THE UNIVERSITY OF CHICAGO

DETERMINANTS OF THE EARLY LIFE MICROBIOME

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM: MICROBIOLOGY

BY
ALYSON LANGE YEE

CHICAGO, ILLINOIS
AUGUST 2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to start by thanking Jack Gilbert, who was an inspiring advisor and always ready to seek out a new collaboration or deliver a pep talk. My research relied heavily on large clinical data sets and I am indebted to our tireless collaborators at the University of South Florida (Maureen Groer, Thao Ho, Samia Dutra, Larry Dishaw, Bradley Kane, Elizabeth Miller, and their research nurses, Judy Zaritt and Marcia Kneusel) and Singapore (Jia Xu, Gerard Wong, Neerja Karnani). I am also thankful for the entire Gilbert lab, who were supportive and would always go out of their way to help me, especially the incomparable Neil Gottel for spearheading the processing and sequencing of all my samples. I am deeply appreciative of my thesis committee, Eugene Chang, Sean Crosson, Sarah Cobey, and Erika Claud, for their valuable insight and encouragement along the way.

I was lucky to benefit from the resources and institutional knowledge of the administrators and fellow students in the Medical Scientist Training Program, the Committee on Microbiology, and the Department of Surgery. This research was funded by the National Institutes of Health award T32GM007281.

ABSTRACT

The genomes of microorganisms that reside in and on the human body vastly outnumber human genes, implying that the human microbiome may be important for precision medicine. Humans have been canonically considered sterile *in utero*, with the majority of colonization occurring during, and immediately following, birth. Factors that have been shown to affect the ecology and succession of the microbiome include host genetic factors, maternal transmission, and environmental exposures. Furthermore, the timing of microbial exposure in relation to immune system development may have lifelong consequences.

In this thesis, I investigated the early life microbiome through two lenses: that of microbiome assembly and that of human genotype. I first analyzed a longitudinal data set of preterm infants, who underwent abnormal and premature microbial exposure, characterized by delayed maturation and higher abundance of potential opportunistic pathogens. I identified associations between microbiome dynamics and infant growth outcomes during their first weeks of life, implying that there may be a link between the microbiota and preterm infant development. I also used human fucosyltransferase 2 (FUT2) genotype to probe the interplay between host genetics and the microbiome in a cross-sectional cohort of post-weaning toddlers. FUT2 secretor status plays an important role in formation of immune complexes, secretion of antigens in body fluid, and susceptibility to pathogenic bacteria. I found differences in microbial community composition and structure between children with different FUT2 genotypes. Bacteria from the same genera were differentially abundant in FUT2 secretors and FUT2 non-secretors, providing evidence for niche differentiation. As laid out in this thesis, both events surrounding birth and individualized human genotypes seem to associate with the microbiome in early life, indicating that these factors require consideration in future study.

CHAPTER 1

INTRODUCTION

1.1 Definition of the human microbiome

The consortium of microorganisms (bacteria, archaea, viruses, fungi) that populate the human body are collectively referred to as the human microbiota and the complement of genomes the human microbiome. Microbes make up about half the cells in the human body, and the genetic content of the microbiota dwarfs the human genome[1]. For example, the inter-individual variability of the microbiome is much greater than that of the human genome, with different people harboring different microbial communities[2]. This all suggests that a better understanding of the human microbiome may be important for precision medicine.

1.2 Acquisition of the microbiome

Humans have canonically been considered sterile *in utero* and undergo significant colonization during, and immediately following, birth. Some studies have isolated bacteria from the umbilical cord, placenta, and meconium of healthy deliveries, while other studies and the ability to rear germ free animals argue against these findings [3–8]. Regardless of whether live bacteria can cross the placenta, it seems clear that microbial products, or their influence on the maternal immune system, might affect fetal development. However, the majority of microbial exposure begins peripartum, with vertical transmission from the maternal microbiota during passage through the birth canal [4]. Infants born by Caesarean section display altered patterns of colonization, with their gut microbiota more similar to their mothers' skin, compared with vaginally delivered infants, who receive more maternal gut and vaginal microbiota[9,10]. Caesarean delivered infants show an overall lower diversity of microbes. These differences by mode of delivery persist, with microbes associated with Caesarean delivery detectable up to two years following birth[11]. Notably, this effect of delivery mode

has been explored in healthy, term infants, while in preterm infants, the effect of delivery mode is eclipsed by other factors that contribute to abnormal development[12,13].

In the context of vertical transmission, the decreased microbial exposure associated with Westernized diet and hygiene likely compounds over generations. That is, each subsequent generation loses key members of the microbiota, so decreased microbial diversity and its potential immune consequences accumulate in highly developed societies[14]. In a murine model, humanized mice fed a diet low in microbially-accessible carbohydrates showed alterations to the microbiota that were reversible within the same generation. However, over several generations, the low carbohydrate diet caused a progressive loss of diversity that was not reversed with the reintroduction of the dietary carbohydrates[15]. Indeed, a high-fiber maternal diet was demonstrated to be protective against offspring developing allergic airway disease in mice, implying that diet is strongly linked to microbiome function and can have intergenerational effects[16]. Notably, the time frame during pregnancy when maternal diet affects the offsprings immune phenotype remains to be clarified[17]. The effects of maternal diet may be transient, via the gut microbiome and inflammation, or longer lasting, such as by epigenetic programming.

Other sources of early-life exposure, such as environmental influences and sharing a home with other children and/or pets, also appear to be important[14,18]. Bacterial communities in residences with dogs or cats present are significantly richer and more diverse compared to those without pets [18]. Cohabiting adults have been shown to share a greater degree of their microbiota with each other if they own a dog, implying that pets can act as vectors. The presence of older siblings in the home has also been shown to increase a child's gut microbial diversity, and diversity correlates with number of people in the household [19]. The mechanisms by which exposure to other children and/or pets alters microbiome composition remain unclear, but it is generally thought that these would increase a child's contact with environmental microbes, promoting a diverse and healthy gut microbiota[18].

Antibiotic exposure has been shown to perturb the microbiome. Early life courses of

antibiotics correlate with decreased diversity and lifelong higher incidence of allergy and atopy[20]. Exposure to antibiotics, even beginning in the intrapartum time frame, can affect a child's microbiome. Prophylactic antibiotics during delivery have been shown to alter colonization in the infant gut[21]. Taken together, these studies have identified sources of environmental exposure that contribute to differences in microbial composition in early life and may affect the immune system.

1.3 Succession and dynamics of the microbiome over time

After birth, maternal and environmental microbes continue to exert influence over the development of the microbiome, particularly through diet. Breastfeeding versus formula feeding has been shown to influence the child's gut microbiome and provide another route of vertical transmission of maternal microbes[22,23]. Bacteria acquired during birth include bacteria capable of digesting human milk oligosaccharides (HMOs), which are indigestible to humans[24–27]. These HMOs act as prebiotics, leading to selective expansion of those pioneer bacteria that can utilize HMOs as a carbon source[28,29]. The early bacteria prime the development of the microbiota and the immune system. Polymorphisms in the mother's fucosyltransferase gene FUT2 affect the HMO composition in her breast milk and can affect her breastfed infants microbial community structure[30,31].

Preterm birth and host genetics are two factors that have been shown to affect the ecology and succession of the microbiome. Microbiome succession has been well documented in healthy, full-term infants, and shows a patterned progression from birth to an adult-like state around 2.5 years of life[9,32,33]. The microbiota during early life is highly plastic and more susceptible to infection, antibiotics, or dietary change. The adult microbiota is more stable and resilient to such perturbations [34].

The initial colonizers of the infant gut are typically facultative anaerobes, shifting to obligate anaerobes within days or weeks[35]. In preterm infants, the microbiome undergoes a similar pattern of development, but after significant delay[34,36]. The preterm infant

microbiome is characterized by low diversity and low stability, as well as an abundance of opportunistic pathogens[37]. Preterm infants are exposed to high levels of antibiotics and low levels of maternally derived microbes, leaving them particularly susceptible to colonization from an environmental source, such the neonatal intensive care unit (NICU) built environment, which may carry antibiotic resistance genes[37,38]. Furthermore, a recent study of late preterm infants, who are likely to receive similar postnatal care to full-term infants, determined that late preterm birth independently affected gut microbiome development during the first six months of life[39].

Beyond the massive public health burden of preterm birth (1 in 10 US births occurs before 37 weeks' gestation, according to the March of Dimes), children born preterm represent a fascinating ecological model. Several studies have demonstrated that early life microbiome composition and structure is associated with lifelong risk of allergic disease, adiposity, and neurological disorders[40–42]. However, the role of the gut microbiome in long term health, particularly in children born preterm, remains poorly understood.

1.4 Host selection on the microbiome

The composition of the microbiome is highly unique to a particular individual. While over 99.5% of the human genome is identical among individuals, there has been evidence that two humans can share zero common microbial species within their microbiome[2]. Interestingly, it seems that there is no "core" human microbiota in terms of shared taxa, but that there may be a shared core of functional genetic potential within the microbial community. For example, in a study of monozygotic and dizygotic twin pairs, a subset of microbial genes were shared between individuals, while species were not[43]. Vastly different microbial communities yielded similar functional pathways[1]. The idea that identical twins' microbiomes are only slightly more similar to each other than are fraternal twins' suggests that the human genome may play less of a role in selecting for particular taxa but may exert more pressure on the functional potential of the microbial community[44]. Therefore, studies examining

the effect of host genetics on the microbial communities should not just compare taxonomic abundances.

The human fucosyltransferase 2 (FUT2) gene represents an important interface in host-microbe coevolution. It encodes the enzyme responsible for H antigen synthesis, an important step in ABO histo-blood group antigen display in body fluids and on the intestinal mucosa[45,46]. Individuals who are homozygous for a nonfunctional FUT2 allele are called nonsecretors, and they do not present ABO antigens in secretions or on epithelial cells[45]. About 20 % of the population are non-secretors, lacking functional FUT2 due to single nucleotide polymorphisms[45]. The secretor phenotype plays an important role in formation of immune complexes, secretion of antigens in body fluid, and susceptibility to pathogenic bacteria. In mice, fucosyltransferase 2 (FUT2) has been identified as a key gene for host-microbiome interaction[47]. Genome wide association studies have found that nonsecretors have increased susceptibility to chronic conditions, such as Type 1 diabetes and Crohn's disease[48–51]. However, secretor status has been suggested to be protective against infection by both viral and bacterial pathogens, implying that the availability of fucosylated glycans plays a role in host-pathogen interaction[52–54]. Indeed, signals from the microbiota are required for FUT2 expression in mice[55]. Secreted ABO antigens are present in breast milk, and have therefore been investigated as a potential prebiotic that selectively enriches the gut microbiome of the infant for microbes that can metabolize fucosylated sugars[24,27,30,48]. Some *Bifidobacterium* species have been singled out as using human milk oligosaccharides in culture and in mice. Therefore, FUT2 represents a key gene that affects the carbohydrate landscape of the gut, particularly in early life, and alters the composition and function of the resident microbes. The functional potential of the microbial community that varies by FUT2 genotype has not yet been examined.

The effect of secretion genotype on microbial community assembly in infants has been examined previously. For example, secretor infants, who are able to produce fucosylated milk oligosaccharides, had increased relative abundance of *Bifidobacterium*[56], and secretor

mothers had significantly increased microbial community stability in breast milk, which was transferred to the gut of breastfed infants[30]. However, the associations between host FUT2 genotype and the functional metabolic potential of the microbiome remain unclear. This has implications for potential precision medicine interventions targeted to the early life microbiome.

1.5 Interplay between the microbiota and the immune system

Just as the host exerts pressure to shape the resident microbiota, the microbiota interacts with both innate and adaptive immune systems in the intestinal mucosa via multiple mechanisms. Early work in germ free mouse models demonstrated that microbial colonization has a profound effect on the organization of lymphoid tissues[57]. Germ free animals are also deficient in antibody production[58]. Commensal microbes can directly act on macrophages, inducing cytokine production that can act downstream to regulate other immune cell populations or affect the epithelial barrier[59]. Microbial ligands, among them lipopolysaccharide (LPS) from the outer membrane of Gram-negative bacteria, have long been known to have immunomodulatory effects in germ free mice[60]. TLR4 is an important receptor for microbial products, including LPS, and mice lacking TLR4 signaling show increased allergic pathology[43], implying that intestinal microbes can modulate immune responses.

In addition to interacting directly with the immune system, the commensal microbiota also produces small molecules that affect immune system development. Primary metabolites are produced by the catabolic and anabolic reactions and are required for cellular growth and homeostasis[63]. Specialized metabolites, including virulence factors, secondary metabolites, and natural products, are produced by accessory genetic elements that can be swapped between bacteria via horizontal gene transfer, and exert direct effects on host cells and other microbes[63]. Examples of microbial metabolites that interface with the host immune system include short chain fatty acids (SCFAs) and histamine, which act on mucosal dendritic cells via G-protein coupled receptors. SCFAs act as an energy source for coloncytes and can

regulate tight junction organization, promoting barrier function[64]. In a germ free mouse model, adding SCFAs to drinking water increased the abundance of colonic T regulatory cells and was protective against inflammation[65]. Another murine study demonstrated that high fiber diet increased the release of SCFAs, while mice lacking receptors for SCFAs had lower levels of CD103+ dendritic cells[66]. SCFAs can bind 'metabolitesensing' Gproteincoupled receptors, which influence Treg biology, epithelial integrity, gut homeostasis, DC biology, and IgA antibody responses[67]. CD103+ DCs metabolize vitamin A to retinoic acid, which promotes Treg cell differentiation and homing to the gut[67]. SCFAs also inhibit histone deacetylase, which affects gene transcription in many cells and tissues, and can lead to epigenetic modification of DNA[67].

In addition to SCFAs, the microbiota can secrete a range of other metabolites that influence mucosal immune responses, such as long chain fatty acids, histamine, and vitamins. Long chain fatty acids (LCFAs) are generally derived from dietary components and act both as energy sources and in the regulation of immune responses[68]. Germ-free animals exhibit alterations of composition in their lipid metabolites and commensal microbes play a role in LCFA metabolism[69,70].The ω3 and ω6 FAs are essential FAs that mammals cannot produce, and ω3 FAs have been shown to have anti-inflammatory effects[71–73]. Histamine is a biogenic amine that induces smooth muscle and vasodepressor activity during anaphylaxis[64]. Cells of both the innate and adaptive immune systems harbor histamine receptors. Histamine can be produced by a number of different bacterial strains, and these bacteria have been shown to be more abundant in patients with asthma compared to controls[64].Vitamins act as antioxidants, transcription factors, and cofactors for metabolism, so they are indispensable for many biological processes. They are obtained from dietary sources and produced by the commensal microbiota. In addition to directly affecting metabolism, microbial metabolites of some vitamins act as ligands for immune cell signaling. Therefore, the commensal microbiota can induce cytokine production by multiple cell types which have downstream effects on adaptive immune cells and epithelial barrier function, or directly stimulate B and

T cells.

The timing of microbial exposure in relation to immune system development may have lifelong consequences. Microbial exposure early in life begins a dynamic process, activating the immature immune system and leading to selection of particular microbes. At birth, the immune system lacks B and T cells in the intestinal mucosa and the innate immune response to the presence of bacteria is crucial for the transition from fetal to postnatal life[74,75]. Recent evidence suggests that there are time-dependent windows of development that are contingent upon microbial signals to trigger maturation of mucosal immunity. For example, childrens microbiome at 2 months of age was predictive of their interleukin expression levels at 2 years of age[76]. Indeed, children with altered microbiome development show increased incidence of immune-mediated disorders in adulthood[40–42,77]. Furthermore, a classic study in a germ free mice showed that colonization with a normal microbiota shortly after birth rescued the mice from inflammatory and allergic processes, but mice colonized as adults showed the same immune sensitivities as germ free mice[78,79]. Germ free mice develop with an abnormal immune system, including smaller lymphoid structures, and decreased activation of T and B cells[80]. Seeding a germ-free mouse with a microbiota as an adult results in gene transcriptional differences compared with germ free mice that get a microbiota at birth, suggesting that some of these immune cell abnormalities are regulated by the microbiota during the early life time window[81]. Exposure to diverse microorganisms during infancy, when immunoregulatory systems are developing, increases the repertoire of organisms that can be tolerated and may pattern memory immune mechanisms to recognize pathogens[14]. Taken together, this implies that perturbations or abnormal assembly during this early life interval may have lifelong health consequences.

1.6 Conclusions

In this thesis, I will investigate the early life microbiome through two lenses: that of preterm infants, which represent an unusual ecological problem of microbiome assembly, and

that of post-weaning children with respect to the human genotype.

CHAPTER 2

LONGITUDINAL MICROBIOME COMPOSITION AND STABILITY IN VERY LOW BIRTH WEIGHT INFANTS

2.1 Abstract

Compared with full term infants, preterm infants are at greater risk of microbial insult, including reduced exposure to maternal vaginal and enteric microbes, higher rates of formula feeding, invasive procedures, antibiotics, and medications that alter gastrointestinal pH. The microbiome of 83 preterm very-low birth weight (VLBW) infants and clinical covariates were analyzed weekly over the course of their initial neonatal intensive care unit (NICU) stay, with infant growth as the primary clinical outcome. Birth weight significantly correlated with increased rate of weight gain in the first 6 weeks of life, while no significant relationship was observed between rate of weight gain and feeding type. Microbial diversity increased with age and was significantly correlated with weight gain and percentage of mothers own milk. As expected, infants who received antibiotics during their NICU stay had significantly lower alpha diversity than those who did not. Infants who showed improved length over their NICU stay had significantly more volatile microbial beta diversity and a significantly decreased microbial maturity index compared with infants who did not. Of the cohort, 25 were followed into childhood. At 2 and 4 years old, the microbiota of these VLBW infants became similar to the mothers microbiota. The number of microbial taxa shared between the infant, toddler, and mother varied, with least overlap between infants and mothers. Microbial beta diversity was significantly different between NICU stay and all other time points (2-, 4- years-old and mothers when their children were 2 and 4 years old). Overall, there was a significant association between microbial community diversity, structure, and infant weight and length gain in an at-risk childhood population.

2.2 Introduction

Humans have canonically been considered sterile *in utero*, with the majority of microbial colonization occurring during, and immediately following birth. This early microbial exposure begins a dynamic process, activating the immature immune system and leading to the selection of particular microbes. At birth, gut-associated lymphoid tissue is uneducated and the innate immune responses to the presence of bacteria is crucial for the transition from fetal to postnatal life[74,75]. Recent evidence suggests that there are time-dependent windows of development that are contingent upon microbial signals to trigger maturation of mucosal immunity. For example, the microbiome at 2 months of age has been shown to be predictive of a childs interleukin expression levels at 2 years of age[76]. Indeed, children with altered microbiome development show increased incidence of immune-mediated disorders in adulthood[40,41,42,77]. This implies that perturbations or abnormal assembly during this early life interval may have lifelong health consequences.

The ecological succession of the human microbiome has been well documented in healthy, full-term infants, and shows a patterned progression from birth to an adult-like state around 2.5 years postpartum[32,33]. Initial colonizers of the infant gut are typically facultative anaerobes, shifting to obligate anaerobes within days or weeks [35] as the maturing colonocytes start to scrub oxygen from the intestinal lumen [83]. In preterm infants, the microbiome undergoes a similar pattern of development, but after significant delay[34]. The preterm infant microbiome is characterized by low diversity and low stability, as well as an abundance of opportunistic pathogens[37]. Microbiota development in premature infants appear to be associated with gestational age but also shaped by maternal environment and lifestyle, the unique environment and clinical practices in the neonatal intensive care unit (NICU)[84–87]. Preterm infants are exposed to high levels of antibiotics and low levels of maternally derived microbes, leaving them particularly susceptible to colonization from an environmental source, such as the NICU built environment, which may carry antibiotic resistance genes [37,38]. Furthermore, a recent study of late preterm infants, who are likely to receive similar

postnatal care to full-term infants, determined that late preterm birth independently affected gut microbiome development during the first six months of life [39].

The goal of this study was to analyze the gut microbiome during the first 6 weeks after birth for very low birth weight (VLBW) Infants, and to determine the associations between the variance in microbial profile and infant health outcomes. Subsequently 25 of the original 83 children were re-examined at ages 2 and 4, and the stool microbial profile was assessed with health, development and growth (weight and length) outcomes. We hypothesized that microbial community structure in the first 6 weeks of NICU stay is statistically associated with mode of delivery, gestational age, weight and length gain, type of feeding, human milk cytokines, fecal calprotectin, and adverse prenatal events; and that the microbial community structure of the infant gut microbiome can be used to predict the microbial community structure in the same child at 2 and 4 years of age.

2.3 Materials and Methods

The study was approved by both the hospital and the university Institutional Review Board and parents gave informed consent for the study in the NICU and agreed to be followed up at a later time for further research.

2.3.1 Patient demographics and clinical information

A total of 83 preterm, very low and extremely low birthweight (VLBW and ELBW) infants (gestational age = 28.44 ± 2.39 weeks, birth weight = 1086.71 ± 218.49 grams) were enrolled as soon as possible after admission to the neonatal intensive care unit (NICU) at Tampa General Hospital, an academic Level III center with a single-patient room floor plan, during the period May 2012–December 2013. Mothers who were drug abusers or HIV positive were excluded, as were infants with major congenital anomalies or who were moribund. Maternal and neonatal clinical information was obtained from electronic medical

records and an investigator-developed demographic form. Long term morbidity data were extracted from the Vermont Oxford Network database. Data were collected for a total of six weeks after entry into the study unless participants were discharged or transferred earlier. Exact volumes of mothers own milk (MOM), formula, fortifier, and donor milk (DM) were collected from the electronic medical record.

2.3.2 Stool sampling, DNA extraction, and amplicon sequencing

Stool samples were scooped aseptically from a diaper at approximately the end of each week (with some variation), stored in a sterile tube, and brought to the lab and frozen at -80° C until processing. DNA was extracted using the MoBio PowerFecal DNA kit (Qiagen, Carlsbad, CA) with modifications based on the Earth Microbiome Project protocols (www.earthmicrobiome.org). The V4 region of 16S rRNA gene was amplified using polymerase chain reaction with modified 515F and 806R primers, followed by amplicon sequencing using the MiSeq platform (Illumina, San Diego, CA) based on existing protocols to generate ≈100,000 250 bp paired-end reads per sample [88]. Raw sequence reads are available through the Sequence Read Archive under accession number phs001578.

2.3.3 Sequence data processing

The V4 region 16S rRNA gene amplicon data were analyzed using the DADA2 plugin [89] in QIIME2[90]. Data from 6 separate sequencing runs were imported into QIIME2-2018.2 and demultiplexed with demux emp-paired. The DADA2 plugin was used for quality control, including filtering phiX reads and chimeric sequences. Trimming was performed to trunclen 140 and trim 20, then the resulting feature tables were merged. Taxonomy was assigned against the Greengenes v13.8 database.

2.3.4 Statistics and machine learning

We calculated the correlation between the relative abundances of the sequence variants and NICU health outcomes by regression analysis in R. We employed UniFrac distances to examine microbiome structure between samples to test differences between groups using multivariate statistical methods such as principal coordinate analysis (PCoA) and permutational multivariate analysis of variance (PERMANOVA). We performed two-sample t-tests with FDR correction to examine differences between binary variables. Volatility analysis was performed by comparing unweighted UniFrac distances on exact sequence variants (ESVs) between subgroups. To identify the predictive value of subgroups on microbiome community composition, we applied random forest machine learning (after rarefying to 5000 sequences/sample, 1000 trees) and Analysis of Composition of Microbiomes (ANCOM;[91]). For continuous clinical variables, we ran the Random Forest regressor model (1000 estimators, bootstrapping, mean square error (MSE)). Maturity analysis was applied using the sample classifier plugin in QIIME2 [92]. UpSet plots were generated using the R package[93].

2.3.5 Infant growth metrics

Infant weight and length were measured at birth and weekly until discharge. Raw weight and length measurements were standardized against the 2013 revision of the Fenton Preterm Growth Chart using the actual age calculator tool [94]. This process converts the raw measurement to z-scores based on the mean and standard deviation of the reference population at each age. This resulted in a weight-for-age and length-for-age z-score at each time point. Standardizing weights and lengths in this manner allow infants of different sexes and gestational ages to be compared as one population. In addition, because the Fenton is intended to represent the ideal growth of preterm infants, it can be used to detect changes in growth status as they age. Therefore, a change in z-score over time can be a useful tool to detect growth faltering (a z-score that becomes smaller) and catch-up growth (a z-score that becomes larger).

Growth (weight gain and length-for-age z-score) over the NICU stay was assessed with the following measurements: a) total average growth rate (difference in birth weight to discharge weight divided by length of stay), growth rate between birth and 6 weeks of life (difference in birth weight to weight at 6 weeks divided by 6 weeks), and growth rate between 6 weeks of life and discharge (difference in weight at 6 weeks to discharge weight divided by length of stay minus 6 weeks), b) Weight for Gestational Age (categorical variable: Small for Gestational Age, Average for Gestational Age, or Large for Gestational Age), and 3) improved length (binary variable of whether or not the infant had greater than expected length-for-age z-scores during the NICU stay). Complete growth metadata were available for only 78 out of the total 83 infants.

2.3.6 Data availability

The study was registered in dbGaP under accession phs001578.v1.p1. De-identified metadata and raw forward and reverse sequence reads were associated with each sample via SRA.

2.4 Results

2.4.1 Overview of the preterm infant cohort

Preterm infant clinical variables (Table 2.1) were determined for a population cohort of 83 preterm infants delivered at Tampa General Hospital between May 2012 and Dec 2013. Each child was sampled for fecal matter approximately weekly for up to 6 weeks during NICU stay, and then followed up at 2 years old and 4 years old. Of the original sample, only about 57 were locatable at 2 years of age, and from that group 25 were recruited for follow up, which significantly impacted our statistical power to develop a predictive model of microbial community structure between the NICU stay and childhood. The longitudinal analysis comprised 425 fecal samples, which were processed for 16S rRNA amplicon sequencing to characterize the microbiome.

Table 2.1: Patient characteristics. Data represent maternal demographics and interventions during NICU stay.

	Overall	**Min**	**Max**
n	83		
Maternal Age(years) (mean (sd))	28.13 (6.89)	15.2	46.1
Received any antibiotics (%)	76 (91.6)		
Days on Antibiotics (mean (sd))	11.73 (18.45)	0	114
Paternal Ethnicity (n (%))			
African American	34 (43.0)		
Asian Pacific Islander	3 (3.8)		
Caucasian	25 (30.9)		
Hispanic Black	1 (1.2)		
Hispanic White	15 (18.5)		
Native American	1 (1.2)		
Other	3 (3.7)		
Parental Education (n (%))			
College Graduate	17 (20.7)		
Grammar/Elementary School	4 (4.9)		
High School	47 (57.3)		
Middle School	8 (9.8)		
Post Graduate Degree	6 (7.3)		
Parental Marital Status (n (%))			
Divorced Separated	6 (7.4)		
Married/Living With Partner	32 (39.5)		
Single	42 (51.9)		
Widowed	1 (1.2)		

Table 2.1 – Continued

Family Income (n (%))			
< 4,999	20 (27.8)		
5,000-14,999	27 (37.5)		
15,000–24,999	12 (16.7)		
40,000–69,999	3 (4.2)		
> 70,000	9 (12.5)		
Unknown	1 (1.4)		
Working (n (%))			
Full Time	17 (21.2)		
Part Time	7 (8.8)		
No	56 (70.0)		
Working hours per week (mean (sd))	34.80 (18.10)	0	80
Maternal Height (inches) (mean (sd))	64.60 (2.80)	59	69
Maternal pre-pregnancy weight (lbs) (mean (sd))	167.20 (46.68)	79	303
Number in house (mean (sd))	3.15 (1.34)	1	7
Total pregnancies (mean (sd))	3.09 (2.46)	1	11
Full term pregnancies (mean (sd))	0.66 (0.95)	0	4
Preterm pregnancies (mean (sd))	0.43 (0.69)	0	2
Induced Abortions (mean (sd))	0.26 (0.64)	0	3
Spontaneous Abortions (mean (sd))	0.74 (1.43)	0	7
Ectopic Pregnancies (mean (sd))	0.07 (0.26)	0	1
Multiple Pregnancies (mean (sd))	0.04 (0.19)	0	1
Living children (mean (sd))	0.91 (1.15)	0	5
Received prenatal care (n(%))	70 (93.3)		
Maternal BMI (mean (sd))	28.08 (7.45)		

Table 2.1 – Continued

Depression (n(%))	10 (12.5)		
Depression treatment received (n (%))			
None	2 (20.0)		
Medication alone	1 (10.0)		
Medication and counseling	7 (70.0)		
Anxiety (n(%))	10 (12.5)		
Anxiety treatment received (n (%))			
None	3 (3.7)		
Counseling alone	1 (1.2)		
Medication alone	3 (3.7)		
Medication and counseling	6 (7.3)		
Any breastfeeding (n(%))	76 (93.8)		
Exclusive breastfeeding (n(%))	51 (64.6)		
Months of breastfeeding (mean (sd))	8.58 (4.25)	0	20
Smoking (n (%))			
No	60 (74.1)		
Not currently	7 (8.6)		
Yes	14 (17.3)		
Smoking during pregnancy = yes (n(%)	10 (12.8)		
Alcohol during pregnancy = yes (n(%)	1 (1.2)		
Maternal drug use = yes (n(%)	2 (2.5)		
Maternal drug use during pregnancy = yes (n(%)	4 (5.0)		
Steroids = Yes (n(%))	53 (67.9)		
Maternal antibiotic = Yes (n(%))	36 (48.0)		
Born at TGH (n (%))	75 (94.9)		

Table 2.1 – Continued

Vaginal delivery (n(%))	18 (22.2)		
Labor length (hours) (n(%))			
Unknown	48 (58.5)		
< 1 hr	7 (8.5)		
1–10 hrs	18 (22.0)		
11–20 hrs	4 (4.9)		
21–30 hrs	2 (2.4)		
41–50 hrs	1 (1.2)		
51–60 hrs	1 (1.2)		
71–80 hrs	1 (1.2)		
Maternal weight at delivery (lbs) (mean (sd))	194.17 (49.93)	102	313
Anesthesia (n(%))			
None	9 (12.2)		
Epidural	10 (13.5)		
General	29 (39.2)		
Spinal	26 (35.1)		
Delivery complications (n (%))	18 (23.7)		
Male infant gender (n(%))	40 (48.8)		
Infant weight at delivery (g) (mean (sd))	1086.71 (218.49)	600	1485
Gestational age at delivery (weeks) (mean (sd))	28.44 (2.39)	24	
Multiples (n(%))			
singleton	54 (74.0)		
twin	16 (21.9)		
triplet	3 (4.1)		
APGAR 1 min (mean (sd))	6.09 (1.98)	1	9

Table 2.1 – Continued

APGAR 5MIN (mean (sd))	7.53 (1.51)	3	9
SNAPPEII FINAL (mean (sd))	18.99 (16.51)	0	81
SNAPII FINAL (mean (sd))	10.08 (9.51)	0	54
SNAPII week1 (mean (sd))	11.73 (11.70)	0	66
SNAPII week2 (mean (sd))	3.53 (6.19)	0	32
SNAPII week3 (mean (sd))	2.47 (5.83)	0	35
SNAPII week4 (mean (sd))	0.79 (3.72)	0	24
SNAPII week5 (mean (sd))	0.53 (2.34)	0	14
SNAPII week6 (mean (sd))	0.84 (3.42)	0	19
Time to enteral feed (days) (mean (sd))	12.25 (5.06)	5	39
Infant weight at 6 weeks (g) (mean (sd))	1887.91 (335.23)	1210	2700
Infant weight at discharge (g) (mean (sd))	2678.96 (900.13)	1101	6500
Retinopathy of prematurity (n (%))	12 (15.2)		
Bronchopulmonary dysplasia (n (%))	5 (6.3)		
Sepsis (n (%))	10 (12.7)		
Necrotizing enterocolitis (n (%))	3 (3.8)		
Intraventricular hemorrhage (n (%))	9 (11.9)		
Received transfusions (n (%))	34 (43.0)		
Days on Oxygen (mean (sd))	14.29 (20.82)	0	101
Length of Stay (days) (mean (sd))	69.84 (36.74)	20	215
Feed intolerance (n (%))	15 (19.5)		
Deceased(n (%))	1 (1.4)		

Relevant available age metrics included gestational age at birth, chronological age, and postmenstrual age. Infants in our cohort had a wide range of gestational ages at birth (24–37 weeks). Therefore, chronological age was not employed as the independent variable; instead

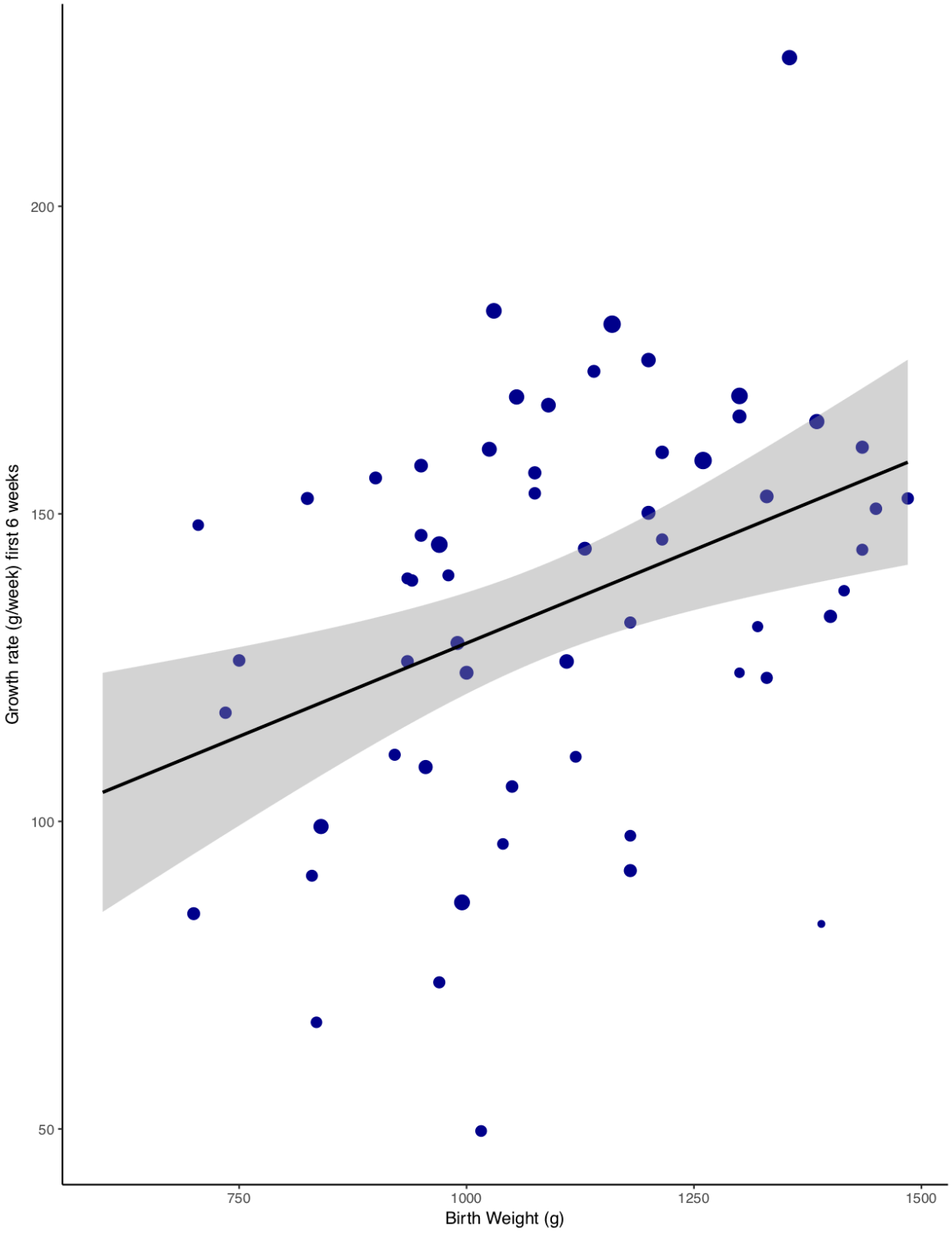
we used postmenstrual age (chronological age + gestational age; also known as corrected gestational age) for all subsequent analyses. 91.6% started antibiotics on first day of life (regardless of gestational age). The mean postmenstrual age for starting antibiotics was 28.2 ± 2.5 (SD) weeks. Most infants were on antibiotics for 3–4 days at a time with a median course of antibiotics lasting 3.5 days. Only 7 infants did not receive any antibiotics. The mean total days on antibiotics was 11.88 ± 18.8; range 0 to 114 days.

In total, 6,383,544 16S rRNA amplicon reads were generated across all 425 samples, with a range of $\approx 1,000$ to 40,483 (median 13,092; mean 15,020.1). Sequence depth was rarified to 1,275 reads per sample resulting in 415 samples used in subsequent analysis. Reads were classified into Exact Sequence Variants (ESVs) using DADA2, creating a total of 2213 ESVs, with a range of 1 to 1,844,477 reads per sample (median 48; mean 2,884.5). All ESVs were annotated to known microbial taxa.

2.4.2 Infant growth as a clinical outcome

We were statistically underpowered for each morbidity (only 5 infants with chronic lung disease (CLD), 3 with necrotizing enterocolitis (NEC), 10 with blood culture positive sepsis, and 9 with any stage of intraventricular hemorrhage (IVHn)); therefore we focused on growth (increased weight and length) as a primary clinical outcome during the NICU stay. We tested the independence of the measured growth variables (Materials and Methods). Birth weight (g) was significantly correlated with growth rate over the first 6 weeks of life (g/week) (Fig. 2.1). Overall there was a weak but significant association between higher initial birth weight and faster initial weight gain (R2= 0.146, $p = 0.001703$), however, this relationship was not significant for birth weight and weight gain between 6 weeks and discharge ($p = 0.8141$). No significant relationships were observed between weight at discharge or weight gain with feeding (total volume of milk received and ratio of mothers breast milk to total enteral feeding volume received; $p \geq 0.2$).

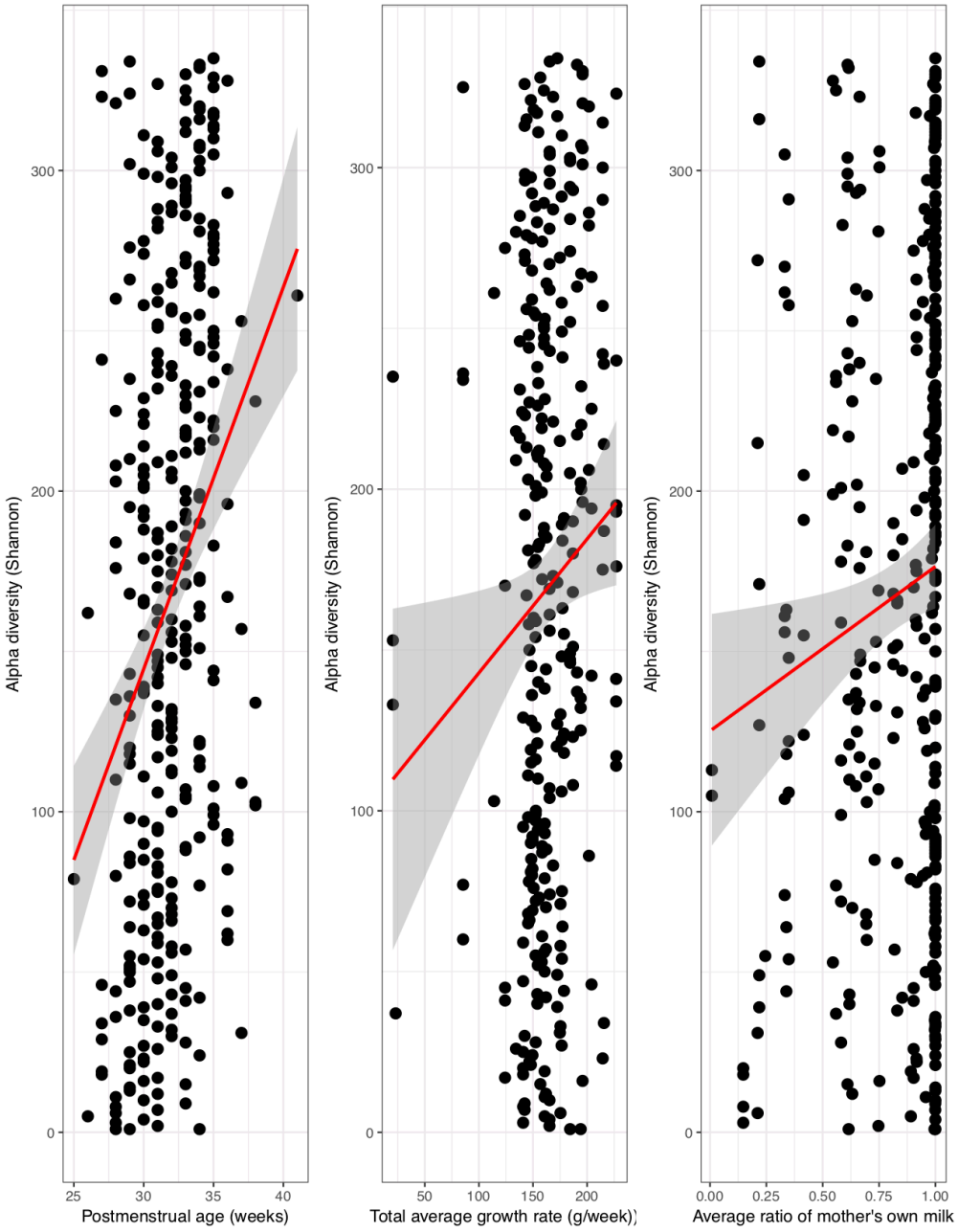
Figure 2.1: Birth weight (grams) correlates with growth per week initially (first 6 weeks). $R2 = 0.146$; $p = 0.001687$.



2.4.3 Microbial diversity increased with age, and is statistically associated with clinical variables

The relationship between weight and length gain outcomes and microbial alpha diversity over the course of the NICU stay was determined. Shannon index significantly increased with increasing postmenstrual age (Multiple R2: 0.09221, $p = 1.28e-08$) (Fig. 2.2a). Adjusting for postmenstrual age during the NICU stay, the alpha diversity also showed weak but significant correlations with overall rate of weight gain (Multiple R2: 0.144, $p = 0.00208$; Fig. 2.2b) and total percentage of mothers own milk (Multiple R2: 0.1097, $p = 0.024918$ Fig 2.2c), but no statistically significant associations were observed with the other clinical variables: mode of delivery, length of hospitalization in the NICU, time to full enteral feeding, days on antibiotics, infant weight at birth or discharge, days on oxygen, APGAR scores at 1 or 5 minutes post birth, duration of breast-feeding, or maternal age. We found no significant associations with microbial alpha diversity and common co-morbidities, although we lacked statistical power within the cohort for each morbidity. As expected, infants who received antibiotics during the NICU stay had significantly lower alpha diversity than those who did not receive antibiotics (t-test, FDR corrected, $p = 0.009$). Shannon alpha diversity difference per infant between first NICU stool sample and last stool sample ranged from -3.3520 to 3.1720 with a median 0.8259. There was no significant difference between delta Shannon alpha diversity between children who showed improved length and those who did not, or by mode of delivery ($p \geq 0.3$, t-test, FDR correction). A significant increase in Shannon index was also observed within each child when comparing between their last time-point of NICU stay and 2 and 4 years of age ($p = 0.01$; t-test); the Shannon index was also significantly different between 2 and 4 years of age. While the alpha diversity during the NICU stay and at 2 years of age was significantly lower compared to the alpha diversity of mothers stool-associated microbiota ($p = 0.01$; t-test), the alpha diversity at 4 years of age was not significantly different to that of the mother, suggesting that the diversity normalizes to that of an adult by 4 years.

Figure 2.2: Correlation of alpha diversity with (left panel) postmenstrual age (Shannon: multiple $R2 = 0.1508$, $p = 1.507e-06$), (middle panel) total growth rate (grams/week), and (right panel) ratio of mothers own milk volume to total milk volume.



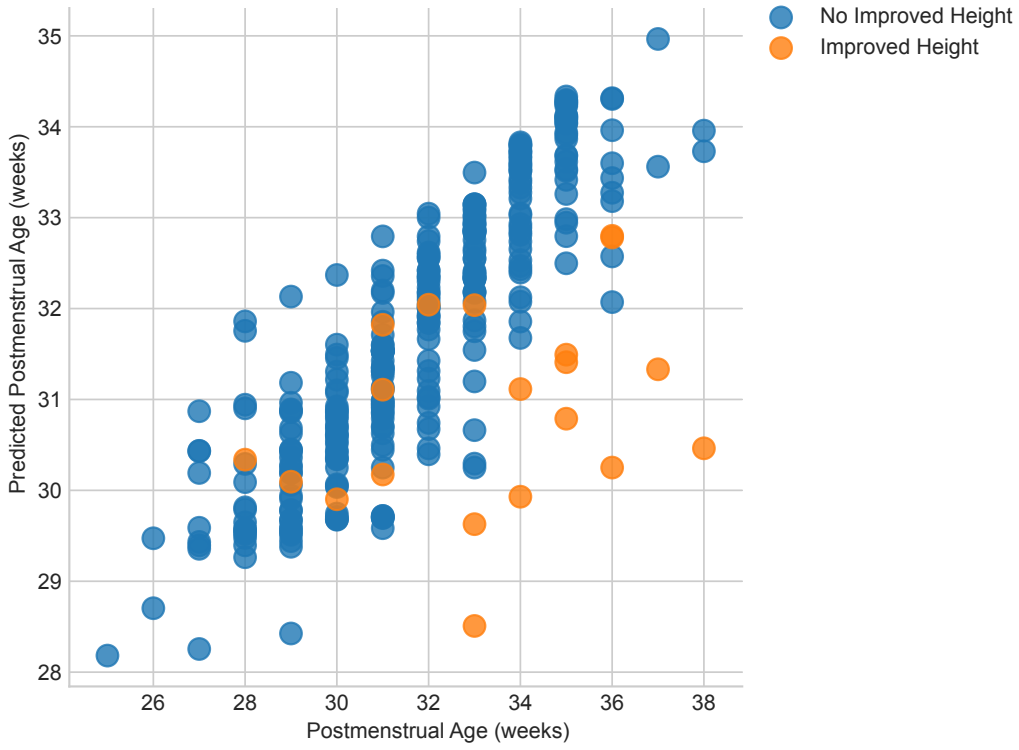
2.4.4 Microbial beta diversity, maturity and microbial stability

The distribution of between-sample unweighted UniFrac distances was calculated pairwise to determine the volatility of microbial community structure over time during the NICU stay. This was then associated with the clinical variables, including both categorical variables (delivery method, sex, gestational age, weight for gestational age, improved length, multiples, weight gain rate, ratio of volume of breastmilk to total milk received, length of stay in NICU) and binary variables (antibiotics treatment, sepsis, NEC, CLD, transfusions, feeding intolerance). Beta-diversity volatility was only significantly correlated with improved length during the NICU stay, whereby infants who showed improved length had significantly more inter-sample variability in beta diversity, and hence had greater microbial volatility over their NICU stay ($p = 0.02805$; FDR corrected). This suggests that infants with a less-volatile microbiota over time may have reduced catchup growth (length) during the NICU stay. While mode-of-delivery alone did not correlate with volatility, all the infants with improved length (n=6) were delivered by Caesarean section.

We ran maturity analysis (Materials and Methods), which uses a regression model used to predict postmenstrual age as a function of the composition of the microbiota. Therefore, if a sample has a microbiota which falls outside of the confidence interval based on the average microbiota composition for infants of each postmenstrual age it is considered to be immature. We correlated the maturity index against improved length during their NICU stay (Fig. 2.3). Those infants who did have improved length during their NICU stay had a significantly reduced level of predicted microbial maturity compared to infants whose length did not significantly improve ($p < 0.0018$). As maturity is inversely correlated with beta-diversity volatility, this result agrees with the beta diversity volatility analysis above.

We also examined the beta diversity distribution (unweighted UniFrac) to determine the microbial community structure variance between the final NICU time point, against 2 and 4-years-old, and maternal samples. The maternal, 2- and 4-year-old microbiota also clustered together, and were significantly different from the samples collected at the end of the NICU

Figure 2.3: Maturity analysis showing actual versus predicted gestational age at birth for infants showing or not showing improved length.



stay (Fig. 2.4; PCoA; $p < 0.05$ after FDR).

2.4.5 Microorganisms associated with clinical outcomes and age

The stool samples collected during the NICU stay maintained a predominance of Gammaproteobacteria, specifically Enterobacteriaceae, members of which are often associated with LPS mediated inflammation. The most abundant Exact Sequence Variant (ESV) in infants during the NICU stay was closely related to the genus *Klebsiella*, which belongs to the Enterobacteriaceae. ANCOM analysis was applied against all categorical and binary clinical variables to determine which bacteria were significantly differentiated in relative abundance. Only two clinical variables showed bacteria that were significantly differentially abundant. Infant weight gain from birth to discharge was significantly negatively correlated with the relative abundance of two ESVs, one associated with *Klebsiella* and the other related to *Staphylococcus* ($p < 0.01$). Additionally, antibiotic administration was associated with a significant enrichment in *Proteus* and a significant proportional decrease of *Streptococcus* and *Bifidobacterium*, compared to infants who did not receive antibiotics ($p < 0.01$). The predominance of Proteobacteria was greater in infants with lower beta-diversity volatility than those with higher volatility. A Random Forest classifier was used to determine whether the microbiota was predictive of categorical clinical variables (clinical factors that could be grouped into discrete components), which suggested that only the use of antibiotics (presence/absence) could be accurately predicted with a baseline to random ratio of > 2 (2.26) suggesting a marginal relationship. The Random Forest features identified as having the most predictive value for the presence or absence of antibiotics were ESVs corresponding to Enterobacteriaceae, *Citrobacter*, *Escherichia*, and Peptostreptococcaceae. For continuous clinical variables, we also ran the Random Forest regressor model (1000 estimators, bootstrapping, MSE), which only identified the duration of antibiotic administration as significantly predictable (R2 of 0.16259; $p = 0.000651$). The top most important features for the regressor-model accuracy were ESVs corresponding to *Veillonella*, Enterobacteriaceae,

Figure 2.4: Unweighted Unifrac beta diversity PCoA of infant samples (ie those from the last NICU time point) and samples from 2-year-old children, 4-year-old children, and mothers.

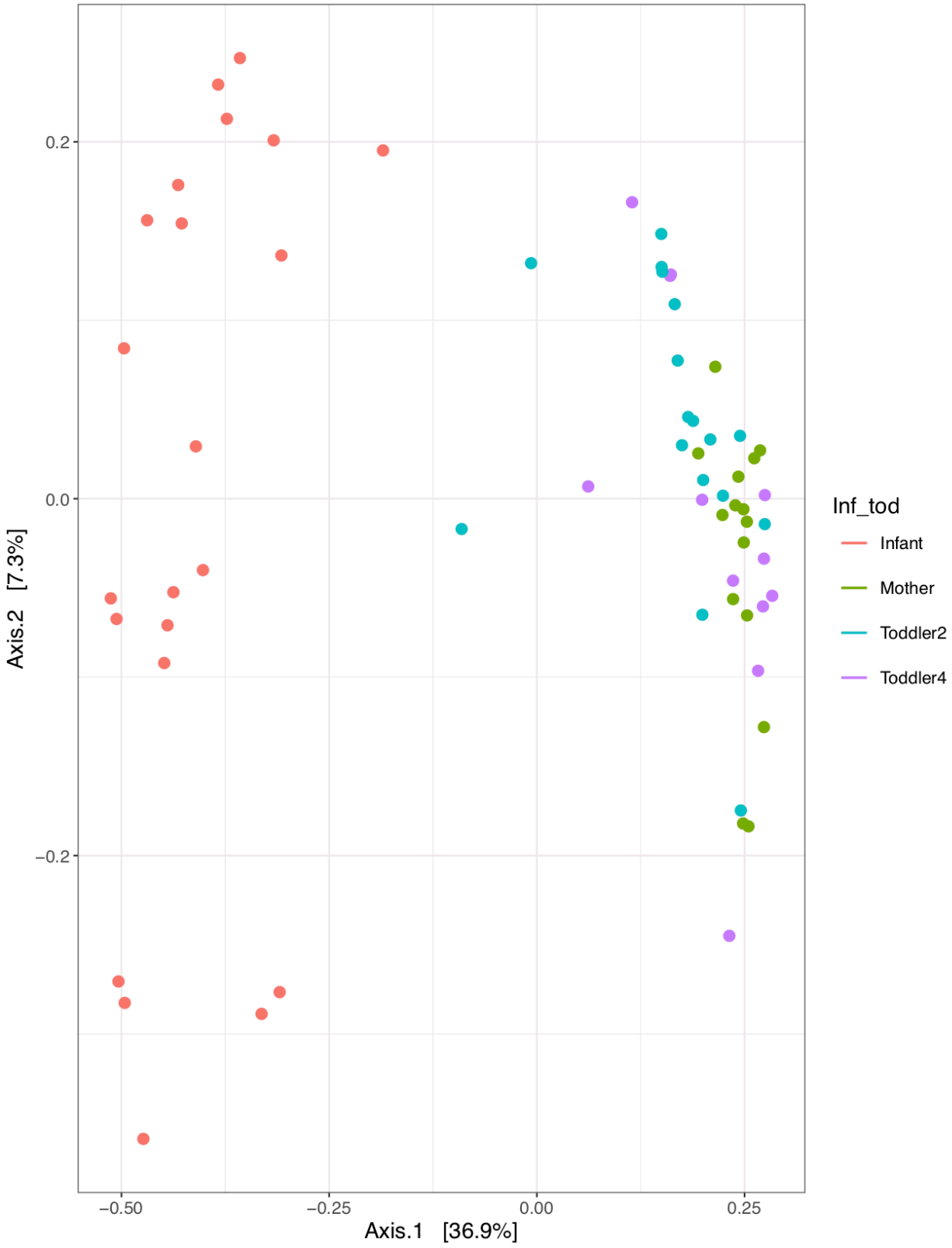
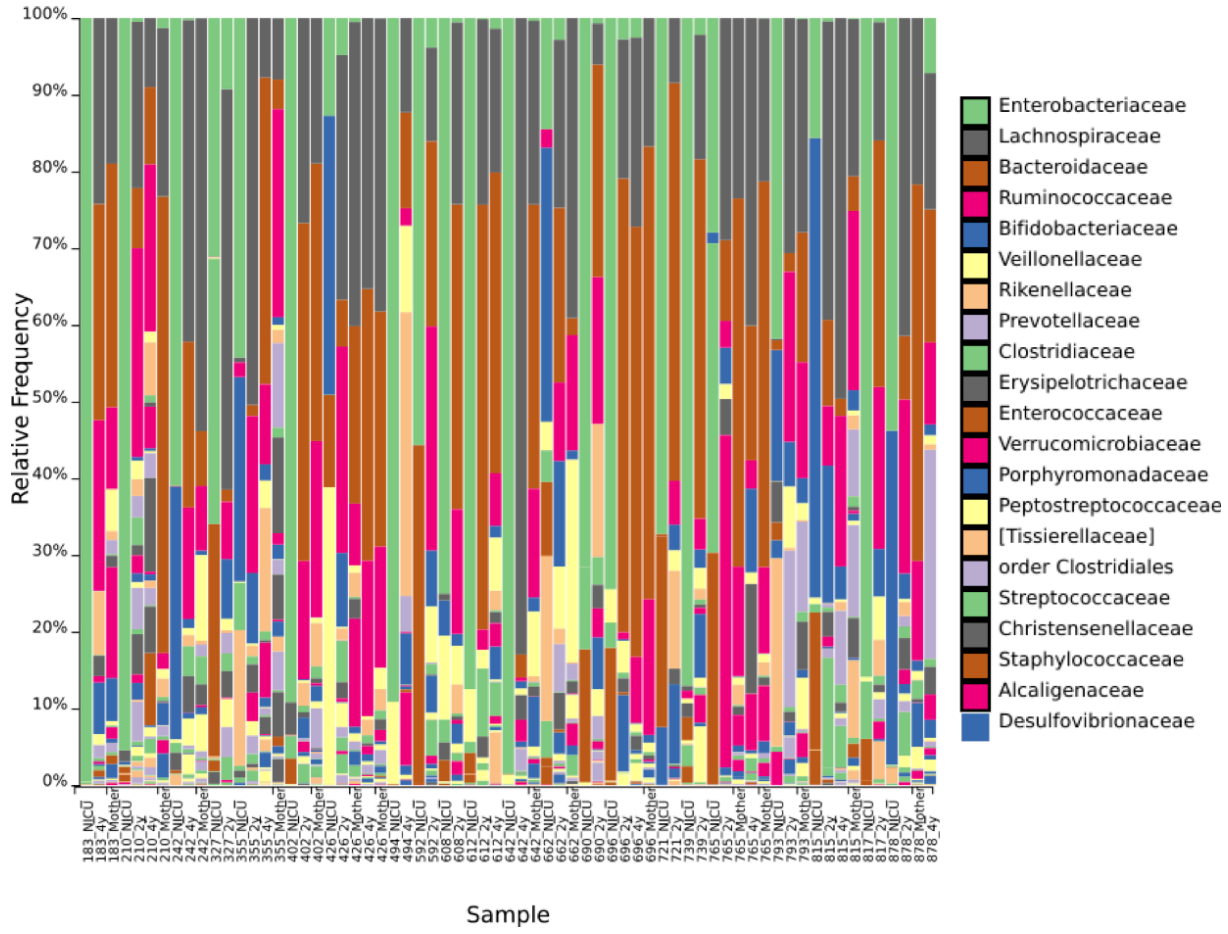


Figure 2.5: Relative abundance of taxa at the family level. Each bar represents a single sample. Bars are sorted by patient identifier (ID) and ordered by age for each patient.



Proteus, and *Escherichia* (importance scores 0.15 to 0.05). While the infant samples at six weeks of life had a significantly greater abundance of ESVs related to *Enterococcus* and *Staphylococcus* (Kruskal-Wallace; Bonferroni corrected $p < 0.0001$), toddler samples at both 2 and 4 years old maintained ESVs associated with Clostridiales and Bacteroidales that were not detected in infant samples for respective individuals. However, infants as a group shared 26 ESVs in total with 2-year-old toddlers as a group, 26 ESVs in common with 4-year-old toddlers and 18 ESVs in common with the mothers as group (Fig. 2.6). Interestingly, 2-year-old toddlers shared 155 ESVs with their mothers, while 4-year-olds shared 210 ESVs with their mothers. This suggests that the number of ESVs shared with the mother group increases with increasing age, likely due to the changing gut environment and increase in food similarity. In total, only 14 ESVs were shared between infants, 2- and 4-year-olds and mothers, which comprise the core ESVs. These included genera such as *Ruminococcus*, *Oscillospira*, *Bacteroides*, *Streptococcus*, *Bifidobacterium*, *Escherichia*, *Eggerthella*, and *Dorea*, and families such as Lachnospiraceae, Peptostreptococcaceae, Rikenellaceae (Table 2.2, Figure 2.5). Interestingly, one of the core ESVs found in all age groups was related to *Haemophilus parainfluenzae*, which is a potential pathogen. Overall, the *H. parainfluenzae* ESV was found in $\approx 10\%$ of the samples, from 29% (24) of the participants. Some infants maintained the *H. parainfluenzae* ESV over multiple NICU timepoints, but it was not found in the corresponding 2- and 4-year old samples for those infants. However, it was found in 3 children at both 2-years-old and 4-years-old, and also in their mothers. As mother stool samples were collected at the same time as the 4-year-old stool sample, it is possible that this *H. parainfluenzae* ESV was endemic in these 3 families during that time.

Table 2.2: Taxonomy of ESVs in common between all age groups (last NICU time point, 2 years, 4 years) and maternal samples.

Order	**Family**	**Genus**	**Species**
Clostridiales	Lachnospiraceae	Ruminococcus	gnavus
Pasteurellales	Pasteurellaceae	Haemophilus	parainfluenzae

Table 2.2 – Continued

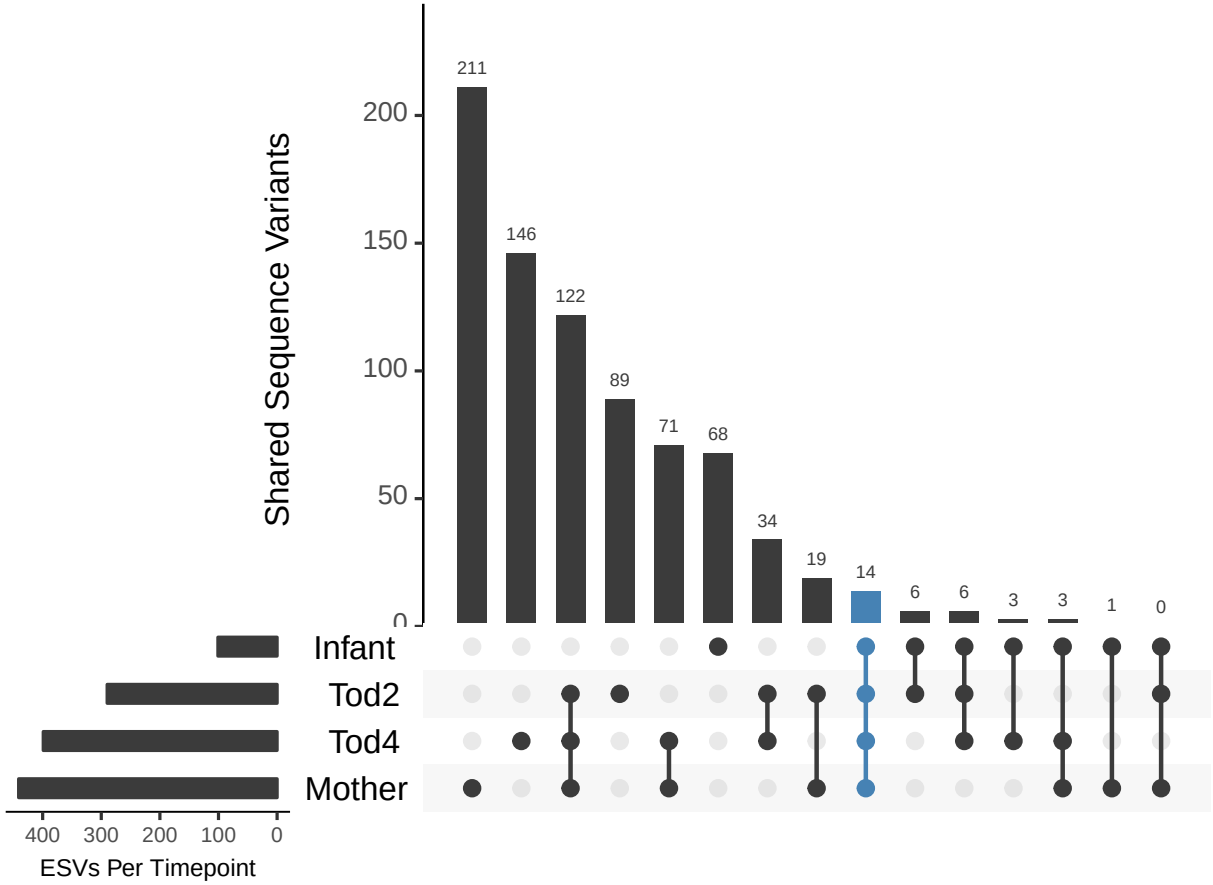
Clostridiales	Ruminococcaceae	Oscillospira	
Bacteroidales	Bacteroidaceae	Bacteroides	
Lactobacillales	Streptococcaceae	Streptococcus	
Clostridiales	Peptostreptococcaceae		
Clostridiales	Lachnospiraceae	Ruminococcus	torques
Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	
Enterobacteriales	Enterobacteriaceae	Escherichia	coli
Coriobacteriales	Coriobacteriaceae	Eggerthella	lenta
Bacteroidales	Rikenellaceae		
Clostridiales	Lachnospiraceae	Dorea	
Lactobacillales	Streptococcaceae	Streptococcus	
Clostridiales	Lachnospiraceae		

2.5 Discussion

Preterm infants are particularly vulnerable to perturbations in gut microbiome development due to their abnormal delivery and prolonged NICU stay, which is associated with many insults. Our study complements prior observations that the preterm infant microbiome is characterized by an abundance of Proteobacteria [13,95]. Bacterial alpha diversity was significantly positively correlated with postmenstrual age, and when we adjusted for postmenstrual age during the NICU stay, the alpha diversity also showed weak but significant correlations with overall rate of weight gain and total volume of mothers own milk.

While we did not identify significant differences between infants microbial beta diversity by mode of delivery or feeding type, we did, as expected, find an effect of antibiotic administration, which correlated with a significant enrichment of *Proteus*. *Proteus*, a gammaproteobacterial genus, can be an opportunistic pathogen in humans and is frequently found in hospital settings [96,97]. It has been associated with high rates of biofilm formation and

Figure 2.6: UpSet plot showing ESVs shared between infant samples and samples from 2-year-old children, 4-year-old children, and mothers. The numbers of unique ESVs shared between groups or intersections of groups are plotted as vertical bars. The 14 core ESVs shared between all four sample groups are indicated in blue.



antibiotic resistance, which may explain its increased abundance following antibiotic perturbation in our study. Furthermore, postmenstrual age (gestational age + chronological age, also known as corrected gestational age) described the majority of variance in beta diversity over time. Infant weight gain during the NICU stay was negatively correlated with the relative abundance of *Klebsiella* and *Staphylococcus*. As both of these taxa are associated with known pathogens, it is possible that their enrichment may be indicative of dysbiosis, which in turn appears to be associated with reduced infant weight and length gain, although to the best of our knowledge this specific association has not been demonstrated previously. However, a previous study has demonstrated a significant association between bacterial alpha diversity and body weight and gestational age, which also associated with an increase in the staphylococci [98].

We also demonstrated that catchup growth in length z-score (improved length) was correlated with increased beta diversity volatility (between sample beta diversity distance) and decreased maturity (predictive power of regressor modeling) during the NICU stay. Catch up growth is an important consideration for preterm infants: VLBW infants are vulnerable to extrauterine growth restriction, and caloric and protein deficits result in slower growth velocity and are associated with major neonatal morbidities, including chronic lung disease and retinopathy of prematurity [99]. Clinically, it is believed that early catchup growth (<2 years, and especially in the first 24 weeks of life) is beneficial to preterm delivered child health outcomes [100,101]. That microbial beta diversity was more volatile for children who demonstrated greater catchup growth could be related to the overall microbial profile for this community. Infants with lower volatility had significantly greater predominance of Proteobacteria, i.e. they had greater week to week stability due to the dominance of a group of potential pathogens (e.g. *Klebsiella*). Infants with more volatility had reduced prevalence of these organisms and as such may have reduced dysbiosis and improved growth. This study is among the first to document a relationship between microbiome and VLBW infant growth. Previous work found that preterm infants with prolonged microbial transitions had

smaller changes in weight-for-age z-scores across time [87]. However, little is known about the relationship between preterm/VLBW length gain and microbiome development. Poor length gain, as an indicator of long-term nutritional stress, indicates that microbiome development can have a long-term effect on infant phenotype. Researchers have noted that the preterm microbiome has similar developmental signatures to undernourished term infants [102], suggesting that therapeutic approaches may be similar for both groups. More research is needed to establish the relationship between VLBW length and weight gain and microbiome development.

Due to the limited size of this cohort, it is not possible to determine if this trend is absolutely significant, but it suggests that further studies using larger cohorts would be useful, as would studies involving the humanization of murine models to demonstrate reduced growth as a transferable phenotype. Those infants who demonstrated reduced microbial maturity also showed improved catchup growth, is likely a result of the population against which these infants are compared. Our maturity analysis is based on the average 'maturity' for a population of VLBW infants who are essentially quite sick, as such infants who show improved catch up growth would be statistical outliers against this group. Having a healthy control population may provide more accurate interpretation of the standard maturity analysis.

In our cohort, we followed the same children to ages 2 and 4 years. We had insufficient statistical power to identify clinical outcomes that correlated with microbiome composition, but we did find that the these VLBW infants did develop a microbiome comparable to that of their adult mothers by ages 2- to 4- years old. As previously demonstrated, the microbiome of children undergoes patterned progression to become more similar to the maternal microbiome as the children increase in age, with an 'adult-like' microbiome established by 4 years [34]. In our cohort of VLBW preterm infants, we also observed shifts in the microbiome and increasing similarity to the maternal microbiome with increasing age. Importantly, many of the ESVs found in infants were also found in 2- and 4-year olds, as well as the mothers, suggesting conservation of microbial composition between generations, ages and environments.

Yet, the ESV overlap between infants and mothers was significantly lower than the overlap between mothers and 4-year-olds, suggesting a gradual shift of the microbial community likely as a result of the changing ecological drivers that shape the niche structure of the developing gastrointestinal tract. For example, as the gut matures, and the immune system becomes educated, and the child starts to eat food more similar to the adult mother, there will be a selective ecological pressure towards a more similar microbial structure[9,83,103–106].

We have demonstrated a significant association between microbial community diversity, structure and infant growth in a significantly at-risk childhood population. While the population had inherent variability, which limits the potential to identify associations between the microbiota and clinical outcomes, compelling correlations between microbial structure, volatility, maturity, composition, and infant weight and length increases were determined that suggest potential biomarkers of dysbiosis in this at-risk population.

CHAPTER 3

NICHE DIFFERENTIATION IN THE EARLY LIFE GUT MICROBIOME BY FUCOSYLTRANSFERASE 2 SECRETOR STATUS

3.1 Abstract

Fucosyltransferase2 (FUT2) is a key gene in host-microbe interactions and is inherited in approximately Mendelian ratios worldwide, implying there are balancing selection pressures. Functional FUT2 is responsible for altering the carbohydrate landscape of the human intestine, so it may contribute to inter-individual variation in the microbiome, which could have implications for precision medicine. The objective of this study was to identify associations in the taxonomic makeup and functional potential of the gut microbiome with host genotype. We performed sequencing on stool samples from 399 children at 24 months of age, of whom 112 were FUT2+/+, 211 were FUT2+/-, and 76 were FUT2-/-. 16S rRNA and ITS marker gene sequencing was performed to examine microbial community composition and metagenomic sequencing was performed to elucidate functional potential. At 24 months of age, the bacterial taxonomic composition and community structure varied with FUT2 genotype. However, these differences were not observed in the fungal community. Interestingly, 16S sequence variants annotated to the genera *Bifidobacterium* showed differing abundance trends in secretors and nonsecretors. These patterns corresponded to differences in the species level taxonomy, as assessed by metagenomic profiling. The effect of host genetics on the human gut microbiome is incompletely understood. FUT2 is a key gene regulating host-microbial interactions. We expanded on previous studies using 16S rRNA marker genes to include fungal taxonomic profiling and shotgun metagenomics. Interindividual variation in the human microbiome may be attributable in part to differences in the genotype of host genes such as FUT2 and the functional response of the microbial community.

3.2 Introduction

The fucosyltransferase 2 (FUT2) gene can regulate host-microbe interaction. This gene is responsible for expressing ABO histo-blood group antigens, which are α1,2-linked fucose containing glycans, in mucus and other secretions[56]. About 20% of the population are 'non-secretors', lacking a functional copy of FUT2 due to nonsense single nucleotide polymorphisms [45,107]. While the exact nonsense mutation varies by population, with different SNPs found in different ethnic groups, worldwide, the FUT2 secretor phenotype is maintained at approximately Mendelian ratios, suggesting balancing selection [45,107]. The secretor phenotype plays an important role in formation of immune complexes, secretion of antigens in body fluids, as such non-secretors can have greater susceptibility to pathogenic bacteria. It has been shown that secretor status contributes to inter-individual microbial variation[30]. The effect of secretion genotype on microbial community assembly in infants has been examined previously, which demonstrated that secretor infants, who are able to produce fucosylated milk oligosaccharides, had an increased proportion of *Bifidobacterium*[56]; additionally, secretor mothers had significantly increased microbial community stability in breast milk, which was transferred to the gut of breastfed infants.[30] This is presumably because nonsecretor mothers are unable to secrete α1,2-linked fucosylated glycans in their breast milk, and these glycans act as a prebiotic to feed the neonatal microbiota, facilitating the selective expansion of mutualist microorganisms and inhibiting growth and adhesion of opportunistic and obligate pathogens at the gut mucosal surface. The microbiota itself can stimulate the induction of mucin expression and FUT2 expression [55,109]. Nonsecretors are known to be resistant to certain strains of rotavirus, *Helicobacter pylori*, and norovirus, but have an increased risk for Crohns disease and type 1 diabetes [50,51,54,56,110-112].Many symbiotic gut bacteria can and do utilize fucose, and it is protective in several disease models, it can also be taken advantage of by harmful microbes [55]. Therefore, mucosal glycans support a robust, stable microbiota, which benefits the mature host through conversion of dietary components into critical metabolites (e.g., short chain fatty acids) and competitive exclusion

of enteric pathogens and occupation of mucosal niches [48,109]. Notably, bifidobacteria can produce 1,2-α-fucosidase to hydrolyse α1,2-fucosyl linkages present in various glycans, such as the histo-blood group antigens [25]. Furthermore, Morrow et al. found differences in survival between premature infants of differing secretor statuses; low secretion or nonsecretion was associated with more severe morbidity [113]. Studies have shown that, in adults, bifidobacterial abundance varies with secretor status[56,114]. However, other recent work has failed to find significant differences in the gut microbiota [115,116] or clinical outcomes [117] by secretor status.

Despite attempts to identify differences in the microbiota of secretors and non-secretors, to date, no attempt has been made to determine if non-secretion has a significant impact on the functional potential of the microbiome. The goal of this study is to investigate the interaction between the childs FUT2 genotype and the microbiome and mycobiome, using both amplicon sequencing of the 16S and ITS rRNA gene, and shotgun metagenomics. These findings have implications for potential precision medicine biomarkers to predict disease endpoints associated with secretor status mediated microbial outcomes in a mature host.

3.3 Materials and Methods

3.3.1 Patient enrollment and sample collection

The Growing Up in Singapore Towards healthy Outcomes (GUSTO) parent-offspring cohort study recruited pregnant women aged 18 years and above, attending their first trimester antenatal dating ultrasound scan clinic at Singapores two major public maternity units, namely National University Hospital (NUH) and KK Women's and Children's Hospital (KKH) between June 2009 and September 2010. The participants approached were Singapore citizens or permanent residents who were of Chinese, Malay or Indian ethnicity with homogeneous parental ethnic background and who had the intention of eventually delivering in NUH or KKH and residing in Singapore for the next 5 years. Informed written consent was

obtained from each participant and children were followed up at the study clinic at 18, 24 and 36 months.[118]For each participant, maternal and child FUT2 genotype was determined, and stool samples were taken using standard protocols. Of the 965 children enrolled in the entire GUSTO cohort who had completed 24-month old visits, 399 children were selected. All of the children were of Chinese ethnicity, due to the difference in inactivating SNP by ethnic group, and had provided stool samples at 24 months. Participant characteristics are summarized in Table 3.1.

Table 3.1: Overview of the FUT2 childhood cohort characteristics.

Patient Characteristic	**Secretor (n=323)**	**Non-secretor (n=76)**
Sex, M/F	165/158	37/39
Ethnicity, % Chinese	100	100
Maternal age, years	32.2 ± 4.5	32.9 ± 4.5
Maternal genotype (SeSe/Sese/ sese/Unknown)	103/156/54/10	0/43/32/1
Child genotype (SeSe/Sese/sese)	112/211/0	0/0/76
Gestational age at birth, weeks	38.9 ± 1.2	39.1 ±1.2

3.3.2 Stool sampling, DNA extraction and amplicon sequencing

DNA was extracted using the MoBio PowerFecal DNA kit (Qiagen, Carlsbad, CA) with modifications based on the Earth Microbiome Project protocols (www.earthmicrobiome.org). The V4 region of 16S rRNA gene was amplified using polymerase chain reaction with modified 515F and 806R primers, followed by amplicon sequencing using the MiSeq platform (Illumina,

San Diego, CA) based on existing protocols to generate ≈17,000 250 bp paired-end reads per sample[88].

The ITS gene was amplified using polymerase chain reaction with ITS1f and ITS2 reverse barcoded primers, followed by amplicon sequencing using the MiSeq platform (Illumina, San Diego, CA) based on existing protocols to generate 230 bp reads [88].

The metagenome sequencing libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA, USA). Per manufacturers protocol genomic DNA underwent fragmentation by tagmentation, followed by limited cycle PCR to add platform specific sequencing adapters and unique barcodes to each sample. Amplicons were size selected and purified by AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA) followed by quantification using the Picogreen assay (Thermo Fisher Scientific, Waltham, MA, USA) on the Infinite 200 Pro plate reader (Tecan GmbH, Grödig, Austria) and pooled at equimolar concentrations for sequencing.

3.3.3 Sequence data processing

The V4 region 16S rRNA gene amplicon data were analyzed using the DADA2 plugin [89] in QIIME2[90]. Data were imported into QIIME2-2018.6 and demultiplexed with demux emp-paired. The DADA2 plugin was used for quality control, including filtering phiX reads and chimeric sequences. Trimming was performed to trunclen 140 and trim 20, then the resulting feature tables were merged. Taxonomy was assigned against the Greengenes v13.8 database. The ITS gene amplicon data were imported into QIIME2-2018.6 and forward reads only were demultiplexed with emp-single. The DADA2 plugin was used for quality control, including filtering phiX reads and chimeric sequences. Trimming was performed to trunclen 200 and trim 0. Taxonomy was assigned against the UNITE version 7 (01/12/2017) database, dynamic classifier.

Metagenomic sequence reads were clipped using nesoni (https://github.com/victorian-bioinformatics-consortium/nesoni), then annotated to known taxa and functional genes using

MetaPhlan2[119] and HumanN2[120], respectively. Draft genomes were assembled using metaSPAdes[121] and the de novo assembled contigs were aligned against NCBI reference genomes using bowtie2[122].

3.3.4 Statistics and data analysis

We calculated the correlation between the relative abundances of the sequence variants and FUT2 genotype by regression analysis in R. We employed UniFrac distances to examine microbiome structure between samples to test differences between groups using multivariate statistical methods such as principal coordinate analysis (PCoA) and permutational multivariate analysis of variance (PERMANOVA). To identify the predictive value of genotype on 16S rRNA and ITS microbiome community composition, we applied random forest machine learning (after rarefying to 5000 sequences/sample, 1000 trees) and Analysis of Composition of Microbiomes (ANCOM;[91]). To probe 16S rRNA and ITS co-occurrence patterns, we sub-selected the samples containing a minimum of 1000 16S reads and 1000 ITS reads from the 24-month stool samples to generate a network using SparCC[123]. We then applied partial CCA analysis to identify co-abundant groups that were associated with secretion status after FDR correction, controlling for delivery mode, duration of breastfeeding, gender, antibiotic exposure in the first year and birth weight, using Monte Carlo Permutation tests, 9999 permutations. Next we regressed the proportions of co-abundant groups against genotype using MaAsLin[124].

To analyze the metagenomic short reads, we performed MetaPhlaN2[119] and HumanN2[120] to annotate taxonomic classification and putative functions, respectively. We used LDA Effective Size (LEfSe) to identify differentially abundant taxa based on MetaPhlaN results[125]. We performed pairwise T-tests with FDR correction to examine differences between percent of metagenomics contigs mapping to NCBI reference genomes.

3.4 Results

3.4.1 Overview of the childhood cohort

Child characteristics (Table 3.1) were determined for the population of 399 24-month-old toddlers, whose metrics and stool samples were acquired as part of the GUSTO study[118]. Of these 399, 133 had an additional stool sample at 48 months.

In total, 8,333,749 16S rRNA and 5,773,371 ITS rRNA amplicon reads were generated across all 532 samples, with a range of 0 to 47,033 (median 16,637; mean 16,601) reads for 16S rRNA and 1 to 73,291 (median 7,997; mean 12,053) reads for ITS rRNA. Sequence depth was rarified to 2500 for 16S rRNA and 1000 for ITS rRNA, which resulted in 500 samples for 16S rRNA analysis and 389 samples used for ITS analysis. Reads were classified into Exact Sequence Variants (ESVs) using DADA2, creating a total of 3277 16S rRNA ESVs, with a range of 2 to 539,761 reads per ESV (median 42; mean 2,543), and 1117 ITS rRNA ESVs, with a range of 1 to 2,075,216 reads per ESV (median 65; mean 5,168). All ESVs were annotated to known microbial taxa against the GreenGenes database for 16S, UNITE version 7 for ITS. We also generated a total of 5,180,189,689 shotgun metagenomic sequence reads, with a range of 309 to 12.77 million reads (median 3.66; mean 3.74 million). Following assembly, there were a range of 6 to 50933 contigs/sample (median 18570; mean 19436), median total length was 42636269 bp, with a mean total length was 44578782 bp. The GC content (%) ranged from 41.5% to 58.3% and N50 ranged from 551 to 89154 with a median of 6504 and mean value of 8796 (statistics calculated in QUAST[126]).

3.4.2 Alpha and beta diversity

No significant differences in 16S rRNA or ITS alpha diversity were observed by secretor status using Shannon, Simpson, or phylogenetic distance metrics (Figure 3.1). Similarly, when comparing the childrens microbiota (16S and ITS) between either maternal genotype (secretor status of the mother), the sex of the child, or the duration of breast feeding,

Figure 3.1: Shannon alpha diversity between a) 16S and b) ITS samples did not differ by FUT2 genotype at 24 months. $p > 0.05$.

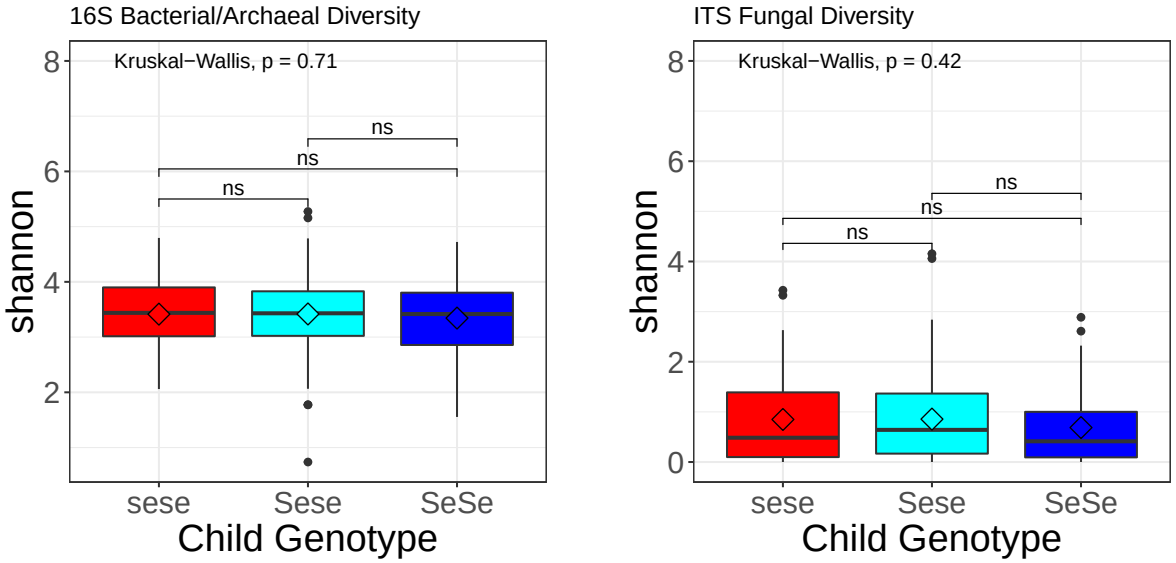


Figure 3.2: PCoA plot of Bray-Curtis dissimilarity on 16S taxonomy.

no significant difference in alpha diversity was observed. However, there was a significant increase in alpha diversity at 48 months of age compared to 24 months of age ($p = 2.926446e-16$).

Metagenomic gene alpha diversity was also calculated, which showed that the Shannon index of gene families did not differ significantly by maternal genotype (secretor status of the mother), the sex of the child, or the duration of breast feeding, but did differ between infant genotypes ($p = 0.0058$), with nonsecretors (sese) and heterozygotes (Sese) having significantly greater gene family alpha diversity (Shannon index) when compared to homozygous secretors (SeSe). However, the number of observed gene families was not significantly different by infant genotype or secretor status, implying that the differences stemmed from the distribution of gene family evenness. Indeed, sese had significantly greater gene family evenness ($p = 0.0377$) compared to SeSe, who had significantly lower evenness compared to Sese ($p = 0.0017$) or sese children ($p = 0.0005$).

As with alpha diversity, microbial beta diversity was significantly different between the

Table 3.2: p-values of beta diversity metrics by maternal genotype, PERMANOVA, 9999 permutations per test.

Secretor status	**Child genotype**	**Maternal genotype**	**Bray-Curtis**	**Weighted Unifrac**	**Unweighted Unifrac**
Non-secretors	sese	Sese vs. sese	0.7665	0.4785	0.0840
Secretors	Sese	Sese vs. sese	0.9175	0.9925	0.8763
	Sese	SeSe vs. Sese	0.8562	0.6539	0.9978
	Sese	SeSe vs. sese	0.6848	0.5193	0.7196
	SeSe	Sese vs. SeSe	0.4742	0.3045	0.3533

Table 3.3: Other characteristics affecting the child gut microbiome at 24 months. p-values were obtained using the Monte-Carlo permutation test in partial CCA. Additive FUT2 genotype was used as the covariate.

Characteristic	**p-value**
Duration of any breastfeeding	0.001
Delivery mode (Vaginal vs. Caesarean section)	0.027
Birth weight	0.033
Gender (Female vs. Male)	0.097
Gestational age (weeks)	0.197

different populations at both time points. Bray Curtis distance using either 16S rRNA and ITS rRNA data were significantly different between the non-secretors and secretors (either homozygous or heterozygous; p¡0.05, PERMANOVA)(Figure 3.2). There was no significant difference in beta diversity between the homozygotic and heterozygotic secretors. No difference in 16S or ITS rRNA beta diversity was observed with mothers secretor status, child gender, or duration of breast feeding (Table 3.2, Table 3.3).

Pairwise PERMANOVA on functional gene beta-diversity using Bray-Curtis dissimilarity showed significant differences between secretors and nonsecretors ($p = 0.018$). Furthermore, a significant difference was found between homozygous secretors (SeSe) and nonsecretors (sese; $p = 0.001$) but not between heterozygous secretors (Sese) and nonsecretors (sese) ($p =$

0.08). Using Jaccard dissimilarities, the same pattern was detected (significant difference between secretors and nonsecretors, $p = 0.007$; significant difference between SeSe and sese groups, $p = 0.004$).

3.4.3 Microbiota taxonomy associates with secretor status

Using ANCOM we compared the proportions of 16S rRNA and ITS ESVs between secretor genotypes to determine if specific taxa were significantly differentially abundant. ANCOM employs multiple-hypothesis correction and allows for co-variates to be controlled for. Here we controlled for maternal genotype, duration of breastfeeding, child gender, delivery mode, and the number of courses of antibiotics received in the first year. For 16S rRNA, ANCOM identified 29 ESVs whose proportions were significantly different between infants when controlling for relevant co-variates (Table 3.4). Specifically, ESVs annotated to *Ruminococcus gnavus*, *Escherichia coli*, and genera *Streptococcus*, *Blautia*, *Bifidobacterium*, *Enterococcus*, and families Erysipelotrichaceae and Peptostreptococcaceae were at a greater proportion in secretors. However, different ESVs of *Streptococcus*, *Blautia*, *Bifidobacterium*, *Ruminococcus gnavus*, and Erysipelotrichaceae were also relatively enriched in non-secretors, implying that the 16S rRNA gene provided inadequate resolution to distinguish between these organisms. For ITS rRNA, ANCOM identified no fungal ESVs significantly different between child FUT2 genotypes.

At 48 months, only a single ESV differed significantly between child genotypes: a *Blautia* ESV, which was found to be more abundant in secretors. At both time points, Random Forest models on 16S rRNA or ITS gene abundance were no better than chance at predicting secretor status or child genotype. This implies that despite certain taxa that differ significantly between secretor status or genotypes, they are insufficiently predictive in the context of the whole community structure.

Table 3.4: ESVs significantly different between genotypes by ANCOM.

Taxonomy	**Lowest in**	**Highest in**
Bifidobacterium	sese	SeSe
Bifidobacterium	sese	SeSe
Bifidobacterium	SeSe	sese
Blautia	sese	SeSe
Blautia	SeSe	sese
Collinsella aerofaciens	sese	SeSe
Dorea	SeSe	sese
Eggerthella lenta	SeSe	sese
Eggerthella lenta	sese	SeSe
Enterobacteriaceae	SeSe	sese
Enterococcus	sese	SeSe
Enterococcus	sese	SeSe
Erysipelotrichaceae	SeSe	sese
Erysipelotrichaceae	SeSe	sese
Erysipelotrichaceae	sese	SeSe
Escherichia coli	Sese	SeSe
Escherichia coli	sese	SeSe
Eubacterium dolichum	SeSe	sese
Haemophilus parainfluenzae	SeSe	sese
Peptostreptococcaceae	sese	SeSe
Peptostreptococcaceae	SeSe	Sese
Peptostreptococcaceae	SeSe	Sese
Peptostreptococcaceae	Sese	SeSe
Pseudoramibacter Eubacterium	sese	SeSe
Ruminococcus gnavus	SeSe	sese
Ruminococcus gnavus	Sese	SeSe
Streptococcus	SeSe	sese
Streptococcus	SeSe	sese
Streptococcus	sese	SeSe

3.4.4 16S-ITS co-occurrence

We sub-selected the samples containing a minimum of 1000 16S reads and 1000 ITS reads from the 24-month stool samples to generate a network using SparCC[123]. 220 16S ESVs (present in 5% of samples) and 31 ITS ESVs (present in ≤2% of samples) clustered into 25 co-abundant groups (CAGs) (SparCC and PERMANOVA, $p < 0.001$). (Figure 3.3a). At 24-months, none of the CAGs maintained significant co-association patterns between 16S and ITS ESVs. Partial CCA analysis identified seven co-abundant groups that were associated with secretion status after FDR correction, while controlling for delivery mode, duration of breastfeeding, gender, antibiotic exposure in the first year and birth weight (Figure 3.3b). The most abundant ESV in each group was: CAG25_Bifidobacterium, CAG24_Blautia, CAG22_Enterococcus, CAG23_Akkermansia, CAG1_Blautia, CAG7_Bifidobacterium, CAG8_Dorea. CAG25_Bifidobacterium was positively associated with secretors (coefficient 0.116), while CAG7_Bifidobacterium was negatively associated with secretors (coefficient -0.082). Similarly, CAG1_Blautia and CAG24_Blautia (Figure 3.3c) showed differential trends, which recapitulated the results from the ANCOM analysis.

3.4.5 Functional genetic potential associates with secretor status

For the shotgun metagenome short reads, we ran Linear Discriminate Effect Size analysis (LefSe) on the MetaPhlan annotations, which demonstrated a significant enrichment of bacterial genomic data annotated to *Bacteroides uniformis*, *Haemophilus parainfluenzae*, *Collinsella tanakaei*, *Ruminococcus gnavus*, *Clostridia*, and *Bifidobacterium longum*; meanwhile similar to the 16S data, *Enterococcus* organisms were significantly enriched in secretors (Fig. 3.4a). When we performed LDA on the HumanN2 annotations to determine which gene families were enriched in each genotype, there were no significant differences in gene families annotated to fucose-related functions. Instead, the only gene families that were differentially abundant were housekeeping genes specific to the taxonomic differences observed by MetaPhlaN (Fig. 3.4b). Therefore, the short read metagenomics analysis did not identify

Figure 3.3: a) SparCC co-abundance network. Positive correlations are indicated by grey edges; negative correlations are in red. 16S taxa are circles and ITS taxa are squares. b) Partial CCA identified significant co-abundant groups (CAGs) that associate with child FUT2 genotype and secretor status, $p = 0.001$ by Monte Carlo permutation, 9999 permutations. c) Heatmap of the abundance of co-abundant groups with significant MaAsLin coefficients. CAG30_Blautia is more abundant in SeSe; CAG1_Blautia is more abundant in sese; CAG31_Bifidobacterium is most abundant in SeSe; CAG8_Bifidobacterium is most abundant in sese.

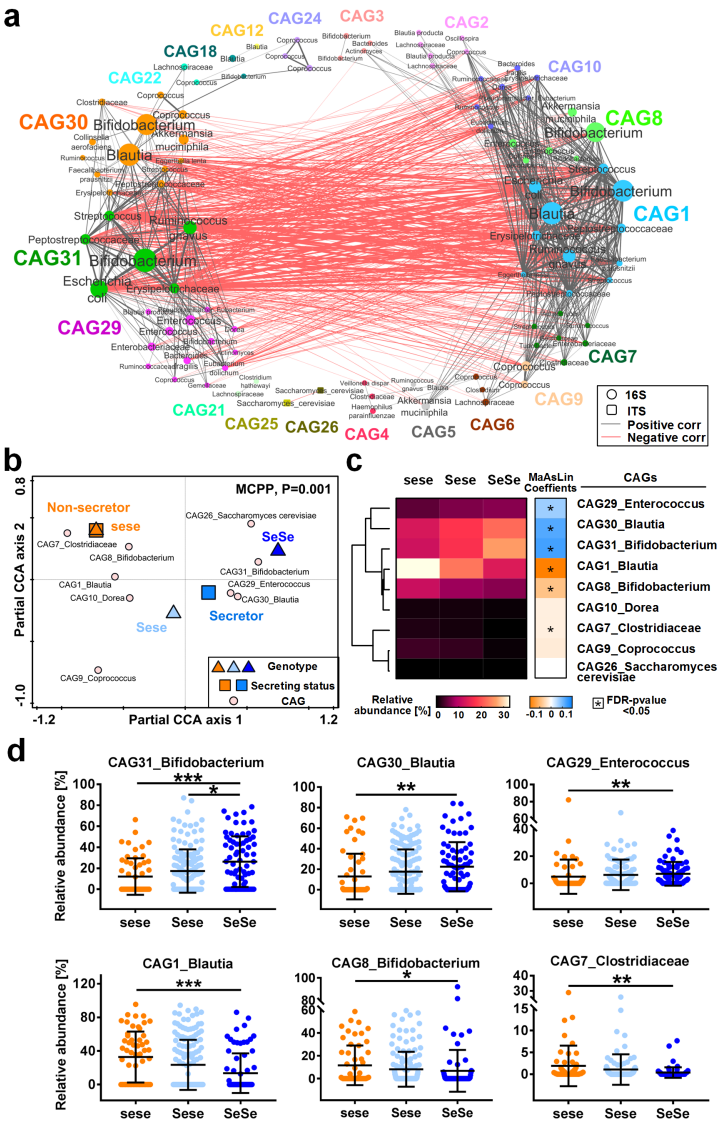
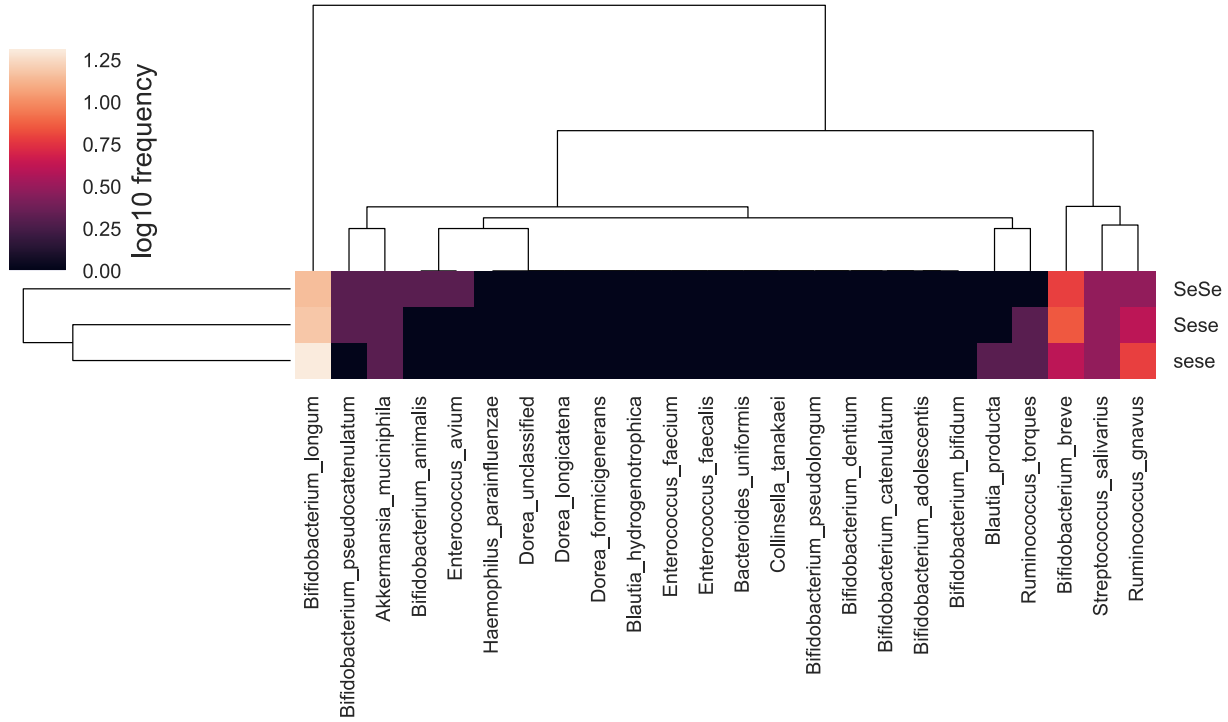


Figure 3.4: Heatmap of top 25 most abundant MetaPhlaN species by child FUT2 genotype.

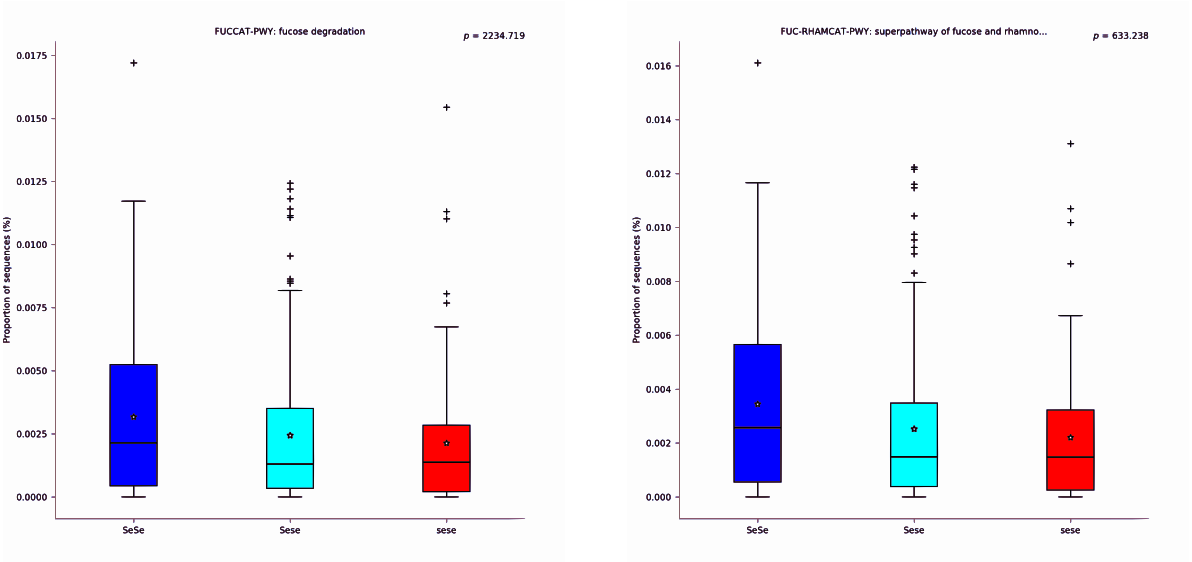


significant functional differences among the FUT2 genotype. Furthermore, a Random Forest model on HumanN2 gene families demonstrated a predictive accuracy of 0.78, suggesting that functional gene profile was predictive of secretor status. The five most important features were proteins of unknown function mapped to *Clostridioides difficile*, *Blautia obeum*, *Ruminococcus gnavum*, and *Dorea* spp.

3.4.6 Comparing genomes by ecological niche

As ESVs and co-abundant groups annotated to the same genus showed opposite associations with secretor status, we originally hypothesized that different species or strains of each genus would show marked differences in their ecological response to fucose availability. We tested this hypothesis by using MetaPhlan on the metagenomics short reads to yield species-level taxonomic annotation. We found different abundances of taxa within genera *Bifidobacterium*, *Blautia*, *Ruminococcus*, and *Streptococcus*. Different species of each

Figure 3.5: The proportion of metagenome short reads mapping to a) fucose degradation pathway and b) superpathway of fucose and rhamnose degradation did not differ between the genotypes $p > 0.05$.



genus showed significant differences in proportion by secretor status (Figure 3.3). The top 25 species assignments showed a significantly greater proportion of *Blautia producta*, *Ruminococcus torques*, and *Ruminococcus gnavus* in nonsecretors. Meanwhile, secretors had a greater abundance of reads mapping to *Bifidobacterium pseudocatenulatum*, *Bifidobacterium animalis*, *Enterococcus avium*, and *Bifidobacterium breve*. However, the overall distribution of reads mapping to annotated fucose pathways did not differ between child FUT2 genotypes (Figure 3.5). Therefore, the opposite associations of certain ESVs of *Bifidobacterium* and *Blautia* observed in 16S rRNA ESV abundance data was probably attributable to species-level taxonomic differences. To further explore the metagenomics species-level associations, we assembled the reads and aligned assembled contigs against NCBI reference genomes of *Bifidobacterium* species, including *B. adolescentis*, *B. animalis*, *B. bifidum*, *B. breve*, *B. catenulatum*, *B. longum*, and *B. pseudocatenulatum*. A significantly greater proportion of metagenomic contigs (2.2%-3.8%; mean 3.04%; median 2.91%) mapped to these genomes using data from homozygous secretors (SeSe), compared to heterozygotes and nonsecretors (Sese=1.9-3.0%; mean 2.46%; median 2.44%; sese=1.8-3.1%; mean 2.47%; median 2.42%;

Table 3.5: Percent of assembled contigs from each genotype mapping to NCBI reference genomes.

Reference Genome	**SeSe**	**Sese**	**sese**
Bifidobacterium adolescentis	3.09	2.62	2.59
Bifidobacterium animalis	2.21	1.85	1.84
Bifidobacterium bifidum	2.91	2.44	2.42
Bifidobacterium breve	3.67	2.94	2.96
Bifidobacterium catenulatum	2.75	2.17	2.14
Bifidobacterium longum	3.77	2.98	3.11
Bifidobacterium pseudocatenulatum	2.9	2.27	2.24
Blautia hydrogenotrophica	1.22	1.2	1.23
Blautia producta	3.15	3.33	4.34
Ruminococcus gnavus	4.09	3.7	3.89
Ruminococcus torques	1.88	2.05	2.13

Table 3.5; $p < 0.001$). The nonsecretor and heterozygous secretor genotypes showed no significant differences in mapping percentage. Interestingly, even for species that were identified as enriched in nonsecretor metagenomes (using MetaPhlan), the percentage of secretor (SeSe) contigs that mapped was still greater than nonsecretor (sese) contigs. We did not observe greater alignment of any bifidobacterial genomes with the nonsecretors, even for those species that were observed to have higher proportions by 16S rRNA and short read sequencing. This implies that the mapping technique did not fully capture functional differences, and that the majority of contigs mapped to conserved bifidobacterial sequences rather than species-specific genes.

3.5 Discussion

With the move toward personalized medicine, it is important to understand how individual differences, including host genetics, influence the functional potential of the human microbiome. Fucosyltransferase 2 (FUT2) is a key gene that has been identified as playing a role in host-microbial interactions, and shows conserved distribution in global human populations. However, previous studies have presented conflicting evidence on whether the microbiome shows a significant association with differences in the FUT2 genotype. These

previous studies employed only 16S rRNA marker gene analysis; we expanded on this by using both 16S and ITS rRNA analysis to further target the fungal populations, as well as using shotgun metagenomic analysis to unpack the impact of FUT2 host genotype on the functional potential of the microbiome.

We identified a few bacterial taxa that varied by FUT2 secretor genotype in 24-month old children. However, most of these differences were abrogated in the 48-month old samples, with the exception of a single variant mapping to *Blautia*. No significant differences between secretor status were observed for any fungal taxa at either timepoint.

Interestingly, we identified the same bacterial genera that presented with opposite proportional trends by secretor status. We hypothesized that this could represent differences in species abundance within these genera. This would suggest that different species maintain different functional potential in response to fucose availability. The metagenomic short read analysis showed species-level differences; however, we were unable to identify any differences in functional pathways between the differentially abundant species. We had a sufficient sample size in our cohort such that the taxonomic differences in the microbial community composition observed between the FUT2 child genotypes are unlikely to be stochastic founder taxa effects. We conclude that the taxonomic differences may contribute to differences in functional potential outside of known fucose pathways. Future work will capitalize on the assemblies from the metagenomics samples to compare *Bifidobacterium* and *Blautia* pangenomes from this cohort and determine if the metabolic functional potential differs across these samples. Furthermore, experiments in genetic knockout animal models can help uncover the mechanistic explanation for the *Bifidobacterium* niche differentiation observed.

A strength of this study was the deep sequencing done on a relatively large cohort. All of our subjects had the same ethnicity and therefore the same FUT2 polymorphism. Limitations of this study included its cross-sectional nature (one or two timepoints instead of longitudinal sampling) and the samples were all taken post-weaning. The literature indicates that maternal secretor phenotype affects bifidobacterial abundance in breastfed infants[30],

whereas we did not see any effect of maternal FUT2 genotype on the childrens gut microbiome. This implies that the early expansion of bifidobacteria in secretor-fed infants may not persist.

CHAPTER 4

DISCUSSION

4.1 Key findings

Altogether, the projects described in this thesis highlight the importance of studying microbiome dynamics, not just composition, and the potential role of host genetics in selecting for microbial function. Preterm infants are an interesting ecological problem with abnormal microbial colonization and succession, likely because they are exposed to microbes and the extra uterine environment earlier in gestation than term infants, and they are frequently given antibiotics and other medical interventions. Similar to our study, others have noted that the preterm infant gut microbiome is highly personalized[127]. In our study, there was high intra-individual variability and postmenstrual age described the majority of variance in beta diversity over time. We identified associations between microbiome composition and infant growth outcomes, implying that there may be a link between the microbiota and preterm infant development. For example, we found that infants with higher proportions of *Klebsiella* and *Staphylococcus* had less weight gain during the NICU stay. Both of these taxa include known pathogens, so it is possible that a higher proportion of these taxa indicates dysbiosis, a state that has been previously associated with poor growth outcomes. We also found an association between microbiome beta diversity volatility and infant growth trajectory (change in length z-scores). Catch up growth is particularly important for preterm infants, and it is believed that growth in certain windows of time can be beneficial to later health outcomes [100,101]. Poor weight and length gain over time can be seen as indicators of nutritional stress, so the associations we observed suggest that microbiome development may be related to infant nutrition, which can have long lasting effects on childhood health. Additional research should target the relationship between VLBW length and weight gain and microbiome development.

The interplay between host genetics and the microbiome is another area in which many

questions remain. The effect of individual differences on the microbiome is an important factor to consider as the field moves towards personalized medicine. Specific host genes such as FUT2 have been identified as playing a role in host-microbe interactions. In septic infants, FUT2 genotype has been shown to affect mortality risk[117], and FUT2 non-secretor status is protective against some pathogenic infections[50,52-54,129]. However, a number of studies have failed to find differences in the gut microbiota composition[115,116]. These opposing results raise the question of what interactions might occur between the microbiota and the immune system. We identified differences in microbial community composition and structure between children with different FUT2 genotypes. Interestingly, bacteria from the same genus were differentially abundant in FUT2 secretors and FUT2 non-secretors, providing evidence for niche differentiation.

The early life window is a crucial time to study the microbiome and it is critical to consider the myriad factors that can contribute to microbial development. Plenty of questions remain about the human microbiome in the first few years of life. Further prospective studies are needed to assess microbiome maturation in preterm infants, as we still do not understand what normal microbiome development looks like in this population. Understanding changes and shifts in the microbiome over time may be equally, if not more important, than solely focusing on the identity of microbial functions and taxa. Focusing on functions, rather than taxa, may be critical in addressing some clinical questions[130], because microbial functions seem to be conserved even when there are no core taxa present between individuals. Our work suggests that taxonomic surveys of the microbiome may be insufficient to reveal associations between microbial niches and human genotype, which would be more appropriately suited to functional microbial surveys.

4.2 Limitations of the current work

Each aim of this thesis had unique limitations. For example, in Chapter 2, for the preterm infant cohort we had a relatively small sample size overall, and lost many patients to follow

up by the 2 and 4 year time points. This left us underpowered to address many comorbidities: conditions like necrotizing enterocolitis, intraventricular hemorrhage, retinopathy of prematurity, and others, could reasonably be hypothesized to have significant effects on the microbiome (either directly or through the associated therapeutic interventions) but we only had a few infants per condition. The goal of this aim was to be longitudinal on a long time-scale, but we only had consistent sampling over the NICU stay for most of the individuals. In Chapter 3, the FUT2 cohort, we had a cross-sectional study design with a larger sample size, but ultimately only a single sample per time point for each individual. While we did have several individuals with a 48-month time point, which approximately recapitulated the genotype ratios, we had a smaller sample size and therefore lower power. Also, with only two time points two years apart we lacked sufficient temporal resolution, so the study was still primarily cross-sectional.

Microbiome dynamics in the early life time period are just beginning to be studied, with much prior work representing cross-sectional single time point birth cohorts. Such study design can be harnessed for statistical associations[2,131] but fail to establish causal relationships. Key principles in ecology that apply to microbial communities are those of the stability (resistance to change) and resilience (the ability to return to initial state following perturbation)[2]. Longitudinal studies with denser temporal sampling and large sample sizes are important to improve our understanding of the microbiome in early life.

The gut microbiome has been shown to be highly personalized to individuals, and furthermore, temporal variability is also personalized[132]. Variability in microbial community composition among individuals generally exceeds intra-individual variability over time, but temporal variability within an individual is considerable[132]. Therefore, studies that collect samples at a single time point may be insufficient to characterize each individuals gut microbiome. In studies that examine large changes in the microbiome in response to interventions or disease states, inter-individual and intra-individual variability may be irrelevant, but for studies like those presented in this thesis, when health (and not disease) is the primary

outcome, shifts in the microbial community might be more subtle and could be eclipsed by inter and intra-individual variability. Some of these variations can be abrogated by using each individual as their own control (before and after treatment, or sampling over time), but the tradeoff would be a necessary reduction in the number of subjects, in part due to the cost of sample processing. Therefore, longitudinal vs. cross-sectional study design is an important consideration for future studies, especially about the early life time period, when intra-individual variability is higher than in adults.

In addition to the constraints of the sampling design, the type of sequencing can also affect the types of conclusions that can be drawn. 16S rRNA gene sequencing is rapid and cost-effective and allows inferences about the taxonomic identity and relative proportions of the bacterial and archaeal populations present in a sample, while whole-metagenome shotgun analyses generate short reads of the genomes of all microorganisms present in a sample[133,134]. 16S rRNA is a useful tool to get a glimpse at who is present, while shotgun metagenomics reveals information about what they might be doing, by providing information about the functional genetic potential of the community. Other approaches, such as metatransciptomics, highlight which functional genes are actually being transcribed, while metabolomics can characterize the products of microbial metabolism. The future of microbiome research lies in integrating multiple 'omics technologies to paint a more complete picture. In Chapter 2, we leveraged 16S rRNA sequencing to generate a detailed survey of the relative abundances of bacteria present in each sample. However, we were limited by an absence of functional gene information. Furthermore, because marker gene analysis relies on amplification, it cannot yield absolute abundance data and only gives relative abundance information. A variety of statistical tools have been developed to overcome this limitation, by taking into account the compositional and zero-inflated nature of microbiome datasets[91,135,136]. In Chapter 3, we used both 16S rRNA and ITS marker gene sequencing as well as shotgun metagenomics sequencing to characterize each sample. This gave us a picture of taxonomic abundances as well as a gene-based inventory and analysis to ex-

plore functional mechanisms of action. Within shotgun metagenomics analysis, differences in methods and approach can also affect the results. For example, we used primarily short read data throughout this thesis. An alternative approach that will be a major component of future work is metagenome assembly. The pros of assembly over read mapping include providing genomic context, lower risks of false positives, and the potential to discover new taxa, while cons are that assembly is computationally expensive, and imposes a higher limit of detection (a lack of assembly does not imply absence from a sample)[137].

There are some challenges to metagenome assembly. In a genomic study or a sample from a single culture, it may be assumed that all reads are from the same original genome, while in metagenomic studies, it is difficult to distinguish closely related species or strains[138]. Related species or strains can introduce overlaps, rendering assembly more complex, and can result in chimeric contigs[138]. The assembly tool we have used, metaSPAdes, does not use the same scaffolding process as genomic assembly, but instead uses paired reads to resolve chimeras[121,139]. The set of contigs output from the assembler is then grouped into sets (bins) to represent species or strains. The binning methods we employed, MetaBAT[140] and CONCOCT[141], leverage sequence composition and coverage across multiple samples to power their binning algorithms. We then usedCheckM, which uses the presence of marker genes to assess assembly quality and also uses the position of a genome within a reference genome tree to improve accuracy[142]. There are a plethora of metagenome assembly approaches and new tools constantly under development, all in the hopes of increasing the precision of reconstituting thousands of genomes from short reads. By using both short reads and assembly-based approaches, we hope to maximize our understanding of the functional genetic profile of our samples.

4.3 Future directions

A natural extension of the work presented in this thesis is to do further metagenomics analysis on each of these cohorts. For Chapter 2, the preterm infant cohort, a smaller subset

of infant samples (three infants at five NICU timepoints and an additional five infants at discharge, 2 year, and 4 year) have been sent for metagenomic sequencing at high enough depth to allow assembly. For Chapter 3, the FUT2 cohort, all of the samples have been sequenced via shotgun and assemblies have been completed. Ongoing work includes a pangenomic approach to compare the coverage of genes between the different bifidobacterial species. This type of analysis will be useful to determine enzymes and metabolites that could be targeted with metabolomics.

Another experimental approach that could establish a mechanistic link between FUT2 genotype and the microbiome would take advantage of a mouse model. FUT2-/- knockout mice have been bred and other groups have successfully rederived them in germ free isolators[49,55]. Such germ free mice could be humanized or gavaged with microbiota samples from our FUT2 cohort. One experiment could be to gavage stool from FUT2+/+ or FUT2-/- donors into mice of each background. We would hypothesize that the FUT2+/+ stool in a FUT2+/+ mouse would be most similar to the original inoculate (bearing in mind that any humanized mouse naturally selects for strains best adapted to the murine host[143]), and FUT2-/- stool in a FUT2-/- mouse would also recapitulate the stool population. However, we would expect to see loss of bifidobacteria and other bacterial populations with the functional capacity to use fucose from the FUT2+/+ stool in FUT2-/- mice over time, presuming there would not be intestinal fucose present to act as a prebiotic. Likewise, FUT2-/- stool in a FUT2+/+ mouse might lead to the selective expansion of minority bifidobacterial populations that do use fucose as a carbon source. Humanized mice do have limitations; for example, their gut-associated immune systems are altered, with deficits in antibody production, fewer and smaller Peyer's patches and mesenteric lymph nodes, and altered patterns of microvilli formation, among other developmental defects[58]. Most importantly, the question of which transferred microbes are well adapted to the mouse to engraft in the gut, might bias the results, since presumably bifidobacterial genera with opposite abundance trends in FUT2+/+ and FUT2-/- humans could also differ in their ability to persist in mice. How-

ever, humanized mice used in this type of experiment would furnish convincing evidence of a causal link between FUT2 genotype and microbiome composition.

4.4 Conclusions

Overall, early life is a critical window of time in microbiome development and deserves further study. It appears that exposure to microbes during and immediately following birth is vital to priming the immune system, which can have lifelong health effects. Furthermore, from an ecological standpoint, the earliest colonizer taxa establish niches and can persist long term. As laid out in this thesis, both events surrounding birth and individualized human genotypes seem to associate with the microbiome in early life, indicating that these factors require consideration in future study.

REFERENCES

[1] Luke K Ursell, Jessica L Metcalf, Laura Wegener Parfrey, and Rob Knight. Defining the human microbiome. *Nutrition reviews*, 70(Suppl 1):S38–S44, 8 2012. ISSN 0029-6643. doi: 10.1111/j.1753-4887.2012.00493.x. PMID: 22861806 PMCID: PMC3426293.

[2] Jack A. Gilbert, Martin J. Blaser, J. Gregory Caporaso, Janet K. Jansson, Susan V. Lynch, and Rob Knight. Current understanding of the human microbiome. *Nature Medicine*, 24(4):392–400, 4 2018. ISSN 1546-170X. doi: 10.1038/nm.4517.

[3] Noel T. Mueller, Elizabeth Bakacs, Joan Combellick, Zoya Grigoryan, and Maria G. Dominguez-Bello. The infant microbiome development: mom matters. *Trends in molecular medicine*, 21(2):109–117, 2 2015. ISSN 1471-499X 1471-4914. doi: 10.1016/j.molmed.2014.12.002. PMID: 25578246 PMCID: PMC4464665.

[4] Martin J. Blaser and Maria G. Dominguez-Bello. The human microbiome before birth. *Cell Host & Microbe*, 20(5):558–560, 11 2016. ISSN 1934-6069. doi: 10.1016/j.chom.2016.10.014. PMID: 27832586.

[5] Kjersti Aagaard, Jun Ma, Kathleen M. Antony, Radhika Ganu, Joseph Petrosino, and James Versalovic. The placenta harbors a unique microbiome. *Science Translational Medicine*, 6(237):237ra65–237ra65, 5 2014. doi: 10.1126/scitranslmed.3008599.

[6] Abigail P. Lauder, Aoife M. Roche, Scott Sherrill-Mix, Aubrey Bailey, Alice L. Laughlin, Kyle Bittinger, Rita Leite, Michal A. Elovitz, Samuel Parry, and Frederic D. Bushman. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome*, 4(1):29, 6 2016. ISSN 2049-2618. doi: 10.1186/s40168-016-0172-3. PMID: 27338728 PMCID: PMC4917942.

[7] Jacob S. Leiby, Kevin McCormick, Scott Sherrill-Mix, Erik L. Clarke, Lyanna R. Kessler, Louis J. Taylor, Casey E. Hofstaedter, Aoife M. Roche, Lisa M. Mattei, Kyle Bittinger, Michal A. Elovitz, Rita Leite, Samuel Parry, and Frederic D. Bushman. Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome*, 6(1):196, 10 2018. ISSN 2049-2618. doi: 10.1186/s40168-018-0575-4.

[8] Laura Moles, Marta Gomez, Hans Heilig, Gerardo Bustos, Susana Fuentes, Willem de Vos, Leonides Fernandez, Juan M. Rodriguez, and Esther Jimenez. Bacterial diversity in meconium of preterm neonates and evolution of their fecal microbiota during the first month of life. *PloS one*, 8(6):e66986, 2013. ISSN 1932-6203 1932-6203. doi: 10.1371/journal.pone.0066986. PMID: 23840569 PMCID: PMC3695978.

[9] Fredrik Bckhed, Josefine Roswall, Yangqing Peng, Qiang Feng, Huijue Jia, Petia Kovatcheva-Datchary, Yin Li, Yan Xia, Hailiang Xie, Huanzi Zhong, MuhammadTanweer Khan, Jianfeng Zhang, Junhua Li, Liang Xiao, Jumana Al-Aama, Dongya Zhang, YingShiuan Lee, Dorota Kotowska, Camilla Colding, Valentina Tremaroli, Ye Yin, Stefan Bergman, Xun Xu, Lise Madsen, Karsten Kristiansen, Jovanna Dahlgren, and Jun Wang. Dynamics and stabilization of the human gut microbiome during the

first year of life. *Cell Host & Microbe*, 17(6):852, 6 2015. ISSN 1931-3128. doi: 10.1016/j.chom.2015.05.012.

[10] Maria G. Dominguez-Bello, Elizabeth K. Costello, Monica Contreras, Magda Magris, Glida Hidalgo, Noah Fierer, and Rob Knight. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences*, 107(26):11971–11975, 6 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1002601107. PMID: 20566857.

[11] Hedvig E. Jakobsson, Thomas R. Abrahamsson, Maria C. Jenmalm, Keith Harris, Christopher Quince, Cecilia Jernberg, Bengt Bjorksten, Lars Engstrand, and Anders F. Andersson. Decreased gut microbiota diversity, delayed bacteroidetes colonisation and reduced th1 responses in infants delivered by caesarean section. *Gut*, 63(4):559–566, 4 2014. ISSN 1468-3288 0017-5749. doi: 10.1136/gutjnl-2012-303249. PMID: 23926244.

[12] Fabien Aujoulat, Laurent Roudiere, Jean-Charles Picaud, Aurelien Jacquot, Anne Filleron, Dorine Neveu, Thierry-Pascal Baum, Helene Marchandin, and Estelle Jumas-Bilak. Temporal dynamics of the very premature infant gut dominant microbiota. *BMC microbiology*, 14:325, 2014. ISSN 1471-2180 1471-2180. doi: 10.1186/s12866-014-0325-0. PMID: 25551282 PMCID: PMC4326509.

[13] Cecilie Dahl, Hein Stigum, Jrgen Valeur, Nina Iszatt, Virissa Lenters, Shyamal Peddada, Jrgen V Bjrnholt, Tore Midtvedt, Siddhartha Mandal, and Merete Eggesb. Preterm infants have distinct microbiomes not explained by mode of delivery, breastfeeding duration or antibiotic exposure. *International Journal of Epidemiology*, 4 2018. ISSN 0300-5771, 1464-3685. doi: 10.1093/ije/dyy064. URL `https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyy064/4980972`. [Online; accessed 2018-08-09].

[14] Sally F Bloomfield, Graham AW Rook, Elizabeth A Scott, Fergus Shanahan, Rosalind Stanwell-Smith, and Paul Turner. Time to abandon the hygiene hypothesis: new perspectives on allergic disease, the human microbiome, infectious disease prevention and the role of targeted hygiene. *Perspectives in Public Health*, 136(4):213–224, 7 2016. ISSN 1757-9139. doi: 10.1177/1757913916650225.

[15] Erica D. Sonnenburg, Samuel A. Smits, Mikhail Tikhonov, Steven K. Higginbottom, Ned S. Wingreen, and Justin L. Sonnenburg. Diet-induced extinctions in the gut microbiota compound over generations. *Nature*, 529(7585):212–215, 1 2016. ISSN 1476-4687. doi: 10.1038/nature16504.

[16] Alison N. Thorburn, Craig I. McKenzie, Sj Shen, Dragana Stanley, Laurence Macia, Linda J. Mason, Laura K. Roberts, Connie H. Y. Wong, Raymond Shim, Remy Robert, Nina Chevalier, Jian K. Tan, Eliana Mario, Rob J. Moore, Lee Wong, Malcolm J. McConville, Dedreia L. Tull, Lisa G. Wood, Vanessa E. Murphy, Joerg Mattes, Peter G. Gibson, and Charles R. Mackay. Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. *Nature Communications*, 6:7320, 6 2015. ISSN 2041-1723. doi: 10.1038/ncomms8320.

[17] Derrick M. Chu, Kristen M. Meyer, Amanda L. Prince, and Kjersti M. Aagaard. Impact of maternal nutrition in pregnancy and lactation on offspring gut microbial composition and function. *Gut microbes*, 7(6):459–470, 11 2016. ISSN 1949-0984 1949-0976. doi: 10.1080/19490976.2016.1241357. PMID: 27686144.

[18] Meghan B. Azad, Theodore Konya, Heather Maughan, David S. Guttman, Catherine J. Field, Malcolm R. Sears, Allan B. Becker, James A. Scott, and Anita L. Kozyrskyj. Infant gut microbiota and the hygiene hypothesis of allergic disease: impact of household pets and siblings on microbiota composition and diversity. *Allergy, Asthma, and Clinical Immunology: Official Journal of the Canadian Society of Allergy and Clinical Immunology*, 9(1):15, 2013. ISSN 1710-1484. doi: 10.1186/1710-1492-9-15. PMID: 23607879 PMCID: PMC3655107.

[19] Martin Frederik Laursen, Gitte Zachariassen, Martin Iain Bahl, Anders Bergstrm, Arne Hst, Kim F. Michaelsen, and Tine Rask Licht. Having older siblings is associated with gut microbiota development during early childhood. *BMC microbiology*, 15:154, 8 2015. ISSN 1471-2180. doi: 10.1186/s12866-015-0477-6. PMID: 26231752 PMCID: PMC4522135.

[20] Farzana Yasmin, Hein Min Tun, Theodore Brian Konya, David S. Guttman, Radha S. Chari, Catherine J. Field, Allan B. Becker, Piush J. Mandhane, Stuart E. Turvey, Padmaja Subbarao, Malcolm R. Sears, CHILD Study Investigators, James A. Scott, Irina Dinu, Anita L. Kozyrskyj, S. S. Anand, M. B. Azad, A. B. Becker, A. D. Befus, M. Brauer, J. R. Brook, E. Chen, M. M. Cyr, D. Daley, S. D. Dell, J. A. Denburg, Q. L. Duan, T. Eiwegger, H. Grasemann, K. HayGlass, R. G. Hegele, D. L. Holness, P. Hystad, M. Kobor, T. R. Kollmann, A. L. Kozyrskyj, C. Laprise, W. Y. W. Lou, J. Macri, P. J. Mandhane, G. Miller, T. J. Moraes, P. Par, C. Ramsey, F. Ratjen, A. Sandford, J. Scott, J. A. Scott, M. R. Sears, F. Silverman, E. Simons, P. Subbarao, T. Takaro, S. J. Tebbutt, T. To, and S. E. Turvey. Cesarean section, formula feeding, and infant antibiotic exposure: Separate and combined impacts on gut microbial changes in later infancy. *Frontiers in Pediatrics*, 5, 2017. ISSN 2296-2360. doi: 10.3389/fped.2017.00200. URL `https://www.frontiersin.org/articles/10.3389/fped.2017.00200/full`. [Online; accessed 2019-03-14].

[21] Jennifer C. Stearns, Julia Simioni, Elizabeth Gunn, Helen McDonald, Alison C. Holloway, Lehana Thabane, Andrea Mousseau, Jonathan D. Schertzer, Elyanne M. Ratcliffe, Laura Rossi, Michael G. Surette, Katherine M. Morrison, and Eileen K. Hutton. Intrapartum antibiotics for gbs prophylaxis alter colonization patterns in the early infant gut microbiome of low risk infants. *Scientific Reports*, 7(1):16527, 11 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-16606-9. PMID: 29184093 PMCID: PMC5705725.

[22] Ted Jost, Christophe Lacroix, Christian P. Braegger, Florence Rochat, and Christophe Chassard. Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environmental Microbiology*, 16(9):2891–2904, 9 2014. ISSN 1462-2920. doi: 10.1111/1462-2920.12238. PMID: 24033881.

[23] J. Penders, C. Vink, C. Driessen, N. London, C. Thijs, and E. E. Stobberingh. Quantification of bifidobacterium spp., escherichia coli and clostridium difficile in faecal samples of breast-fed and formula-fed infants by real-time pcr. *FEMS Microbiol Lett*, 243, 2005. doi: 10.1016/j.femsle.2004.11.052. URL http://dx.doi.org/10.1016/j.femsle.2004.11.052.

[24] Sharon M. Donovan, Mei Wang, Min Li, Iddo Friedberg, Scott L. Schwartz, and Robert S. Chapkin. Host-microbe interactions in the neonatal intestine: role of human milk oligosaccharides. *Advances in nutrition (Bethesda, Md.)*, 3(3):450S–5S, 5 2012. ISSN 2156-5376 2161-8313. doi: 10.3945/an.112.001859. PMID: 22585924 PMCID: PMC3649482.

[25] S. Asakuma, E. Hatakeyama, T. Urashima, E. Yoshida, T. Katayama, and K. Yamamoto. Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. *J Biol Chem*, 286, 2011. doi: 10.1074/jbc.M111.248138. URL http://dx.doi.org/10.1074/jbc.M111.248138.

[26] Kyunghun Jeong, Vi Nguyen, and Jaehan Kim. Human milk oligosaccharides: the novel modulator of intestinal microbiota. *BMB reports*, 45(8):433–441, 8 2012. ISSN 1976-670X 1976-6696. PMID: 22917027.

[27] A. M. Zivkovic, J. B. German, C. B. Lebrilla, and D. A. Mills. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *Proc Natl Acad Sci U S A*, 108, 2011. doi: 10.1073/pnas.1000083107. URL http://dx.doi.org/10.1073/pnas.1000083107.

[28] Daniela Barile and Robert A. Rastall. Human milk and related oligosaccharides as prebiotics. *Current opinion in biotechnology*, 24(2):214–219, 4 2013. ISSN 1879-0429 0958-1669. doi: 10.1016/j.copbio.2013.01.008. PMID: 23434179.

[29] Lars Bode. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology*, 22(9):1147–1162, 9 2012. ISSN 1460-2423 0959-6658. doi: 10.1093/glycob/cws074. PMID: 22513036 PMCID: PMC3406618.

[30] Zachery T. Lewis, Sarah M. Totten, Jennifer T. Smilowitz, Mina Popovic, Evan Parker, Danielle G. Lemay, Maxwell L. Van Tassell, Michael J. Miller, Yong-Su Jin, J. Bruce German, Carlito B. Lebrilla, and David A. Mills. Maternal fucosyltransferase 2 status affects the gut bifidobacterial communities of breastfed infants. *Microbiome*, 3(1):1–21, 2015. ISSN 2049-2618. doi: 10.1186/s40168-015-0071-z.

[31] Zachery T. Lewis and David A. Mills. Differential establishment of bifidobacteria in the breastfed infant gut. *Nestle Nutrition Institute workshop series*, 88:149–159, 2017. ISSN 1664-2147. doi: 10.1159/000455399. PMID: 28346936 PMCID: PMC5535791.

[32] C. Palmer, E. M. Bik, D. B. DiGiulio, D. A. Relman, and P. O. Brown. Development of the human infant intestinal microbiota. *PLoS Biol*, 5, 2007. doi: 10.1371/journal.pbio.0050177. URL http://dx.doi.org/10.1371/journal.pbio.0050177.

[33] Pl Trosvik, Nils Christian Stenseth, and Knut Rudi. Convergent temporal dynamics of the human infant gut microbiota. *The ISME Journal*, 4(2):151, 2 2010. ISSN 1751-7370. doi: 10.1038/ismej.2009.96.

[34] P. S. La Rosa, B. B. Warner, Y. Zhou, G. M. Weinstock, E. Sodergren, C. M. Hall-Moore, H. J. Stevens, W. E. Bennett, N. Shaikh, L. A. Linneman, J. A. Hoffman, A. Hamvas, E. Deych, B. A. Shands, W. D. Shannon, and P. I. Tarr. Patterned progression of bacterial populations in the premature infant gut. *Proc Natl Acad Sci U S A*, 111, 2014. doi: 10.1073/pnas.1409497111. URL `http://dx.doi.org/10.1073/pnas.1409497111`.

[35] Jeremy E. Koenig, Ayme Spor, Nicholas Scalfone, Ashwana D. Fricker, Jesse Stombaugh, Rob Knight, Largus T. Angenent, and Ruth E. Ley. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl 1:4578–4585, 3 2011. ISSN 1091-6490 0027-8424. doi: 10.1073/pnas.1000081107. PMID: 20668239 PMCID: PMC3063592.

[36] Isobel Leake. Gut microbiota: Ordered development of gut microbiota observed in premature infants. *Nature reviews. Gastroenterology & hepatology*, 11(10):579, 10 2014. ISSN 1759-5053 1759-5045. doi: 10.1038/nrgastro.2014.152. PMID: 25157624.

[37] B. Brooks, B. A. Firek, C. S. Miller, I. Sharon, B. C. Thomas, R. Baker, M. J. Morowitz, and J. F. Banfield. Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. *Microbiome*, 2, 2014. doi: 10.1186/2049-2618-2-1. URL `http://dx.doi.org/10.1186/2049-2618-2-1`.

[38] Brandon Brooks, Matthew R. Olm, Brian A. Firek, Robyn Baker, Brian C. Thomas, Michael J. Morowitz, and Jillian F. Banfield. Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. *Nature Communications*, 8(1):1814, 11 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-02018-w. PMID: 29180750 PMCID: PMC5703836.

[39] M. Forsgren, E. Isolauri, S. Salminen, and S. Rautava. Late preterm birth has direct and indirect effects on infant gut microbiota development during the first six months of life. *Acta Paediatrica (Oslo, Norway: 1992)*, 106(7):1103–1109, 7 2017. ISSN 1651-2227. doi: 10.1111/apa.13837. PMID: 28316118.

[40] Hans Bisgaard, Nan Li, Klaus Bonnelykke, Bo Lund Krogsgaard Chawes, Thomas Skov, Georg Paludan-Muller, Jakob Stokholm, Birgitte Smith, and Karen Angeliki Krogfelt. Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. *The Journal of allergy and clinical immunology*, 128(3):646–652.e1–5, 9 2011. ISSN 1097-6825 0091-6749. doi: 10.1016/j.jaci.2011.04.060. PMID: 21782228.

[41] J. Blustein, T. Attina, M. Liu, A. M. Ryan, L. M. Cox, M. J. Blaser, and L. Trasande. Association of caesarean delivery with child adiposity from age 6 weeks to 15 years.

International journal of obesity (2005), 37(7):900–906, 7 2013. ISSN 1476-5497 0307-0565. doi: 10.1038/ijo.2013.49. PMID: 23670220.

[42] Yuliya E. Borre, Gerard W. O'Keeffe, Gerard Clarke, Catherine Stanton, Timothy G. Dinan, and John F. Cryan. Microbiota and neurodevelopmental windows: implications for brain disorders. *Trends in molecular medicine*, 20(9):509–518, 9 2014. ISSN 1471-499X 1471-4914. doi: 10.1016/j.molmed.2014.05.002. PMID: 24956966.

[43] Peter J. Turnbaugh, Micah Hamady, Tanya Yatsunenko, Brandi L. Cantarel, Alexis Duncan, Ruth E. Ley, Mitchell L. Sogin, William J. Jones, Bruce A. Roe, Jason P. Affourtit, Michael Egholm, Bernard Henrissat, Andrew C. Heath, Rob Knight, and Jeffrey I. Gordon. A core gut microbiome in obese and lean twins. *Nature*, 457 (7228):480–484, 1 2009. ISSN 0028-0836. doi: 10.1038/nature07540. PMID: 19043404 PMCID: PMC2677729.

[44] Julia K. Goodrich, Jillian L. Waters, Angela C. Poole, Jessica L. Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, William Van Treuren, Rob Knight, Jordana T. Bell, Timothy D. Spector, Andrew G. Clark, and Ruth E. Ley. Human genetics shape the gut microbiome. *Cell*, 159(4):789–799, 11 2014. ISSN 1097-4172. doi: 10.1016/j.cell.2014.09.053. PMID: 25417156 PMCID: PMC4255478.

[45] A. Ferrer-Admetlla, M. Sikora, H. Laayouni, A. Esteve, F. Roubinet, and A. Blancher. A natural history of fut2 polymorphism in humans. *Mol Biol Evol*, 26, 2009. doi: 10.1093/molbev/msp108. URL `http://dx.doi.org/10.1093/molbev/msp108`.

[46] Alline R. Pacheco, Daniela Barile, Mark A. Underwood, and David A. Mills. The impact of the milk glycobiome on the neonate gut microbiota. *Annual review of animal biosciences*, 3:419–445, 2015. ISSN 2165-8110 2165-8102. doi: 10.1146/annurev-animal-022114-111112. PMID: 25387230 PMCID: PMC4349412.

[47] Joseph M. Pickard and Alexander V. Chervonsky. Intestinal fucose as a mediator of hostmicrobe symbiosis. *The Journal of Immunology*, 194(12):5588–5593, 6 2015. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1500395. PMID: 26048966.

[48] Purna C. Kashyap, Angela Marcobal, Luke K. Ursell, Samuel A. Smits, Erica D. Sonnenburg, Elizabeth K. Costello, Steven K. Higginbottom, Steven E. Domino, Susan P. Holmes, David A. Relman, Rob Knight, Jeffrey I. Gordon, and Justin L. Sonnenburg. Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. *Proceedings of the National Academy of Sciences*, 110(42):17059–17064, 10 2013. doi: 10.1073/pnas.1306070110.

[49] M. Tong, I. McHardy, P. Ruegger, M. Goudarzi, P. C. Kashyap, and T. Haritunians. Reprogramming of gut microbiome energy metabolism by the fut2 crohns disease risk polymorphism. *ISME J*, 8, 2014. doi: 10.1038/ismej.2014.64. URL `http://dx.doi.org/10.1038/ismej.2014.64`.

[50] Dermot P. B. McGovern, Michelle R. Jones, Kent D. Taylor, Kristin Marciante, Xiaofei Yan, Marla Dubinsky, Andy Ippoliti, Eric Vasiliauskas, Dror Berel, Carrie Derkowski,

Deb Dutridge, Phil Fleshner, David Q. Shih, Gil Melmed, Emebet Mengesha, Lily King, Sheila Pressman, Talin Haritunians, Xiuqing Guo, Stephan R. Targan, Jerome I. Rotter, and International IBD Genetics Consortium. Fucosyltransferase 2 (fut2) non-secretor status is associated with crohn's disease. *Human Molecular Genetics*, 19(17): 3468–3476, 9 2010. ISSN 1460-2083. doi: 10.1093/hmg/ddq248. PMID: 20570966 PMCID: PMC2916706.

[51] Deborah J. Smyth, Jason D. Cooper, Joanna M. M. Howson, Pamela Clarke, Kate Downes, Trupti Mistry, Helen Stevens, Neil M. Walker, and John A. Todd. Fut2 nonsecretor status links type 1 diabetes susceptibility and resistance to infection. *Diabetes*, 60(11):3081–3084, 11 2011. ISSN 1939-327X. doi: 10.2337/db11-0638. PMID: 22025780 PMCID: PMC3198057.

[52] B. . M. Imbert-Marcille, L. Barbe, M. Dupe, B. Moullac-Vaidye, B. Besse, and C. Peltier. A fut2 gene common polymorphism determines resistance to rotavirus a of the p[8] genotype. *J Infect Dis*, 209, 2013.

[53] Daniel C. Payne, Rebecca L. Currier, Mary A. Staat, Leila C. Sahni, Rangaraj Selvarangan, Natasha B. Halasa, Janet A. Englund, Geoffrey A. Weinberg, Julie A. Boom, Peter G. Szilagyi, Eileen J. Klein, James Chappell, Christopher J. Harrison, Barbara S. Davidson, Slavica Mijatovic-Rustempasic, Mary D. Moffatt, Monica McNeal, Mary Wikswo, Michael D. Bowen, Ardythe L. Morrow, and Umesh D. Parashar. Epidemiologic association between fut2 secretor status and severe rotavirus gastroenteritis in children in the united states. *JAMA pediatrics*, 169(11):1040–1045, 11 2015. ISSN 2168-6203. doi: 10.1001/jamapediatrics.2015.2002. PMID: 26389824 PMCID: PMC4856001.

[54] Joseph M. Pickard, Corinne F. Maurice, Melissa A. Kinnebrew, Michael C. Abt, Dominik Schenten, Tatyana V. Golovkina, Said R. Bogatyrev, Rustem F. Ismagilov, Eric G. Pamer, Peter J. Turnbaugh, and Alexander V. Chervonsky. Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. *Nature*, 514(7524):638–641, 10 2014. ISSN 1476-4687. doi: 10.1038/nature13823. PMID: 25274297 PMCID: PMC4214913.

[55] P. Wacklin, H. Mkivuokko, N. Alakulppi, J. Nikkil, H. Tenkanen, and J. Rbin. Secretor genotype (fut2 gene) is strongly associated with the composition of bifidobacteria in the human intestine. *PLoS One*, 6, 2011. doi: 10.1371/journal.pone.0020113. URL `http://dx.doi.org/10.1371/journal.pone.0020113`.

[56] L. V. Hooper, D. R. Littman, and A. J. Macpherson. Interactions between the microbiota and the immune system. *Science*, 336, 2012. doi: 10.1126/science.1223490. URL `http://dx.doi.org/10.1126/science.1223490`.

[57] June L. Round and Sarkis K. Mazmanian. The gut microbiome shapes intestinal immune responses during health and disease. *Nature reviews. Immunology*, 9(5):313–323, 5 2009. ISSN 1474-1733. doi: 10.1038/nri2515. PMID: 19343057 PMCID: PMC4095778.

[58] Hugh A. Sampson, Liam O'Mahony, A. Wesley Burks, Marshall Plaut, Gideon Lack, and Cezmi A. Akdis. Mechanisms of food allergy. *The Journal of Allergy and Clinical Immunology*, 141(1):11–19, 1 2018. ISSN 1097-6825. doi: 10.1016/j.jaci.2017.11.005. PMID: 29307410.

[59] M. J. Wannemuehler, H. Kiyono, J. L. Babb, S. M. Michalek, and J. R. McGhee. Lipopolysaccharide (lps) regulation of the immune response: Lps converts germfree mice to sensitivity to oral tolerance induction. *Journal of Immunology (Baltimore, Md.: 1950)*, 129(3):959–965, 9 1982. ISSN 0022-1767. PMID: 6980928.

[60] Catherine H. Plunkett and Cathryn R. Nagler. The influence of the microbiome on allergic sensitization to food. *Journal of Immunology (Baltimore, Md.: 1950)*, 198(2): 581–589, 2017. ISSN 1550-6606. doi: 10.4049/jimmunol.1601266. PMID: 28069753 PMCID: PMC5228404.

[61] Mohamed Elfatih H. Bashir, Steve Louie, Hai Ning Shi, and Cathryn Nagler-Anderson. Toll-like receptor 4 signaling by intestinal microbes influences susceptibility to food allergy. *Journal of Immunology (Baltimore, Md.: 1950)*, 172(11):6978–6987, 6 2004. ISSN 0022-1767. PMID: 15153518.

[62] Jack A. Gilbert, Robert A. Quinn, Justine Debelius, Zhenjiang Z. Xu, James Morton, Neha Garg, Janet K. Jansson, Pieter C. Dorrestein, and Rob Knight. Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature*, 535(7610):94–103, 7 2016. ISSN 1476-4687. doi: 10.1038/nature18850.

[63] Weronika Barcik, Marcin Wawrzyniak, Cezmi A Akdis, and Liam OMahony. Immune regulation by histamine and histamine-secreting bacteria. *Current Opinion in Immunology*, 48:108–113, 10 2017. ISSN 0952-7915. doi: 10.1016/j.coi.2017.08.011.

[64] Patrick M. Smith, Michael R. Howitt, Nicolai Panikov, Monia Michaud, Carey Ann Gallini, Mohammad Bohlooly-Y, Jonathan N. Glickman, and Wendy S. Garrett. The microbial metabolites, short-chain fatty acids, regulate colonic treg cell homeostasis. *Science*, 341(6145):569–573, 8 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1241165. PMID: 23828891.

[65] Jian Tan, Craig McKenzie, Peter J. Vuillermin, Gera Goverse, Carola G. Vinuesa, Reina E. Mebius, Laurence Macia, and Charles R. Mackay. Dietary fiber and bacterial scfa enhance oral tolerance and protect against food allergy through diverse cellular pathways. *Cell Reports*, 15(12):2809–2824, 2016. ISSN 2211-1247. doi: 10.1016/j.celrep.2016.05.047. PMID: 27332875.

[66] Craig McKenzie, Jian Tan, Laurence Macia, and Charles R. Mackay. The nutrition-gut microbiome-physiology axis and allergic diseases. *Immunological Reviews*, 278(1): 277–295, 7 2017. ISSN 1600-065X. doi: 10.1111/imr.12556.

[67] So-Ichiro Hirata and Jun Kunisawa. Gut microbiome, metabolome, and allergic diseases. *Allergology International: Official Journal of the Japanese Society of Allergology*, 66(4):523–528, 10 2017. ISSN 1440-1592. doi: 10.1016/j.alit.2017.06.008. PMID: 28693971.

[68] Naoko Shibata, Jun Kunisawa, and Hiroshi Kiyono. Dietary and microbial metabolites in the regulation of host immunity. *Frontiers in Microbiology*, 8, 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.02171. URL `https://www.frontiersin.org/articles/10.3389/fmicb.2017.02171/full`. [Online; accessed 2018-08-17].

[69] Shigenobu Kishino, Michiki Takeuchi, Si-Bum Park, Akiko Hirata, Nahoko Kitamura, Jun Kunisawa, Hiroshi Kiyono, Ryo Iwamoto, Yosuke Isobe, Makoto Arita, Hiroyuki Arai, Kazumitsu Ueda, Jun Shima, Satomi Takahashi, Kenzo Yokozeki, Sakayu Shimizu, and Jun Ogawa. Polyunsaturated fatty acid saturation by gut lactic acid bacteria affecting host lipid composition. *Proceedings of the National Academy of Sciences*, 110(44):17808–17813, 10 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1312937110. PMID: 24127592.

[70] Jun Kunisawa, Makoto Arita, Takahiro Hayasaka, Takashi Harada, Ryo Iwamoto, Risa Nagasawa, Shiori Shikata, Takahiro Nagatake, Hidehiko Suzuki, Eri Hashimoto, Yosuke Kurashima, Yuji Suzuki, Hiroyuki Arai, Mitsutoshi Setou, and Hiroshi Kiyono. Dietary 3 fatty acid exerts anti-allergic effect through the conversion to 17,18-epoxyeicosatetraenoic acid in the gut. *Scientific Reports*, 5:9750, 6 2015. doi: 10.1038/srep09750.

[71] Clara M. Yates, Philip C. Calder, and G. Ed Rainger. Pharmacology and therapeutics of omega-3 polyunsaturated fatty acids in chronic inflammatory disease. *Pharmacology & Therapeutics*, 141(3):272–282, 3 2014. ISSN 0163-7258. doi: 10.1016/j.pharmthera.2013.10.010.

[72] Lefkothea-Stella Kremmyda, Maria Vlachava, Paul S. Noakes, Norma D. Diaper, Elizabeth A. Miles, and Philip C. Calder. Atopy risk in infants and children in relation to early exposure to fish, oily fish, or long-chain omega-3 fatty acids: a systematic review. *Clinical Reviews in Allergy & Immunology*, 41(1):36–66, 8 2011. ISSN 1559-0267. doi: 10.1007/s12016-009-8186-2. PMID: 19997989.

[73] J. L. Kaplan, H. N. Shi, and W. A. Walker. The role of microbes in developmental immunologic programming. *Pediatr Res*, 69, 2011. doi: 10.1203/PDR.0b013e318217638a. URL `http://dx.doi.org/10.1203/PDR.0b013e318217638a`.

[74] Harm Wopereis, Raish Oozeer, Karen Knipping, Clara Belzer, and Jan Knol. The first thousand days - intestinal microbiology of early life: establishing a symbiosis. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology*, 25(5):428–438, 8 2014. ISSN 1399-3038 0905-6157. doi: 10.1111/pai.12232. PMID: 24899389.

[75] Maria A. Johansson, Shanie Saghafian-Hedengren, Yeneneh Haileselassie, Stefan Roos, Marita Troye-Blomberg, Caroline Nilsson, and Eva Sverremark-Ekstrom. Early-life gut bacteria associate with il-4-, il-10- and ifn-gamma production at two years of age. *PloS one*, 7(11):e49315, 2012. ISSN 1932-6203 1932-6203. doi: 10.1371/journal.pone.0049315. PMID: 23185315 PMCID: PMC3502541.

[76] Amy D. Proal, Paul J. Albert, and Trevor G. Marshall. The human microbiome and autoimmunity. *Current opinion in rheumatology*, 25(2):234–240, 3 2013. ISSN 1531-6963 1040-8711. doi: 10.1097/BOR.0b013e32835cedbf. PMID: 23370376.

[77] Torsten Olszak, Dingding An, Sebastian Zeissig, Miguel Pinilla Vera, Julia Richter, Andre Franke, Jonathan N. Glickman, Reiner Siebert, Rebecca M. Baron, Dennis L. Kasper, and Richard S. Blumberg. Microbial exposure during early life has persistent effects on natural killer t cell function. *Science*, 336(6080):489–493, 4 2012. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1219328. PMID: 22442383.

[78] Fernando D. Martinez. The human microbiome. early life determinant of health outcomes. *Annals of the American Thoracic Society*, 11(Supplement 1):S7–S12, 1 2014. ISSN 2329-6933. doi: 10.1513/AnnalsATS.201306-186MG.

[79] Thomas Gensollen and Richard S. Blumberg. Correlation between early-life regulation of the immune system by microbiota and allergy development. *Journal of Allergy and Clinical Immunology*, 139(4):1084–1091, 4 2017. ISSN 0091-6749. doi: 10.1016/j.jaci.2017.02.011.

[80] Sahar El Aidy, Guido Hooiveld, Valentina Tremaroli, Fredrik Bckhed, and Michiel Kleerebezem. The gut microbiota and mucosal homeostasis. *Gut Microbes*, 4(2):118–124, 1 2013. ISSN 1949-0976. doi: 10.4161/gmic.23362. PMID: 23333858.

[81] Mariana X. Byndloss, Sandy R. Pernitzsch, and Andreas J. Bumler. Healthy hosts rule within: ecological forces shaping the gut microbiota. *Mucosal Immunology*, 5 2018. ISSN 1933-0219, 1935-3456. doi: 10.1038/s41385-018-0010-y. URL `http://www.nature.com/articles/s41385-018-0010-y`. [Online; accessed 2018-08-09].

[82] Silvia Arboleya, Borja Sanchez, Christian Milani, Sabrina Duranti, Gonzalo Solis, Nuria Fernandez, Clara G. de los Reyes-Gavilan, Marco Ventura, Abelardo Margolles, and Miguel Gueimonde. Intestinal microbiota development in preterm neonates and effect of perinatal antibiotics. *The Journal of pediatrics*, 166(3):538–544, 3 2015. ISSN 1097-6833 0022-3476. doi: 10.1016/j.jpeds.2014.09.041. PMID: 25444008.

[83] Christopher J. Stewart, Nicholas D. Embleton, Emma C. L. Marrs, Daniel P. Smith, Andrew Nelson, Bashir Abdulkadir, Tom Skeath, Joseph F. Petrosino, John D. Perry, Janet E. Berrington, and Stephen P. Cummings. Temporal bacterial and metabolic development of the preterm gut reveals specific signatures in health and disease. *Microbiome*, 4, 12 2016. ISSN 2049-2618. doi: 10.1186/s40168-016-0216-8. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5200962/`. PMID: 28034304 PMCID: PMC5200962.

[84] Aifric O'Sullivan, Marie Farver, and Jennifer T. Smilowitz. The influence of early infant-feeding practices on the intestinal microbiome and body composition in infants. *Nutrition and Metabolic Insights*, 8(Suppl 1):1–9, 2015. ISSN 1178-6388. doi: 10.4137/NMI.S29530. PMID: 26715853 PMCID: PMC4686345.

[85] Alex Grier, Xing Qiu, Sanjukta Bandyopadhyay, Jeanne Holden-Wiltse, Haeja A. Kessler, Ann L. Gill, Brooke Hamilton, Heidie Huyck, Sara Misra, Thomas J. Mariani, Rita M. Ryan, Lori Scholer, Kristin M. Scheible, Yi-Horng Lee, Mary T. Caserta, Gloria S. Pryhuber, and Steven R. Gill. Impact of prematurity and nutrition on the developing gut microbiome and preterm infant growth. *Microbiome*, 5, 12 2017. ISSN 2049-2618. doi: 10.1186/s40168-017-0377-0. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5725645/`. PMID: 29228972 PMCID: PMC5725645.

[86] William Walters, Embriette R. Hyde, Donna Berg-Lyons, Gail Ackermann, Greg Humphrey, Alma Parada, Jack A. Gilbert, Janet K. Jansson, J. Gregory Caporaso, Jed A. Fuhrman, Amy Apprill, and Rob Knight. Improved bacterial 16s rrna gene (v4 and v4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems*, 1(1), 2 2016. ISSN 2379-5077. doi: 10.1128/mSystems.00009-15. PMID: 27822518 PMCID: PMC5069754.

[87] Benjamin J. Callahan, Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. Dada2: High-resolution sample inference from illumina amplicon data. *Nature Methods*, 13(7):581, 7 2016. ISSN 1548-7105. doi: 10.1038/nmeth.3869.

[88] J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, and E. K. Costello. Qiime allows analysis of high-throughput community sequencing data. *Nat Methods*, 7, 2010. doi: 10.1038/nmeth.f.303. URL `http://dx.doi.org/10.1038/nmeth.f.303`.

[89] Siddhartha Mandal, Will Van Treuren, Richard A. White, Merete Eggesb, Rob Knight, and Shyamal D. Peddada. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease*, 26:27663, 2015. ISSN 0891-060X. PMID: 26028277 PMCID: PMC4450248.

[90] Nicholas Bokulich, Matthew Dillon, Evan Bolyen, Benjamin D. Kaehler, Gavin A. Huttley, and J. Gregory Caporaso. q2-sample-classifier: machine-learning tools for microbiome classification and regression. *bioRxiv*, page 306167, 4 2018. doi: 10.1101/306167.

[91] Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. Upset: Visualization of intersecting sets. *IEEE transactions on visualization and computer graphics*, 20(12):1983–1992, 12 2014. ISSN 1941-0506. doi: 10.1109/TVCG.2014.2346248. PMID: 26356912 PMCID: PMC4720993.

[92] Tanis R. Fenton and Jae H. Kim. A systematic review and meta-analysis to revise the fenton growth chart for preterm infants. *BMC Pediatrics*, 13(1):59, 4 2013. ISSN 1471-2431. doi: 10.1186/1471-2431-13-59.

[93] Mohan Pammi, Julia Cope, Phillip I. Tarr, Barbara B. Warner, Ardythe L. Morrow, Volker Mai, Katherine E. Gregory, J. Simon Kroll, Valerie McMurtry, Michael J. Ferris, Lars Engstrand, Helene Engstrand Lilja, Emily B. Hollister, James Versalovic, and

Josef Neu. Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis. *Microbiome*, 5(1):31, 2017. ISSN 2049-2618. doi: 10.1186/s40168-017-0248-8. PMID: 28274256 PMCID: PMC5343300.

[94] Dominika Drzewiecka. Significance and roles of proteus spp. bacteria in natural environments. *Microbial Ecology*, 72(4):741–758, 2016. ISSN 0095-3628. doi: 10.1007/s00248-015-0720-6. PMID: 26748500 PMCID: PMC5080321.

[95] Jann-Tay Wang, Pei-Chen Chen, Shan-Chwen Chang, Yih-Ru Shiau, Hui-Ying Wang, Jui-Fen Lai, I-Wen Huang, Mei-Chen Tan, and Tsai-Ling Yang Lauderdale. Antimicrobial susceptibilities of proteus mirabilis: a longitudinal nationwide study from the taiwan surveillance of antimicrobial resistance (tsar) program. *BMC Infectious Diseases*, 14, 9 2014. ISSN 1471-2334. doi: 10.1186/1471-2334-14-486. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4162950/. PMID: 25192738 PMCID: PMC4162950.

[96] A. Jacquot, D. Neveu, F. Aujoulat, G. Mercier, H. Marchandin, E. Jumas-Bilak, and J. C. Picaud. Dynamics and clinical evolution of bacterial gut microflora in extremely premature patients. *J Pediatr*, 158, 2011. doi: 10.1016/j.jpeds.2010.09.007. URL http://dx.doi.org/10.1016/j.jpeds.2010.09.007.

[97] Fabio Mosca, Maria Lorella Giann, Paola Roggero, Camilla Menis, Laura Morlacchi, Nadia Liotto, and Beatrice Bracco. Critical questions on nutrition of preterm infants. *Journal of Pediatric and Neonatal Individualized Medicine (JPNIM)*, 6(2):e060203, 6 2017. ISSN 2281-0692. doi: 10.7363/060203.

[98] Anke Raaijmakers and Karel Allegaert. Catch-up growth in former preterm neonates: No time to waste. *Nutrients*, 8(12), 12 2016. ISSN 2072-6643. doi: 10.3390/nu8120817. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5188471/. PMID: 27999303 PMCID: PMC5188471.

[99] M. Hack, B. Weissman, and E. Borawski-Clark. Catch-up growth during childhood among very low-birth-weight children. *Arch Pediatr Adolesc Med*, 150, 1996. doi: 10.1001/archpedi.1996.02170360012002. URL http://dx.doi.org/10.1001/archpedi.1996.02170360012002.

[100] Ruairi C. Robertson, Amee R. Manges, B. Brett Finlay, and Andrew J. Prendergast. The human microbiome and child growth first 1000 days and beyond. *Trends in Microbiology*, 10 2018. ISSN 0966-842X. doi: 10.1016/j.tim.2018.09.008. URL http://www.sciencedirect.com/science/article/pii/S0966842X1830204X. [Online; accessed 2018-11-08].

[101] Kei E. Fujimura, Alexandra R. Sitarik, Suzanne Havstad, Din L. Lin, Sophia Levan, Douglas Fadrosh, Ariane R. Panzer, Brandon LaMere, Elze Rackaityte, Nicholas W. Lukacs, Ganesa Wegienka, Homer A. Boushey, Dennis R. Ownby, Edward M. Zoratti, Albert M. Levin, Christine C. Johnson, and Susan V. Lynch. Neonatal gut microbiota associates with childhood multisensitized atopy and t cell differentiation. *Nature*

Medicine, 22(10):1187–1191, 2016. ISSN 1546-170X. doi: 10.1038/nm.4176. PMID: 27618652 PMCID: PMC5053876.

[102] Juliana Durack, Nikole E. Kimes, Din L. Lin, Marcus Rauch, Michelle McKean, Kathryn McCauley, Ariane R. Panzer, Jordan S. Mar, Michael D. Cabana, and Susan V. Lynch. Delayed gut microbiota development in high-risk for asthma infants is temporarily modifiable by lactobacillus supplementation. *Nature Communications*, 9 (1):707, 2 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-03157-4.

[103] Yasmine Belkaid and Timothy W. Hand. Role of the microbiota in immunity and inflammation. *Cell*, 157(1):121–141, 3 2014. ISSN 1097-4172. doi: 10.1016/j.cell.2014.03.011. PMID: 24679531 PMCID: PMC4056765.

[104] G. P. Donaldson, M. S. Ladinsky, K. B. Yu, J. G. Sanders, B. B. Yoo, W.-C. Chou, M. E. Conner, A. M. Earl, R. Knight, P. J. Bjorkman, and S. K. Mazmanian. Gut microbiota utilize immunoglobulin a for mucosal colonization. *Science (New York, N.Y.)*, 360(6390):795–800, 2018. ISSN 1095-9203. doi: 10.1126/science.aaq0926. PMID: 29724905 PMCID: PMC5973787.

[105] Mikiko Soejima and Yoshiro Koda. Distinct single nucleotide polymorphism pattern at the fut2 promoter among human populations. *Annals of Hematology*, 87(1):19–25, 1 2008. ISSN 1432-0584. doi: 10.1007/s00277-007-0362-y. PMID: 17805536.

[106] Anuhya Gampa, Phillip A. Engen, Rima Shobar, and Ece A. Mutlu. Relationships between gastrointestinal microbiota and blood group antigens. *Physiological Genomics*, 49(9):473–483, 9 2017. ISSN 1531-2267. doi: 10.1152/physiolgenomics.00043.2017. PMID: 28710295 PMCID: PMC5625272.

[107] Jess Rodrguez-Daz, Izaskun Garca-Mantrana, Susana Vila-Vicent, Roberto Gozalbo-Rovira, Javier Buesa, Vicente Monedero, and Maria Carmen Collado. Relevance of secretor status genotype and microbiota composition in susceptibility to rotavirus and norovirus infections in humans. *Scientific Reports*, 7:45559, 3 2017. ISSN 2045-2322. doi: 10.1038/srep45559.

[108] Isobel Franks. Gut microbiota: Fut2 genotype influences the gut microbiota in patients with crohn's disease and healthy individuals. *Nature Reviews. Gastroenterology & Hepatology*, 9(1):2, 12 2011. ISSN 1759-5053. doi: 10.1038/nrgastro.2011.237. PMID: 22158379.

[109] S. Lindn, J. Mahdavi, C. Semino-Mora, C. Olsen, I. Carlstedt, and T. Born. Role of abo secretor status in mucosal innate immunity and h. pylori infection. *PLoS Pathog*, 4, 2008. doi: 10.1371/journal.ppat.0040002. URL `http://dx.doi.org/10.1371/journal.ppat.0040002`.

[110] Ardythe L. Morrow, Jareen Meinzen-Derr, Pengwei Huang, Kurt R. Schibler, Tanya Cahill, Mehdi Keddache, Suhas G. Kallapur, David S. Newburg, Meredith Tabangin, Barbara B. Warner, and Xi Jiang. Fucosyltransferase 2 non-secretor and low secretor status predicts severe outcomes in premature infants. *The Journal of pediatrics*, 158

(5):745–751, 5 2011. ISSN 1097-6833 0022-3476. doi: 10.1016/j.jpeds.2010.10.043. PMID: 21256510 PMCID: PMC3412418.

[111] Pirjo Wacklin, Jarno Tuimala, Janne Nikkil, Sebastian Tims, Harri Mkivuokko, Noora Alakulppi, Pia Laine, Mirjana Rajilic-Stojanovic, Lars Paulin, Willem M. de Vos, and Jaana Mtt. Faecal microbiota composition in adults is associated with the fut2 gene determining the secretor status. *PLoS ONE*, 9(4):e94863, 4 2014. doi: 10.1371/journal.pone.0094863.

[112] Emily R. Davenport, Julia K. Goodrich, Jordana T. Bell, Tim D. Spector, Ruth E. Ley, and Andrew G. Clark. Abo antigen and secretor statuses are not associated with gut microbiota composition in 1,500 twins. *BMC genomics*, 17(1):941, 11 2016. ISSN 1471-2164 1471-2164. doi: 10.1186/s12864-016-3290-1. PMID: 27871240 PMCID: PMC5117602.

[113] Williams Turpin, Larbi Bedrani, Osvaldo Espin-Garcia, Wei Xu, Mark S. Silverberg, Michelle I. Smith, David S. Guttman, Anne Griffiths, Paul Moayyedi, Remo Panaccione, Hien Huynh, Hillary Steinhart, Guy Aumais, Konstantin Shestopaloff, Levinus A. Dieleman, Dan Turner, Andrew D. Paterson, and Kenneth Croitoru. Fut2 genotype and secretory status are not associated with fecal microbial composition and inferred function in healthy subjects. *Gut Microbes*, pages 1–12, 3 2018. ISSN 1949-0984. doi: 10.1080/19490976.2018.1445956. PMID: 29533703.

[114] Martin Demmert, Anne Schaper, Julia Pagel, Corinna Gebauer, Michael Emeis, Friedhelm Heitmann, Angela Kribs, Jens Siegel, Dirk Muller, Annette Keller-Wackerbauer, Hubert Gerleve, Christian Wieg, Egbert Herting, Wolfgang Gopel, and Christoph Hartel. Fut 2 polymorphism and outcome in very-low-birth-weight infants. *Pediatr Res*, 77(4):586–590, 4 2015. ISSN 0031-3998.

[115] Shu-E. Soh, Mya Thway Tint, Peter D. Gluckman, Keith M. Godfrey, Anne Rifkin-Graboi, Yiong Huak Chan, Walter Stnkel, Joanna D. Holbrook, Kenneth Kwek, Yap-Seng Chong, and Seang Mei Saw. Cohort profile: Growing up in singapore towards healthy outcomes (gusto) birth cohort study. *International Journal of Epidemiology*, 43(5):1401–1409, 10 2014. ISSN 0300-5771. doi: 10.1093/ije/dyt125.

[116] Nicola Segata, Levi Waldron, Annalisa Ballarini, Vagheesh Narasimhan, Olivier Jousson, and Curtis Huttenhower. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, 9(8):811–814, 6 2012. ISSN 1548-7105. doi: 10.1038/nmeth.2066. PMID: 22688413 PMCID: PMC3443552.

[117] Eric A. Franzosa, Lauren J. McIver, Gholamali Rahnavard, Luke R. Thompson, Melanie Schirmer, George Weingart, Karen Schwarzberg Lipson, Rob Knight, J. Gregory Caporaso, Nicola Segata, and Curtis Huttenhower. Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, 15(11):962–968, 11 2018. ISSN 1548-7105. doi: 10.1038/s41592-018-0176-y. PMID: 30377376 PMCID: PMC6235447.

[118] Sergey Nurk, Dmitry Meleshko, Anton Korobeynikov, and Pavel A. Pevzner. metaspades: a new versatile metagenomic assembler. *Genome Research*, 27(5):824–834, 2017. ISSN 1549-5469. doi: 10.1101/gr.213959.116. PMID: 28298430 PMCID: PMC5411777.

[119] Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9(4):357–359, 3 2012. ISSN 1548-7105. doi: 10.1038/nmeth.1923. PMID: 22388286 PMCID: PMC3322381.

[120] Jonathan Friedman and Eric J. Alm. Inferring correlation networks from genomic survey data. *PLOS Computational Biology*, 8(9):e1002687, 9 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002687.

[121] Xochitl C. Morgan, Timothy L. Tickle, Harry Sokol, Dirk Gevers, Kathryn L. Devaney, Doyle V. Ward, Joshua A. Reyes, Samir A. Shah, Neal LeLeiko, Scott B. Snapper, Athos Bousvaros, Joshua Korzenik, Bruce E. Sands, Ramnik J. Xavier, and Curtis Huttenhower. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology*, 13(9):R79, 4 2012. ISSN 1474-760X. doi: 10.1186/gb-2012-13-9-r79. PMID: 23013615 PMCID: PMC3506950.

[122] Nicola Segata, Jacques Izard, Levi Waldron, Dirk Gevers, Larisa Miropolsky, Wendy S. Garrett, and Curtis Huttenhower. Metagenomic biomarker discovery and explanation. *Genome Biology*, 12(6):R60, 6 2011. ISSN 1474-760X. doi: 10.1186/gb-2011-12-6-r60. PMID: 21702898 PMCID: PMC3218848.

[123] Quast: quality assessment tool for genome assemblies. - pubmed - ncbi. URL `https://www.ncbi.nlm.nih.gov/pubmed/23422339`. [Online; accessed 2019-01-31].

[124] Stephen Wandro, Stephanie Osborne, Claudia Enriquez, Christine Bixby, Antonio Arrieta, and Katrine Whiteson. The microbiome and metabolome of preterm infant stool are personalized and not driven by health outcomes, including necrotizing enterocolitis and late-onset sepsis. *mSphere*, 3(3):e00104–18, 6 2018. ISSN 2379-5042. doi: 10.1128/mSphere.00104-18. PMID: 29875143.

[125] M. Thorven, A. Grahn, K. Hedlund, H. Johansson, and C. Wahlfrid. A homozygous nonsense mutation (428g3a) in the human secretor (fut2) gene provides resistance to symptomatic norovirus (ggii) infections. *J Virol*, 79, 2005. doi: 10.1128/JVI.79.24.15351-15355.2005. URL `http://dx.doi.org/10.1128/JVI.79.24.15351-15355.2005`.

[126] Maria Gloria Dominguez-Bello, Filipa Godoy-Vitorino, Rob Knight, and Martin J. Blaser. Role of the microbiome in human development. *Gut*, pages gutjnl–2018–317503, 1 2019. ISSN 0017-5749, 1468-3288. doi: 10.1136/gutjnl-2018-317503. PMID: 30670574.

[127] Noel T. Mueller, Hakdong Shin, Aline Pizoni, Isabel C. Werlang, Ursula Matte, Marcelo Z. Goldani, Helena A. S. Goldani, and Maria Gloria Dominguez-Bello. Birth mode-dependent association between pre-pregnancy maternal weight status and the

neonatal intestinal microbiome. *Scientific Reports*, 6:23133, 4 2016. ISSN 2045-2322. doi: 10.1038/srep23133. PMID: 27033998 PMCID: PMC4817027.

[128] Gilberto E Flores, J Gregory Caporaso, Jessica B Henley, Jai Ram Rideout, Daniel Domogala, John Chase, Jonathan W Leff, Yoshiki Vzquez-Baeza, Antonio Gonzalez, Rob Knight, Robert R Dunn, and Noah Fierer. Temporal variability is a personalized feature of the human microbiome. *Genome Biology*, 15(12), 2014. ISSN 1465-6906. doi: 10.1186/s13059-014-0531-y. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4252997/`. PMID: 25517225 PMCID: PMC4252997.

[129] Michael Tessler, Johannes S. Neumann, Ebrahim Afshinnekoo, Michael Pineda, Rebecca Hersch, Luiz Felipe M. Velho, Bianca T. Segovia, Fabio A. Lansac-Toha, Michael Lemke, Rob DeSalle, Christopher E. Mason, and Mercer R. Brugler. Large-scale differences in microbial biodiversity discovery between 16s amplicon and shotgun sequencing. *Scientific Reports*, 7(1):6589, 7 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-06665-3.

[130] Juan Jovel, Jordan Patterson, Weiwei Wang, Naomi Hotte, Sandra O'Keefe, Troy Mitchel, Troy Perry, Dina Kao, Andrew L. Mason, Karen L. Madsen, and Gane K.-S. Wong. Characterization of the gut microbiome using 16s or shotgun metagenomics. *Frontiers in Microbiology*, 7, 4 2016. ISSN 1664-302X. doi: 10.3389/fmicb.2016.00459. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4837688/`. PMID: 27148170 PMCID: PMC4837688.

[131] Sophie Weiss, Zhenjiang Zech Xu, Shyamal Peddada, Amnon Amir, Kyle Bittinger, Antonio Gonzalez, Catherine Lozupone, Jesse R. Zaneveld, Yoshiki Vzquez-Baeza, Amanda Birmingham, Embriette R. Hyde, and Rob Knight. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5(1):27, 3 2017. ISSN 2049-2618. doi: 10.1186/s40168-017-0237-y.

[132] James T. Morton, Jon Sanders, Robert A. Quinn, Daniel McDonald, Antonio Gonzalez, Yoshiki Vzquez-Baeza, Jose A. Navas-Molina, Se Jin Song, Jessica L. Metcalf, Embriette R. Hyde, Manuel Lladser, Pieter C. Dorrestein, and Rob Knight. Balance trees reveal microbial niche differentiation. *mSystems*, 2(1):e00162–16, 2 2017. ISSN 2379-5077. doi: 10.1128/mSystems.00162-16. PMID: 28144630.

[133] Thomas D. Otto. From sequence mapping to genome assemblies. *Methods in Molecular Biology (Clifton, N.J.)*, 1201:19–50, 2015. ISSN 1940-6029. doi: 10.1007/978-1-4939-1438-8_2. PMID: 25388106.

[134] Martin Ayling, Matthew D. Clark, and Richard M. Leggett. New approaches for metagenome assembly with short reads. *Briefings in Bioinformatics*, 2019. doi: 10.1093/bib/bbz020. URL `https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbz020/5363831`. [Online; accessed 2019-04-26].

[135] John Vollmers, Sandra Wiegand, and Anne-Kristin Kaster. Comparing and evaluating metagenome assembly tools from a microbiologists perspective - not only size matters! *PLOS ONE*, 12(1):e0169662, 1 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0169662.

[136] Dongwan Kang, Feng Li, Edward S Kirton, Ashleigh Thomas, Rob S Egan, Hong An, and Zhong Wang. Metabat 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. Technical report, 2 2019. URL `https://peerj.com/preprints/27522`. DOI: 10.7287/peerj.preprints.27522v1.

[137] Johannes Alneberg, Brynjar Smri Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z. Ijaz, Leo Lahti, Nicholas J. Loman, Anders F. Andersson, and Christopher Quince. Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11(11):1144–1146, 11 2014. ISSN 1548-7105. doi: 10.1038/nmeth.3103.

[138] Donovan H. Parks, Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7):1043–1055, 7 2015. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.186072.114. PMID: 25977477.

[139] Marie-Claire Arrieta, Jens Walter, and B. Brett Finlay. Human microbiota-associated mice: A model with challenges. *Cell Host & Microbe*, 19(5):575–578, 5 2016. ISSN 1931-3128. doi: 10.1016/j.chom.2016.04.014.