methods section of the manuscript. MD'S and DS collected and prepared clinical data for analysis from the University of Chicago Comer Children's Hospital's electronic medical records system. PDH and AZW were the acting physicians that recruited patients and provided expert medical consultation for this manuscript. BX processed the 16S rRNA gene sequencing data and deposited it into the NCBI SRA repository. MEM provided expert consultation in neurodevelopmental assessment for this clinical study and manuscript. EC and BA are the grant holders that designed the clinical study, obtained IRB approval, and oversaw data interpretation and editing of the manuscript. All authors critically reviewed and approved of the final manuscript.

Acknowledgements

The authors would like to thank the Digestive Disease Research Core Center at the University of Chicago (P30DK42086) for its support on this study.

References

1. Sayre RK, Devercelli AE, Neuman MJ, Wodon Q. Investing in Early Childhood Development: Review of the World Bank's Recent Experience [Internet]. The World Bank; 2015 [cited 2021 Mar 26]. Available from: <https://elibrary.worldbank.org/doi/abs/10.1596/978-1-4648-0403-8>
2. Ertem IO, Organization WH. Developmental difficulties in early childhood: prevention, early identification, assessment and intervention in low- and middle-income countries: a review. Developmental difficulties in early childhood: prevention, early identification, assessment and intervention in low- and middle-income countries: a review [Internet]. 2012 [cited 2021 Mar 26]; Available from: <https://apps.who.int/iris/handle/10665/97942>
3. Rice D, Barone S. Critical periods of vulnerability for the developing nervous system: evidence from humans and animal models. Environmental Health Perspectives. Environmental Health Perspectives; 2000;108:511–33.
4. Lenroot RK, Giedd JN. Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. Neuroscience & Biobehavioral Reviews. 2006;30:718–29.
5. Oliphant K, Lu J. Chapter 6 - Neurodevelopment and the gut microbiome. In: Claud EC, editor. The Developing Microbiome [Internet]. Academic Press; 2020 [cited 2020 Jul 20]. p. 115–43. Available from: <http://www.sciencedirect.com/science/article/pii/B9780128206027000064>
6. CDC. Facts About Developmental Disabilities | CDC [Internet]. Centers for Disease Control and Prevention. 2019 [cited 2021 Mar 26]. Available from: <https://www.cdc.gov/ncbddd/developmentaldisabilities/facts.html>
7. Olusanya BO, Davis AC, Wertlieb D, Boo N-Y, Nair MKC, Halpern R, et al. Developmental disabilities among children younger than 5 years in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Global Health. 2018;6:e1100–21.
8. You D, Hug L, Ejdemyr S, Idele P, Hogan D, Mathers C, et al. Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. Lancet. 2015;386:2275–86.
9. Lyaal K, Croen L, Daniels J, Fallin MD, Ladd-Acosta C, Lee BK, et al. The Changing Epidemiology of Autism Spectrum Disorders. Annu Rev Public Health. 2017;38:81–102.
10. Danielson ML, Bitsko RH, Ghandour RM, Holbrook JR, Kogan MD, Blumberg SJ. Prevalence of Parent-Reported ADHD Diagnosis and Associated Treatment Among U.S. Children and Adolescents, 2016. J Clin Child Adolesc Psychol. 2018;47:199–212.

11. Raghuram K, Yang J, Church PT, Cieslak Z, Synnes A, Mukerji A, et al. Head Growth Trajectory and Neurodevelopmental Outcomes in Preterm Neonates. *Pediatrics*. 2017;140.
12. Kuban KCK, Allred EN, O'Shea TM, Paneth N, Westra S, Miller C, et al. Developmental Correlates of Head Circumference at Birth and Two Years in a Cohort of Extremely Low Gestational Age Newborns. *J Pediatr*. 2009;155:344-9.e1-3.
13. Neubauer V, Fuchs T, Griesmaier E, Kager K, Pupp-Peglow U, Kiechl-Kohlendorfer U. Poor postdischarge head growth is related to a 10% lower intelligence quotient in very preterm infants at the chronological age of five years. *Acta Paediatrica*. 2016;105:501–7.
14. Cordova EG, Cherkerzian S, Bell K, Joung KE, Collins CT, Makrides M, et al. Association of Poor Postnatal Growth with Neurodevelopmental Impairment in Infancy and Childhood: Comparing the Fetus and the Healthy Preterm Infant References. *The Journal of Pediatrics* [Internet]. 2020 [cited 2020 Jul 13]; Available from: <http://www.sciencedirect.com/science/article/pii/S0022347620307034>
15. Hack M, Breslau N, Weissman B, Aram D, Klein N, Borawski E. Effect of very low birth weight and subnormal head size on cognitive abilities at school age. *N Engl J Med*. 1991;325:231–7.
16. Brants C, van Tienoven TP, Rayyan M, Allegaert K, Raaijmakers A. Earlier achievement of full enteral feeding in extremely low birth weight neonates is not associated with growth improvement in the first 2 years of life. *Eur J Pediatr*. 2018;177:1247–54.
17. Travers CP, Wang T, Salas AA, Schofield E, Dills M, Laney D, et al. Higher or Usual Volume Feedings in Very Preterm Infants: A Randomized Clinical Trial. *The Journal of Pediatrics* [Internet]. 2020 [cited 2020 Jul 13]; Available from: <http://www.sciencedirect.com/science/article/pii/S0022347620306247>
18. Cormack BE, Jiang Y, Harding JE, Crowther CA, Bloomfield FH. Relationships between Neonatal Nutrition and Growth to 36 Weeks' Corrected Age in ELBW Babies–Secondary Cohort Analysis from the Provide Trial. *Nutrients*. 2020;12:760.
19. Mariani E, Biasini A, Marvulli L, Martini S, Aceti A, Faldella G, et al. Strategies of Increased Protein Intake in ELBW Infants Fed by Human Milk Lead to Long Term Benefits. *Front Public Health*. 2018;6:272.
20. Darlow BA, Graham PJ, Rojas-Reyes MX. Vitamin A supplementation to prevent mortality and short- and long-term morbidity in very low birth weight infants. *Cochrane Database Syst Rev* [Internet]. 2016 [cited 2020 Jul 20];2016. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7038719/>
21. Salas AA, Woodfin T, Phillips V, Peralta-Carcelen M, Carlo WA, Ambalavanan N. Dose-response Effects of Early Vitamin D Supplementation on Neurodevelopmental and Respiratory Outcomes of Extremely Preterm Infants at 2 years of age: A Randomized Trial. *Neonatology*. 2018;113:256–62.
22. Torsvik IK, Ueland PM, Markestad T, Midttun Ø, Monsen A-LB. Motor development related to duration of exclusive breastfeeding, B vitamin status and B12 supplementation in infants with a birth weight

- between 2000-3000 g, results from a randomized intervention trial. *BMC Pediatr* [Internet]. 2015 [cited 2020 Jul 20];15. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4683944/>
23. Miller J, Tonkin E, Damarell RA, McPhee AJ, Suganuma M, Suganuma H, et al. A Systematic Review and Meta-Analysis of Human Milk Feeding and Morbidity in Very Low Birth Weight Infants. *Nutrients* [Internet]. 2018 [cited 2020 Apr 8];10. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6024377/>
24. La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM, et al. Patterned progression of bacterial populations in the premature infant gut. *Proc Natl Acad Sci U S A*. 2014;111:12522–7.
25. Korpela K, Blakstad EW, Moltu SJ, Strømmen K, Nakstad B, Rønnestad AE, et al. Intestinal microbiota development and gestational age in preterm neonates. *Sci Rep* [Internet]. 2018 [cited 2020 Jul 20];8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5802739/>
26. Tauchi H, Yahagi K, Yamauchi T, Hara T, Yamaoka R, Tsukuda N, et al. Gut microbiota development of preterm infants hospitalised in intensive care units. *Benef Microbes*. 2019;10:641–51.
27. Grier A, Qiu X, Bandyopadhyay S, Holden-Wiltse J, Kessler HA, Gill AL, et al. Impact of prematurity and nutrition on the developing gut microbiome and preterm infant growth. *Microbiome* [Internet]. 2017 [cited 2020 Jul 20];5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5725645/>
28. Slykerman RF, Thompson J, Waldie KE, Murphy R, Wall C, Mitchell EA. Antibiotics in the first year of life and subsequent neurocognitive outcomes. *Acta Paediatrica*. 2017;106:87–94.
29. Slykerman RF, Coomarasamy C, Wickens K, Thompson JMD, Stanley TV, Barthow C, et al. Exposure to antibiotics in the first 24 months of life and neurocognitive outcomes at 11 years of age. *Psychopharmacology*. 2019;236:1573–82.
30. Gur TL, Palkar AV, Rajasekera T, Allen J, Niraula A, Godbout J, et al. Prenatal stress disrupts social behavior, cortical neurobiology and commensal microbes in adult male offspring. *Behavioural Brain Research*. 2019;359:886–94.
31. Farshim P, Walton G, Chakrabarti B, Givens I, Saddy D, Kitchen I, et al. Maternal Weaning Modulates Emotional Behavior and Regulates the Gut-Brain Axis. *Scientific Reports*. Nature Publishing Group; 2016;6:21958.
32. Cerdó T, Diéguez E, Campoy C. Early nutrition and gut microbiome: interrelationship between bacterial metabolism, immune system, brain structure, and neurodevelopment. *American Journal of Physiology-Endocrinology and Metabolism*. American Physiological Society; 2019;317:E617–30.
33. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*. Nature Publishing Group;

2018;562:583–8.

34. Lu J, Claud EC. Connection between gut microbiome and brain development in preterm infants. *Dev Psychobiol.* 2019;61:739–51.
35. Pulikkan J, Mazumder A, Grace T. Role of the Gut Microbiome in Autism Spectrum Disorders. *Adv Exp Med Biol.* 2019;1118:253–69.
36. Dam SA, Mostert JC, Szopinska-Tokov JW, Bloemendaal M, Amato M, Arias-Vasquez A. The Role of the Gut-Brain Axis in Attention-Deficit/Hyperactivity Disorder. *Gastroenterol Clin North Am.* 2019;48:407–31.
37. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems.* 2016;1.
38. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 2012;6:1621–4.
39. Wemheuer F, Taylor JA, Daniel R, Johnston E, Meinicke P, Thomas T, et al. Tax4Fun2: prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. *Environmental Microbiome.* 2020;15:11.
40. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller A-B, Narwal R, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *The Lancet.* 2012;379:2162–72.
41. Bourke J, Wong K, Srinivasjois R, Pereira G, Shepherd CCJ, White SW, et al. Predicting Long-Term Survival Without Major Disability for Infants Born Preterm. *The Journal of Pediatrics.* 2019;215:90-97.e1.
42. Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatrics.* 2013;13:59.
43. Jobe AH, Bancalari E. Bronchopulmonary dysplasia. *Am J Respir Crit Care Med.* 2001;163:1723–9.
44. Schmidt B, Asztalos EV, Roberts RS, Robertson CMT, Sauve RS, Whitfield MF, et al. Impact of bronchopulmonary dysplasia, brain injury, and severe retinopathy on the outcome of extremely low-birth-weight infants at 18 months: results from the trial of indomethacin prophylaxis in preterms. *JAMA.* 2003;289:1124–9.
45. Martin CR, Dammann O, Allred EN, Patel S, O'Shea TM, Kuban KCK, et al. Neurodevelopment of extremely preterm infants who had necrotizing enterocolitis with or without late bacteremia. *J Pediatr.* 2010;157:751-756.e1.

46. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*. Nature Publishing Group; 2019;37:852–7.
47. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
48. Murali A, Bhargava A, Wright ES. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome*. 2018;6:140.
49. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*. 2018;36:996–1004.
50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
51. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*. 2016;7:1451–6.
52. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*. 2014;2:15.
53. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
54. Killick R, Eckley IA. changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software*. 2014;58:1–19.
55. Lavielle M. Using penalized contrasts for the change-point problem. *Signal Processing*. 2005;85:1501–10.
56. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics*. 2010;26:1340–7.
57. Subramanian S, Huq S, Yatsunenkov T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature*. 2014;510:417–21.
58. Layeghifard M, Li H, Wang PW, Donaldson SL, Coburn B, Clark ST, et al. Microbiome networks and change-point analysis reveal key community changes associated with cystic fibrosis pulmonary exacerbations. *NPJ Biofilms Microbiomes* [Internet]. 2019 [cited 2021 Apr 7];5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6341074/>

59. Kortsch S, Primicerio R, Beuchel F, Renaud PE, Rodrigues J, Lønne OJ, et al. Climate-driven regime shifts in Arctic marine benthos. *PNAS. National Academy of Sciences*; 2012;109:14052–7.
60. Samhouri JF, Andrews KS, Fay G, Harvey CJ, Hazen EL, Hennessey SM, et al. Defining ecosystem thresholds for human activities and environmental pressures in the California Current. *Ecosphere*. 2017;8:e01860.
61. Masi AC, Stewart CJ. The role of the preterm intestinal microbiome in sepsis and necrotising enterocolitis. *Early Hum Dev*. 2019;138:104854.
62. Wang S, Egan M, Ryan CA, Boyaval P, Dempsey EM, Ross RP, et al. A good start in life is important—perinatal factors dictate early microbiota development and longer term maturation. *FEMS Microbiol Rev. Oxford Academic*; 2020;44:763–81.
63. Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, et al. Microbiota Modulate Behavioral and Physiological Abnormalities Associated with Neurodevelopmental Disorders. *Cell*. 2013;155:1451–63.
64. Louis P, Young P, Holtrop G, Flint HJ. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environmental Microbiology*. 2010;12:304–14.
65. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalog established by metagenomic sequencing. *Nature*. 2010;464:59–65.
66. den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud D-J, Bakker BM. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of Lipid Research*. 2013;54:2325–40.
67. Steinmeyer S, Lee K, Jayaraman A, Alaniz RC. Microbiota metabolite regulation of host immune homeostasis: a mechanistic missing link. *Curr Allergy Asthma Rep*. 2015;15:24.
68. Furusawa Y, Obata Y, Fukuda S, Endo TA, Nakato G, Takahashi D, et al. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature*. 2013;504:446–50.
69. Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-Y M, et al. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science*. 2013;341:569–73.
70. Arpaia N, Campbell C, Fan X, Dikiy S, van der Veeken J, deRoos P, et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature*. 2013;504:451–5.
71. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, et al. Induction of Colonic Regulatory T Cells by Indigenous Clostridium Species. *Science*. 2011;331:337–41.

72. Stolp HB, Johansson PA, Habgood MD, Dziegielewska KM, Saunders NR, Ek CJ. Effects of neonatal systemic inflammation on blood-brain barrier permeability and behaviour in juvenile and adult rats. *Cardiovasc Psychiatry Neurol*. 2011;2011:469046.
73. Heida FH, Beydüz G, Bulthuis MLC, Kooi EMW, Bos AF, Timmer A, et al. Paneth cells in the developing gut: when do they arise and when are they immune competent? *Pediatric Research*. Nature Publishing Group; 2016;80:306–10.
74. Kelsey CM, Prescott S, McCulloch JA, Trinchieri G, Valladares TL, Dreisbach C, et al. Gut microbiota composition is associated with newborn functional brain connectivity and behavioral temperament. *Brain, Behavior, and Immunity* [Internet]. 2020 [cited 2020 Nov 26]; Available from: <http://www.sciencedirect.com/science/article/pii/S0889159120323771>
75. Loughman A, Ponsonby A-L, O'Hely M, Symeonides C, Collier F, Tang MLK, et al. Gut microbiota composition during infancy and subsequent behavioural outcomes. *EBioMedicine*. 2020;52:102640.
76. Aguilar-Lopez M, Dinsmoor AM, Ho TTB, Donovan SM. A systematic review of the factors influencing microbial colonization of the preterm infant gut. *Gut Microbes*. 13:1–33.
77. Feng L, Raman AS, Hibberd MC, Cheng J, Griffin NW, Peng Y, et al. Identifying determinants of bacterial fitness in a model of human gut microbial succession. *Proc Natl Acad Sci U S A*. 2020;117:2622–33.
78. Shao Y, Forster SC, Tsiliki E, Vervier K, Strang A, Simpson N, et al. Stunted microbiota and opportunistic pathogen colonisation in caesarean section birth. *Nature*. 2019;574:117–21.
79. Gregory KE, LaPlante RD, Shan G, Kumar DV, Gregas M. Mode of Birth Influences Preterm Infant Intestinal Colonization With *Bacteroides* Over the Early Neonatal Period. *Adv Neonatal Care*. 2015;15:386–93.
80. Combellick JL, Shin H, Shin D, Cai Y, Hagan H, Lacher C, et al. Differences in the fecal microbiota of neonates born at home or in the hospital. *Sci Rep* [Internet]. 2018 [cited 2021 Jun 12];8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6199260/>
81. Ruoss JL, Bazaclicu C, Russell JT, Cruz D de la, Li N, Gurka MJ, et al. Routine Early Antibiotic Use in Symptomatic Preterm Neonates: A Pilot Randomized Controlled Trial. *The Journal of Pediatrics*. 2021;229:294-298.e3.
82. Kim CS, Grady N, Derrick M, Yu Y, Oliphant K, Lu J, et al. Effect of Antibiotic Use Within First 48 Hours of Life on the Preterm Infant Microbiome: A Randomized Clinical Trial. *JAMA Pediatr*. 2020;
83. Walsh C, Lane JA, van Sinderen D, Hickey RM. Human milk oligosaccharides: Shaping the infant gut microbiota and supporting health. *J Funct Foods*. 2020;72:104074.

84. Pichler MJ, Yamada C, Shuoker B, Alvarez-Silva C, Gotoh A, Leth ML, et al. Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways. Nat Commun [Internet]. 2020 [cited 2020 Dec 10];11. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7335108/>
85. Stinson LF, Sindi ASM, Cheema AS, Lai CT, Mühlhäusler BS, Wlodek ME, et al. The human milk microbiome: who, what, when, where, why, and how? Nutrition Reviews [Internet]. 2020 [cited 2020 Dec 10]; Available from: <https://doi.org/10.1093/nutrit/nuaa029>
86. Shiozaki A, Yoneda S, Yoneda N, Yonezawa R, Matsubayashi T, Seo G, et al. Intestinal microbiota is different in women with preterm birth: results from terminal restriction fragment length polymorphism analysis. PLoS One. 2014;9:e111374.
87. Hyman RW, Fukushima M, Jiang H, Fung E, Rand L, Johnson B, et al. Diversity of the vaginal microbiome correlates with preterm birth. Reprod Sci. 2014;21:32–40.
88. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science. American Association for the Advancement of Science; 2016;352:560–4.

Tables

Table 1. *Metabolic KEGG pathways containing the most significantly differentially abundant KOs by infant head circumference growth.*

KEGG Pathway	AHCGT vs. SHCGT
1. CARBOHYDRATE METABOLISM	16
1.1 Starch and sucrose metabolism:	8
<i>glycogen synthase [EC:2.4.1.11]</i>	Overall: $p=0.0006$, $R^2=0.10$; $p=ns$, 75% vs. 60% 31-36: $p=0.00007$, $R^2=0.14$; $p=0.004$, 72% vs. 28%
<i>glucan 1,4-α-glucosidase [EC:3.2.1.3]</i>	Overall: $p=0.001$, $R^2=0.10$; $p=ns$, 86% vs. 77% 31-36: $p=0.0003$, $R^2=0.14$; $p=ns$, 80% vs. 44%
<i>glycogen phosphorylase/synthase [EC:2.4.1.1 2.4.1.11]</i>	Overall: $p=0.0004$, $R^2=0.10$; $p=ns$, 75% vs. 63% 31-36: $p=0.0001$, $R^2=0.13$; $p=0.004$, 72% vs. 28%
<i>α-amylase [EC:3.2.1.1]</i>	Overall: $p=0.009$, $R^2=0.07$; $p=ns$, 79% vs. 67% 31-36: $p=0.005$, $R^2=0.09$; $p=0.009$, 80% vs. 40%
<i>hexokinase [EC:2.7.1.1]</i>	Overall: $p=0.003$, $R^2=0.07$; $p=ns$, 64% vs. 37% 31-36: $p=0.002$, $R^2=0.09$; $p=0.007$, 56% vs. 16%
1.2 Amino sugar and nucleotide sugar metabolism:	5
<i>fucokinase [EC:2.7.1.52]</i>	Overall: $p=0.001$, $R^2=0.09$; $p=ns$, 75% vs. 57% 31-36: $p=0.0005$, $R^2=0.12$; $p=0.002$, 72% vs. 24%
<i>N-acetylglucosamine 2-epimerase [EC:5.1.3.8]</i>	Overall: $p=0.004$, $R^2=0.08$; $p=ns$, 79% vs. 67% 31-36: $p=0.003$, $R^2=0.10$; $p=0.009$, 80% vs. 40%
<i>N-acetyl-α-D-muramate 1-phosphate uridylyltransferase [EC:2.7.7.99]</i>	Overall: $p=0.005$, $R^2=0.04$; $p=ns$, 100% vs. 100%

	31-36: $p=0.002$, $R^2=0.11$; $p=ns$, 100% vs. 92%
<i>mannose-1-phosphate guanylyltransferase [EC:2.7.7.13]</i>	Overall: $p=0.01$, $R^2=0.06$; $p=ns$, 100% vs. 100%
	31-36: $p=ns$, $R^2=0.04$; $p=ns$, 100% vs. 100%
<u>1.3 Fructose and mannose metabolism:</u>	3
<u>1.4 Butanoate metabolism:</u>	3
2. GLYCAN BIOSYNTHESIS AND METABOLISM	11
3. METABOLISM OF COFACTORS AND VITAMINS	9
4. LIPID METABOLISM	6
4. ENERGY METABOLISM	6
7. AMINO ACID METABOLISM	5
8. BIOSYNTHESIS OF OTHER SECONDARY METABOLITES	4
9. NUCLEOTIDE METABOLISM	3
9. METABOLISM OF TERPENOID AND POLYKETIDES	3

Legend: Study groups defined by difference in head circumference z-score from birth to 36 weeks postmenstrual age (PMA) as calculated by the Fenton growth curve: appropriate head circumference growth trajectory (≥ 0.5 ; AHCGT) and suboptimal head circumference growth trajectory (< 0.5 ; SHCGT). For KEGG pathway classifications, the total number of significantly differentially abundant or prevalent KEGG database orthologies (KOs) is indicated with at least 3 being requisite for listing. Significance for abundance was evaluated by ANOVA on multivariate regression with infant head circumference growth trajectory and PMA as fixed effects and patient as a random effect, and for prevalence by the Fisher's exact test ($p < 0.05$; $FP < 1\%$). The p values and R^2 values (abundance); p values and percent prevalence per study group (prevalence) are reported for the KOs that are more abundant or prevalent amongst infants with AHCGT. Significance was found both overall and during the 31-36 completed weeks PMA time window. Abbreviations: ns = non-significant.

Table 2. *Clinical characteristics of the MIND infant cohort at the University of Chicago Comer Children's Hospital.*

Demographics	AHCGT	Mildly SHCGT	Moderately SHCGT	Severely SHCGT	<i>p</i> value
Mode of delivery, VD	42.9% (12)	12.5% (2)	12.5% (1)	16.7% (1)	0.1
EGA at birth, completed weeks	28.3 ± 2.6	27.1 ± 2.2	27.0 ± 3.4	26.2 ± 3.1	0.3
Sex, male	46.4% (13)	56.3% (9)	25.0% (2)	66.7% (4)	0.4
BW, kg	1.02 ± 0.38	0.98 ± 0.33	1.07 ± 0.55	0.96 ± 0.53	0.9
Birth HC, cm	24.9 ± 2.9	24.7 ± 2.8	24.6 ± 4.1	24.3 ± 3.8	0.9
Outcomes	AHCGT	Mildly SHCGT	Moderately SHCGT	Severely SHCGT	<i>p</i> value
BPD	53.6% (15)	56.3% (9)	62.5% (5)	83.3% (5)	0.6
NEC	0.0% (0)	0.0% (0)	0.0% (0)	50.0% (3)	0.0006
SBI	7.1% (2)	0.0% (0)	12.5% (1)	50.0% (3)	0.01
Seizures	7.1% (2)	12.5% (2)	12.5% (1)	16.7% (1)	0.7
Sepsis	0.0% (0)	0.0% (0)	25.0% (2)	0.0% (0)	0.03
SRP	3.6% (1)	12.5% (2)	25.0% (2)	16.7% (1)	0.2
Any listed morbidity	60.7% (17)	56.3% (9)	62.5% (5)	83.3% (5)	0.8
2+ listed morbidities	10.7% (3)	18.8% (3)	25.0% (2)	66.7% (4)	0.03
Length of NICU stay, days	77.9 ± 34.8	83.7 ± 40.6	126.8 ± 90.3	124.2 ± 52.3	0.2
PMA at discharge, completed weeks	39.0 ± 3.8	38.5 ± 4.3	44.5 ± 10.0	43.5 ± 5.0	0.1

Legend: Head circumference growth groups are defined in legend for **Figure 1**. Binary variables are reported as the percentage of patients (number of patients) per group, with *p* values calculated by the Fisher's exact test. Numerical variables are reported as the mean ± standard deviation per group, with *p* values calculated by Welch's ANOVA. Abbreviations: VD = Vaginal delivery; EGA = Estimated gestational age; BW = Birthweight; HC = Head circumference; BPD = Bronchopulmonary dysplasia; NEC = Necrotizing enterocolitis; SBI = Severe brain injury; SRP = Severe retinopathy of prematurity; NICU = Neonatal intensive care unit; PMA = Postmenstrual age.

Table 3. *Infant head circumference growth associated faecal microbial taxa that significantly moderated vaginal delivery effects.*

Faecal microbial taxon	AHCGT/ Mildly SHCGT	AHCGT/ Moderately SHCGT	AHCGT/ Severely SHCGT	<i>p</i> value	McFadden's R ²
Acutalibacteraceae	0.859	0.563	0.629	<2x10 ⁻¹⁶	0.03
Bacteroidaceae	1.28	0.576	0.778	<2x10 ⁻¹⁶	0.02
<i>Bacteroides_B</i>	1.06	0.803	1.08	<2x10 ⁻¹⁶	0.03
<i>dorei</i>	0.403	0.0359	0.506	<2x10 ⁻¹⁶	0.02
<i>vulgatus</i>	0.880	0.849	0.893	<2x10 ⁻¹⁶	0.02
Dialisteraceae	ns				
<i>Dialister</i>	ns				
<i>invisus</i>	0.496	0.398	0.437	<2x10 ⁻¹⁶	0.02
Erysipelatoclostridiaceae	ns				
<i>Coprobacillus</i>	-0.282	-0.889	-0.899	<2x10 ⁻¹⁶	0.02
<i>Erysipelatoclostridium</i>	0.686	0.750	0.796	<2x10 ⁻¹⁶	0.03
Lachnospiraceae	1.28	0.309	0.478	<2x10 ⁻¹⁶	0.03
<i>Eubacterium_E</i>	0.453	0.426	0.568	<2x10 ⁻¹⁶	0.02
<i>Hungatella</i>	ns				
<i>effluvii</i>	0.489	0.0441	0.463	<2x10 ⁻¹⁶	0.03
Megasphaeraceae	ns				
<i>Megasphaera</i>	0.380	0.494	0.557	<2x10 ⁻¹⁶	0.03
Mycobacteriaceae	0.756	0.588	0.634	<2x10 ⁻¹⁶	0.02
Rikenellaceae	ns				
<i>Alistipes</i>	ns				
<i>putredinis</i>	0.251	0.172	0.390	<2x10 ⁻¹⁶	0.02
Ruminococcaceae	ns				
<i>Faecalibacterium</i>	0.335	0.677	0.599	<2x10 ⁻¹⁶	0.03
<i>prausnitzii</i>	0.718	0.573	0.731	<2x10 ⁻¹⁶	0.03

Weeksellaceae	ns				
<i>Cloacibacterium</i>	0.112	0.0749	0.0586	3×10^{-11}	0.02

Legend: Significance of moderation was evaluated by the Wald statistic ($p < 0.05$; FP < 1%) and McFadden’s R^2 on cumulative link mixed regression of infant head circumference growth trajectories for the interaction of delivery mode and abundance of a given faecal microbial taxon, with PMA, and delivery mode and the given faecal microbial taxon abundance individually as fixed effects, plus patient as a random effect. The Cohen’s D effect size of the centre-log ratio transformed microbial taxon abundances is indicated for each pairwise head circumference growth group comparison specifically for vaginally delivered infants. Head circumference growth groups are defined in legend for **Figure 1**. Abbreviations: ns = non-significant.

Figures

Figure 1

Changes in infant head circumference growth and gut microbiome β -diversity over completed weeks postmenstrual age. Head circumference growth group is defined by the difference in head circumference z-score (HCZ) from birth to 36 weeks postmenstrual age as calculated by the Fenton growth curve: appropriate head circumference growth trajectory (≥ 0.5 ; AHCGT), mildly suboptimal head circumference growth trajectory ($< 0.5-1$; mildly SHCGT), moderately suboptimal head circumference growth trajectory ($< 1-1.5$; moderately SHCGT) and severely suboptimal head circumference growth trajectory (< 1.5 ; severely SHCGT). (a) Redundancy analysis (RDA) of 16S rRNA gene sequencing data generated from faecal samples collected weekly. Samples are coloured by head circumference growth group, with the centroids for each group indicated by crosshairs. (b) Number of change points, as identified by non-parametric analysis, in mean percent abundance pattern of each microbial taxon by head circumference growth group for each completed postmenstrual age week. Data is shown for all examined taxonomic levels from phylum to species. (c) Difference in HCZ from birth as calculated by the Fenton growth curve for each completed postmenstrual age week by head circumference growth group. Box-plot centre line, median; limits, first and third quartiles; whiskers, 1.5x interquartile range; points, outliers. * $p < 0.05$, Welch's ANOVA.

Figure 2

Mean percent abundance of faecal microbial taxa that significantly differ by infant head circumference growth. Head circumference growth groups are defined in legend for Figure 1. Left panel displays mean percent abundance pattern over completed weeks postmenstrual age for the limited morbidity subset. Middle and right panels display mean percent abundance with standard deviation bars for both the

complete and limited morbidity datasets, respectively. * $p < 0.05$; $FP < 1\%$, ANOVA on multivariate regression.

(a) Firmicutes. (b) Actinobacteriota. (c) Bacteroidota. (d) Lachnospiraceae.

Figure 3

Influence of clinical factors on faecal microbiome and infant head circumference growth relationships. Random forest classifiers were built for predicting appropriate head circumference growth (HCG) trajectory (AHCGT) versus any suboptimal HCG trajectory (SHCGT) for infants as defined in legend for Figure 1, at the distinct key time points of 24-30 completed weeks postmenstrual age (PMA) (a) and 31-36 completed weeks PMA (b). The relative importance of features was ranked by permutation importance, or the number of permutations yielding lower importance than observed out of 1001. Faecal microbiome features (purple, with shading by bacterial phylum) out ranked most clinical factors, including antibiotics (red), birth (i.e., patient demographic) factors (blue), enteral feeding (green) and morbidity (orange). The exception to this rule was delivery mode, which was examined further by moderation analysis (c + d). Vaginal delivery (VD) significantly (solid black) increased the abundance of faecal Bacteroidota (mean and standard deviation percent abundance indicated), and the abundance of faecal Bacteroidota was both significantly directly associated with infant HCG trajectories and significantly moderated the effect of delivery mode on infant HCG trajectories (c). The abundances of other faecal microbial taxa (see Table 3) were also both significantly directly associated with infant HCG trajectories and significantly moderated the effect of delivery mode on infant HCG trajectories but were not significantly (dashed grey) increased in abundance by VD. Several clinical factors significantly

moderated the effect of delivery mode on infant HCG trajectories (d); these clinical factors impacted more specifically VD infants and not Caesarean-section (C/S) delivered infants as indicated by the large differences in Cohen's D effect sizes (AHC GT/any SHCGT) by delivery mode. That would explain why a significant direct effect of these clinical factors on infant HCG trajectories was mostly not observed. Abbreviations: EGA = Estimated gestational age; BPD = Bronchopulmonary dysplasia; NEC = Necrotizing enterocolitis; SBI = Severe brain injury; SRP = Severe retinopathy of prematurity.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xlsx](#)
- [Additionalfile10.xlsx](#)
- [Additionalfile11.xlsx](#)
- [Additionalfile12.xlsx](#)
- [Additionalfile2.xlsx](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile6.xlsx](#)
- [Additionalfile7.xlsx](#)
- [Additionalfile8.xlsx](#)
- [Additionalfile9.xlsx](#)