I, Diana H Taft, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Epidemiology (Environmental Health).

It is entitled:
Host Genotype, Intestinal Microbial Phenotype, and Late-Onset Sepsis in the Premature Infant

Student's name: Diana H Taft

This work and its defense approved by:

Committee chair: Ardythe Luxion Morrow, Ph.D.

Committee member: Christina Valentine, Ph.D.

Committee member: Ranjan Deka, Ph.D.

Committee member: John Harley, M.D., Ph.D.

Committee member: Marepalli Rao, Ph.D.
Host Genotype, Intestinal Microbial Phenotype, and Late-Onset Sepsis in the Premature Infant

A dissertation submitted to the
Graduate School
Of the University of Cincinnati
In partial fulfillment of the
Requirements for the degree of

Doctor of Philosophy

In the Division of Epidemiology and Biostatistics
In the Department of Environmental Health
Of the College of Medicine

by

Diana Hazard Taft
B.A. Cornell University
Submitted July 2014

Committee Chair: A.L. Morrow, Ph.D.
Committee Members: R. Deka, Ph.D.
J. Harley, M.D., Ph.D.
M.B. Rao, Ph.D.
C. Valentine, M.D.
Abstract

Background: Small studies have found a connection between intestinal microbial colonization and risk of late-onset sepsis, but leave open questions about the generalizability of these findings. Also largely unstudied is the role of the host genotype in shaping the intestinal microbiome of the premature infant.

Methods: The Novel Biomarkers Cohort is a large, prospective cohort of premature infants (born ≤32 weeks gestational age) enrolled from Neonatal Intensive Care Units (NICUs) in Birmingham, Alabama and Cincinnati, Ohio. Birmingham infants were compared Cincinnati infants to determine the degree to which findings regarding the microbiota may be generalizable between NICUs. Then, a nested case-control study of 66 infants was conducted to determine colonization patterns associated with late-onset sepsis. Finally, the microbiota of Cincinnati infants were analyzed to explore the contribution of host genetics to shaping the intestinal microbial community.

Results: Of the NICUs, Birmingham preterm infants tend to have more stable bacterial communities during the first three weeks of life, with more Proteobacteria and Bacteriodes, and less Firmicutes than Cincinnati preterm infants. The composition of infant samples varied over calendar months and between postnatal weeks of life. Streptococcaceae and Clostridia were associated with sepsis in Birmingham and the Bacilli and the lack of Actinobacteria were associated with sepsis in Cincinnati. SNPs in specific fucosyltransferase and mucin genes were associated with higher or lower relative abundance of specific bacteria. However, these differences were not relevant to the patterns observed prior to sepsis.
Table of Contents

Abstract ........................................................................................................................................... 2

Table of Contents .......................................................................................................................... 3

Definitions ..................................................................................................................................... 4

Chapter 1: Background, Aims, and Innovation ............................................................................ 5

Chapter 2: Overview of Methods ................................................................................................. 11

Chapter 3: Differences in Intestinal Colonization of Infants in Geographically Separate Neonatal Intensive Care Units over time ............................................................................. 22

Chapter 4: The Intestinal Microbiota Preceding Late Onset Neonatal Sepsis compared to Controls: A Nested Case-Control Study ......................................................................................... 51

Chapter 5: Infant Genotype and Intestinal Microbiota Colonization ........................................... 71

Chapter 6: Conclusions and Future Research ............................................................................. 88

Appendix 1: Bayesian Network Analysis ...................................................................................... 94

Appendix 2: Human Subjects Protection ..................................................................................... 102

Acknowledgements ..................................................................................................................... 114
Definitions

16s rRNA gene: A gene encoding the small ribosomal subunit in bacteria that has both highly conserved and hypervariable regions

Alpha Diversity: The within individual bacterial diversity

Beta Diversity: The between individuals bacterial diversity

GA: Gestational Age at time of delivery

LEfSe: Linear Effect Size Estimator, an online tool designed to test for differences in microbiota at all levels of taxonomy between user specified comparison groups

Microbiome: Collective genomes and gene products of the microbial community residing in and on humans

Microbiota: Organisms that comprise the microbial community residing within and on humans

NEC: Necrotizing enterocolitis

NICU: Neonatal Intensive Care Unit

NMDS: Non-metric Multidimensional Scaling, an ordination technique well suited to non-normal data

OTU: Operational Taxanomic Unit, sequence reads clustered to 97% similarity and then assigned taxonomy

OTU table: A table of count data showing the number of sequences from each sample that clustered with each OTU

Rarefy: Random selection of reads from each sample so that each sample is represented by an equal number of reads for comparison purposes

SNP: Single Nucleotide Polymorphism
Chapter 1: Background, Aims, and Innovation

Background

Dysregulated intestinal microbial colonization may underlie various clinical conditions or diseases of premature neonates, including late-onset sepsis (1). Late onset sepsis (LOS) is a clinical syndrome of neonates that is characterized by systemic signs of infection and accompanied by bacteremia occurring more than 72 hours after birth (2). Almost one quarter of very low birth weight (VLBW) infants develop late-onset sepsis (3). Late-onset sepsis case fatality is 7% to 11% in infants with gram-positive sepsis and 20% to 74% in infants with gram-negative sepsis (2). Of late-onset sepsis cases, more than 70% are caused by Gram-positive organisms (4). The greatest risk factors of late-onset neonatal sepsis are low birthweight and younger gestational age (4). The clinical signs and symptoms of late-onset neonatal sepsis are non-specific and include apnea, bradycardia, increased oxygen requirement, gastrointestinal problems, lethargy, hypotonia, temperature instability, and hypotension (5). In addition, late-onset sepsis carries a high long-term risk of delayed motor and mental development and cerebral palsy (2).

The gastrointestinal tract is an important reservoir of organisms that can cause late onset sepsis in the premature infant. The sepsis causing organism often detected in rectal swabs prior to disease (6, 7). A recent small study of late onset sepsis in six premature infants found that sepsis was preceded by a less diverse microbial community early in life and further characterized by a predominance of Staphylococcus (1). Another recent study of 10 cases of late onset sepsis in neonates compared to 18 controls reported that cases had a less diverse microbiota in the weeks prior to sepsis, but that cases were not enriched in any particular taxa, while controls were enriched in Bifidobacteria (8).

The composition of the intestinal microbiome has important health implications for the premature infant. Small studies have found differences in colonization patterns associated with
both late-onset sepsis and necrotizing enterocolitis (NEC) (8-11), but the degree to which these findings are generalizable remains an open question. The findings relating the intestinal microbiome to sepsis are particularly inconsistent (8-10), despite evidence suggesting the causative organism of sepsis often is first detected in the intestinal tract (6, 12).

The intestinal microbiota evolves quickly in early life and has a large impact on health throughout life (13, 14). Evidence is emerging about the cross-talk between host and microbiome. Host genotype may contribute to shaping some aspects of microbial community composition. Host genotype may also be important in shaping host response to the microbiota (15). Genes thought to influence intestinal microbial composition include fucosyltransferase and mucin genes (16-20). Membrane-bound mucins act as an important part of membrane barrier function while soluble mucins help provide nutrients to specific gut colonizing bacteria. This suggests that mucin genotype is likely to have an impact on the development of the intestinal microbiota.

Mucus, formed primarily of mucins, covers the entire gastrointestinal tract and protects the intestinal epithelium from penetration by bacteria (21). There are two mucus layers, a loose outer layer which bacteria can penetrate and firm inner layer which excludes bacteria (21). Mucins are highly glycosylated, this helps protect them from degradation by proteases in the gastrointestinal tract (22). The expression of mucin genes and fucosyltransferase genes are related to each other (23). Fucosyltransferases are responsible for the addition of fucose to mucins (24). One example of how the mucus is layer is influenced by individual mucin genotype is from the mucin 2 (muc2) knock out mouse. Muc2 is a major component of both the inner and outer mucin layers, and mice lacking muc2 are unable to exclude bacteria from the intestinal epithelium (21). MUC2 can also be degraded by commensal bacteria, providing both binding sites and a food source for bacteria (22). The commensal bacteria appear to have species specific compositions, and the constant repertoire of mucin glycans in the colon is one
suggested mechanism for the maintenance of commensal communities (22). Given the importance of both mucins and fucosyltransferases to the composition of gastrointestinal mucus, these genes are logical targets for studies of host genotype and intestinal microbiota colonization phenotype.

Several studies have shown that individuals homozygous for a null mutation in the Fucosyltransferase 2 (FUT2) gene tend to have less intestinal colonization with Bifidobacteria, have a less diverse microbiota, and have other differences in microbial community composition compared to individuals with at least one active FUT2 allele(25, 26). The intestinal microbiota colonization differences associated with FUT2 may also have implications for disease risk, as individuals homozygous for the null mutation in FUT2 are more likely to develop Crohn’s Disease than other FUT2 genotypes (26). Other fucosyltransferase genes may also influence the intestinal microbiota. For example, Fucosyltransferase 3 (FUT3) contributes to the production of an α(1,4)-linked fucose that can act as a binding site for Helicobacter pylori (27).

Studies directly comparing mucin genotype and intestinal microbiota phenotype are lacking. There is some evidence that changes in mucin genotype is associated with changes in intestinal microbiome phenotype (28). In particular, Inflammatory Bowel Disease is associated with altered Mucin 1 (MUC1), MUC2, Mucin 3A (MUC3A), Mucin 3B (MUC3B), Mucin 12 (MUC12), and Mucin 17 (MUC17) single nucleotide polymorphisms (SNPs) as well as with an altered microbial community structure (28, 29). Additional evidence of genetic difference in mucin genotype influencing bacteria exists in that H. pylori adheres significantly better to cell lines with larger MUC1 variable number tandem repeat domains (30).

The central purpose of this study is to determine the relationship between intestinal microbial composition and late-onset sepsis, and explore the relationship between microbial composition and host genetics (Figure 1). I hypothesize that distinct microbial patterns predate sepsis and that host genotype influences microbial colonization phenotype. To test this hypothesis, I will accomplish the following specific aims:
Aim 1: Compare microbial community composition over the first three weeks of life between premature infants who later develop sepsis and matched control infants who remain healthy.

Previous studies of the intestinal microbiota in late onset sepsis in premature infants have been limited by small sample sizes, with studies to date including only 6 to 10 cases \((1, 8)\). The research described here involves a substantially larger sample size than published studies to date, and uses infants from two cities, and thus, provides greater power to detect differences and improved generalizability. Before combining data from the two cities into a single analysis, I will compare the intestinal microbiota of control infants from the two cities to check that background colonization patterns are similar between the two sites.

Aim 2: Assess differences in the microbial community composition of premature infants in relation to their mucin and fucosyltransferase genotypes.

Emerging evidence suggests that host genotype can influence the type and abundance of bacteria capable of colonizing the human gut \((19, 25, 26)\), however this question is not well studied. Thus, the question posed here is innovative.

**Figure 1:** Conceptual model and proposed relationship between host genetics, intestinal microbial colonization, and late-onset neonatal sepsis. Aim 1 focuses on the association between intestinal microbial colonization and sepsis. Aim 2 focuses on the relationship of infant fucosyltransferase and mucin genes on microbial colonization.

**Significance to Environmental Health**
The types and amount of bacterial colonization represent an important part of the environment experienced by the host. This study is highly significant because of the critical role the microbiome plays in maintaining health and the dearth of information about how genotype interacts with microbial colonization patterns. Understanding how host genetics interacts with microbiome composition is critical to understanding which probiotics and prebiotics may ultimately prove most useful to prevent disease.

Innovation

This innovative study is larger than any previous study of the microbiome and sepsis in premature infants. It is the first study to look at how host mucin and fucosyltransferase genotype impact initial colonization of the premature infant gut.

Chapter 1 References

Chapter 2: Overview of Methods

Rational

The principal purpose of this study was to assess patterns of intestinal microbial colonization associated with sepsis in premature infants. Analysis of the preterm microbiome included infants who were enrolled in the study in level III Neonatal Intensive Care Units (NICUs) in Cincinnati, Ohio and Birmingham, Alabama. Intestinal microbial colonization was assessed by collection of stool samples over the first 3 weeks of life, and analysis of the 16s microbial community using Illumina Mi-Seq. The following are descriptions of subject enrollment and data collection processes and methods for sample collection, storage, and analysis. The next section of this chapter details methodologic analyses undertaken to determine the impact of collection, storage, and extraction conditions on analysis of the microbiome.

Clinical Population and Data Collection

Study population: All subjects included in my dissertation were enrolled in the Novel Biomarkers Cohort. This prospective cohort study was funded by the National Institute of Child Health and Human Development (NICHD; R01 HD059140) and designed to determine biomarkers of necrotizing enterocolitis. Enrolled infants were born at less than or equal to 32 weeks gestational age and were free of congenital anomalies. Four neonatal intensive care units (NICUs) were involved in enrolling infants in the study: one NICU was in Birmingham, AL and three NICUs were in Cincinnati, OH.

A total of 581 infants were enrolled in the Novel Biomarkers Cohort. The Birmingham NICU enrolled 175 infants. The Cincinnati NICUs enrolled at the largest site 283 infants, at the next largest site 120 infants, and at the smallest site 3 infants. Infants from the smallest Cincinnati NICU were subsequently excluded from analysis in my dissertation because of the very small number of infants enrolled at that site.
For the Birmingham NICU infants, only stool samples from infants born at less than 29 weeks gestational age (GA) were extracted for sequencing, limiting the intestinal microbiota analyses that used Birmingham data to infants less than 29 weeks GA. As a result, data from at least one sequenced stool sample was available for 77 of the 175 Birmingham infants. There were no GA restrictions placed on stool samples extracted for sequencing in Cincinnati. Data from at least one sequenced stool was available for 237 of 283 infants at the larger Cincinnati NICU and 103 of 120 infants at the next largest Cincinnati NICU. Because microbiome data from Birmingham infants was available only for infants less than 29 weeks GA, any portion of my dissertation that looked at both Birmingham and Cincinnati infants was restricted to infants less than 29 weeks GA. For each of the analyses conducted, the specific infants and samples included in that portion of the study are as described in relevant chapter. Subsets of infants were selected for each analysis rather than using the entire cohort to minimize the potential for confounding.

**Data collection methods.** Clinical data was collected in accordance with the NICHD’s Neonatal Research Network (NRN) protocols. The NRN began in 1987 as an initiative to conduct multicenter studies of therapeutic interventions and as such created a “generic” database to measure practice at baseline by creating generic data forms and a compendium of definitions (1). NRN protocols collect extensive data relevant to this dissertation (1), including timing of onset of sepsis; causative organism of sepsis; timing, type, and duration of antibiotic use; maternal history of antibiotic use during delivery; infant race, gender, and ethnicity; infant GA, birthweight, and birth length; delivery mode; and infant discharge or death. The Novel Biomarkers Cohort also collected extensive information on infant feeding including information on formula or human milk and date of first enteral feeding.

**Stool Collection:** The Novel Biomarkers Cohort also collected biological samples from enrolled subjects. Collected samples include maternal and infant DNA samples, infant saliva samples,
infant urine samples, and infant stool samples. The scheduled collection of stool samples was targeted for infant days of life 5, 8, 11, 14, and 21 plus or minus two days for each sample.

Because early analysis found a connection between week 1 of life (infant days of life 4 to 9) and NEC and between week 2 of life (infant days of life 10 to 16) and NEC (2), stool samples from infants enrolled later during the study were collected more intensively from days of life 4 to 14, but the day of life 21 stool collection was ended along with the sequencing of DNA from banked week 3 stool samples. This limited the time in which week 3 samples could be compared between the Birmingham and Cincinnati NICUs. If an infant stooled during a target collection window, nurses placed the dirty diaper in a biohazard bag and stored it in a NICU refrigerator at 4°C until laboratory staff could retrieve the sample. Samples were transported on ice to the laboratory, where the stool was scraped from the diaper into a storage vial and either frozen directly at -80°C or covered in thioglycollate and then frozen at -80°C.

**Bacterial DNA Extraction**: Bacterial DNA was extracted using one of three methods:

1. a phenol-chloroform extraction
2. a QiaGen QiaAmp Stool Kit
3. a QiaGen AllPrep DNA/RNA Mini Kit.

All methods included a beading beating step, for additional details please see Chapter 4.

**16s Analysis Methods**

**Sequencing and Bioinformatics**: All sequencing and the bioinformatics pipeline was completed at the Broad Institute (Boston, MA.) Using extracted DNA, 180nt paired-end reads were generated using established primers and protocols, with samples allocated across multiple Illumina MiSeq runs (3). Sequence results from all extracted stool samples in the Novel Biomarkers Cohort were run through the bioinformatics pipeline concurrently. Spurious reads
can occur with polymerase chain reaction (PCR), therefore the first step was to filter out poorly
copied sequences and any reads resulting from non-specific binding of the primers to organelles
or regions of DNA other than the 16s rRNA gene in bacteria. Read pairs were merged to create
amplicon-spanning sequences that were then filtered to remove those with less than 70%
identity to any read in the rRNA16S.gold.fasta reference set
(http://drive5.com/uchime/uchime_download.html) using "usearch -usearch_global -id 0.70".
79,076,883 sequences were processed by the UPARSE pipeline (4), software version
usearch7.0.959_i86linux64. The following commands were used with default settings unless
otherwise specified. Dereplication resulted in 35,605,130 sequences (-derep_fulllength);
removal of singleton reads in 2,206,563 sequences (-sortbysize -minsize 2), and clustering
yielded 7,249 OTU representative sequences (-cluster_otus). The OTU table was constructed
by mapping reads to OTUs (-usearch_global -strand plus -id 0.97) and applying the python
script uc2otutab.py (http://drive5.com/python/). Additional chimera filtering was not applied.
Chimera checking removes OTUs that form as the result of overlap from two or more actual
sequences to form a new hybrid sequence (5). QIIME (6) version 1.6 was used to provide
classifications of the OTU representative sequences using the gg_13_5 GreenGenes taxonomy
and representative sequences constructed at 99% ID. A phylogenetic tree was constructed
within the Qiime package using fasttree and filtered pynast alignments of the OTU
representative sequences. OTUs were removed from the OTU table using a filter with a
minimum count fraction of 0.0002 in Qiime, resulting in 525 unique OTUs. Since sequencing
data is inherently noisy, removal of OTUs that occur infrequently further reduces the chance of
including OTUs that resulted from sequencing error (7). The resulting data was structured into
an OTU table, consisting of one row for each sample and one column for each OTU. Values in
each cell represented the number of reads belonging to a given OTU in a given sample.

**Statistical Analysis:** The statistical tools used most frequently in this dissertation are non-metric
multidimensional scaling (NMDS) and Linear Effect Size Estimator (LEfSe). NMDS is a non-
parametric ordination technique that uses ranked distances and is compatible with any distance measure (8). The distance measures I have chosen to use are weighted and unweighted UniFrac, which are measures that account for the phylogenetic relatedness of samples. Weighted UniFrac also accounts for the relative abundance of taxa while unweighted UniFrac only accounts for the presence or absence of taxa (9). When using NMDS stress and the number of dimensions should be minimized and a stable solution should be sought (8). A stress less than 2.5 is considered excellent (8). For all NMDS plot reported here, stress was less than 2.5 and the number of dimensions was 2 and a stable solution was found.

LEfSe screens for potential biomarkers of user specified groups (10). LEfSe first uses a Kruskal-Wallace test to compare all levels of taxa between the groups, and then checks if those differences reach a specified effect size using linear discriminant analysis (10). Unless otherwise specified, analyses conducted with LEfSe used a p-value of <0.05 for the Kruskal-Wallace test and the default effect size of 2. LEfSe produces plots of differences between groups ranked by effect size and cladograms (10). Cladograms are essentially phylogenetic trees presented in a circle, with the central circle representing kingdom and each successive ring representing one step lower on the phylogenetic tree (e.g., the first ring from the center is phyla and the next ring after that is class.) Phylogenetic differences between groups are indicated by highlighting in different colors different portions of the phylogenetic tree. For more details on LEfSe and examples of how to use LEfSe please see the study by Segata et. al. (10)

Alpha diversity was estimated using the Simpson Diversity Index and the Chao1 index. The Simpson Diversity Index represents the probability that two randomly selected reads belong to the different OTUs, and as such allows for both richness (the total number of species present in a sample) and evenness (the distribution of species in a sample) (8). The Chao1 index estimates richness by dividing the number of species with only a single read present to the number of species with exactly two reads present and adding that value to the total number of
species present. The concept is that if species with only a single exemplar present are still being found, there are likely still more species to find but if every species has at least two exemplars, there are likely not more species to find (11). This means that given two sample populations with 100 reads each of 10 different species, with population A having 10 reads of each species and population B having 82 reads of 1 species and 2 reads each of the remaining 9 species, the Chao1 index would find that populations A and B have the same diversity but that the Simpson Diversity Index would find that population A is more diverse than population B.

In some analyses, I used Jaccard index to estimate the rate of change in the intestinal microbiota between samples. Jaccard index is a proportion coefficient that measures the overlap in OTUs between two samples (8). The greater the degree of similarity between two samples, the lower the value of the Jaccard index.

**Methodologic Analyses.**

To examine the impact of study methods on resulting analysis of the microbiome, we analyzed:

1. the time between sample collection and freezing, that is, the length of time samples were refrigerated in study NICUs prior to freezing in our research lab
2. Storage buffer use
3. DNA extraction methods.

**Time between sample collection and freezing:** Consistent with a studies by Wu et. al. and Lauber et. al. (12, 13) the length of time stool samples were stored at 4°C prior to freezing had only a minimal impact on results. In our samples from Cincinnati, there was no difference in alpha diversity between samples that were frozen in less than 24 hours and those that were frozen 24 hours or more after initial collection by a nurse. The only difference detected between samples frozen in less than 24 hours and samples frozen after 24 hours or more was an increase in *Lactococcus* in the samples frozen within 24 hours (Figure 1).
Figure 1: Results of running Linear Effect Size Estimator (LEfSe) comparing samples collected within one day and later than one day. LEfSe uses a Kruskal-Wallis test to search for differences between groups of samples on all taxonomic levels. Significance was set at p<0.05. The only difference detected between the samples frozen within 24 hours and those frozen after 24 hours or more was an increase in *Lactococcus* in the samples frozen within 24 hours.

**Storage buffer**: Stool samples frozen in thioglycollate were also compared to those frozen without any buffer. A subset of 7 samples from Cincinnati in 2009 were stored without any thioglycollate, these were frequency matched to samples stored in thioglycollate on infant day of life at collection and infant history of necrotizing enterocolitis (NEC). Unlike all other analyses in my thesis, this analysis was conducted using 454 pyrosequencing data that was originally used on samples from the Novel Biomarkers Cohort. This data was used instead of the Illumina MiSeq data because only 2 of the 7 samples frozen without thioglycollate were successfully run on the Illumina MiSeq platform. 454 pyrosequencing produces longer reads that can be successfully classified to the genus level, but also provides less total coverage of the infant microbiota. No differences between the samples were identified using non-metric multidimensional scaling (NMDS) based on either weighted on unweighted UniFrac distance. A comparison using LEfSe found that stool samples frozen without thioglycollate (stool normal
samples) were more likely to contain *Weissella, Flavobacteriales, Flavobacteriaceae, Chryseobacterium, Flavobacteria, Bacteroidetes, Leuconostoc, Leuconostocaceae, Propionibacterineae, Propionibacteriaceae, and Propionibacterium* (Figure 2). The samples stored in thioglycollate also showed a trend towards lower alpha diversity (Figure 3), though this did not reach statistical significance in this small sample.

**Figure 2**: Comparison of bacteria found in samples frozen without any buffer (stool normal) and samples frozen in thioglycollate (stool thioglycollate). Red bars indicate bacteria taxa enriched in stool normal samples and the green bar indicates bacteria taxa enriched in stool thioglycollate samples. OTUs are identified by the letter X followed by the OTU number.
Figure 3: Alpha diversity in samples stored without buffer (stool normal) and with thioglycollate (stool thioglycollate). Panel A: Boxplot showing that Simpson Diversity Index trends lower in samples stored in thioglycollate. Panel B: Boxplot showing that Chao1 Index trends lower in samples stored in thioglycollate.

DNA extraction methods: Extraction protocol did influence what bacteria were detected by 16s rRNA sequencing (Figure 4), necessitating either restriction to a single protocol or careful
balancing of extraction protocol between comparison groups. This is consistent with the findings of Wu et. al. (12)

Figure 4: LEfSe comparing the three extraction protocols in week 2 Cincinnati infants. Phenol-chloroform extracted samples (in red) were enriched in *Pseudomonadales, Leuconostocaceae, Leuconostoc*, and *Viellonella*. QiaAmp Stool Kit extracted samples (in green) were enriched in
Citrobacter. AllPrep DNA/RNA Mini Kit extracted samples (in blue) were enriched in
Bacteroidaceae, Bacteroides, Proteus, Streptococcaceae, Streptococcus, and Serratia
Marcescens.

Chapter 2 References

Chapter 3: Differences in Intestinal Colonization of Infants in Geographically Separate Neonatal Intensive Care Units Over Time

Rationale

The primary aim of this dissertation was to identify differences between sepsis cases and controls. However, careful examination of the data revealed that overall differences in the microbiome of infants in Birmingham and Cincinnati overwhelmed any potential differences that could be observed between cases and controls, even with careful matching of hospital of origin. This indicated the need for a thorough analysis comparing intestinal colonization patterns of infants in Birmingham and Cincinnati NICUs, which led to my initial manuscript from this work. The remaining text of this chapter (below) represents the manuscript I submitted to a peer-reviewed journal.
Intestinal Microbiota of Preterm Infants Differ Over Time and Between Hospitals

Short title: Preterm microbiota between sites and over time

Authors

Diana H. Taft\textsuperscript{1,2}, Namasivayam Ambalavanan\textsuperscript{3}, Kurt R. Schibler\textsuperscript{1}, Zhuoteng Yu\textsuperscript{4}, David S. Newburg\textsuperscript{4}, Doyle V. Ward\textsuperscript{5}, Ardythe L. Morrow\textsuperscript{1,2,6}

\textsuperscript{1} Perinatal Institute, Cincinnati Children’s Hospital Medical Center, Dept. Pediatrics, 3333 Burnet Ave, Cincinnati, Ohio 45229, USA
\textsuperscript{2} Department of Environmental Health, University of Cincinnati College of Medicine, 3223 Eden Ave, Cincinnati, Ohio 45267, USA
\textsuperscript{3} Department of Pediatrics, University of Alabama at Birmingham, 1700 6\textsuperscript{th} Ave S, 9380 176F WIC, Birmingham, Alabama 35249, USA
\textsuperscript{4} Boston College, Department of Biology, Higgins Hall, 140 Commonwealth Ave, Chestnut Hill, Massachusetts 02467, USA
\textsuperscript{5} Broad Institute, 320 Charles St, Cambridge, Massachusetts 02141, USA
\textsuperscript{6} Division of Biostatistics and Epidemiology, Cincinnati Children’s Hospital Medical Center, 3333 Burnet Ave, Cincinnati, Ohio 45229, USA

Corresponding author:

Ardythe L. Morrow, PhD
Professor of Pediatrics, Nutrition, and Environmental Health
Perinatal Institute, Cincinnati Children’s Hospital Medical Center
3333 Burnet Ave., MLC 7009
Cincinnati, OH 45229
Phone: 513-636-7626, Fax: 513-803-0969, Email: Ardythe.Morrow@cchmc.org
Abstract

Background: Intestinal microbiota are implicated in risk of necrotizing enterocolitis (NEC) and sepsis, major diseases of preterm infants in neonatal intensive care units (NICUs). Rates of these diseases vary over time and between NICUs, but comparison of the intestinal microbiota of preterm infants over time and between NICUs is lacking.

Methods: We included 66 singleton infants <29 weeks gestational age with stool samples collected in postnatal weeks 1 to 3 of life who survived free of NEC and sepsis. Infants were enrolled during 2010 and 2011. Twenty-six infants were enrolled at hospital-1 in Cincinnati, OH and 40 infants were enrolled at hospital-2 in Birmingham, AL. Samples from postnatal weeks 1, 2, and 3 were analyzed separately. DNA extracted from stool was used to sequence the 16s rRNA gene by Illumina MiSeq using universal primers. Resulting operational taxonomic unit tables were analyzed for differences between years and hospitals using linear discriminant analysis effect size algorithm (p<0.05).

Results: Significant variation was observed in infant microbiota by year and hospital. Among hospital-1 infants, week 1 samples had more phylum Firmicutes (class Bacilli) in 2010 and more phylum Proteobacteria (family Enterobacteriaceae) in 2011. While week 2 samples did not significantly vary over time, week 3 samples from hospital-1 had more Bacilli in 2010 than in 2011. However, among hospital-2 infants, the week 1 shift was nearly opposite, with more Proteobacteria (family Enterobacteriaceae) in 2010 and more Firmicutes (class Bacilli) in 2011. Week 2 samples at hospital-2 exhibited the same pattern. Regression analysis of clinical covariates found that antibiotic use of mothers and infants had an important influence, but did not explain these observed shifts in microbiota over time within and between hospitals. Microbial succession from week 1 to 3 also differed by hospital, with greater compositional change in microbiota observed in hospital-1 than hospital-2 infants (p<0.01, Kruskal Wallis test of Jaccard distance metric).
**Conclusion:** Colonizing microbiota differ over time and between NICUs in ways that could be relevant to disease. These data suggest the need for multi-site, longitudinal studies to reliably define the impact of intestinal microbiota on adverse outcomes of preterm infants.

**Word count:** 350

**Keywords**

Infants, premature; Microbiome; Geo-temporal analysis; Microbial succession
Background

John Snow’s investigation of cholera in 1854 pioneered the understanding of disease risk in relation to not only individual characteristics (person), but also geographic patterns (place), and temporal patterns (time) (1). Since then, epidemiologic investigations have advanced our understanding of the carriage frequency, pathogenicity, and transmission dynamics of enteric organisms in relation to person, place, and time. With the advent of high throughput sequencing methods, research on the human intestinal microbiome has led to important insights regarding the impact of personal characteristics such as age, gender, diet, and disease state on the colonizing microbial communities. By comparison, few studies have investigated geographic and temporal influences on human microbial communities. Understanding temporal and geographic influences on the colonizing human microbiota is a critical step towards understanding the transmission dynamics of microbial communities. Such understanding may also help understand the extent to which studies undertaken in a population defined by time and location can be generalized to the same population at a later time or populations at other locations.

Geographic differences in the human intestinal microbiota have been more extensively studied than temporal differences within populations. For example, Bacteroidetes are known to dominate the microbiomes of African children, while the microbiota of European children are enriched in Firmicutes and Proteobacteria (2). Similarly, Lee et al. reported differences in the microbiota of Korean and U.S. twins, particularly in taxa belonging to Firmicutes and Bacteroidetes (3). Other comparative studies The microbiota of built environments can differ significantly (4); such differences could be attributed to the broader local environment or to the inhabitants of the built environment. A study of the colonization patterns of preterm infants in three distinct Florida neonatal intensive care units (NICUs) demonstrated significant differences in the relative abundance of the phyla Bacteroidetes, Proteobacteria, and Firmicutes in the inhabitants of these NICUs (5). Variations in infant microbiota can lead to differences in
environmental exposures and vice versa, as bacteria in the NICU environment are also found to colonize the preterm infant gut (6).

It is not known whether colonization patterns of intestinal microbial communities shift within human populations over calendar time, but temporal patterns are well documented with many enteric pathogens(7). Cholera outbreaks have one or two annual peaks, in spring and fall. Helicobacter pylori also shows peak colonization in children in spring and fall (8, 9). Although it is unclear if commensals also exhibit such temporal patterns, one study found that the microflora of buildings varied over a single year of sampling, and in one of the studied buildings there was a shift from Actinobacteria to Proteobacteria during the year (4).

To address the question of temporal and geographic differences in the intestinal microbiota, we conducted a prospective study of the intestinal microbiota of preterm infants at two large, geographically separated level III NICUs. We used two years of data from this cohort study to test the hypothesis that initial microbiota colonization of disease-free survivors changes over time and differs between birth hospitals.

**Methods**

**Subjects**: Study infants were a subset of the infants enrolled from one level III NICU in Cincinnati, OH (hospital-1) and one level III NICU in Birmingham, AL (hospital-2) in 2010 and 2011 as part of an ongoing cohort study of novel biomarkers for necrotizing enterocolitis (NEC). Study infants were ≤32 weeks gestational age. The average daily census of the NICU in hospital-1 was 46 with a capacity of 60 and average daily census in the NICU at hospital-2 was 85 with a capacity of 120. From the overall cohort, this study was restricted to infants <29 weeks gestational age, singleton births, who survived free of proven NEC or proven sepsis until discharge, and were free of congenital anomalies. These inclusion criteria were applied to establish fundamental comparability between the two NICUs studied, and remove the influence of disease outcomes per se, or other potentially extraneous differences. In addition, all infants had at least one stool sample from week 1 or week 2 that was successfully extracted and
sequenced using the methods described below. The Institutional Review Boards at Cincinnati Children’s Hospital and the two participating hospitals approved the study. Parent or guardian consent was obtained for all infants included in this study.

Sample collection and storage: Serial stool samples were collected from infants during the first weeks of life on the following schedule: Postnatal days 5, 8, 11, 14, and 21 plus or minus 2 days. Due to intermittency of infant stooling, we categorized samples as having been collected in week 1 (3-9 days of life), week 2 (10-16 days of life), and week 3 (17-23 days of life) in order to optimize the number of infant samples included for analysis at each time point. These time windows were selected because no study infant stoole prior to day of life 3, which is typical for premature infants (10). Samples were collected from soiled diapers, immediately refrigerated in the NICU and transported to the laboratory where they remained in the refrigerator until processing for storage at -80°C. Cryogenic storage in hospital-1 included thioglycollate buffer except for samples with target day of life of collection 11, by chance more day of life 11 samples were included in week 2 of 2011 than in week 2 of 2010 from hospital-1. Storage in hospital-2 never used thioglycollate.

To determine the impact of thioglycollate buffer on the microbiome, we compared samples from Cincinnati in 2009 - 7 stored without and 7 stored with thioglycollate - matched on day of life of collection and infant clinical history. We found that samples stored without thioglycollate were enriched in Bacteroidetes, Leuconostocaceae, and Propionibacterineae. As samples from the Birmingham NICU were stored without thioglycollate and samples from the Cincinnati NICU were stored with thioglycollate during the study period, identified differences between the Birmingham and Cincinnati sites in these taxa were considered a potential artifact of the storage protocol. Our analyses therefore excluded these taxa from further consideration, and were restricted to taxa not influenced by storage protocol.

Stool extraction and 16s rDNA sequencing: If stored in thioglycollate, stool was thawed and centrifuged 10 minutes at 4000xg and the supernatant was removed. For all samples, 100 µL of
TE buffer with lysozyme and proteinase K were added to 0.24 g of thawed stool and vortexed for 10 minutes. 1.2 mL of buffer RLT with beta-mercaptoethanol were added to the sample and transferred to sterile bead beating tubes containing 0.3 g of 0.1 mm glass beads. Samples were homogenized for 3 minutes in a bead beater and centrifuged at 4000xg for 5 minutes to pellet debris. The supernatant was transferred to a clean microcentrifuge tube and spun at 4000xg for an additional 2 minutes to remove remaining debris. Supernatant was then transferred to a QiaGen AllPrep DNA spin column and DNA and RNA were isolated using the QiaGen AllPrep DNA/RNA mini kit (Valencia, CA).

Using extracted DNA, 180nt paired-end reads were generated using established primers and protocols, with samples allocated across multiple Illumina MiSeq runs (11). Read pairs were merged to create amplicon-spanning sequences that were then filtered to remove those with less than 70% identity to any read in the rRNA16S.gold.fasta reference set (http://drive5.com/uchime/uchime_download.html) using "usearch -usearch_global -id 0.70". 79,076,883 sequences were processed by the UPARSE pipeline (12), software version usearch7.0.959_i86linux64. The following commands were used with default settings unless otherwise specified. Dereplication resulted in 35,605,130 sequences (-derp_fulllength); removal of singleton reads in 2,206,563 sequences (-sortbysize -minsize 2), and clustering yielded 7,249 OTU representative sequences (-cluster_otus). The OTU table was constructed by mapping reads to OTUs (-usearch_global -strand plus -id 0.97) and applying the python script uc2otutab.py (http://drive5.com/python/). Additional chimera filtering was not applied.

QIIME (13) version 1.6 was used to provide classifications of the OTU representative sequences using the gg_13_5 GreenGenes taxonomy and representative sequences constructed at 99% ID. A phylogenetic tree was constructed within the Qiime package using fasttree and filtered pynast alignments of the OTU representative sequences. OTUs were removed from the OTU table using a filter with a minimum count fraction of 0.0002 in Qiime, resulting in 525 unique OTUs.
Statistical Analysis: Differences in clinical characteristics among groups by week were tested using Fisher’s Exact test for categorical variables and t-test for continuous variables. To standardize comparisons of microbiota, we rarefied the OTU table to 2000 reads per sample. Rarefaction randomly selects reads from the complete set obtained for each sample until the specified number of reads is obtained. This means that each sample has an equal chance of including rare OTUs so that the samples can be compared.

Alpha diversity was calculated for weeks 1 and 2 using two metrics; Simpson Diversity Index (\(1-D\)) and Chao1. Kruskal-Wallis (KW) was used to test for differences in alpha diversity by year and hospital.

To examine beta diversity, we used non-metric dimensional scaling (NMDS) to ordinate the microbial communities based on both the unweighted and weighted UniFrac distance calculated in QIIME as described in Morrow et al. (14). The unweighted UniFrac examines presence/absence only while the weighted UniFrac accounts for abundance differences. Significant differences in specific taxa between hospitals and by year were identified by Linear Discriminant Analysis Effect Size (LEfSe) (15). Generalized estimating equations (GEE) models were used assuming linear and logistic relationships to test the association of taxa identified by LEfSe with hospitals after adjustment with other potential confounders. A backwards elimination approach was used to remove non-significant covariates from the models. Samples from the same infant in different weeks were included in the same model.

Differences in degree of succession between hospitals were tested using the Jaccard index. Values for the Jaccard index were calculated between the weeks 1 and 3 samples from 28 infants who had samples collected across the first three weeks of life. The Jaccard index is a distance metric used to show similarity over time. Identical communities will have a Jaccard index value of 0 while completely non-overlapping communities will have a Jaccard index value of 1. KW was used to test for differences in the intra-subject Jaccard index values by hospital and year.
Results

Subjects: Sequence data were generated from a total of 66 infants with 51 samples in week 1 (22 from hospital-1 and 29 from hospital-2, both years), 60 samples in week 2 (24 from hospital-1 and 36 from hospital-2, both years), and 40 samples in week 3 (23 from hospital-1 and 17 from hospital-2, both years.) A total of 28 infants (18 from hospital-1 and 10 from hospital-2, both years) had samples available from all three weeks for use in the analysis of microbial succession. The infant characteristics by hospital and by week are presented in Table 1. Infants at both hospitals were generally well matched demographically in week 1 and week 2, with more black infants born and more infants born to slightly younger mothers at hospital-2. In week 3, only a single sample was collected in 2011 from a hospital-2 infant, limiting the analysis that could be conducted during week 3. Otherwise, week 3 infant characteristics were comparable to weeks 1 and 2 infant characteristics. Infant characteristics within a single hospital in weeks 1 and 2 between 2010 and 2011 were similar with the exception that hospital-2 mothers were younger in 2010 than in 2011 (weeks 1 and 2, p<0.05, data not shown). Infants with week 3 stool sample from hospital-1 tended to have a younger gestational age (p<0.05) and later day of first enteral feeding (p<0.05) in 2010 compared to 2011. Similar types of antibiotics were used at both hospitals, a combination of ampicillin and gentamicin for the majority of infants. Infants at hospital-2 were also more likely to be exposed to antibiotics and for longer durations in the first 14 days of life. Also, a higher percentage of mothers received antibiotics at the time of delivery at hospital-2. At hospital-1, three infants received no antibiotics and one infant received nafcillin and gentamycin. At hospital-2, eight infants received ampicillin and tobramycin. The first day of life of enteral feeding was similar at both hospitals but the hospitals differed in their use of formula. Hospital-1 infants received exclusively either mother’s own milk or pasteurized human donor milk for the first 14 days of life, while approximately 20% of hospital-2 infants were fed formula (Table 1).
Table 1: Characteristics of infants with samples included in analysis, by week. Hospital-2 week 3 infants are not shown because no hospital-2 week 3 comparison was conducted. There was only a single week 3 sample from an infant born at hospital-2 in 2011.

<table>
<thead>
<tr>
<th></th>
<th>Hospital 1 Week 1</th>
<th>Hospital 2 Week 1</th>
<th>Hospital 1 Week 2</th>
<th>Hospital 2 Week 2</th>
<th>Hospital 1 Week 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2010</td>
<td>2011</td>
<td>2010</td>
<td>2011</td>
<td>2010</td>
</tr>
<tr>
<td>n=9</td>
<td>n=12</td>
<td>n=17</td>
<td>n=11</td>
<td>n=13</td>
<td>n=15</td>
</tr>
<tr>
<td></td>
<td>n=15</td>
<td>n=21</td>
<td>n=11</td>
<td>n=12</td>
<td></td>
</tr>
<tr>
<td>Mean birthweight ± sd (g)</td>
<td>963±216</td>
<td>965±261</td>
<td>938±234</td>
<td>985±193</td>
<td>914±216</td>
</tr>
<tr>
<td></td>
<td>93±213</td>
<td>97±261</td>
<td>95±223</td>
<td>99±261</td>
<td>96±207</td>
</tr>
<tr>
<td></td>
<td>914±216</td>
<td>962±207</td>
<td>932±227</td>
<td>1016±261</td>
<td></td>
</tr>
<tr>
<td>Mean birthlength ± sd (cm)</td>
<td>35.8±3.0</td>
<td>34.0±2.9</td>
<td>34.0±3.1</td>
<td>34.8±2.9</td>
<td>35.8±3.5</td>
</tr>
<tr>
<td></td>
<td>35.2±2.9</td>
<td>35.5±2.9</td>
<td>35.6±2.9</td>
<td>35.7±3.5</td>
<td>35.4±3.1</td>
</tr>
<tr>
<td></td>
<td>37.0±2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26.2 (25.5-27.5)</td>
<td>27.1 (26-28)</td>
<td>27.1 (26-28)</td>
<td>27.0 (26-28)</td>
<td>25.8 (24.5-26.5)</td>
</tr>
<tr>
<td></td>
<td>27.3 (26.8-28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (%)</td>
<td>5 (56%)</td>
<td>6 (46%)</td>
<td>5 (42%)</td>
<td>7 (41%)</td>
<td>4 (36%)</td>
</tr>
<tr>
<td></td>
<td>7 (54%)</td>
<td>4 (40%)</td>
<td>11 (52%)</td>
<td>7 (58%)</td>
<td>4 (36%)</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>0 (0%)</td>
<td>1 (7.7%)</td>
<td>1 (8.3%)</td>
<td>0 (0%)</td>
<td>1 (7.7%)</td>
</tr>
<tr>
<td></td>
<td>0 (0%)</td>
<td>1 (7.7%)</td>
<td>1 (6.7%)</td>
<td>0 (0%)</td>
<td>1 (8.3%)</td>
</tr>
<tr>
<td>Black (%)</td>
<td>5 (56%)</td>
<td>6 (46%)</td>
<td>8 (67%)</td>
<td>6 (35%)</td>
<td>10 (67%)</td>
</tr>
<tr>
<td></td>
<td>3 (27%)</td>
<td>2 (15%)</td>
<td>8 (38%)</td>
<td>6 (35%)</td>
<td>10 (67%)</td>
</tr>
<tr>
<td></td>
<td>4 (36%)</td>
<td>2 (17%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Csection (%)</td>
<td>5 (56%)</td>
<td>5 (38%)</td>
<td>6 (50%)</td>
<td>10 (59%)</td>
<td>6 (55%)</td>
</tr>
<tr>
<td></td>
<td>3 (27%)</td>
<td>2 (15%)</td>
<td>7 (47%)</td>
<td>10 (48%)</td>
<td>7 (55%)</td>
</tr>
<tr>
<td></td>
<td>5 (38%)</td>
<td>2 (17%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maternal abx at delivery (%)</td>
<td>6 (67%)</td>
<td>10 (77%)</td>
<td>11 (92%)</td>
<td>17 (100%)</td>
<td>14 (93%)</td>
</tr>
<tr>
<td></td>
<td>8 (73%)</td>
<td>10 (77%)</td>
<td>14 (93%)</td>
<td>20 (95%)</td>
<td>7 (64%)</td>
</tr>
<tr>
<td>Mean maternal age ± sd (years)**</td>
<td>27.4±1.6</td>
<td>28.0±1.4</td>
<td>22.5±4.9</td>
<td>27.8±6.1</td>
<td>21.9±5.1</td>
</tr>
<tr>
<td></td>
<td>28.5±6.7</td>
<td>29.1±5.6</td>
<td>21.9±5.1</td>
<td>27.7±5.7</td>
<td>28.4±6.8</td>
</tr>
<tr>
<td></td>
<td>29.5±5.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal parity (25th-75th percentile)</td>
<td>2.0 (1-3)</td>
<td>2.3 (1-3)</td>
<td>2.5 (1-3)</td>
<td>1.9 (1-2)</td>
<td>2.5 (1.5-3)</td>
</tr>
<tr>
<td></td>
<td>2.3 (1-3)</td>
<td>2.5 (1-3)</td>
<td>1.9 (1-2)</td>
<td>2.5 (1.5-3)</td>
<td>2.1 (1-3)</td>
</tr>
<tr>
<td></td>
<td>2.1 (1-3)</td>
<td>1.9 (1-3)</td>
<td>2.5 (1.5-3)</td>
<td>2.4 (1-3)</td>
<td></td>
</tr>
<tr>
<td>Preeclampsia (%)</td>
<td>1 (11%)</td>
<td>2 (15%)</td>
<td>3 (25%)</td>
<td>1 (5.8%)</td>
<td>3 (20%)</td>
</tr>
<tr>
<td></td>
<td>1 (9.1%)</td>
<td>3 (2.3%)</td>
<td>3 (20%)</td>
<td>1 (4.8%)</td>
<td>2 (18%)</td>
</tr>
<tr>
<td></td>
<td>1 (4.8%)</td>
<td>3 (25%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days on abx in first 7 days ± sd</td>
<td>4.0±2.6</td>
<td>3.1±2.1</td>
<td>4.7±2.0</td>
<td>3.8±1.7</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Days on abx in first 14 days ± sd</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>NA</td>
<td>4.4±3.1</td>
<td>3.6±3.8</td>
<td>7.6±4.4</td>
</tr>
<tr>
<td></td>
<td>5.4±3.4</td>
<td>4.7±3.3</td>
<td>2.8±2.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day of life first enteral feed, median (25th-75th percentile)***</td>
<td>4.6 (3-5)</td>
<td>2.5 (2-3)</td>
<td>3 (2-3.5)</td>
<td>2.8 (2-3)</td>
<td>3.6 (2-4.5)</td>
</tr>
<tr>
<td></td>
<td>3 (2-4)</td>
<td>3.8 (3-5)</td>
<td>2.6 (2-3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formula feeds in first 7 days (% total feeds)</td>
<td>0 (0%, 45)</td>
<td>0 (0%, 81)</td>
<td>18 (26%, 69)</td>
<td>18 (19%, 96)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>18 (19%, 96)</td>
<td>18 (19%, 96)</td>
<td>18 (19%, 96)</td>
<td>18 (19%, 96)</td>
<td>18 (19%, 96)</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Formula feeds in first 14 days (% total feeds)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0 (0%, 59)</td>
<td>21 (26%, 81)</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>NA</td>
<td>0 (0%, 59)</td>
<td>24 (21%, 117)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>NA</td>
<td>21 (26%, 81)</td>
<td>24 (21%, 117)</td>
<td>21 (26%, 81)</td>
</tr>
<tr>
<td>Formula feeds in first 21 days (% total feeds)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>12 (5.5%, 218)</td>
<td>13 (4.5%, 289)</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>12 (5.5%, 218)</td>
<td>13 (4.5%, 289)</td>
</tr>
</tbody>
</table>

*Significantly different between years, hospital-1 week 3, p=0.012; **Significantly different between years hospital-2 week 1, p=0.015 and hospital-2 week 2, p=0.0031; ***Significantly different between years hospital-1 week 3, p=0.028
**Microbial Diversity:** We observed no significant differences in alpha diversity in week 1 or week 2 samples, by hospital or year, using Simpson or Chao1 metrics. Furthermore, clear separation between hospitals or years was not observed in weighted ordinations (not shown). However, unweighted UniFrac ordinations stratified by hospital showed a trend towards separation based on year (Figure 1.) Weighted UniFrac is a distance measure that takes into account both relative abundance of the detected OTUs and the phylogenetic relatedness of the detected OTUs. Unweighted UniFrac is a distance measure that uses the presence/absence of OTUs and their phylogenetic relatedness. The difference in the unweighted UniFrac ordination combined with the lack of difference in the weighted UniFrac ordination suggest that there are significant differences in the OTUs present between years at each of the two hospitals, but that the taxa comprising the majority of the reads are similar between years.
Figure 1: NMDS ordination of week 1 and week 2 samples for each hospital based on unweighted UniFrac. Samples from 2010 are indicated as red dots and samples from 2011 are indicated as black dots. In each ordination the 2010 samples (red dots) tend to be more on the left side while the 2011 samples (black dots) tend to be more on the right suggesting a shift in microbial community composition between the years. Panel A: Week 1 results from hospital-1. Number of dimensions was 2 and stress was 0.17. Panel B: Week 1 results from hospital-2. Number of dimensions was 2 and stress was 0.19. Panel C: Week 2 results from hospital-1. Number of dimensions was 2 and stress was 0.20. Panel D: Week 2 results from hospital-2. Number of dimensions was 2 and stress was 0.19.

In week 1 samples, comparison of infants within each hospital by year revealed that for hospital-1 the 2010 infants were enriched in *Firmicutes* while the 2011 infants were enriched in *Proteobacteria* (Figure 2A). Differences also occurred at lower phylogenetic levels between
years, with Clostridiaceae and Enterococcaceae (families of phylum Firmicutes) enriched in 2010. In 2011, infants were enriched in Enterobacteriaceae, the largest family of phylum Proteobacteria. In week 1 samples of hospital-2, infants showed the opposite pattern by year on the phylum level, such that the 2010 infants were enriched in Proteobacteria while the 2011 infants were enriched in Firmicutes (Figure 2B). Hospital-2 infants also showed differences at lower phylogenetic levels. In 2010, they were enriched in family Enterobacteriaceae while in 2011 they were enriched in class Bacilli. As a correlate of these shifts within hospitals, in 2010 the infant microbiota differed between hospitals with hospital-1 infants having more Firmicutes and hospital-2 infants having more Proteobacteria. In 2011, infants from both hospitals had similar microbiota, with only OTU level differences detected (not shown.) Overall in hospital-1, Firmicutes was more abundant in 2010 than in 2011 (Figure 3A) whereas Proteobacteria (Figure 3B) was less abundant in 2010 and more abundant in 2011. In hospital-2 samples, phylum Proteobacteria (family Enterobacteriaceae) was slightly more abundant in 2010 than 2011 (Figure 3A), and phylum Firmicutes was slightly less abundant in 2010 and more abundant in 2011 (Figure 3B).

In week 2 samples, comparing infants within hospital-1 by year revealed that colonization patterns were relatively stable between 2010 and 2011 demonstrating only OTU level (Figure 2C). In hospital-2 infants, colonization patterns changed between years with decreased Proteobacteria and increased Firmicutes in 2011 (Figure 2D). In hospital-2, there were also differences at lower phylogenetic levels, with infants in 2010 enriched in Staphylococcaceae and infants in 2011 enriched in Enterobacteriaceae. Week 2 samples were different between the two hospitals, with hospital-2 enriched in Proteobacteria and hospital-1 enriched in Firmicutes. In 2011, the week 2 infant colonization patterns were similar between the hospitals, consistent with the observation of week 1 colonization patterns. Since week 2 samples were similar at hospital-1 between years, but different at hospital-2 between years, this
suggests that the similarity of week 2 samples in 2011 is due to hospital-2 shifting to have colonization patterns more similar to those of hospital-1.

In week 3 samples, only a single sample was available from hospital-2 in 2011. Therefore comparisons were restricted to between hospitals in 2010 and between years in hospital-1. In 2010, hospital-1 was enriched in Bacilli, a class of phylum Firmicutes. Week 3 samples in hospital-1 were relatively stable between the two years (not shown), similar to the week 2 pattern.
Figure 2: Cladograms generated by LEfSe indicating differences in taxa between hospitals in samples from weeks 1 and 2. The central yellow dot in each cladogram represents kingdom; each successive circle is one step lower phylogenetically (phylum, class, order, family, and OTU.) Regions in red indicate taxa that were enriched in 2010 compared to 2011, while regions in green indicate taxa that were enriched in 2011 compared to 2010. Panel A: Week 1 sample from hospital-1. In 2010, samples were enriched for Firmicutes, with a shift in microbiota towards an increase in Proteobacteria in 2011. Panel B: Week 1 sample from hospital-2. In 2010, infants at hospital-2 were enriched in Proteobacteria while in 2011 they were enriched in Firmicutes. Panel C: Week 2 sample from hospital-1. Only OTU level differences were detectable between 2010 and 2011. Panel D: Week 2 sample from hospital-2. In 2010, samples were enriched for Proteobacteria. In 2011, the infant microbiota were enriched in Firmicutes, particularly, in Bacilli.
To examine the changes in infant colonization over time in greater detail, we plotted the median relative abundance of phyla *Firmicutes* and *Proteobacteria* by quarter in weeks 1 and 2 (Figure 3). Quarters with fewer than 3 study infants enrolled from each hospital were excluded from the graph. In week 1 samples, hospital-1 infants showed a sharp decline over time in median *Firmicutes* abundance and a corresponding increase in *Proteobacteria* abundance; hospital-2 infants showed fluctuating amounts of *Firmicutes* and a modest decline in *Proteobacteria* colonization (Figure 3A and B). We then examined the median relative abundance of Staphylococcaceae, the most abundant family in phylum *Firmicutes*, and the median relative abundance of Enterobacteriaceae, the most abundant family in phylum *Proteobacteria* (Figure 3C and D). These major bacterial families similarly demonstrated time trends, though differed from the patterns of their respective phyla in a few quarters.

Regarding time trends by quarter in week 2 samples, infants at hospital-1 again showed fluctuations in *Firmicutes* and *Proteobacteria* relative abundance, but different from week 1 sample, there was no clear distinction between study years. Similar to week 1 sample, infants at hospital-2 showed a peak in *Firmicutes* and a drop in *Proteobacteria* during the second quarter of 2011, likely explaining the difference in the abundance of these phyla between the two years in week 2 (Figure 4A and B). For most quarters, the relative abundance of Staphylococcaceae again corresponded with the relative abundance of *Firmicutes* and the relative abundance of Enterobacteriaceae again corresponded with the relative abundance of *Proteobacteria* (Figure 4C and D).
Figure 3: Week 1 sample: Line plots indicating shifts in the median relative abundance of selected taxa by quarter for hospital-1 and hospital-2 in week 1. Red lines indicate hospital-1 and blue lines indicate hospital-2. Quarters 1, 7, and 8 are excluded because fewer than three infants per hospital had samples during those quarters. Panel A: Median relative abundance of *Firmicutes*. In hospital-1, *Firmicutes* started high in 2010 and then fell in 2011. In hospital-2, *Firmicutes* started low in 2010 and then rose slightly in 2011. Panel B: Median relative abundance of *Proteobacteria*. At both sites, *Proteobacteria* followed a pattern opposite that of
Firmicutes. Panel C: Median relative abundance of *Staphylococcaceae*, the most commonly detected family of *Firmicutes*. Pattern in the relative abundance of *Staphylococcaceae* was similar to that of *Firmicutes* for most quarters. Panel D: Median relative abundance of *Enterobacteriaceae*, the most commonly detected family of *Proteobacteria*. Pattern in relative abundance of *Enterobacteriaceae* was similar to that of *Proteobacteria* for most quarters.

![Graph A](image1)

![Graph B](image2)
Figure 4: Week 2 sample: Line plots indicating shifts in the median relative abundance of selected taxa by quarter for hospital-1 and hospital-2. Red lines indicate hospital-1 and blue lines indicate hospital-2. Quarters 1, 7, and 8 are excluded because fewer than three infants per hospital had samples during those quarters. Panel A: Median relative abundance of *Firmicutes*. In hospital-1, *Firmicutes* again began high in 2010 and then fell in 2011 before rebounding slightly. In hospital-2, relative abundance of *Firmicutes* peaked in early 2011. Panel B: Median relative abundance of *Proteobacteria*. Relative abundance of *Proteobacteria*
appears to be inversely related to relative abundance of *Firmicutes*. Panel C: Median relative abundance of *Staphylococcaceae*, the most commonly detected family of *Firmicutes*. The abundance of *Staphylococcaceae* follows the same general trend as *Firmicutes*. Panel D: Median relative abundance of *Enterobacteriaceae*, the most commonly detected family of *Proteobacteria*. The relative abundance of *Enterobacteriaceae* follows the same general trend as *Proteobacteria*.

To determine the influence of differences in clinical factors on the differences observed over study years within each NICUs, we used GEE models of the relative abundance of *Proteobacteria* including samples from both weeks 1 and 2. *Proteobacteria* was examined as an outcome using two different models: first, as a continuous variable in a linear model and second, in logistic model with the relative abundance of *Proteobacteria* dichotomized into high and low relative abundance based on a natural cut-point in the data distribution. The low category was defined as 25% or lower relative abundance of *Proteobacteria* and the high category was defined as more than 25% relative abundance of *Proteobacteria*. Percentage of feeds that were formula was not included in the Cincinnati model because infants there received no formula prior to day of life 14. Included covariates were birth year, maternal antibiotic use, infant antibiotic use in the first 14 days of life, infant gestational age, maternal age at delivery, delivery mode, and in Birmingham percentage of feeds that were formula. In Cincinnati, whether or not a sample was stored in thioglycollate was included in the models. Backwards elimination was used to determine the final model. Any variable with a p-value of greater than 0.10 that did not change the coefficient of the year variable by more than 10% was eliminated starting with the covariate with the highest p-value. Both models at both hospital found that birth year remained significant, but in most models, most other covariates dropped out of most of the models (Table 2.)

**Table 2:** Results of GEE models of Proteobacteria relative abundance within each hospital NICU.
<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Hospital-1</th>
<th>Hospital-2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>P Value</td>
</tr>
<tr>
<td><strong>Linear models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year (2011)</td>
<td>0.313</td>
<td>0.00256</td>
</tr>
<tr>
<td>Maternal Antibiotic Use</td>
<td>NA</td>
<td>NS</td>
</tr>
<tr>
<td>Infant gestational age</td>
<td>NA</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Logistic models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year (2011)</td>
<td>1.63</td>
<td>0.033</td>
</tr>
<tr>
<td>Maternal Age at Delivery</td>
<td>NA</td>
<td>NS</td>
</tr>
<tr>
<td>Infant Gestational Age</td>
<td>NA</td>
<td>NS</td>
</tr>
</tbody>
</table>

We then analyzed microbial succession patterns within each hospital over the first 3 weeks of life. There were no significant differences in Jaccard Distance between repeated infant samples in infants born in 2010 compared to infants born in 2011 at hospital-1. Since there was only a single infant with sample from all three weeks in 2011, a similar comparison by time could not be conducted. Infants were therefore included in a single analysis comparing the hospitals regardless of birth year.

At both hospital-1 and hospital-2, most infants had samples in week 1 of life with *Proteobacteria* accounting for >50% of the reads (distribution of infants by quarter was not even, more infants were included in this study were born in quarters with higher relative abundance of *Proteobacteria* during week 1). At hospital-1, infant samples that were not dominated by *Proteobacteria* were instead dominated by either *Firmicutes* or *Fusobacteria*, or lacked any single phyla that accounted for more than half the reads. At hospital-2, however, infant samples that were not dominated by *Proteobacteria* were instead dominated by *Bacteroidetes*. The microbial dominance pattern identified in hospital-2 infants appeared to persist between weeks, but hospital-1 infant microbiomes appeared to shift between weeks (Figure 5). Comparing the Jaccard index calculated from OTU tables confirmed this difference; the Jaccard index between week 1 and week 2, week 1 and week 3, and week 2 and week 3 samples was significantly lower for hospital-2 infants (Kruskal-Wallis, p<0.01 for each of the three time points). A lower
value for the Jaccard index indicates a greater degree of similarity between samples, indicating that infants at hospital-2 experienced less change in microbial colonization from week to week than did infants at hospital-1. The greater rate of change observed between week 1 and week 2 samples in hospital-1 may be accounted for by the longer time interval between week 1 and week 2 samples in hospital-1. However, there was a shorter interval between the week 2 and week 3 samples in hospital-1 and the number of days between week 1 and week 3 samples was comparable at both sites. Since a greater rate of change was still observed in hospital-1 in a shorter period of time and in a period of time comparable to that of hospital-2, the conclusion that hospital-1 infants are experiencing a more rapid turnover remains.

**Figure 5:** Changes in relative abundance of phyla by week for infants with samples collected in weeks 1, 2, and 3. Each set of 3 bars represent samples from a single infant with samples from week 1, then week 2, and finally week 3 going from left to right. Samples are grouped by hospital and dominant phyla (>50% of reads) in week 1. The other category also contains an infant who had no single phyla accounting for >50% of the reads. Infants from Birmingham had significantly more similar stool samples from week to week compared to infants from Cincinnati.
as measured by Jaccard distance calculated at the OTU level and tested by Kruskal-Wallis 
(p<0.01.)

**Discussion**

Consistent with previous reports, the microbiota of the preterm infants in our study were 
most frequently dominated by the phylum *Proteobacteria*; *Firmicutes* was the next most frequent 
phylum. Nevertheless, our observations support the hypothesis that temporal and geographic 
factors influence the intestinal microbial colonization of preterm infants. These temporal and 
geographic differences were not explained by factors known to influence the intestinal 
microbiome, such as antibiotics. Differences in *Proteobacteria* colonization between 2010 and 
2011 remained even after modeling to adjust for the effect of antibiotic use on colonization 
patterns. Microbial shifts observed at the phylum level involved multiple taxa. Differences within 
NICUs over time of phylum *Firmicutes* largely included organisms of the class *Bacilli*, while 
differences within NICUs over time of phylum *Proteobacteria* largely involved organisms of the 
family *Enterobacteriaceae*.

We speculate that the observed patterns could be due to shifts in the microbiota of the 
NICU environment itself. Longer term studies are needed to determine the extent to which 
temporal shifts in colonization recur in regular patterns. Studies with concurrent sampling of 
both infants and the hospital environment are needed to determine the extent to which hospital 
resident microbes influence infant colonization patterns and the changes in infant colonization 
patterns over time.

Differences in hospital environments over time could be due to changes in 
environmental management, clinical practices, ongoing quality improvement initiatives, patient 
population, or other factors. Hospital-1 has a NEC task force which implemented changes to 
the standard of care in its NICU over the course of this study to reduce the NEC incidence rate. 
Changes included the introduction of a standard feeding protocol and use of donor milk when
mother’s own milk is unavailable, both of which could potentially influence the developing microbiome of the infant. In February 2010, the hospital-2 NICU moved to a new location. Part of the change in infant colonization during this time could have resulted from the change in location and its evolving microflora. While we report these events as qualitative background information, our study was not designed to analyze the impact of specific events quantitatively.

This report presents the microbial colonization patterns of relatively healthy infants and does not address disease outcomes per se. Nevertheless, the differences that we observed could have important clinical consequences. During the period of this study, the hospital NICUs varied in their rates of late onset sepsis and NEC, two conditions in which intestinal microbiota are implicated. The NICU of hospital-2 had double the rate of late onset sepsis observed at hospital-1, though sepsis rates remained constant between 2010 and 2011. Rates of NEC at hospital-1 were approximately double those of hospital-2 in 2010, but in 2011, the rates in the two hospitals were indistinguishable.

Nationally, 10% of preterm infants born <29 weeks gestational age develop necrotizing enterocolitis (NEC). This risk has been attributed in part to immature response to LPS-bearing organisms, specifically, excessive TLR4 signaling and hyper-inflammatory response to Proteobacteria. Investigators have noted that a surge or “bloom” in Proteobacteria or a decline in Firmicutes precedes NEC (14, 16, 17). The tendency towards surge in Proteobacteria and equivalent decline in Firmicutes may differ between NICUs. Indeed, in our study, in addition to observing differences between intestinal colonization between the hospitals by week of life, we also observed differences in microbial succession within infants over the first few weeks of postnatal life. Hospital-1 infants experienced a more rapid turnover of bacterial species than did hospital-2 infants. This could have important implications for disease risk.

**Conclusion**

Time of birth and hospital correlate with distinct premature infant intestinal colonization patterns. Identifying these differences in various institutions and over years may provide a
biomarker for monitoring disease risk. Our data provide additional evidence for caution regarding the degree to which studies of the microbiome and diseases of prematurity can be generalized to other hospitals or even to the same hospital later in time. Indeed, our data supports the need for large, multi-site, multi-year epidemiological studies in preterm infants and other patient populations to understand the association between intestinal microbial colonization and disease risk. The cohort study from which infants were selected for this study contained too few subjects with NEC and sepsis to effectively assess temporal changes in case infants. This study also clearly indicates a need for careful matching strategies on time and site to examine disease risk in studies of the intestinal microbiota.

Chapter 3 References

Chapter 4: The Intestinal Microbiota Preceding Late Onset Neonatal Sepsis compared to Controls: A Nested Case-Control Study

Rationale

Observed differences between microbial communities of preterm infants in Birmingham and Cincinnati resulted in stratifying on city to analyze the microbiota of sepsis cases and controls. This stratified analysis enabled the detection of site-specific microbial colonization patterns that correlate with the later development of sepsis. The following pages of this chapter describe the analysis of the microbiota in advance of sepsis, and represent the second manuscript from this work. The association between intestinal colonization and late-onset sepsis was further explored using Bayesian network analysis. That analysis was considered supplementary, and is provided in Appendix 1.

Background

Late-onset neonatal sepsis is a major cause of morbidity and mortality in preterm infants. Late-onset sepsis occurs in more than 20% of very low birth weight infants (1) and is associated with an increased risk of mortality (2). Furthermore, survivors of late-onset neonatal sepsis exhibit worse long-term outcomes, with increased risk of cerebral palsy, vision, and hearing impairment (3). Despite the seriousness of late-onset neonatal sepsis, diagnosis is a challenge as diagnostic and predictive biomarkers are needed (4).

There is evidence that sepsis is often caused by organisms that originate in the gut and translocate across the intestinal lining into the bloodstream. Supporting this concept, several studies demonstrated that the microbe identified in the bloodstream that ultimately caused neonatal sepsis (in particular, group B Streptococcus, Serratia marcescens, or Escherichia coli strains) was detectable in stool samples collected prior to the occurrence of sepsis (5, 6). Furthermore, a study of six infants found that those who later developed sepsis had an intestinal
microbiota distinct from that of healthy infants (7). Samples from infants who later developed sepsis tended to have more *Staphylococcus* and *Proteobacteria* than healthy infants who tended to exhibit greater microbial diversity and more anaerobic bacteria (7). Another larger study also found a decrease in diversity in infants who developed sepsis but found that cases had lower *Bifidobacteria* counts than did control infants (8). A third study found an increase in *Staphylococcaceae* associated with sepsis but did not report a difference in diversity between cases and controls (9).

The microflora of the gut is a complex community, and disruptions to that community can result in disease. The clearest example of this process is *Clostridium difficile* infection following antibiotic treatment: Antibiotic treatment may remove microbial community members that normally keep *C. difficile* in check, thereby allowing *C. difficile* to overgrow and cause disease. Restoration of the microbial community through fecal transplant is highly successful in treating refractory disease (10). Another example of microbial community disruption predisposing to disease is found in the different colonization patterns of preterm infants who develop necrotizing enterocolitis (NEC) compared to infants who remain healthy (11, 12). The ability of a healthy microbial community to prevent disease may have implications for late-onset sepsis; one possibility is that infants who have healthy, diverse microbial communities are less likely to develop sepsis even if intestinally colonized with potentially pathogenic organisms.

We therefore hypothesize that the composition of the intestinal microbiome influences the risk of late-onset sepsis in premature infants. Understanding the microbial community structures that place infants at risk may also lay the ground work for novel predictive biomarkers or probiotic treatments of sepsis.

**Methods**
Subjects: Study infants were selected from the infants <29 weeks gestational age who were enrolled in an ongoing cohort study of novel biomarkers for NEC. Infants were enrolled from three level III Neonatal Intensive Care Units (NICUs), including one in Birmingham, AL and two in Cincinnati, OH. Late onset sepsis was defined as infants who developed sepsis more than 72 hours after birth with at least one stool sample collected and successfully processed prior to sepsis onset. Controls born during the same time period were frequency matched on study site and gestational age. All controls survived free of NEC and sepsis. The Cincinnati Children’s Hospital and Medical Center IRB and the IRBs of the participating hospitals approved this study. Written consent was obtained from parents or guardians of study subjects.

Sample Collection: Serial stool samples were collected from infants on scheduled collection days of life 5, 8, 11, 14, and 21 plus or minus two days. For this study, the first successfully sequenced stool sample and the last sample prior to sepsis (control samples were frequency matched on collection day of life) from each infant were included. Samples were collected from soiled diapers, immediately refrigerated at 4°C in the NICU and transported to the laboratory where they remained in the refrigerator until processing with thioglycollate and storage at -80°C (most Cincinnati samples), or storage at -80°C without additional buffer (Birmingham, earliest Cincinnati samples, late Cincinnati samples, and day of life 11 Cincinnati samples.)

Stool Extraction and 16s rDNA Sequencing: Stool extractions were completed using the methods described in Morrow et al. 2013 (11).

Using extracted DNA, 180nt paired-end reads were generated using established primers and protocols and samples were allocated across multiple Illumina MiSeq runs (13). Read pairs were merged to create amplicon-spanning sequences and sequences then filtered to remove those with less than 70% identity to any read in the rRNA16S.gold.fasta reference set (http://drive5.com/uchime/uchime_download.html) using "usearch -usearch_global -id 0.70".
79,076,883 sequences were processed utilizing the UPARSE pipeline (14), software version usearch7.0.959_i86linux64. The following commands were used with default settings unless otherwise specified. Dereplication resulted in 35605130 sequences (-derep_fulllength); removal of singleton reads in 2206563 sequences(-sortbysize -minsize 2), and clustering yielded 7249 OTU representative sequences (-cluster_otus). The OTU table was constructed by mapping reads to OTUs (-usearch_global -strand plus -id 0.97) and applying the python script uc2otutab.py (http://drive5.com/python/). Additional chimera filtering was not applied. QIIME (15) version 1.6 was used to provide classifications of the OTU representative sequences using the gg_13_5 GreenGenes taxonomy and representatives sequences constructed at 99% ID. A phylogenetic tree was constructed using fasttree and filtered pynast alignments of the OTU representative sequences, also within the Qiime package. OTUs were removed from the OTU table using a filter with a minimum count fraction of 0.0002 in Qiime. This resulted in an OTU table containing 525 unique OTUs. The OTU table was then rarified to 2000 reads per sample for analysis.

**Statistical Analysis:** Due to differences in stool storage protocols associated with differences in colonization patterns, all analysis was stratified by the NICU city (Cincinnati vs Birmingham, see chapter 2). Two separate analyses were conducted, one using the first available stool sample form each infant and one using the last sample prior to sepsis. Non-metric multidimensional scaling ordination (NMDS) was conducted using the weighted and unweighted UniFrac results for the selected samples. The linear discriminant analysis effect size tool (LEfSe) was then used to screen for differences at all taxonomic levels between the case and the control samples (16). Classification trees were generated to provide additional information on differences in microbial composition between the two sites. To generate trees, the rpart package in R was used with default settings for the Cincinnati analysis. Due to the small number of subjects in Birmingham, the minimum number to split changed from 20 to 10 for that location. Logistic
regression using a backwards elimination approach conducted in R was used in combination
with the taxa differences identified by the classification tree to test for significance of the models
created by the classification trees. Alpha diversity was calculated using the Chao1 index and
the Simpson index with the Vegan package in R, the Kruskal-Wallis test was used to test for
differences in alpha diversity between cases and controls.

Results

Study Subjects: There were a total of 13 infants in Birmingham and 22 infants in Cincinnati who
developed late-onset sepsis with at least one successfully sequenced stool sample collected
prior to disease onset. Therefore, 13 Birmingham and 22 Cincinnati control infants were
selected for inclusion. In both Birmingham and Cincinnati, sepsis cases and controls were well
matched (Table 1). Of the 13 Birmingham cases, 11 had Gram-positive sepsis (median day of
life 12 for sepsis onset), 1 had Gram-negative sepsis (sepsis occurred on day of life 12), and 1
infant was blood culture positive for both a Gram-negative and a Gram-positive organism
(sepsis occurred on day of life 21). Of the 22 Cincinnati cases, 14 were Gram-positive (median
day of life 12) and 8 were Gram-negative (median day of life 33) (Figure 1). The sequenced
reads from the stool samples could be reliably classified only to the family level. Genetic data
on the causative strain of bacteria in sepsis cases was unavailable. Therefore, DNA from stool
could not be tested to confirm the presence of an identical strain in stool during the days prior to
sepsis onset. Because OTUs were generally classified to the family level, we checked instead
for a match between OTUs in stool samples and the causative organism of sepsis at the family
level. Reads belonging to the same family as the causative organism of sepsis were found in
the first stool samples for 8 of the 13 Birmingham sepsis cases; one infant who had sepsis
positive for multiple organisms matched on the family level for one causative organism but not
for the other. For the last stool sample prior to sepsis, 9 of 13 infants in Birmingham matched
causative organisms and reads at the family level, including the infant with multiple organisms
causing sepsis. In Cincinnati, reads belonging to the same family as the family of the causative organism of sepsis were found in both the first and last stool sample for 15 of 21 cases. One case infant had the causative organism recorded as Gram-negative rods and so could not be used to check concordance between the stool samples and causative organism of sepsis.

Table 1: Comparison of cases and controls in Birmingham and Cincinnati

<table>
<thead>
<tr>
<th></th>
<th>Birmingham</th>
<th></th>
<th>Cincinnati</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases n=13</td>
<td>Controls n=13</td>
<td>Cases n=22</td>
<td>Controls n=22</td>
</tr>
<tr>
<td>Multiple Birth</td>
<td>3 (23%)</td>
<td>2 (15%)</td>
<td>8 (36%)</td>
<td>7 (32%)</td>
</tr>
<tr>
<td>C-section</td>
<td>5 (38%)</td>
<td>6 (46%)</td>
<td>11 (50%)</td>
<td>12 (55%)</td>
</tr>
<tr>
<td>Gender, female</td>
<td>7 (54%)</td>
<td>7 (54%)</td>
<td>7 (32%)</td>
<td>12 (55%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1 (8%)</td>
<td>0 (0%)</td>
<td>2 (9%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Race, Black, other</td>
<td>7 (54%), 0 (0%)</td>
<td>7 (54%), 0 (0%)</td>
<td>8(36%), 1(5%)</td>
<td>9(41%), 1(5%)</td>
</tr>
<tr>
<td>Maternal antibiotics given</td>
<td>12 (92%)</td>
<td>12 (92%)</td>
<td>18 (82%)</td>
<td>14 (64%)</td>
</tr>
<tr>
<td>Infant birthweight, g</td>
<td>736</td>
<td>804</td>
<td>801</td>
<td>856</td>
</tr>
<tr>
<td>Infant gestation, weeks (median)</td>
<td>24</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Days on antibiotics in first 14 (median)</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>
Figure 1: Sepsis causative organisms in Birmingham and Cincinnati

Microbial Diversity: First Sample Analysis. In Birmingham, all samples were stored without thioglycollate. In Cincinnati, 4 control samples and 7 case samples were stored without thioglycollate (Fisher's Exact Test, p=0.49.)

For Birmingham, LEfSe detected a significantly higher relative abundance and greater frequency of colonization with *Clostridia* and *Clostridiales* in infants who later developed sepsis compared to control infants. Other differences between cases and controls occurred primarily at the OTU level and followed no clear patterns (Figure 2A). The classification tree model confirmed that *Clostridia* differed between cases and controls, but also found a signal from *Streptococcaceae* (Figure 3A). The misclassification rate of the tree was 15%. Logistic regression models did not find a significant association between the cut points identified in the classification tree and sepsis. The final model did not include any other covariates, all dropped out of the model as non-significant (not shown).
For Cincinnati, LEfSe detected a significantly higher relative abundance and greater frequency of colonization of phylum *Actinobacteria* in control infants compared to case infants. In addition, control infants were more likely to be colonized with order *Pseudomonadales* and family *Moraxellaceae*, but this difference appears to be primarily driven by differences in colonization with *Actinobacter guillouiae* (Figure 2B). The classification tree confirmed that phylum *Actinobacteria* differed between cases and controls, and detected a further signal in the infants whose stool samples lacked *Actinobacteria*. In this subset, infants whose stool samples where at least 65 of 2000 rarefied reads classified as *Bacilli* were more likely to be cases than were infants whose stool samples had fewer reads (Figure 3B). The misclassification rate of the tree in Cincinnati was 23%. Logistic regression was again used to determine the significance of the cut points at p<0.05 identified in the classification tree, both the *Actinobacteria* and the *Bacilli* cut point were significant, but the *Bacilli* cut point was significant only when the term accounting for the interaction of the *Actinobacteria* and the *Bacilli* cut points was included in the model. All other covariates dropped out of the model as non-significant (Table 2).
**Figure 2:** First sample analysis LEfSe results comparing sepsis cases and controls (A) Birmingham results, cases are enriched in *Clostridia*. (B) Cincinnati results, controls are enriched in *Actinobacteria*
**Figure 3:** First sample classification tree results. (A) Birmingham results, showing infants colonized with *Streptococcaceae* and infants with higher levels of *Clostridia* are at greater risk of sepsis. (B) Cincinnati results, showing that infants colonized with *Actinobacteria* and infants with lower levels of *Bacilli* are protected from sepsis.
Table 2: All potential confounders and effect modifiers tested in first sample analysis were non-significant and did not alter beta. Confounders and effect modifiers that remained in the model for the last sample analysis are listed with p-value. Modelling confirmed the statistical significance of classification tree bacteria cut points except for the Birmingham first sample tree (not shown) and the unknown Bacillales in the Birmingham last sample tree.

**Last Sample Analysis.** All Birmingham samples were stored without thioglycollate. In Cincinnati, 1 control sample and 5 case samples were stored without thioglycollate (Fisher’s Exact Test, p=0.19.)

Birmingham sepsis infants had significantly lower alpha diversity than control infants using the Simpson index (p<0.01) but not using the Chao1 index. This suggests that the total number of species present in cases and controls in the last sample analysis is similar, but that the distribution of those species is more even in controls. Cincinnati infants had no significant differences in alpha diversity for either metric. There was no visibly clustering in weighted UniFrac ordinations based on either Birmingham or Cincinnati samples. However, the unweighted UniFrac analysis for Birmingham did show some separation between sepsis cases and controls (Figure 4A). There was no visible separation between sepsis cases and controls in the Cincinnati unweighted UniFrac analysis (Figure 4B).
Figure 4: Last sample unweighted UniFrac NMDS results for (A) Birmingham which shows some separation between controls in black and cases in red and (B) Cincinnati which shows no separation between cases and controls.

For Birmingham, LEfSe detected only OTU level differences with no clear pattern of differences between cases and controls (Figure 5A). A classification tree found that the presence or absence of Lactobacillales, the presence or absence of an OTU of Bacillales of unknown family, and infant gestational ages were the best predictors of sepsis status (Figure 6A). The misclassification rate was 12%. Logistic regression confirmed the significance of the presence of Lactobacillales as protective against sepsis, but not the unknown Bacillales. Infant
gestational age was borderline significant. No other covariates, including the interaction term between *Lactobacillales* and the unknown *Bacillales* remained in the model (Table 2).
Figure 5: Last sample LEfSe cladogram results for (A) Birmingham where only OTU level differences were observed and (B) Cincinnati where case infants were enriched in

*Proteobacteria, Gammaproteobacteria, Enterobacteriales, and Enterobacteriaceae*
Figure 6: Last sample classification tree results for (A) Birmingham showing infants without \textit{Lactobacillales}, with the unknown OTU of \textit{Bacillales}, or born at younger gestational ages were more at risk of sepsis and (B) Cincinnati showing that infants with high levels of \textit{Enterobacteriales} or high levels of \textit{Firmicutes} were at increased risk of sepsis.

For Cincinnati, LEfSe detected a greater relative abundance of \textit{Proteobacteria}, \textit{Gammaproteobacteria}, \textit{Enterobacteriales}, and \textit{Enterobacteriaceae} in sepsis cases compared to controls (Figure 5B). A classification tree agreed with the \textit{Enterobacteriales} finding, splitting first on \textit{Enterobacteriales} and then on \textit{Firmicutes}, with infants with either abundant \textit{Enterobacteriales} or abundant \textit{Firmicutes} at increased risk of sepsis while those infants without high levels of \textit{Enterobacteriales} or high levels of phylum \textit{Firmicutes} were more likely to be controls (Figure 6B). The misclassification rate was 25%. Logistic regression confirmed that the cut points
created by the classification tree were significant predictors of sepsis. This time the model included the covariates preeclampsia and gravida (Table 2).

**Discussion**

This is the largest study of intestinal microbiota and risk of late-onset sepsis conducted using high-throughput sequencing technology to date. The colonization patterns that correlated with infant risk of sepsis differed both by location of the NICU and by timing of sample collection. This difference in results is consistent with the literature. Several previous studies have reported a microbial community signal prior to sepsis onset but the association does not appear to be consistent across studies, with studies sometimes but not always finding an association with increased *Staphylococcaceae* and sepsis and sometimes finding a reduced alpha diversity in sepsis cases (7-9). The differences in background colonization patterns (See Part 1) necessitate stratification by location to successfully detect community patterns that are associated with sepsis risk. These results also suggest that sepsis is a heterogeneous condition with multiple pathways leading to the same disease. Both locations in this study had risk microbial communities that differed from the findings of a small prior study based at Dartmouth. Like Birmingham, the Dartmouth study found reduced alpha diversity prior to sepsis. Unlike Birmingham, the Dartmouth study found that Clostridium occurred more frequently in the control infants (7). The results in Cincinnati were more similar to the findings of a University of Florida study, with Cincinnati cases lower in phylum *Actinobacteria* and University of Florida cases lower in *Bifidobacteria* (a genus belonging to phylum *Actinobacteria*) (8). The Birmingham infants were similar to both the Dartmouth and the University of Florida infants because cases had lower alpha diversity than did controls (7, 8).

When only the first sample from each infant is considered, the Cincinnati but not the Birmingham infants show a protective association with the presence of bacteria of the phylum
Actinobacteria. As phylum Actinobacteria contains the probiotic family Bifidobacteriaceae, we choose to examine the OTUs of Actinobacteria present in Birmingham and Cincinnati in the first sample from each infant more closely. Although no significant differences were found, Birmingham infants were less likely to be colonized with Actinobacteria in the first stool sample. Only 8 of the 26 infants had any detectable (at least one read) Actinobacteria in the first stool sample, compared with 20 of 44 infants in Cincinnati. Furthermore, the number of different OTUs of Bifidobacteriaceae detected differed between the two locations, with only a single OTU of Bifidobacteriaceae present in Birmingham compared with 6 OTUs of Bifidobacteriaceae present in Cincinnati.

When considering the last sample analysis, Lactobacillales was protective in Birmingham, but not in Cincinnati. Comparisons of Lactobacillales between the two sites found that in Birmingham, 18 (69%) of 26 infants had detectable Lactobacillales in their stool, compared to 35 (80%) of 44 infants in Cincinnati. There were 18 OTUs of Lactobacillales in the stool of Birmingham infants compared to 32 different OTUs in Cincinnati infants. Intriguingly, the single OTU present in Birmingham and not in Cincinnati was a member of family Enterococcaceae, and was detected in 5 of 13 control infants but only 1 of 13 case infants. This suggests a need for technologies such as Whole Genome Sequencing that are capable of classifying organisms at phylogenetic levels lower than family.

The heterogeneity of intestinal microbial communities associated with risk of late-onset neonatal sepsis suggests that functional studies are needed to better understand the role of the microbiome in sepsis. In adults, the functional composition of the intestinal microbiome is similar between individuals despite high phylogenetic diversity (17). Probiotics generally do not appear to have a protective affect against sepsis (18), this may be because the probiotic strains do not contribute to decreasing gut permeability. In support of this, a study in adults found that administration of a probiotic mixture reduced the presence of potential pathogens, but did not
alter markers of gut permeability (19). The microbiome does play a role in intestinal permeability; mouse studies have demonstrated that microbiota are important to the barrier function of gastrointestinal epithelium and that alterations to intestinal microbiota influence permeability (20). Future functional studies could help to identify critical components influencing barrier function that are not dependent on taxonomy.

Chapter 4 References

4. P. C. Ng, H. S. Lam, Biomarkers for late-onset neonatal sepsis: cytokines and beyond. *Clinics in perinatology* 37, 599 (Sep, 2010).


Chapter 5: Infant Genotype and Intestinal Microbiota Colonization

Rationale

In addition to the contribution of intestinal colonization to risk of late-onset neonatal sepsis, genetics may also contribute to disease risk (1). Intriguingly, one gene that has been reported to be associated with sepsis is Fucosyltransferase 2 (FUT2) (2), a gene that is known to alter the intestinal microbiome in adults. Background work on how human genetics impacts intestinal colonization in the preterm infant is entirely lacking. Therefore, Chapter 5 explores the association between fucosyltransferase and mucin genes and bacterial intestinal colonization in preterm infants.

Background

The composition of the human intestinal microbiome is associated with a wide array of diseases including obesity, inflammatory bowel disease (IBD), type 1 diabetes, necrotizing enterocolitis (NEC), and allergy (3-7). The community composition of the intestinal microbiome depends in part on exposure history, including factors such as delivery mode, exposure to antibiotics, or diet (8-10) but is also shaped by host genetics (4, 11). The human genes and the manner in which they alter the microbiome is however largely unexplored.

Early provocative studies of individual genes have suggested gene families that are in need of further study. One study that has looked at intestinal microbiota colonization and the risk of IBD examined the single nucleotide polymorphism (SNP) rs601338 in Fucosyltransferase 2 (FUT2) and found that homozygotes for the minor allele (non-secretors) had colonization patterns when healthy that came closer to those of IBD patients than did individuals with at least one copy of the major allele (secretors) (4). FUT2 catalyzes the expression of α(1,2) linked fucose (“secretor oligosaccharide”) in mucosal tissue (12). These oligosaccharides are both an important food source for commensal bacteria (13) and a binding site for pathogens (14).
SNP rs601338 is a nonsense mutation that causes the loss of expression of secretor oligosaccharide in the mucosal tissue of non-secretors (12) making this SNP a logical candidate for study in relation to the microbiome. In addition to FUT2, other fucosyltransferases also contribute to the expression of different fucoses in mucosal tissue (15). This family of genes is therefore a logical choice for study in relation to the intestinal microbiome.

Another family of genes expressed by intestinal epithelium are the mucin (MUC) genes. These genes help to form the barrier separating bacteria from the intestine, and are likely to also play a part in shaping the human intestinal microbiome. For example, probiotic strains of Lactobacillus can adhere to human mucus, displacing and reducing the viability of a potential pathogen, Staphylococcus aureus (16). Further evidence of the influence of MUC genes on the microbiome is found in mouse models, for example muc2 knockout mice are protected from alcoholic fatty liver disease and resistant to alcohol associated microbiome changes (17). Studying both fucosyltransferase genes and mucin genes was a logical choice, as the oligosaccharide made by the fucosyltransferase gene enzymes is involved in the post-translational modification of the mucin gene products (18). Meaning, that mucins are typically covered with oligosaccharide, and thus, polymorphisms in either gene family would tend to mutually influence oligosaccharide and mucin presentation in the gut.

Methods

Subjects: Study infants were selected from the large prospective Novel Biomarkers Cohort study (IRB#2008-0463). For inclusion in this study, infants had to be enrolled at one of the Cincinnati Neonatal Intensive Care Units (NICUs), have genotype results from the Illumina Immunochip, have had at least one stool sample successfully genotyped for the 16s bacterial rRNA gene by Illumina MiSeq, and belong to one of two homogeneous ancestral groups (called black and white based on the self-identified race of the majority of individuals within each group)
generated from ancestral informative markers on the Immunochip. In cases where twins or triplets all had sample in the same week, one infant was randomly selected for inclusion using random.org (19).

**Genotyping and Ancestry Classification:** All infants and their mothers with sufficient genomic DNA and at least one stool sample collected enrolled in the Novel Biomarkers Cohort were genotyped using the Illumina Immunochip (20). An additional cohort of individuals enrolled in the Novel Vaccines Surveillance Network (NVSN) was genotyped on the Immunochip concurrently. Genotyping was considered successful if a call rate >95% for all SNPs (21) and if gender as listed in the medical record matched the gender called from the Immunochip results. Because of the comparatively small number of subjects in the Novel Biomarkers Cohort, ancestry analysis was conducted combining both the Novel Biomarkers Cohort and the NVSN cohort. The use of the combined cohorts allowed for the successful identification of individuals with Mesoamerican ancestry (e.g. many Hispanics), a group which occurred only rarely in the Novel Biomarkers Cohort. To identify ancestry, Principal Coordinates Analysis (PCA) in Eigenstrat was used based on 1652 ancestral informative markers on the Immunochip (22). The settings for PCA were 3 dimensions, 5 iterations, and outlier removal at 6 standard deviations (22). The same set of SNPs was then used with STRUCTURE (23) with the outlier subjects identified by PCA removed from the data set. STRUCTURE settings specified three ancestral groups with admixture permitted and used a burn in of 5000 and 5000 repetitions. Individuals were considered to belong to an ancestral group if there posterior probability of belonging to that group was ≥80%.

**SNP selection:** Due to the small number of subjects in the Novel Biomarkers Cohort and the complexity of intestinal microbiota examining all SNPs on the Immunochip in relation to the intestinal microbiome was impossible. Therefore a small number of candidate SNPs were selected for inclusion. First a small number of candidate genes were chosen: *FUT2, FUT5,
For FUT6, FUT8, FUT9, MUC1, MUC4, MUC13, MUC15, MUC16, MUC17, MUC19, and MUC21. For FUT2, the SNP rs601338 (428G>A) was selected for inclusion because it is a nonsense mutation resulting in the non-secretor phenotype, which is already known to influence microbial colonization (4, 11). For the remaining genes, any SNP with a call rate of >95% (24) and a minor allele frequency (MAF) of >15% was considered for inclusion. The 13 SNPs which met the minimum call rate and MAF were checked for Hardy-Weinberg equilibrium (HWE) in both the black and the white infant groups using all infants who survived until discharge free of sepsis and necrotizing enterocolitis. A Bonferroni corrected p value of p<0.0038 was used in testing for HWE. All SNPs were in HWE. Table 1 lists the SNPs included in analysis. The genetics package in R was then used to generate a correlation matrix for the SNPs in the white population. The black population had two SNPs that lacked minor allele homozygotes, and so the genetics package could not be used to produce a complete correlation matrix for this population. The correlation matrix from the white population was then used in combination with the web-based program matSpD to calculate the effective number of independent variables (25). The effective number of variables was 10 (26). Since a correlation matrix was unavailable for the black population, 10 was used as an approximation for the effective number of variables in this population as well.
Table 1: SNP name, gene, alleles, and mutation type for SNPs included in genotype to intestinal microbial phenotype analysis

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Major Allele on Immunochip</th>
<th>Minor Allele on Immunochip</th>
<th>Mutation Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12411216</td>
<td>MUC1</td>
<td>C</td>
<td>A</td>
<td>upstream variant 2KB</td>
</tr>
<tr>
<td>rs12435908</td>
<td>FUT8</td>
<td>C</td>
<td>A</td>
<td>intron variant</td>
</tr>
<tr>
<td>rs1862461</td>
<td>MUC16</td>
<td>A</td>
<td>C</td>
<td>synonymous codon</td>
</tr>
<tr>
<td>rs2431795</td>
<td>MUC16</td>
<td>G</td>
<td>A</td>
<td>intron variant</td>
</tr>
<tr>
<td>rs2517425</td>
<td>MUC21</td>
<td>A</td>
<td>C</td>
<td>upstream variant 2KB</td>
</tr>
<tr>
<td>rs2591591</td>
<td>MUC16</td>
<td>C</td>
<td>A</td>
<td>Missense</td>
</tr>
<tr>
<td>rs2591595</td>
<td>MUC16</td>
<td>G</td>
<td>A</td>
<td>synonymous codon</td>
</tr>
<tr>
<td>rs4679166</td>
<td>MUC13</td>
<td>G</td>
<td>A</td>
<td>intron variant</td>
</tr>
<tr>
<td>rs4902408</td>
<td>FUT8</td>
<td>A</td>
<td>T</td>
<td>intron variant</td>
</tr>
<tr>
<td>rs601338</td>
<td>FUT2</td>
<td>G</td>
<td>A</td>
<td>stop gained</td>
</tr>
<tr>
<td>rs6571086</td>
<td>FUT9</td>
<td>G</td>
<td>A</td>
<td>intron variant</td>
</tr>
<tr>
<td>rs886403</td>
<td>MUC21</td>
<td>A</td>
<td>G</td>
<td>UTR variant 3 prime</td>
</tr>
<tr>
<td>rs3096337</td>
<td>MUC4</td>
<td>A</td>
<td>G</td>
<td>intron variant</td>
</tr>
</tbody>
</table>

Statistical Analysis: Stool extraction and 16s rRNA gene DNA sequencing were conducted as described in Part 2. Only samples from week 3 (days of life 17 to 23) were used in analysis. Only the first sample per subject in week 3 was included for analysis. The week 3 samples were selected because fucosyltransferases and mucins are often upregulated on exposure to bacteria (27, 28). Since the week 3 samples are the latest samples available in this data set this gives the maximum possible time for expression of these genes to be upregulated and for the bacteria to respond to that upregulation. Differences in alpha diversity by genotype were tested for using the Simpson Diversity Index and the Chao1 Index and analysis of variance (ANOVA). LEfSe (29) was used to screen for differences in bacteria taxa based on genotype in each of the two population groups by week, using a Bonferroni corrected p-value based on 10 independent variables of 0.005. All three genotypes and each pair of genotypes were compared. Differences at more than the operational taxonomic unit (OTU) level were examined more closely using boxplots.
Results

A total of 97 individuals belonging to the white population had week 3 sample available for analysis. A total of 40 individuals belonging to the black population had week 3 sample available for analysis. There were no significant differences in microbial alpha diversity by host genotype for any of the SNPs.

Differences in colonization patterns at greater than the OTU level associated with genotype were detected by LEfSe in the genes \textit{MUC1, MUC4}, and \textit{MUC13} in the black population (Figure 1). Differences in colonization patterns at greater than the OTU level associated with genotype were detected by LEfSe in the genes \textit{FUT9, MUC4, MUC13}, and \textit{MUC21} (Figure 2). A summary of the SNPs with significant differences above the OTU level and the comparisons that were significant is provided in Table 2.

\textbf{Table 2:} Summary of bacterial taxa that differ by genotype at p<0.0056 including SNP, population, and week.

<table>
<thead>
<tr>
<th>Bacteria taxa</th>
<th>SNP</th>
<th>population</th>
<th>Significant Comparison</th>
<th>Enriched Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Clostridium subterminale}</td>
<td>rs12411216</td>
<td>Black</td>
<td>AA v CC</td>
<td>AA</td>
</tr>
<tr>
<td>\textit{Bacteroides ovatus}</td>
<td>rs12411216</td>
<td>Black</td>
<td>AA v CC</td>
<td>AA</td>
</tr>
<tr>
<td>\textit{Actinobacter rhizosphaerae}</td>
<td>rs3096337</td>
<td>Black</td>
<td>AA v AG v GG</td>
<td>GG</td>
</tr>
<tr>
<td>\textit{Lactobacillales}</td>
<td>rs4679166</td>
<td>Black</td>
<td>GA v GG</td>
<td>GG</td>
</tr>
<tr>
<td>\textit{Actinomycetaceae}</td>
<td>rs6571086</td>
<td>White</td>
<td>AA v GA</td>
<td>AA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AA v GG</td>
<td>AA</td>
</tr>
<tr>
<td>\textit{Bifidobacteria breve}</td>
<td>rs6571086</td>
<td>White</td>
<td>AA v GA</td>
<td>AA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AA v GG</td>
<td>AA</td>
</tr>
<tr>
<td>\textit{Erwinia}</td>
<td>rs2517425</td>
<td>White</td>
<td>AA v CC</td>
<td>CC</td>
</tr>
<tr>
<td>\textit{Bacteroides fragilis}</td>
<td>rs4679166</td>
<td>White</td>
<td>AA v GA</td>
<td>AA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AA v GG</td>
<td>AA</td>
</tr>
<tr>
<td>\textit{Bacteroidaceae}</td>
<td>rs4679166</td>
<td>White</td>
<td>AA v GG</td>
<td>GG</td>
</tr>
<tr>
<td>\textit{Bacteroides}</td>
<td>rs4679166</td>
<td>White</td>
<td>AA v GG</td>
<td>GG</td>
</tr>
<tr>
<td>\textit{Erwinia}</td>
<td>rs886403</td>
<td>White</td>
<td>AA v GG</td>
<td>GG</td>
</tr>
<tr>
<td>\textit{Dermabacter}</td>
<td>rs3096337</td>
<td>White</td>
<td>AG v GG</td>
<td>GG</td>
</tr>
<tr>
<td>\textit{Dermabacteraceae}</td>
<td>rs3096337</td>
<td>White</td>
<td>AG v GG</td>
<td>GG</td>
</tr>
<tr>
<td>\textit{Actinobacteria}</td>
<td>rs3096337</td>
<td>White</td>
<td>AA v AG</td>
<td>AA</td>
</tr>
</tbody>
</table>
Figure 1: Black population week 3 LEfSe results with more than just OTU level differences. (A) SNP rs12411216 genotype AA versus genotype CC. (B) SNP rs4679166 genotype GA versus genotype GG. (C) SNP rs3096337 genotype AA versus genotype AG versus genotype GG.
Figure 2: White population week 3 LEfSe results with more than just OTU level differences. (A) SNP rs6571086 genotype AA versus genotype GA (top) and genotype AA versus genotype GG (bottom). (B) SNP rs2517425 genotype AA versus genotype CC. (C) SNP rs4679166 genotype AA versus genotype GA (top) and genotype AA versus genotype GG (bottom). (D) SNP rs886403 genotype AA versus genotype GG. (E) SNP rs3096337 genotype AG versus genotype GG (top) and genotype AA versus genotype GG (bottom).

Because LEfSe results may be swayed by just one or two outliers, boxplots were examined to determine if the differences detected by LEfSe are plausible or likely to have resulted from an outlier. An example boxplot of a difference caused by an outlier is shown in Figure 3. Boxplots of differences not caused by a single outlier are shown in Figure 4 and Table 3 summarizes these differences. In the black population, differences in colonization by genotype remained for the genes MUC13. In the white population, differences in colonization by genotype remained for the gene FUT9 and MUC4.
**Figure 3:** Boxplot of the number of reads of Erwinia by SNP rs886403 genotype, showing that a single individual of genotype GG with a large number of reads of Erwinia is responsible for the differences detected by LEfSe. Most individuals either totally lacked Erwinia or had only a single read.
**Figure 4:** Boxplots showing bacteria taxa differences not driven by outliers by genotype.
Table 3: Summary of SNPs with differences above the OTU level that on examination of boxplots don’t appear to be driven by outliers.

<table>
<thead>
<tr>
<th>Bacteria taxa</th>
<th>SNP</th>
<th>population</th>
<th>Significant Comparison</th>
<th>Enriched Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactobacillales</td>
<td>rs4679166</td>
<td>Black</td>
<td>GA v GG</td>
<td>GG</td>
</tr>
<tr>
<td>Bifidobacteria breve</td>
<td>rs6571086</td>
<td>White</td>
<td>AA v GA, AA v GG</td>
<td>AA, AA</td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>rs3096337</td>
<td>White</td>
<td>AA v AG</td>
<td>AA</td>
</tr>
</tbody>
</table>

Discussion

Our study demonstrated differences in the intestinal microbiota by genotype for the genes FUT9, MUC4, and MUC13. This strongly suggests that further study of these genes in larger cohorts is worthwhile. The New Vaccine Surveillance Network (NVSN) has collected stool samples and has Immunochip data on participants, including several hundred controls. The healthy controls from this nationwide cohort may provide an excellent population for follow up studies on FUT9, MUC4, and MUC13 as well as continued study of FUT2 in relation to the intestinal microbiome. Future studies should also consider sequencing these genes in participants to examine additional SNPs not included on the Immunochip.

Additional information in the literature provides some evidence suggesting that there is a relationship between FUT9 and the intestinal microbiome. FUT9 is expressed in the human stomach and can synthesize the Lewis X epitopes (30), suggesting that expression relevant to the gastrointestinal microbiome occurs. For example, it is known that the Lewis X epitope can act as a binding site for the bacterium Pseudomonas aeruginosa in the lung (31). And, fut9 knockout mice show a refractory response to lipopolysaccharide challenge compared to wild type mice (32). Of particular interest to the preterm population is a study that found a SNP in FUT9 associated with risk of placental malaria infection (33). The placenta has its own microbiome (34), and that microbiome potentially includes Bifidobacterium (18). This suggests
that the observed differences in intestinal colonization of the preterm infant by FUT9 genotype could potentially have begun in utero.

There is some evidence of the importance of MUC4 to the developing intestinal microbiome. In mice born to mothers colonized with *Lactobacillus acidophilus* there is increased transcription of muc4 on postnatal day 1 of life, but on postnatal day 6 transcription levels have lowered in pups with detectable levels of *L. acidophilus* but remain high in pups who do not yet have detectable levels of *Lactobacillus* (35). But, differences in muc4 transcription levels were not observed in mice colonized with other types of bacteria (35). In pigs, muc4 genotype is associated with vulnerability to enterotoxigenic *Escherichia coli* vulnerability (36).

There is also some evidence in the literature of the association between MUC13 and the intestinal microbiome. *MUC13* is highly expressed in the large intestine and moderately expressed in the small intestine (37). In pigs, *MUC13* plays an important role in susceptibility to enterotoxigenic *Escherichia coli* infection (38), a fact supporting *MUC13*’s importance to the intestinal microbiome. Research with Muc13 knockout mice has shown that knockout mice are more susceptible to DSS colitis than are wild type mice, indicating that Muc13 plays a role in inhibiting epithelial cell apoptosis (39). This study did not find an increase in direct bacterial contact with the epithelium in Muc13 +/- mice (39). However, the study did not look for differences in microbial composition related to Muc13 (39). The authors also suggest that Muc13’s role in inhibiting apoptosis has important implications for intestinal infections (39).

One study limitation is that it was likely underpowered to detect differences by genotype. Initial power estimates assumed the ability to combine infants from Cincinnati and Birmingham to examine the effect of host genetics, but differences in the colonization patterns by site precluded this possibility (see part 1). The lack of difference based on the FUT2 SNP rs601338 genotype was surprising, given its association with the intestinal microbiome in adults (4, 11). It
may be possible to combine the black and white populations to examine the rs601338 SNP because this nonsense mutation occurs in the same frequency in African and European origin populations. This SNP is a major nonsense mutation that results in lack of expression of H-antigens, carbohydrate containing an alpha(1,2)-fucose (2FL) linkage (12). Nevertheless, other complications exist in looking at this gene in our study population, namely that the milk of FUT2 positive (secretor) mothers, as well as donor human milk, contains abundant quantities of 2FL, an H-antigen analog (40). As all infants in this study were fed either mother’s own milk or donor milk for the first 14 days of life, it is likely that 2'FL and other “secretor oligosaccharide” was received by most study infants. About half of non-secretor (AA) infants have GA mothers, whose milk contains 2'FL and other secretor oligosaccharide. Further, donor milk is pooled from multiple mothers. Since approximately 75% of mothers of European descent and African descent are secretors (41), most donors are therefore expected to have 2FL present in their breast milk (40). Donor milk is pooled from 3 individuals enhancing the likelihood of some secretor oligosaccharide being conveyed in donor milk. Thus, it is likely that most AA infants have 2FL in their intestinal tract despite lacking the genetic capability to express the sugar themselves.

In summary, infant fucosyltransferase and mucin genotype do appear to influence infant intestinal microbiota colonization patterns. Additional, larger studies are needed to further characterize these differences and to determine the causal variants associated with these differences and to deal with potential confounding factors such as infant feeding type.

Chapter 5 References


7. H. Bisgaard et al., 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, 646 (Sep, 2011).

8. M. G. Dominguez-Bello et al., Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proceedings of the National Academy of Sciences of the United States of America 107, 11971 (Jun 29, 2010).


16. S. Vesterlund, M. Karp, S. Salminen, A. C. Ouwehand, Staphylococcus aureus adheres to human intestinal mucus but can be displaced by certain lactic acid bacteria. Microbiology 152, 1819 (Jun, 2006).


30. M. Kaneko *et al.*, Alpha1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX. *FEBS letters* **452**, 237 (Jun 11, 1999).


Chapter 6: Conclusions and Future Research

Overview

My dissertation research was focused on understanding the contribution of the intestinal microbiota to health and disease in the preterm infant. Late-onset neonatal sepsis is a major cause of morbidity and mortality in the preterm infant (1-3), and previous studies have found that sepsis-causing organisms often originate in the gut. As such, understanding how the intestinal microbiota contribute to sepsis risk either on the causal pathway to disease or as a biomarker of the disease process has important implications for clinical care of the preterm infant.

Learning from each research chapter

Methodologic lessons learned (chapter 2). The cohort study that supported secondary analysis for my dissertation began in 2009. As the cohort study was launched, decisions were made by the project team regarding storage buffer protocol and DNA extraction protocol. As these were early days of microbiome research methods, there were few methodologic publications and limited understanding that the impact of these choices would have on the measurement and analysis of the microbiome. We then undertook methodologic studies, and similar to others (4-7), found that storage buffer and extraction protocol influence the microbiota efficiency of detection of certain taxa. We also found, similar to others (4), that the time a sample is kept refrigerated at 4°C prior to freezing does not have a significant impact on the microbiota that are detected. This is important information for future studies: Storage
buffer and extraction protocol should be standardized but it is not necessary to assure sample processing within the same day.

Furthermore, for the current study, these findings have meant that it has been necessary to incorporate the storage and extraction protocols into our analysis plan. This is a study limitation, but it has also been possible to address the issue adequately by identifying the taxa influenced and excluding analysis of these taxa, through sample matching, or through our approach to analysis. For example, when comparing the two hospitals, which used different storage buffer and extraction protocols, we first restricted to a single extraction protocol. We then identified taxa that differed within a single hospital based on storage buffer. As a result, we knew that differences between the two hospitals in abundance of *Bacteroidetes*, *Leuconostocaceae*, and *Propriornibacterineae* were potential confounded by storage protocol. Therefore, differences in these taxa between sites were not considered differences between hospitals but rather an artifact of storage protocol. Since our comparison of sepsis cases and controls was stratified on hospital, neither storage protocol nor extraction protocol had a significant influence. For our exploration of genetic factors in relation to the microbiome, all analysis was limited to week 3 Cincinnati samples. This means that all samples were stored with thioglycollate. The limited number of subjects eligible for inclusion made restricting to a single extraction protocol impractical because of the loss of power that would occur. Fortunately, there were no significant differences in extraction protocol by genotype for either ancestral population (Fisher’s Exact Test, \( p < 0.05 \)) which minimizes the chance for confounding due to extraction protocol.
The impact of place and time on the microbiome (chapter 3). In this study, I showed that the intestinal microbiome of the premature infant is influenced by temporal trends (time) and by environmental factors (place), specifically, the location Neonatal Intensive Care Unit (NICU). The findings suggest that studies of the intestinal microbiota and disease risk identified in a single population and time have limited generalizability to other populations and times. Large, multi-site, multi-year studies are needed to determine the degree of variability in intestinal microbiota of comparatively healthy premature infants to provide background for studies of the intestinal microbiome and disease risk. The intestinal microbiota of premature infants is associated with risk of late-onset sepsis, but this association is site and time dependent.

The microbiota colonizing infants differs both between hospitals and over time within a single hospital. The rate at which the intestinal microbial communities of premature infants change also appears to vary between hospitals. Since both community composition and changes in community composition may be associated with risk of diseases such as sepsis and necrotizing enterocolitis (NEC), understanding colonization patterns in premature infants that remain healthy is crucial background to understanding the intestinal microbiome as a biomarker of disease.

Future studies that include more than two NICUs from more than two cities are critical to understand the degree of variability between NICUs. These studies should also be longitudinal in nature so that more can be learned about the frequency with which background colonization changes within NICUs.
The intestinal microbiome and sepsis (chapter 4). The composition of intestinal microbiota does appear to be associated with the risk of late-onset sepsis in premature infants, but in a site dependent manner. Multi-site studies are needed to determine if there are a few patterns of intestinal colonization associated with sepsis and individual sites will cluster into groups or if the intestinal microbiota associated with sepsis is too variable to group sites by disease risk. The functional component of the intestinal microbiome, not measured in this study, might be similar even in individuals where the species composition of the microbiota is highly variable (8). Therefore, future studies should use tools such as whole genome sequencing and RNA sequencing to consider whether the functional aspect of the intestinal microbiome is more consistently associated with the risk of late-onset neonatal sepsis at different sites.

Microbiota and host genetics (chapter 5). Among the premature infants in our study, mucin 4 (MUC4), mucin 13 (MUC13), and fucosyltransferase 9 (FUT9) genotype all appear to be associated with differences in premature infant intestinal microbiome composition. Surprisingly, there was no association with fucosyltransferase 2 (FUT2) genotype. Prior studies in adults have found that FUT2 genotype is associated with intestinal microbiota colonization (9). The lack of association with FUT2 in this population may be because of the small sample size or because of confounding by the presence of α(1,2) linked fucose in human milk (10). The small number of infants included in this study and the complexity of confounding factors present in premature infants also limit the conclusions that can be drawn from this study for all genes. Nevertheless, this is the first study, to our knowledge of these gene polymorphisms, and these preliminary findings are potentially important. Future studies of host mucin and fucosyltransferase
genotype and intestinal microbiota phenotype should be undertaken using larger populations with less confounding, such as would be anticipated in studies of adults, older children or term infants after weaning. Future studies should also consider using whole genome (shotgun) sequencing rather than 16s analysis. As previously noted, 16s analysis is highly reliable for classification of taxa to the level of bacterial family, but this resolution may not be adequate for analyses of the microbiome in relation to host genotype, which is likely to depend on the specific species or strain of bacteria.

Overall summary and conclusions

This study found that the intestinal microbiota of preterm infants is influenced by both time and place of birth and by infant genetic factors. Furthermore, preterm infant intestinal microbiota is associated with risk of late-onset sepsis but in a site dependent manner. This suggests that studies of intestinal microbiota and disease risk have limited generalizability to other populations and times. Large, multi-site, multi-year studies are needed to determine the degree of variability in intestinal microbiota of comparatively healthy premature infants to provide background for studies of the intestinal microbiome and disease risk. These studies should also use techniques such as whole genome sequencing to determine if functional components of the intestinal microbiome of the premature infant also vary with time, location, and disease risk.

Chapter 6 References


Appendix 1: Bayesian Network Analysis of Sepsis Cases and Controls

Methods

Rationale: Bayesian networks create graphic models of probabilistic dependencies of random variables (1). I applied Bayesian network analysis to create a graphical representation of patterns of co-occurrence of bacteria at different taxonomic levels to understand their influences in relation to risk of sepsis.

Subjects: The same set of infants used in part 2 (sepsis cases and controls) were used in this analysis. This analysis used the first sample available from each case and control.

Statistical Analysis: Bayesian network analysis is a statistical tool that represents a set of discrete random variables and their conditional dependencies where results are presented as a directed acyclic graph (1). Each node on the graph represents a variable, and each arrow represents the direct influence of one variable on another (2). The variable an arrow points to is the child node, the variable an arrow originates from is the parent node (2). An example of a Bayesian network is given Kevin Murphy on his website (3). A brief summary of this example follows.

Consider a dataset examining the chance of grass being wet. Other variables in the dataset include whether or not a sprinkler is running and whether or not it is raining. On the graph, the sprinkler variable and the raining variable are parent nodes and the wet grass variable is a daughter node. This means that the posterior probability of having wet grass is altered based on the value of the sprinkler and rain variables. Closer examination of the data in this case showed that the grass is more likely to be
wet when the sprinkler is on and that the grass is more likely to be wet when it is raining. If the sprinkler is on and it is raining, the grass is even more likely to be wet (3).

In the context of microbiome data, this allows for the detection of associations between the presence and absence of different taxa and disease risk. In an imaginary example, knowing that there are bacteria from phyla *Firmicutes* present in a sample might mean that there is a greater chance of detecting bacteria from family *Clostridiaceae* (a family in phylum *Firmicutes*). In turn, the presence of *Clostridiaceae* might be associated with increased risk of antibiotic associated diarrhea. Bayesian network analysis can find connections between unrelated taxa, identifying taxa that tend to co-occur and taxa that tend to exclude one another in large datasets, as well as how the presence and absence of these taxa relate to disease risk.

Bayesian network analysis was conducted in R using the bnlearn package. The `hc` command with default settings was used to run the hill climbing algorithm on the data. The `bn.fit` command with default settings was used to fit the parameters of Bayesian network. Analysis was stratified by city of origin. Included in the model was the presence or absence of sepsis, and the presence or absence of bacterial taxa from phylum through family levels.

**Results**

The network for Birmingham infants found a connection between *Moraxellaceae* and sepsis (Figure 1). In this case, infants who had *Moraxellaceae* in their stool were more likely to develop sepsis. The network for Cincinnati infants also found a connection between *Moraxellaceae* and sepsis (Figure 2), but in this case infants were
less likely to develop sepsis. Many of the additional connections at both Birmingham and Cincinnati were of higher levels of taxonomy predicting lower levels, e.g. infants whose samples lacked order Bifidobacteriales also lacked family Bifidobacteriaceae.
Figure 1: Bayesian network of Birmingham samples. The connection showing Moraxellaceae as a predictor of sepsis is circled in red. The variables encoded by each node are numbered; please see the end of this appendix for table showing what each variable is.
Figure 2: Bayesian network of Cincinnati samples. The connection showing Moraxellaceae as a predictor of sepsis is circled in red. The variables encoded by each node are numbered; please see the end of this appendix for table showing what each variable is.
Discussion

The initial appearance of similarity between the Birmingham and Cincinnati Bayesian networks at first suggested that this technique might be robust to differences in microbial colonization that occur by location, because both Birmingham and Cincinnati had the presence of Moraxellaceae altering the probability of sepsis. Unfortunately, closer examination revealed that Moraxellaceae increased the chance of sepsis in Birmingham but decreased it in Cincinnati. This makes it impossible to generalize the role of Moraxellaceae in sepsis beyond individual sites. Furthermore, Bayesian network analysis is unable to examine the relative abundance of different bacterial taxa in relation to the categorical outcome of late-onset sepsis. The median relative abundance of Moraxellaceae in samples with any Moraxellaceae in Cincinnati (0.85%) is greater than in Birmingham (0.15%). The failure to account for relative abundance may contribute to the opposite effect of Moraxellaceae on sepsis between the two cities and limits the utility of Bayesian network analysis in examining microbiome data.

Tables of Figure Variables

<table>
<thead>
<tr>
<th>Birmingham Variable</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sepsis</td>
<td>X1</td>
</tr>
<tr>
<td>Actinobacteria.x</td>
<td>X2</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>X3</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>X4</td>
</tr>
<tr>
<td>Fusobacteria</td>
<td>X5</td>
</tr>
<tr>
<td>Tenericutes</td>
<td>X6</td>
</tr>
<tr>
<td>Verrucomicrobia</td>
<td>X7</td>
</tr>
<tr>
<td>Actinobacteria.y</td>
<td>X8</td>
</tr>
<tr>
<td>Bacilli</td>
<td>X9</td>
</tr>
<tr>
<td>Bacteroidia</td>
<td>X10</td>
</tr>
<tr>
<td>Betaproteobacteria</td>
<td>X11</td>
</tr>
<tr>
<td>Clostridia</td>
<td>X12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cincinnati Variable Name</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sepsis</td>
<td>X1</td>
</tr>
<tr>
<td>Actinobacteria.x</td>
<td>X2</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>X3</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>X4</td>
</tr>
<tr>
<td>Fusobacteria</td>
<td>X5</td>
</tr>
<tr>
<td>Tenericutes</td>
<td>X6</td>
</tr>
<tr>
<td>Actinobacteria.y</td>
<td>X7</td>
</tr>
<tr>
<td>Bacilli</td>
<td>X8</td>
</tr>
<tr>
<td>Bacteroidia</td>
<td>X9</td>
</tr>
<tr>
<td>Betaproteobacteria</td>
<td>X10</td>
</tr>
<tr>
<td>Clostridia</td>
<td>X11</td>
</tr>
<tr>
<td>Erysipelotrichi</td>
<td>X12</td>
</tr>
<tr>
<td>Birmingham Variable</td>
<td>Number</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Erysipelotrichi</td>
<td>X13</td>
</tr>
<tr>
<td>Fusobacteria</td>
<td>X14</td>
</tr>
<tr>
<td>Mollicutes</td>
<td>X15</td>
</tr>
<tr>
<td>Verrucomicrobiae</td>
<td>X16</td>
</tr>
<tr>
<td>Actinomycetales</td>
<td>X17</td>
</tr>
<tr>
<td>Aeromonadales</td>
<td>X18</td>
</tr>
<tr>
<td>Bacillales</td>
<td>X19</td>
</tr>
<tr>
<td>Bacteroidales</td>
<td>X20</td>
</tr>
<tr>
<td>Bifidobacteriales</td>
<td>X21</td>
</tr>
<tr>
<td>Burkholderiales</td>
<td>X22</td>
</tr>
<tr>
<td>Clostridiales</td>
<td>X23</td>
</tr>
<tr>
<td>Erysipelotrichales</td>
<td>X24</td>
</tr>
<tr>
<td>Fusobacteriales</td>
<td>X25</td>
</tr>
<tr>
<td>Lactobacillales</td>
<td>X26</td>
</tr>
<tr>
<td>Mycoplasmatales</td>
<td>X27</td>
</tr>
<tr>
<td>Pasteurellales</td>
<td>X28</td>
</tr>
<tr>
<td>Pseudomonadales</td>
<td>X29</td>
</tr>
<tr>
<td>Verrucomicrobiae</td>
<td>X30</td>
</tr>
<tr>
<td>Xanthomonadales</td>
<td>X31</td>
</tr>
<tr>
<td>UnknownBacilli3.x</td>
<td>X32</td>
</tr>
<tr>
<td>Tissierellaceae</td>
<td>X33</td>
</tr>
<tr>
<td>Actinomycetaceae</td>
<td>X34</td>
</tr>
<tr>
<td>Aeromonadaceae</td>
<td>X35</td>
</tr>
<tr>
<td>Bacteroidaceae</td>
<td>X36</td>
</tr>
<tr>
<td>Bifidobacteriaceae</td>
<td>X37</td>
</tr>
<tr>
<td>Carnobacteriaceae</td>
<td>X38</td>
</tr>
<tr>
<td>Clostridaceae</td>
<td>X39</td>
</tr>
<tr>
<td>Corynebacteriaceae</td>
<td>X40</td>
</tr>
<tr>
<td>Enterococccaceae</td>
<td>X41</td>
</tr>
<tr>
<td>Erysipelotrichaceae</td>
<td>X42</td>
</tr>
<tr>
<td>Fusobacteriaceae</td>
<td>X43</td>
</tr>
<tr>
<td>Lachnospiraceae</td>
<td>X44</td>
</tr>
<tr>
<td>Leptotrichiaceae</td>
<td>X45</td>
</tr>
<tr>
<td>Moraxellaceae</td>
<td>X46</td>
</tr>
<tr>
<td>Mycoplasmataceae</td>
<td>X47</td>
</tr>
<tr>
<td>Oxalobacteriaceae</td>
<td>X48</td>
</tr>
<tr>
<td>Pasteurellaceae</td>
<td>X49</td>
</tr>
<tr>
<td>Peptostreptococcaceae</td>
<td>X50</td>
</tr>
<tr>
<td>Porphyromonadaceae</td>
<td>X51</td>
</tr>
<tr>
<td>Prevotellaceae</td>
<td>X52</td>
</tr>
</tbody>
</table>
### Birmingham Variable

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudomonadaceae</td>
<td>X53</td>
</tr>
<tr>
<td>Ruminococcaceae</td>
<td>X54</td>
</tr>
<tr>
<td>Staphylococcaceae</td>
<td>X55</td>
</tr>
<tr>
<td>Streptococcaceae</td>
<td>X56</td>
</tr>
<tr>
<td>Veillonellaceae</td>
<td>X57</td>
</tr>
<tr>
<td>Verrucomicrobiaceae</td>
<td>X58</td>
</tr>
<tr>
<td>Xanthomonadaceae</td>
<td>X59</td>
</tr>
<tr>
<td>UnknownBacillales1</td>
<td>X60</td>
</tr>
<tr>
<td>UnknownBacillales2</td>
<td>X61</td>
</tr>
<tr>
<td>UnknownBacillales3</td>
<td>X62</td>
</tr>
<tr>
<td>UnknownBacillales4</td>
<td>X63</td>
</tr>
<tr>
<td>UnknownBacillales5</td>
<td>X64</td>
</tr>
<tr>
<td>UnknownBacillales6</td>
<td>X65</td>
</tr>
<tr>
<td>UnknownBacilli3.y</td>
<td>X66</td>
</tr>
<tr>
<td>UnknownClostridiales1</td>
<td>X67</td>
</tr>
<tr>
<td>UnknownClostridiales2</td>
<td>X68</td>
</tr>
<tr>
<td>UnknownLactobacillales2</td>
<td>X69</td>
</tr>
</tbody>
</table>

### Cincinnati Variable Name

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevotellaceae</td>
<td>X53</td>
</tr>
<tr>
<td>Pseudomonadaceae</td>
<td>X54</td>
</tr>
<tr>
<td>Rikenellaceae</td>
<td>X55</td>
</tr>
<tr>
<td>Ruminococcaceae</td>
<td>X56</td>
</tr>
<tr>
<td>Staphylococcaceae</td>
<td>X57</td>
</tr>
<tr>
<td>Streptococcaceae</td>
<td>X58</td>
</tr>
<tr>
<td>Veillonellaceae</td>
<td>X59</td>
</tr>
<tr>
<td>Xanthomonadaceae</td>
<td>X60</td>
</tr>
<tr>
<td>UnknownBacillales1</td>
<td>X61</td>
</tr>
<tr>
<td>UnknownBacillales2</td>
<td>X62</td>
</tr>
<tr>
<td>UnknownBacillales3</td>
<td>X63</td>
</tr>
<tr>
<td>UnknownBacillales4</td>
<td>X64</td>
</tr>
<tr>
<td>UnknownBacillales5</td>
<td>X65</td>
</tr>
<tr>
<td>UnknownBacillales6</td>
<td>X66</td>
</tr>
<tr>
<td>UnknownBacilli3.y</td>
<td>X67</td>
</tr>
<tr>
<td>UnknownClostridiales1</td>
<td>X68</td>
</tr>
<tr>
<td>UnknownClostridiales2</td>
<td>X69</td>
</tr>
<tr>
<td>UnknownClostridiales10</td>
<td>X70</td>
</tr>
<tr>
<td>UnknownClostridiales11</td>
<td>X71</td>
</tr>
<tr>
<td>UnknownClostridiales12</td>
<td>X72</td>
</tr>
<tr>
<td>UnknownClostridiales3</td>
<td>X73</td>
</tr>
<tr>
<td>UnknownClostridiales4</td>
<td>X74</td>
</tr>
<tr>
<td>UnknownClostridiales5</td>
<td>X75</td>
</tr>
<tr>
<td>UnknownClostridiales9</td>
<td>X76</td>
</tr>
</tbody>
</table>

### Appendix 1 References

Appendix 2: Human Subjects Protection

Prior to the study start, the Novel Biomarkers Cohort received Institutional Review Board (IRB) approval from the Cincinnati Children's Hospital and Medical Center IRB. Prior to subject enrollment, informed consent was obtained from the parent or guardian of each infant. The original IRB approval and an example of the consent form are included in the following pages.
Institutional Review Board - Federalwide Assurance #00002988
Cincinnati Children's Hospital Medical Center

From: IRB Committee
To: Principal Investigator: Ardythe Morrow
Divison: Biostatistics and Epidemiology

Re: Study ID: 2008-0463
Study Title: Novel genetic and salivary glycan biomarkers for risk of NEC in ELBW infants

The above referenced protocol and all applicable additional documentation provided to the CCHMC IRB were reviewed and RE-APPROVED using an EXPEDITED review procedure set forth in 45 CFR 46.110(b)(1), Category(ies)(see below) on 6/9/2009.

This study will be due for continuing review at least 30 days before 6/8/2010.

The following documents were reviewed and approved:
Name
Consent form Version 1
HIPAA form
NEC protocol
UAB.doc

Please note the following requirements:

OTHER APPROVALS: Principal investigators are responsible for maintaining approval from other applicable review committees and performance sites. This includes, but is not limited to, Divisional Scientific Review committee, General Clinical Research Center (GCRC), Radiation Safety, Institutional Biosafety Committee (IBC), Conflict of Interest (COI) Committee, and any sites (i.e. schools, hospitals) where the research may be conducted. Principal investigators are also responsible for maintaining approval from the FDA and a valid contract between the sponsor and CCHMC, as applicable. If any of these entities require changes to the IRB-approved protocol and/or informed consent/assent document(s), the changes must be submitted to and approved by the IRB prior to implementation.

AMENDMENTS: The principal investigator is responsible for notifying the IRB of any changes in the protocol, participating investigators, procedures, recruitment, consent forms, FDA status, or conflicts of interest. Approval is based on the information as submitted. New procedures cannot be initiated until IRB approval has been given. If you wish to change any aspect of this study, please submit an Amendment via ePAS to the IRB, providing a justification for each requested change.

CONTINUING REVIEW: The investigator is responsible for submitting a Continuing Review via ePAS to the IRB at least 30 days prior to the expiration date listed above. Please note that study procedures may only continue into the next cycle if the IRB has reviewed and granted re-approval prior to the expiration date.

UNANTICIPATED PROBLEMS: The investigator is responsible for reporting unanticipated problems promptly to the CCHMC IRB via ePAS according to current CCHMC reporting policy found on CenterLink.

STUDY COMPLETION: The investigator is responsible for notifying the IRB by submitting a Request to Close via ePAS when the research, including data analysis, has completed.
Research Categories

Prospective collection of biological specimens for research purposes by noninvasive means. Examples: (a) hair and nail clippings in a nondisfiguring manner; (b) deciduous teeth at time of exfoliation or if routine patient care indicates a need for extraction; (c) permanent teeth if routine patient care indicates a need for extraction; (d) excreta and external secretions (including sweat); (e) uncanulled saliva collected either in an unstimulated fashion or stimulated by chewing gumbase or wax or by applying a dilute citric solution to the tongue; (f) placenta removed at delivery; (g) amniotic fluid obtained at the time of rupture of the membrane prior to or during labor; (h) supra- and subgingival dental plaque and calculus, provided the collection procedure is not more invasive than routine prophylactic scaling of the teeth and the process is accomplished in accordance with accepted prophylactic techniques; (i) mucosal and skin cells collected by buccal scraping or swab, skin swab, or mouth washings; (j) sputum collected after saline mist nebulization.

Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis). (NOTE: Some research in this category may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b)(4). This listing refers only to research that is not exempt.)

If you have any questions about the information in this letter, please contact the Institutional Review Board office at 513-636-8039.

Thank you for your cooperation during the review process.

Office of Research Compliance & Regulatory Affairs  |  3333 Burnet Avenue  |  Suite F3325  |  MLC 7040
Cincinnati, OH 45229-3039  |  P 513-636-2754  F 513-636-1321
CINCINNATI CHILDREN'S HOSPITAL MEDICAL CENTER

INFORMED CONSENT/PARENTAL PERMISSION
FOR PARTICIPATION IN A RESEARCH STUDY

STUDY TITLE: Novel Genetic and Salivary Glycan Biomarkers for Risk of Necrotizing Enterocolitis (NEC) and Late Onset Sepsis (LOS) in Very Low Birth Weight Infants

SPONSOR NAME: National Institute of Child Health and Human Development (NICHD)

INVESTIGATOR INFORMATION:

Ardythe L. Morrow, PhD  513-636-7626
Principal Investigator Name  Office Telephone Number

Kurt R. Schibler, MD  513-636-3972 / 513-736-5649
Co-Principal Investigator Name  Office Telephone Number / 24 hr Emergency Contact

INTRODUCTION:

You are being asked to give permission for you and your infant to participate in a research study. Before agreeing to give permission for your infant to participate in this study, it is important that you read and understand the following explanation. It describes, in words that can be understood by a lay person, the purpose, procedures, benefits, risks and discomforts of the study and the precautions that will be taken. It also describes the alternatives available and your right to withdraw your infant from the study at any time. No guarantee or assurance can be made as to the results of the study. Participation in this research study is completely voluntary. Refusal to participate will involve no penalty or loss of benefits to which you or your infant are otherwise entitled. You may withdraw your infant from the study at any time without penalty.

WHO IS CONDUCTING THE RESEARCH STUDY?

This study is directed by Dr. Ardythe Morrow and Dr. Kurt Schibler, the researchers at Cincinnati Children's Hospital. Dr. Kurt Schibler is responsible for the medical supervision of this research study.

Funds to conduct this study are being provided by the National Institute of Child Health and Human Development (NICHD).
WHY IS THIS RESEARCH BEING DONE?

The purpose of this research study is to determine whether substances called glycans can be used to predict the risk for developing necrotizing enterocolitis (NEC), a severe disease of the intestines, and infection. Both occur in premature infants. Glycans are present in saliva and in the intestines and they play a role in controlling the number and the types of bacteria living in the intestines. It is normal to have bacteria living in the intestines. Some types of bacteria can be helpful for good health, but other types can be harmful and cause disease. The amount of glycans present in saliva can differ between individuals due to heredity (genetics) and due to the types of bacteria living in the intestines. While the cause of necrotizing enterocolitis is unknown, recent studies suggest that the types of bacteria that live in the intestines of premature infants are different than those that live in the intestines of term infants. Glycan production, types of bacteria living in the intestines, and immaturity of the immune system, in premature infants are all believed to contribute to the development of necrotizing enterocolitis and infection.

WHY HAS YOUR CHILD BEEN ASKED TO TAKE PART IN THIS RESEARCH STUDY?

Your infant is being asked to take part in this research study because your infant was born premature (less than 29 weeks gestational age) and is hospitalized in the newborn intensive care unit.

WHO SHOULD NOT BE IN THE RESEARCH STUDY?

Your infant should not be in this study if he/she has any of the following:

- Gestational age at birth of 29 weeks or greater
- Major birth defect

HOW LONG WILL YOUR INFANT BE IN THE RESEARCH STUDY?

Your infant will be in the research study for the duration he/she is in the neonatal intensive care unit. The researcher may decide to end your infant’s participation in this research study at any time, without your permission, for any of the following reasons: the study doctor determines that it is in your infant’s medical best interest, the study is ended early for any reason, or new information becomes available.

HOW MANY PEOPLE WILL TAKE PART IN THE RESEARCH STUDY?

This study will be conducted at approximately two research centers throughout the country. A total of 600 infants will participate in this research study at all of the research centers.

Approximately 30 infants will take part at the Children’s Hospital study site in Cincinnati. The other Cincinnati sites include the Good Samaritan and University Hospitals.

WHAT IS INVOLVED IN THE RESEARCH STUDY?

If you agree to allow your infant to participate and your infant qualifies for this study, the following procedures will be performed:

- A sample of saliva will be collected from you with a cotton swab.
• Information about your infant will be collected from the hospital medical record. This includes information about your infant's health and hospital course in the newborn intensive care unit.

• Samples of saliva will be collected from your infant with a soft cotton swab. A total of seven samples of saliva will be collected according to the schedule displayed below.

• Stool will be collected from your infant's diaper, daily during the first two weeks of life.

• Urine will be collected from a cotton ball placed in your infant's diaper, daily during the first two weeks of life.

• If you have decided to provide breast milk to your infant, a total of two samples of your breast milk will be collected, around the second to third week of life. You will be asked to provide 6 mL of breast milk collected twice for this study (a total of 12 mL). However, we will not use any milk for the study without making sure your infant has plenty of milk for feeding.

• If the doctor suspects your infant has an infection while in the newborn intensive care unit, a sample of his/her blood may be taken. This sample is sent to the lab to determine whether bacteria are growing in your infant's blood. If bacteria are found in your infant's blood, a sample of the growing bacteria (isolate) will be collected for this study.

• Laboratory analyses will be performed on all samples, to provide information that will assist the researchers in understanding glycan production, types of bacteria living in the intestines, and the immune system in premature infants.

• If available, we will also analyze samples of the placenta that are stored in the hospital's clinical laboratory.

Table 1. Timing of specific procedures and sample collections by study day (postpartum age of infant)

<table>
<thead>
<tr>
<th>PROCEDURES</th>
<th>Enrollment</th>
<th>Study Day</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2±1</td>
</tr>
<tr>
<td>Informed Consent</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Inclusion/Exclusion</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Saliva for DNA (Mom)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Saliva for DNA (Infant)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Saliva for Glycan (Infant)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Stool (Infant)*</td>
<td>If available, collect one sample each day (1-14±1)</td>
<td></td>
</tr>
<tr>
<td>Breast Milk</td>
<td></td>
<td>X±3</td>
</tr>
<tr>
<td>Urine (Infant)</td>
<td>If available, collect one sample each day (1-14±1)</td>
<td></td>
</tr>
</tbody>
</table>

*The timing of the day 2 saliva collection from mother and infant is flexible. If this sample is missed for either you or your infant, we will ask to reschedule this sample collection.

Page 3 of 9
WHAT ARE THE RISKS AND DISCOMFORTS OF THE RESEARCH STUDY?

Potential risks to the participants are minimal. There is a risk of mild discomfort from the collection of saliva. There is no risk to collection of urine or stool. There may be some discomfort from the collection of breast milk.

There is a risk of loss of confidentiality. Every effort will be made to keep your infant’s medical record confidential. There will be no name or patient identification in any study report that may be published. Measures to protect you and your baby’s identity are described in the confidentiality section. In regards to genetic confidentiality the same measures will be taken.

There may be unknown or unforeseen risks associated with study participation.

FUTURE STUDIES:

The samples will be collected from your infant as part of this study and they may be maintained for a long period of time for research purposes or until the samples are gone.

During the course of this study if your infant’s samples are not completely used up, you may choose whether or not to allow other researchers to use your infant’s samples. We may use these samples to help us:

- Learn more about diseases of premature infants
- Learn about genetic causes of diseases that are common to premature infants
- Someday learn how to prevent or cure diseases affecting premature infants.

In such cases, samples will be coded so that the other researchers will have no way to identify your infant. Under no circumstances will we release information that identifies your infant to other people without your written consent. Please initial one of the following choices.

[_____] Yes, I agree to share my infant’s samples.

__________________________
Signature of Infant’s Parent

[_____] No, I do not agree to share my infant’s samples.

ARE THERE DIRECT BENEFITS TO TAKING PART IN THE RESEARCH STUDY?

If you agree to allow your infant to take part in this research study, your infant will not receive a direct medical benefit.

The information learned from this research study may benefit other patients who have been born premature in the future.
WHAT OTHER CHOICES ARE THERE?

Instead of being in this research study, you have the option of not having you or your infant participate in the study.

HOW WILL INFORMATION ABOUT YOUR CHILD BE KEPT PRIVATE AND CONFIDENTIAL?

Cincinnati Children’s Hospital Medical Center and/or the Investigator will take the following precautionary measures to protect your infant’s privacy and confidentiality of your infant’s research and/or medical records: All of your infant’s information, including samples studied in the laboratory, will be coded using a unique identifier. Only members of the study team will be given access to the study records. No individual identifying information will be used in publications or reports that come from this study.

A copy of this consent form will be included in your infant’s medical research record.

Your infant will be registered in the Cincinnati Children’s Hospital Medical Center’s computer system as a research subject.

By signing this consent form you are giving permission for representatives of the Cincinnati Children’s Hospital Medical Center ("CCHMC"), the Investigator, and CCHMC employees involved with the research study, including the Institutional Review Board and the Office for Research Compliance and Regulatory Affairs, and any sponsoring company or their appointed agent to be allowed to inspect sections of your child’s medical and research records related to this study.

WILL THE RESULTS OF MY CHILD’S RESEARCH-RELATED TESTS BE AVAILABLE?

Information regarding any of our results pertaining to your infant are for research purposes only and will not be shared.

WHAT IF NEW INFORMATION BECOMES AVAILABLE DURING THE RESEARCH?

The investigator will tell you and your child about new information from this or other studies that may affect your infant’s health, welfare, or willingness to stay in this study.

The information from the research study may be published; however, your infant will not be identified in such a publication. The publication will not contain information about your child that would enable someone to determine your child’s identity as a research participant without your authorization.

WHAT ARE YOUR COSTS TO BE IN THIS STUDY?

There are no costs to you to be part of this research study.

WILL YOU/YOUR CHILD BE PAID TO PARTICIPATE IN THIS RESEARCH STUDY?

You will not receive reimbursement for your infant’s participation in this study.
The above samples collected during this research project may be used in the development of a product that could be patented or licensed. There are no plans to provide financial compensation to you or your infant should this occur.

WHAT ARE YOUR RIGHTS AS A PARTICIPANT?

You and your infant’s participation in this study are completely voluntary. You may choose either to take part or not to take part in this research study. Your decision whether or not to participate will not result in any penalty or loss of benefits to you or your infant and the standard medical care for your child’s condition will remain available to him/her.

If you decide to participate and to allow your infant to take part in the research study, you are free to withdraw your permission and discontinue your participation and your infant’s participation in this research study at any time. Leaving the study will not result in any penalty or loss of benefits to you or your infant.

If you have questions about the study, you will have a chance to talk to one of the study staff or your infant’s regular doctor. Do not sign this form unless you have had the chance to ask questions and have received satisfactory answers.

Nothing in this parental permission form waives any legal rights you or your infant may have, nor does it release the investigator, the sponsor, the institution, or its agents from liability for negligence.

ABILITY TO CONDITION TREATMENT ON PARTICIPATION IN THIS STUDY:

You have a right to refuse to sign this parental permission form and authorization to use/disclose your infant’s Protected Health Information for research purposes.

If you refuse to sign this consent, your rights and those of your infant concerning treatment, payment for services, enrollment in a health plan, and eligibility for benefits, will not be affected.

WHOM DO YOU CALL IF YOU HAVE QUESTIONS OR PROBLEMS?

For questions, concerns, or complaints about this research study or to report a research-related injury, you can contact the researcher Dr. Kurt Schibler at 513-636-3972. Researchers are available to answer any questions you may have about the research study at any time.

If you have general questions about your infant’s rights as a research participant in this research study, or questions, concerns, or complaints about the research, you can call the Cincinnati Children’s Hospital Medical Center Institutional Review Board at 513-636-8039. You can also call this number if the research staff could not be reached, or if you wish to talk to someone other than the research staff.

HIPAA AUTHORIZATION FOR USE/DISCLOSURE OF PROTECTED HEALTH INFORMATION FOR A RESEARCH STUDY:

We understand that information about you and your health is personal and we are committed to protecting the privacy of that information. Because of our commitment to protect your privacy, we must obtain your written authorization (permission) before we may use or disclose (release) your “protected health information”
IRB #: 2006-0463

Cincinnati Children’s

Approved: 3/18/2014
Do Not Use After: 3/17/2015

(sometimes referred to as “PHI”) related to the study described to you. This form provides that authorization and helps us make sure that you are properly informed of how this information will be used or disclosed. Please read the information below carefully before signing this form either for you, as the participant, or as the personal representative (parent, legal guardian, etc.) for the participant. Note that when we refer to “you” or “your” throughout this document, we are referring to the participant, even when this form is signed by the participant’s personal representative.

Use and disclosure covered by this authorization:

If you sign this document, you give permission to Cincinnati Children’s Hospital Medical Center (“Cincinnati Children’s”) to use or disclose your medical and research information for the purpose of this study. Your PHI that will be used and disclosed in connection with this study consists of:

- Your Cincinnati Children’s medical records
- Your research record for this study
- Results of your laboratory tests
- Clinical and research observations made during your participation in the study
- In the event that your medical record contains such information, information concerning HIV testing or the treatment of AIDS or AIDS-related conditions, drug or alcohol abuse, drug-related conditions, alcoholism, and/or psychiatric/psychological conditions (but not psychotherapy notes).

Who will disclose, receive and/or use the information?

This form authorizes the following to disclose, use and receive your PHI:

- Every research site of the study (including Cincinnati Children’s) and each site’s research staff and medical staff
- Every health care provider who provides services to you in connection with the study
- Any laboratories and other individuals and organizations that analyze your PHI in connection with the study
- The Sponsor and the people and companies they use to oversee, administer and/ or conduct the study
- Federal regulatory agencies, other foreign regulatory agencies, and others as required by law
- The members of the Cincinnati Children’s Institutional Review Board and staff of the Office of Research Compliance and Regulatory Affairs
- The Principal Investigator and members of the study’s research team
- Data Safety Monitoring Board (if applicable)

By signing this document, you are authorizing Cincinnati Children’s to use and/or disclose your PHI for this study. The purpose for the uses and disclosures is to conduct the study explained to you during the informed consent process and to ensure that information relating to the study is available to all parties who may need it for research purposes.

Those persons who receive your information may not be required by Federal privacy laws (such as the Health Insurance Portability and Accountability Act, also known as (“HIPAA”) to protect it and may share the information with others without your permission, if permitted by laws governing them.

You may revoke (choose to withdraw) this authorization at any time after you have signed it by providing the Principal Investigator (listed on the first page of the informed consent document) with a written statement that you wish to revoke it. Your revocation will be effective immediately and your PHI can no longer be used or disclosed for this study by Cincinnati Children’s and the other persons or organizations that are identified
above, except to the extent that Cincinnati Children’s and/or the other persons or organizations identified above have already acted in reliance on the Authorization. In addition, the information may continue to be used and/or disclosed to preserve the integrity of the study.

Unless you notify us in writing of your decision to withdraw this authorization to use and disclose your PHI, it will expire at the end of the study. If the study involves the creation or maintenance of a research database repository, this authorization will not expire.

If you refuse to sign this authorization, you may not be able to receive research-related procedures and may not be able to continue in this study. However, your rights concerning treatment not related to this study, payment for services, enrollment in a health plan or eligibility of benefits will not be affected.

For further information about your rights, please see the Cincinnati Children’s Notice of Privacy Practices on our website at http://www.cincinnatichildrens.org/about/corporate/hipaa.
SIGNATURES:

The research team has discussed this study with you and answered all of your questions. Like any research, the researchers cannot predict exactly what will happen. Once you have had enough time to consider whether your child should participate in this research you will document your permission by signature below.

You will receive a copy of this signed document for your records.

Signature of Mother as Participant:

__________________________
Signature of Participant (or Legally Authorized Representative)*

__________________________
Date

*Complete by a Legally Authorized Representative (parent, legal guardian, etc.)

__________________________
Description of Legally Authorized Representative’s Authority to Sign for Participant

__________________________
Printed Name of Legally Authorized Representative

Signature for Infant’s Participation:

__________________________
Signature of Participant’s Mother

__________________________
Date

__________________________
Printed Name of Research Participant (Infant)

__________________________
Signature of Individual Obtaining Consent

__________________________
Date
Acknowledgements

The completion of this project would not have been possible without the help and support of many individuals. I would like to thank especially my research mentor, Dr. Ardythe Morrow; my academic advisor Dr. Ranjan Deka; and the rest of my committee, Dr. John Harley, Dr. MB Rao, and Dr. Christina Valentine. My thanks also to the Morrow lab group: Barbara Davidson, Maria Hughes, Aimee LeFevers, Kelly Dingess, Myra Johnson, Rebecca Currier, and Olivia Ballard. Thanks also go to the many collaborators who have taken time to help including Dr. Doyle Ward, Dr. Kurt Schibler, Dr. David Newburg, Dr. Zhouteng Yu, Dr. Laurie Nommsen-Rivers, Estelle Fischer, Jeanie Kleiman, and others too numerous to name. My eternal gratitude goes to my family and friends who kept me sane throughout this process, most especially my parents Linda and Arthur Taft and my dear friend Stephanie Contino.

This project and my training has been funded by the National Institute of Environmental Health Sciences (T32 ES010957), the National Institute of Child Health and Human Development, National Institutes of Health (grant numbers HD059140, HD13021, and HD27853); National Center for Research Resources, National Institutes of Health (grant numbers 5UL1RR026314-03, U54HG004969); National Human Genome Research Institute, National Institutes of Health (HG005969); Danone Research (PLF-5972-GD); and the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900018C.