THE IMPACT OF PROMOTER POLYMORPHISMS ON CYTOKINE CONCENTRATION IN PRETERM BREAST MILK AND SUBSEQUENT INFANT OUTCOMES

by

Kelley Lynne Baumgartel

B.S.N., University of Pennsylvania, 2006

Submitted to the Graduate Faculty of

School of Nursing in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

SCHOOL OF NURSING

This dissertation was presented

by

Kelley L. Baumgartel

It was defended on

May 27th, 2015

and approved by

Susan Cohen, DNS, Associate Professor, School of Nursing

Maureen Groer, PhD, Professor, University of South Florida College of Nursing

Dianxu Ren, MD, PhD, Associate Professor, School of Nursing

Diane L. Spatz, PhD, Professor, University of Pennsylvania School of Nursing

Dissertation Advisor: Yvette P. Conley, Professor, School of Nursing

THE IMPACT OF PROMOTER POLYMORPHISMS ON CYTOKINE CONCENTRATION IN PRETERM BREAST MILK AND SUBSEQUENT INFANT OUTCOMES

Kelley L. Baumgartel, BSN, PhD
University of Pittsburgh, 2015

The immune protection offered through breast milk is especially important for premature infants. Interleukins (ILs), found in breast milk but in varying concentrations, may provide preterm infants with protection against prematurity-related complications. Promoter polymorphisms have been associated with variable IL levels, though this relationship has never been investigated in breast milk. The aims of this study were to: 1) examine the relationship between maternal IL genotypes and weekly milk concentrations of IL4, IL6, and IL10, 2) describe the trajectories of milk IL change over the first three weeks postpartum, 3) examine whether maternal IL genotypes predict milk IL trajectories, 4) examine if weekly IL levels and/or IL trajectories predict infant outcomes, and 5) explore a relationship between maternal IL genotypes and infant outcomes. An ancillary study was conducted that extracted maternal DNA from breast milk for genotyping using TaqMan. Trajectory modeling was used to identify IL subgroups. After controlling for gestational age and prepregnancy BMI, there was an inverse association between rs1800796 minor allele absence (MAA) and milk IL6 among African Americans (p=0.0722). Subsequently, higher milk IL6 was also associated with decreased risk of IVH in African Americans (OR=0.32, p=0.1059). Additionally, among African Americans, there was a positive association between IL6 milk levels and calprotectin (week one p=0.0794, week two p=0.0978). Caucasians had an inverse relationship between rs1800795 MAA and milk IL6 (p=0.0966). Subsequently, there is a relationship between milk IL6 and infant calprotectin in Caucasians (p=0.0290). MAA of rs1800896 was associated with milk IL10 levels among African Americans (p=0.0705), though there was no relationship between milk IL10 levels and outcomes. There were no associations

between maternal SNP and IL trajectory groups. Trajectory analysis resulted in linear group shapes, with two distinct subgroups in IL6, and three subgroups in both IL4 and IL10. Infants who received milk from IL4 group 2 were more likely to receive a blood transfusion than infants who received milk from group 3 (OR=4.16, p=0.0712). There was an association between IL6 group 1 membership and both IVH (OR=6.275, p=0.0412) and fecal calprotectin (p=0.0822). Traditionally significant findings ($p \leq 0.05$) included relationships between maternal IL genotypes and NICU outcomes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Kelley would like to acknowledge her outstanding committee members, peers, and the following support that have helped her to complete this dissertation: Targeted Research and Academic Training of Nurses in Genomics (NR00975909), Corrine M. Barnes Award, International Society of Nurses in Genetics Research Award, Sigma Theta Tau Research Award (Pittsburgh chapter), Judith A. Erlen Research Award, and the Dorothy Drake Brookes Endowment Award.

1.0 PROPOSAL INTRODUCTION

Breast milk is considered a medicine in the Neonatal Intensive Care Unit (NICU), as it provides immunological properties that cannot be prepared exogenously. The protective advantages of human milk are well established, and the American Academy of Pediatrics recommends that all preterm infants receive breast milk (American Academy of Pediatrics, 2012). There is great variability in breast milk protein composition among women who deliver at the same gestational age, including women who deliver premature babies. While this variation in breast milk composition has been established at the protein level, there is a dearth of information linking maternal factors and mechanism for variability in composition to level of protection offered through breast milk. This has a clinical application, as evidence shows that variability of immune components in breast milk may contribute to better infant outcomes (Fituch, Palkowetz, Goldman, & Schanler, 2004).

Variability in breast milk composition is influenced by lifestyle factors, including but not limited to: diet (Peng et al., 2009), drug use (Friguls et al., 2010), and exercise (M. W. Groer & Shelton, 2009). Another maternal variable that influences milk composition, and one that is easily measured, is Body Mass Index (BMI). For example, maternal BMI is positively associated with leptin concentration in breast milk (Fields & Demerath, 2012), and overweight mothers have lower levels of TGF-β2 and sCD14 levels in their breast milk when compared with normal weight mothers (Collado, Laitinen, Salminen, & Isolauri, 2012). The mechanism through which

BMI impacts breast milk composition is unknown, but one potential mechanism involves DNA methylation. The role of BMI on DNA methylation has been explored in areas unrelated to lactation. BMI was associated with methylation in the following studies: 1) higher BMIs are associated with hypomethylation of PBMC L1 gene among women with a history of abnormal PAP smears (Piyathilake, Badiga, Alvarez, Partridge, & Johanning, 2013); 2) methylation of the serotonin transporter gene increases by 1% per 0.33 increase in BMI (Zhao, Goldberg, & Vaccarino, 2013); 3) hypermethylation of Wnt signaling genes that are implicated in colorectal cancer (Rawson et al., 2012); and 4) hypermethylation of BRCA1 gene related to breast cancer (Bosviel et al., 2012). This evidence of BMI-influenced DNA methylation, combined with the impact of BMI on milk composition, led to our hypothesis that maternal BMI may influence breast milk variability through an epigenetic mechanism.

This study further hypothesizes that DNA methylation is a driving force behind immune factor composition in breast milk, which is critical to improved infant outcomes. We hypothesize that methylation patterns of DNA from the milk fed to premature infants will be correlated with interleukin levels from those same milk samples, and that these methylation patterns will be associated with infant outcomes (fecal calprotectin levels, [a commonly used marker of gastrointestinal inflammation], weight gain, time to full enteral feeds, and time to discharge).

1.1 PURPOSE AND SPECIFIC AIMS

Because an epigenetic study is needed to better understand breast milk variability among women who deliver preterm, this study will:

1. Determine if maternal BMI influences extent of methylation of immunity genes (IL4, IL6, and IL10) that show variation in the parent study.
2. Describe the relationship between extent of methylation of these immunity genes and interleukin levels in milk.
3. Explore whether extent of methylation of the IL4, IL6, and IL10 genes is associated with infant outcomes.

The candidate's long term research interests involve the environmental impact on gene expression and how a potential variation in milk protective factors influences outcomes for preterm infants. This also has a clinical application in the use of donor breast milk, which is usually full term milk. Results from the proposed dissertation may contribute to the mechanistic understanding of breast milk variability and subsequent infant outcomes.

1.2 BACKGROUND AND SIGNIFICANCE

The United States experiences the 6th highest preterm birth rate in the world, which is the leading cause of death among infants less than four weeks old (World Health Organization, 2014). The World Health Organization promotes breastfeeding as a key strategy to prevent death and complications that often result from prematurity. While universal breastfeeding promotion is certainly saving lives in this vulnerable population, outcome disparities exist among preterm infants who receive breast milk. Breast milk composition varies greatly between women who deliver prematurely and those who deliver at term, perhaps contributing to the protective role of premature breast milk against NEC and other infections. Protective factors in preterm breast milk include an increase in immunogolobulins (Araujo et al., 2005), fatty acids (Berenhauser, Pinheiro

do Prado, da Silva, Gioielli, & Block, 2012), and cytokines (Ustundag et al., 2005). The mechanism of compositional variation between preterm and term breast milk remains unknown, despite increasing breastfeeding rates and preterm births.

1.2.1 Rationale for taking an epigenetic approach

Breast milk composition is influenced by the environment, and epigenetic changes to DNA are influenced by environmental conditions. Lifestyle contributors to milk variation include diet (Peng et al., 2009), exercise (M. W. Groer & Shelton, 2009), and smoking (Etem Piskin, Nur Karavar, Arasli, & Ermis, 2012; Szlagatys-Sidorkiewicz, Martysiak-Zurowska, Krzykowski, Zagierski, & Kaminska, 2013; Szlagatys-Sidorkiewicz, Wos, et al., 2013), among others. Despite this environmental impact on milk composition, the molecular mechanism is not understood and it seems the environmental component of epigenetics is an ideal approach with which to begin.
Breast milk is highly variable among women, even those who deliver at the same gestational age. Protective components found in variable amounts that influence infant outcomes include: retinol (Ribeiro, Araujo, Pereira, & Dimenstein, 2007), IgA (Ballabio et al., 2007; Takahashi et al., 2002), and cytokines (Ustundag et al., 2005). This milk disparity may explain why some infants, despite receiving breast milk, develop complications while in the NICU. Epigenetic influences, including diet, influence breast milk composition (Kelishadi et al., 2012; Peng et al., 2009), and may help explain breast milk variability.

Epigenomic approaches have led to a better understanding of many complex diseases, including: oncology, cardiovascular disease and mental health disorders (Ordovas & Smith, 2010), (Read, Bentall, & Fosse, 2009; Vineis et al., 2011). While methylation analyses of *term* breast milk have been applied to understand breast cancer (Browne et al., 2011; Qin et al., 2012;

Wong et al., 2010), none of these milk methylation studies examined immune-related genes, nor did they examine milk with respect to maternal or infant health. Few genetic studies have been conducted to better understand the dynamic immunologic nature of human milk, and none have taken an epigenetic approach, despite the many environmental factors known to influence milk. A few animal studies have examined methylation of DNA in breast milk, and have confirmed the impact of methylation on gene expression in the lactating mammary gland in the mouse, cow and rat (Platenburg et al., 1996). Their results found that methylation variation exists among lactating animals. This study also noted that lactoferrin cDNA, a protein found in breast milk that protects against infection (Hadsell et al., 2007; Venkatesh & Abrams, 2009), showed variable expression due to methylation status.

Preterm infants who receive breast milk experience fewer complications than those who receive formula, though the disparity among breastfed infants who develop complications is not well established. It seems there is a link between milk composition and infant outcomes, as the milk fed to infants who present with "failure to thrive" resembles weaning milk (Motil, Sheng, & Montandon, 1994). Although this was a case study, this suggests that the contents in this milk may fail to meet the needs of a growing infant and may also explain some of the disparities seen among breastfed infants. For example, long chain polyunsaturated fatty acids (LCPUFA) are present in highly variable amounts in breast milk, and are heavily influenced by maternal environment and diet. Preterm infants with high levels of arachidonic acid (a major type of LCPUFA) experience improved growth outcomes up to one year (Carlson, Werkman, Peeples, Cooke, & Tolley, 1993). Perhaps the most relevant study which addresses specific milk components and subsequent outcomes revealed that IL-10, a regulatory cytokine examined in the

proposed study, was undetectable in milk fed to preterm infants who developed NEC, but was detectable in the milk fed to preterm infants who did not develop NEC (Fituch et al., 2004).

1.2.2 Rationale for examining BMI and its impact on methylation of immunity-related genes in breast milk.

Breast milk from women with higher BMIs has lower milk protein concentration (Bachour, Yafawi, Jaber, Choueiri, & Abdel-Razzak, 2012), higher fat content (Barbosa, Butte, Villalpando, Wong, & Smith, 1997; Nikniaz, Mahdavi, Arefhoesseini, & Sowti Khiabani, 2009; Rocquelin, 1998), and higher linoleic acid content (Villalpando et al., 2001). Milk samples taken from mothers at one and six months postpartum and were measured for immunological components (TGF-β2, soluble CD14 (sCD14), cytokines, and microbiota (Collado et al., 2012). Milk from overweight mothers contained less TGF-β2 and sCD14 when compared with normal weight mothers. This is significant, since TGF-β2 is a pro-immune regulatory type of cytokine and sCD14 is involved in the activation of the innate immune response. While TGF-β2 and sCD14 are present at high amounts in colostrum, they decrease significantly by the time mature milk is produced and failure to follow this pattern is associated with atopic dermatitis and asthma (Snijders et al., 2010). Additionally, overweight mothers have higher milk levels of Staphyloccus and lower milk levels of Bifidobacterium when compared with normal weight mothers, which indicates an imbalance of microbiotia in the breast milk from overweight mothers. These studies, especially those examining the immunological profile of breast milk as it relates to BMi, suggests that not all milk is created equal, and that BMI may be contributing to this disparity.

Leptin, known as the ‘hunger hormone’, is present in elevated amounts in the milk of women with higher BMIs (Fields & Demerath, 2012). This is important, since this relationship

affects the infant, as higher leptin levels in milk is negatively associated with infant weight gain (Fields & Demerath, 2012; Miralles, Sanchez, Palou, & Pico, 2006; Schuster, Hechler, Gebauer, Kiess, & Kratzsch, 2011). IL-6, a cytokine examined in the proposed study, increases leptin levels (Trujillo et al., 2004). Weight gain in the NICU is an important indicator of infant health. If BMI impacts infant weight gain through leptin level intake from breast milk, it is crucial to understand how BMI also impacts immunological components in breast milk that may predispose an infant to infection or complications.

Epigenetic studies have established the influence of BMI on methylation of Wnt signaling genes (Rawson et al., 2012), BRCA1 gene (Bosviel et al., 2012), and PBMCL1 gene (Piyathilake et al., 2013). There also appears to be a dose response to BMI, as Zhao (2013) revealed a 1% methylation increase for every 0.33 increase in BMI (Zhao et al., 2013). The potential link between BMI and methylation of immunity genes and subsequent milk composition that may predispose or protect a preterm infant will be evaluated in the proposed study. If this link is established, improved education efforts could target mothers with higher BMIs to gain an appropriate amount of weight during pregnancy for their health and to optimize their milk

1.2.3 Rationale for examining fecal calprotectin as a biomarker for infant outcome

Fecal calprotectin has been used as a biomarker for inflammation within the preterm infant population, as calprotectin is an accurate indicator of neutrophil migration toward the GI tract (Kapel et al., 2010; Kapel et al., 2005). Preterm infants with NEC symptoms experience a transient rise in fecal calprotectin when compared with preterm infants of the same gestational age without NEC (Campeotto et al., 2007). High fecal calprotectin levels suggest increased

granulocytes in the intestinal lumen from high permeability and/or lymphoid tissue development. Kapel suggests that environmental factors, including feeding, should be evaluated, as they could individually impact this process (Kapel et al., 2010). This continuous variable reflects the inflammatory state of the environment that is being directly exposed to the breast milk being studied. Clinical outcomes will also be available, including: weight gain, time to full enteral feeds, and time to discharge.

1.2.4 Genes selected for investigation

Breast milk is considered medicine in the NICU, as the immunological components provide infants with protection that their immature immune systems are incapable of producing. Cytokines are an integral component of breast milk immunobiology and provide preterm infants with protection from infecting microbes. Interleukins are present in variable amounts and milk concentration of these cytokines is associated with infant outcomes (Fields & Demerath, 2012; Fituch et al., 2004). Il-4, present in breast milk, responds to Antigen Presenting Cells (APC) by inducing lymphocytic antibody production. Il-4 and Il-10 (a down-regulating cytokine), are produced in low levels during early infancy, but are present in breast milk. Il-6, a pro-inflammatory cytokine, is inefficiently regulated in preterm infants (Currie et al., 2011), but is present in human milk. Aim 2 will focus on three of these cytokine genes (Il-4, Il-6, IL-10), which were found by the parent study to be present in variable amounts.

1.2.5 Conceptual Framework

The following figure illustrates the conceptual framework that guided this study, which examined: 1) the association between maternal BMI and methylation of immunity genes, 2) whether methylation is correlate with breast milk cytokine concentration and 3) the relationship between IL gene methylation and infant outcomes.

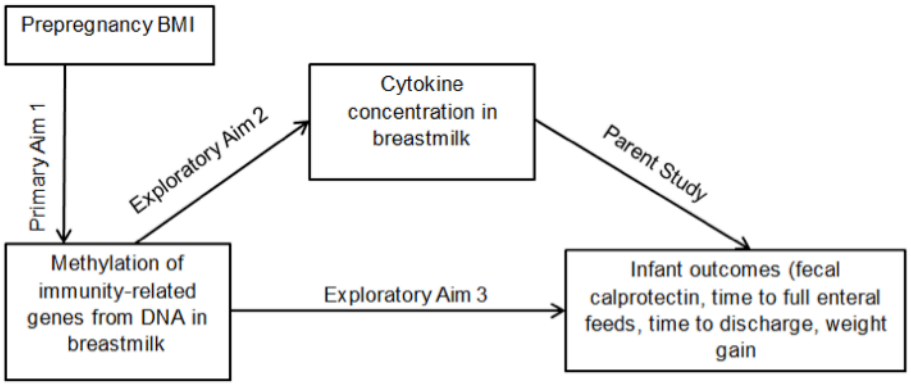

Figure 1: Conceptual Framework

1.2.6 Significance and Innovation

The concept that BMI is related to variability in breast milk composition is established, though the mechanism for this is not understood; therefore this Aim 1 will examine the impact of BMI on DNA methylation extracted from milk. The mechanism for breast milk variability that exists between normal and high BMI mothers has never been explored. Understanding the biological pathways in which BMI influences breast milk could lead to optimization strategies of human milk in the NICU. The potential impact of BMI on breast milk disparity could also inform preconception and prenatal education surrounding healthy weight gain during pregnancy. This

will be the first epigenetic study to evaluate the impact of maternal BMI on the extent of methylation from DNA in breast milk.

Exploratory Aim 2 will explore whether the extent of DNA methylation is correlated with the concentration of immunological components in breast milk known to be important to neonatal outcomes. Findings may provide evidence to support that breast milk is not a uniform substance and reveal mechanisms surrounding disparities that could be used for targets to optimize human milk fed to all infants in the NICU. While epigenetic studies have been applied to breast milk, they were examining methylation of cancer associated genes, and none have used preterm milk. Milk expressed by a woman who delivers preterm is compositionally very different than term milk, particularly among immunological components.

Milk differences between women who deliver preterm may explain some of the disparity seen among infants who develop necrotizing enterocolitis (NEC), despite receiving breast milk. Preterm infants born <1500 grams at birth are more susceptible to infections, particularly necrotizing enterocolitis, which costs an additional $6.5 million to treat per year (Bisquera, Cooper, & Berseth, 2002). Fituch (2004) uncovered a disparity in outcomes associated with IL-10 concentration in breastmilk fed to preterm infants. IL-10 was undetectable in the milk of women whose infants had NEC, while infants who did not develop NEC were fed milk with measurable IL-10 levels (Fituch et al., 2004). Exploratory Aim 3 will link extent of DNA methylation as a mechanism to explain variability in milk composition with a subsequent link to infant outcomes would aid the research community in their efforts to optimize feedings and hence outcomes of infants in the NICU. This study is the first to conduct methylation analyses of immunity genes from breast milk. Methylation analysis of three interleukin genes will provide

valuable information on the variability of protection offered through breast milk that is administered to preterm infants.

1.2.7 Summary

Despite evidence that breast milk is variable between women, this evidence appears to be ignored in clinical practice, as breast milk is administered uniformly. Additionally, the mechanism for breast milk variability is poorly understood, despite evidence that milk variability may contribute to infant outcomes. The proposed study challenges the clinical assumption that breast milk is a uniform substance by examining the variability of DNA methylation levels from human milk and linking it with maternal factors and infant outcomes. Methylation variability, which may contribute to breast milk differences between women, has never been analyzed with respect to infant outcomes. The ability to differentiate methylation profiles of milk that are more protective may aid in better understanding the mechanism for milk variability and subsequent infant outcomes. The knowledge gained from this project has a potential future application in milk optimization strategies for preterm infants that could prevent NICU complications and subsequently improve child health.

1.3 PRELIMINARY STUDIES

The following table lists several milestones that have been achieved since entrance into the BSN to PhD program in May 2009. All milestones listed support the feasibility and scientific merit of

the proposed dissertation project titled “Breast Milk is not a Uniform Substance: Epigenetic Mechanisms”.

Table 1: Milestones

Milestone	Date
Corrine M. Barnes Award	September 2010 September 2011
Sigma Theta Tau International Honor Society of Nursing, Eta Chapter Research Award	May 2012
University of Pittsburgh IRB Approval: pilot study “Epigenomics of Ductal Cells from Breastmilk” (expedited review, PRO11050673)	April 2012
University of Pittsburgh IRB Approval: “Breastmilk is not a Uniform Substance: Epigenetic Mechanisms” (exempt review, PRO13040181)	April 2013
Material Transfer Agreement with USF obtained	July 2013
Comprehensive Exam and Overview	July 2013
Judith A. Erlen Student Research Award	September 2013
International Society of Nurses in Genetics, Research Award	October 2013
Breast milk DNA extraction complete	May 2014

1.4 RESEARCH DESIGN AND METHODS

This ancillary epigenetic study will utilize a retrospective, observational design that focuses on methylation patterns of three cytokine genes (IL-4, IL-6 and IL-10) from DNA in breast milk expressed by mothers who deliver preterm infants. This proposed study will take the efficient approach of conducting an ancillary study to an ongoing clinically-relevant NINR supported project titled “The Association between Preterm Milk Immunobiology and Infant Health”

(referred to as the parent study). Dr. Maureen Groer is the PI of the parent project that will provide banked breast milk samples, demographic and clinical data, fecal calprotectin level measurements and milk cytokine concentrations. Permission to access these samples was granted by the PI of the parent study and IRB approval was obtained for this dissertation study. The timeline for this study is included below.

Table 2: Study timeline

ADJUSTED TIMETABLE FOR PROJECT:	Year 1	Year 2	Year 3	Year 4	Year 5
Coursework					
Summer Genetics Institute					
IRB approval – pilot study					
DNA extraction from breast milk (pilot study)					
IRB approval – dissertation					
DNA extraction from breast milk (dissertation)					
Collection of methylation data					
Analysis of data					
Preparation of manuscripts					
Defend dissertation					

1.4.1 Setting and Sample

"The Association between Preterm Milk Immunobiology and Infant Health", which was conducted in Tampa, FL, examined: 1) the relationship between milk immunobiology and infant

health with attention made to specific protein components in human milk that are most beneficial, and 2) the relationship of preterm infant outcomes to total volume of human milk received in the NICU. The breast milk collected by the parent study, the same milk that will be provided for this proposed study, is a pooled weekly aliquot of the milk ultimately fed to preterm infants over a six week NICU stay (up to six samples per infant). The inclusion criteria for the parent study require that women deliver their infant at the Tampa General Hospital. Infants must weigh <1500 grams at birth and be admitted to the NICU. Exclusion criteria include: mothers with HIV, infants with major congenital anomalies, and moribund infants. This proposed ancillary study does not have any additional inclusion or exclusion criteria.

1.4.1.1 Breast milk samples and DNA extraction

All breast milk samples, which are required to complete each aim, were/will be collected as part of the parent study. Breast milk aliquots from each feeding are collected for up to six weeks on infants who were born weighing less than 1500 grams. Breast milk aliquots from each feeding are pooled weekly. All milk is collected and stored frozen until brought to the laboratory twice each week. The milk is pooled each week for a total of six maximum number of breast milk samples per subject, centrifuged, defatted, and filtered, and the whey is frozen at -80 C. The whey portion from this previously frozen milk sample contains cell-free DNA, and this will be extracted using the QIAmp DNA extraction minikit from Qiagen Corp. Extracted DNA will be stored in 1X TE buffer at 4°C.

1.4.1.2 Data available for this project

Demographic data includes: education, income, race, marital status and employment status. Maternal information includes medical history, mental health history, obstetrical history, and

most recent pregnancy information including labor and delivery. Lifestyle data includes: smoking status, drug, and alcohol use. We will also have data related to the volume of breast milk received by each infant. Infant data includes: SNAP-2 (severity of illness) scores, gender, gestational age, ethnicity, Apgar scores, birthweight, length of stay, infections, NEC, blood transfusions, oxygen requirements, death, and time to full feeds. Table 3 (below) illustrates aim-specific data available.

Table 3: Aim-specific available data

Aim	Data available
1	Prepregnancy BMI
2	Cytokine concentrations using a Luminex platform
3	Fecal calprotectin, weight gain, time to full enteral feeds, and time to discharge

1.4.1.3 Methylation quality assessment and data collection

We will conduct bisfulfite conversion of the DNA followed by pyrosequencing of the promoter regions of IL-4, IL-6, and IL-10 genes to determine the methylation status of each gene. Epitect Bisulfite Kits (Qiagen Corp) will be used to convert unmethylated cytosines to uracils. This kit allows for conversion of previously frozen samples that have limited DNA available. Each sample (maximum of six pools of weekly collected breast milk per subject) will be assessed for each gene. Two internal controls will also be converted and used to normalize the data and assist in data interpretation. PyroMark PCR Kit (Qiagen Corp) will be used to generate the gene specific fragment for sequencing as well as perform the sequencing reaction for evaluation by pyrosequencing. Data will be double called by the student as well as a blinded technician in Dr.

Conley's laboratory, results compared, and discrepancies evaluated using raw data or repeating data collection. Samples with incomplete bisulfite conversion or poor sequencing success will not pass quality checks and will be eliminated from analyses.

1.4.2 Analysis

1.4.2.1 Sample size justification

Our primary aim (Aim 1) will be sufficiently powered. Our sample size of 100 achieves 80% power to detect a small to moderate effect size of 0.28 with a significance level of 0.05 for this aim. Our other aims are exploratory, and therefore may not be adequately powered, but will provide valuable pilot data to inform future studies.

1.4.2.2 Preliminary analysis

Univariate outliers will be assessed three ways: 1) frequency table evaluation (categorical/ordinal variables) and 2) graphical methods including: histograms, box plots, normal probability plots, detrended normal probability plots (continuous variables) and 3) z-scores will be evaluated for large standardized values (continuous variables). Potential multivariate outliers will be assessed three ways: 1) scatterplots, 2) calculating a critical value for Mahalanobis Distance and 3) leverage values >0.05.

Should potential confounders be identified during the preliminary analysis, they will be considered when addressing all aims. Confounders that will be considered include but are not limited to: maternal age, maternal infection, gestational age at delivery and maternal smoking/drug use.

Missing data will be assessed for both amount (percentage) and pattern. The pattern of missing data will be assessed to determine if missing data is random or nonrandom. At each time point (1-6 weeks), we will generate missing value indicators; if breast milk data is missing, we will dummy code it as "1" (otherwise it will be coded as 0). We will develop a contingency table to describe the amount of missing data across time by subject ID. SPSS Missing Values Analysis (MVA) will be used to highlight patterns of missing data. To test whether missingness is random, t-tests will be performed on the variable(s) of interest with percent of missing data. If only a few cases are missing, and they are missing at random, we will consider deleting the case(s). Another option, multiple imputation, does not assume randomness of missing data. If the missing data is not randomly distributed, we will consider multiple imputation to estimate missing data. PROC TRAJ model in SAS will be used with missing independent and dependent variable values at one or more follow-up time points. PROC TRAJ requires that missing data be missing completely at random (Arrandale, 2006) therefore, if there is a pattern to the missingness we will implement multiple imputation.

Normality of the distribution for each variable will be assessed graphically and statistically. The censored normal model will be used to analyze the trajectory groups for methylation, interleukin concentration, and outcomes. Values of skewness and kurtosis (peak) will be determined, and values of both should be close to zero and tests should be nonsignificant. The distribution will also be evaluated graphically at each time point, since underestimates of variance associated with positive kurtosis disappear with samples of $\geq$ 100 samples. IBM SPSS FREQUENCIES will be used to evaluate skewness and kurtosis. Frequency histograms will be used to graphically assess that the normality assumption is met. Additionally, normal probability plots and detrended expected normal probability plots will be evaluated to compare deviations

from normality. These plots are available in the IBM SPSS MANOVA. Residual plots will also be evaluated for normality and should be independently distributed. This diagnostic technique is available through IBM SPSS REGRESSION. If the normality assumption is met among the residuals, then both the normal probability plot and detrended normal probability plot will look the same. Lastly, the Shapiro-Wilks test of normality will be used to statistically examine normality. If the p-value is >0.05, we will conclude that the data is normally distributed.

Linearity will be assessed by looking at the residual plots of the dependent variables for each aim. Nonlinearity will be diagnosed when most of the residuals are above the zero line at some values and below the zero line at other predicted value. Ideally, the scatterplot will be oval-shaped, which would indicate normal distribution and linearity. Bivariate scatterplots are available through the IBM SPSS GRAPH function. These scatterplots will also be used to assess independence. Ungrouped data will be used in this analysis; therefore, homoscedasticity of the data should show variability in scores that is approximately the same at all values of another variable. Homoscedasticity will be screened using bivariate scatterplots between the continuous variables.

Data transformations will be considered as a remedial measure in the following circumstances: 1) non-normality, 2) nonlinearity, 3) heteroscedasticity. Additionally, transformations with higher order effects will be considered (nonlinearity), as will weighted transformations (heteroscedasticity). Transformations will be checked after application to ensure that the appropriate transformation was applied.

1.4.2.3 Analyses

Trajectory modeling will be used to examine all aims. This approach is based on a semiparametric, group-based modeling strategy that is a mixture of probability distributions

specified to describe the data (Jones, 2001). Trajectories of methylation changes to DNA in breast milk during the first six weeks postpartum will be examined using the PROC TRAJ procedure in SAS. This procedure allows for estimation of multiple groups within a population, as opposed to a traditional regression that models only one mean within the population (Arrandale, 2006). Ultimately, PROC TRAJ allows for identification of distinct subgroup memberships within a population and estimates a regression model for each identified subgroup. Using contingency tables and chi-square test of independence, we will test if the independent variable is associated with the dependent variable for each aim. The table below outlines how each aim's independent and dependent variables will be subjected to trajectory analyses:

Table 4: Aim specific analyses

Aim	Independent Variable(s)	Dependent Variables
Determine if maternal BMI influences extent of methylation of immunity genes	Maternal BMI	Methylation trajectory group for each gene (IL-4, IL-6, IL-10)
Describe the relationship between extent of methylation of IL-4, IL-6 and IL-10 genes and interleukin concentration in breast milk	Methylation trajectory group of each gene (IL-4, IL-6, IL-10)	Milk concentration of IL-4, IL-6 and IL-10
Explore if methylation of immunity genes is associated with infant outcomes	Methylation trajectory group of each gene (IL-4, IL-6, IL-10)	NICU outcomes and fecal calprotectin

As suggested by Nagin (2005), we will set all group orders to second order when fitting the maximum number of groups to our a priori group number estimate (Nagin, 2005). We will decide on the maximum number of trajectory groups based on prior knowledge. A Bayesian Information Criteria (BIC) will be compared between the models to determine the appropriate number of groups and trajectory weights. The change in BIC scores between the two models is a

measure of the evidence against the null model, which is always simpler (Arrandale, 2006). The best fit model will be the one with the lowest BIC score.

After the number of groups has been determined, trajectory shapes for each group will be evaluated in a step-wise manner. PROC TRAJ can model up to a fourth order polynomial of both linear and non-linear trajectories (Arrandale, 2006). We will evaluate shapes by comparing the BIC change, since prior knowledge is limited and all patients in this study are expressing preterm milk.

We will generate graphic displays of the fitted model groups from the estimated group membership probabilities using the TRAJ PLOT command in SAS. We also intend to use the average posterior probabilities to explore between group differences in covariates not in the model. The posterior probability values measure the probability that a subject with a specific methylation profile belongs to a specific trajectory group. Individuals are assigned to a group based on their highest posterior group probability.

Output from the PROC TRAJ command includes: 1) group parameter estimates, 2) group membership probabilities, and 3) model fit statistics. Parameter estimates will be used to construct group regression equations and a system of equations to describe the population. We will be able to make inferences about group differences by using the relative difference between estimates for the same covariates.

1.5 POTENTIAL LIMITATIONS OF PROPOSED PROCEDURES AND ALTERNATIVE APPROACHES TO ACHIEVE THESE AIMS

While conducting this ancillary study is efficient and cost conscience, using data and samples from a parent study poses some limitations: 1) In the event that we fail to find distinct methylation patterns over time and/or at least 5% of the sample is not represented in each trajectory group, we will use traditional linear regression analyses with a time-dependent covariate to examine associations. Residual analysis will be performed for all models to identify model misspecifications or influential observations; 2) Milk samples are pooled weekly, and we will not be able to distinguish any daily and diurnal immunological changes in milk. Breast milk gene expression changes throughout the day (Maningat et al., 2009), and milk composition varies throughout the lactation period. We do not believe that this will compromise our ability to answer our research questions, since the pooled breastmilk samples will reflect the weekly immunological exposure the infant has received and these are the same pooled samples used for the parent study that identified differences in cytokine protein levels; 3) Maternal BMI for the parent study is self-reported, and this could introduce increased error. We will consider grouping subjects into BMI category (underweight, normal, overweight, obese), as this may be more reliable than using BMI as a continuous variable. An additional limitation is that BMI for this project will reflect pre-pregnancy BMI that is available through the parent study. Design of future studies will benefit from these data but will also attempt to collect BMI across the data collection timepoints; 4) The parent study does not exclude examination of donor milk samples, which is overwhelmingly term milk (Dempsey & Miletin, 2010), and pooled from multiple donors. This may influence methylation results, since preterm breast milk is very different from term breast milk, particularly among immunological components. Currently, of the 65 recruited

subjects, only eight mothers have required donor milk. For this ancillary study we may exclude evaluation of the donor samples or, if enough are available, we may analyze them separately; 5) Fecal calprotectin, while a common biomarker of infant gut inflammation, may not provide a global view of infant outcomes. The infants in this study are receiving breast milk, and are therefore receiving protection from common neonatal complications. With a sample size of 100, it is unlikely that we will have enough "sick" babies to test significantly (it is noteworthy to mention that there has been only one NEC case, to date). To further address infant outcomes we are including weight gain, time to full enteral feeds, and time to discharge as additional infant outcomes. 6) The whey portion of milk is being used from an existing study. This milk, which was frozen, will contain DNA from cells that lysed in the freezing process. It will therefore be impossible to determine the cell type that contributed the DNA for this study. This is not a big limitation as the DNA evaluated in this study will represent the combined DNA from the sample; however, future studies that allow for the possibility of isolating cells prior to DNA extraction would add value to the interpretation of our findings.

1.6 HAZARDOUS MATERIALS AND PROCEDURES

Although human milk is not a biohazard, the student will be exposed to bodily fluids and universal precautions will be implemented at all times when working with breastmilk. The student has received blood borne pathogen training and will complete chemical hygiene training prior to initiating laboratory experiments. All experiments will be conducted in an appropriately equipped laboratory.

1.7 RESEARCH PARTICIPANT RISK AND PROTECTION

1.7.1 Human subjects

IRB approval has been obtained for the proposed study (IRB PRO11050673). The specimens that will be utilized for the proposed study have been collected, or are currently being collected, for the ongoing trial Breast milk and the Health of Infants Study. No further involvement from the participants will occur under the proposed research study.

Inclusion criteria for the parent study require the live birth of an infant who is admitted to the NICU at Tampa General Hospital and who weighs less than 1500 grams at birth. Mothers with HIV, and infants with major congenital anomalies, and moribund infants are excluded. The parent study collects 0.5 mL of breast milk from each feeding, which is aliquoted and pooled weekly. The proposed study will investigate DNA methylation from cells in breast milk collected as part of this study.

1.7.2 Sources of materials

All of the DNA samples (extracted from breast milk) are available from the “Breastmilk and the Health of Infants” study. In addition, clinical data and protein levels from the study are housed in secure databases. Neonatal outcome data, including fecal calprotectin levels, are collected or measured as part of the parent study and are available to the student. The parent study’s IRB has provided permission to share breast milk samples and data to conduct the proposed study. The genetic and clinical data that is obtained will be used solely for research purposes and will be continuously safeguarded by the student and her advisor. For the proposed study, the banked

breast milk samples obtained from the “Breast milk and the Health of Infants” study will be used to collect DNA methylation data.

1.7.3 Potential risks, benefits, and protection against risks

All subjects to be included in the proposed research study have consented to the utilization of their breast milk for research purposes. In addition, all data generated form the proposed study would be reported as aggregate data and the results will not be revealed to participants. Furthermore, data generated from the proposed research will be stored in a database secured on a password-protected computer. While breach of confidentiality is of great concern with genetic studies, the above precautions mitigate this risk. In reference to potential benefits, there is no direct benefit to the research subjects; however the results generated may reveal a predisposition or protection from the development of NICU complications, which could ultimately direct future scientific inquiry and impact clinical practice by suggesting novel interventions and therapies.

1.7.4 Data and safety monitoring plan

This is not a clinical trial; however we do have a plan to monitor data collection and protection. All data obtained will be used for research purposes only and will be safeguarded by the student and her advisor. Data collection and analyses will be performed using a unique numerical identifier assigned to each specimen from the “Breast milk and the Health of Infants” study. Data collected will be entered using these unique numerical identifiers into a password-protected computer. Regular meetings will allow for the discussion of project progress and data monitoring

to assure that data collection and the analysis are conducted in a manner that maintains the anonymity of the samples and data.

2.0 SUMMARY OF STUDY

The purpose of this dissertation research was to examine: 1) the relationship between maternal BMI and methylation of interleukin genes, 2) whether methylation drives interleukin concentration, and 3) if methylation of interleukin genes is associated with neonatal outcomes. One article written during the course of this dissertation is provided in Appendix F. Published in *Breastfeeding Medicine*, this article highlights the use of breast milk in genetic/genomic studies. All of the studies included in this review extracted maternal DNA or RNA from breast milk, and focused on the uses of breast milk for genetic studies. This article highlights that few studies have examined breast milk with an epigenetic approach, and the few that have do not explore milk properties with respect to infant outcomes.

2.1 PROPOSAL CHANGES

Several changes were made to the approved dissertation proposal. Described below are the steps taken to assess primer validity, DNA integrity, and modifications to the proposal to reflect the limited utility of the DNA. Specific modifications and the rationale for these changes are highlighted below.

2.1.1 MOLECULAR APPROACH TO EXAMINE INTERLEUKIN VARIABILITY

CpG islands of interest were originally identified in the literature for IL4 (Kwon, Kim, Lee, Oh, & Choi, 2008), IL6 (Nile, Read, Akil, Duff, & Wilson, 2008), and IL10 (Fu et al., 2011). Primers were designed for use with the PyroMark CpG Assays for methylation array validation (Qiagen). DNA extraction was performed on the whey portion of preterm breast milk using the QIAGEN QIAamp® DNA Mini Kits (QIAGEN, Valencia, CA). To assess DNA yield, both TaqMan® allelic discrimination (Applied Biosystems Inc., Carlsbad, CA) and PCR was used. Following confirmed DNA extraction, DNA was converted using the Epitect Bisulfite Kit (QIAGEN, Valencia, CA). Following bisulfite conversion, the three representative IL promoter regions were subjected to amplification using Pyromark PCR kits. *Less than 10% of the samples were successfully amplified for pyrosequencing.* Described below are the steps taken to troubleshoot the amplification for each of the three targeted areas:

1. Breast milk DNA concentration is variable, so DNA volume in the PCR master mix was added at variable amounts (from 1.0 µL to 2.2 µL).

2. Initial PCR amplification using the recommended annealing temperatures yielded no PCR product. We conducted a temperature gradient for each gene, and discovered that annealing temperatures required for amplification were variable: IL4 at 49°C, IL6 at 60°C, and IL10 at 56°C. Unfortunately, we were unable to obtain at least 90% PCR product for each plate, despite using modified annealing temperatures.

3. We performed volume gradients on the PCR products when running the gel (1 µL to 14 µL), but this did not yield a noticeable difference.

4. The number of cycles was increased from 45 to 60, and this provided more PCR products. We used 60 cycles for all subsequent PCR reactions.

Despite these PCR protocol modifications, we were unable to obtain any more than 10% success per run. The issue was not with the bisulfite conversions or the CpG island amplifications with the pyromark assays given that the lab control DNAs and methylated/unmethylated controls were successfully amplified.

2.1.2 CIDR amplifications

The DNA extracted from breast milk samples was sent to the Center for Inherited Disease Research (CIDR) (Johns Hopkins University, Baltimore, MD). CIDR was to do the pyrosequencing after we had completed the bisulfite conversions and amplifications of the CpG islands; however because we were unable to accomplish this we decided to have CIDR give the entire process a try and sent them unmodified DNA samples to work with. CIDR had the exact same experience that we had in-house. They were able to bisulfite convert and amplify a subset of samples but the majority of them were not amplifiable. CIDR then further assessed the DNA integrity by subjecting thirty-nine samples that represented different subjects' milk to gDNA analysis using a fragment analyzer. None of the samples subjected to the fragment analyzer with peaks (1K-2K range, 4-14 ng/uL) amplified, despite successful amplification of the methylated controls. The fact that our in-house attempts failed, CIDRs attempts failed, and the DNA was noted to not be of high enough quality to proceed with evaluation of methylation status of these genes, we decided to attempt a polymorphism based approach to assess the variability in these genes.

2.1.3 Polymorphism based assessments of the IL4, IL6, and IL10 genes

Fortunately the DNA extracted from whey was of high enough quality to allow for polymorphism based data collection using TaqMan allelic discrimination for data collection. This meant that the genes could be assessed for this project, just not from a DNA methylation point of view, but from a polymorphism point of views. Data collection was successfully conducted on the same breast milk samples to genotype seven functional promoter SNPs of IL4, IL6, and IL10. It is well documented that interleukin variability is at least partially attributed to SNPs (Nguyen et al., 2004; Qaddourah et al., 2014; Velez, Fortunato, Williams, & Menon, 2008); however, this relationship has never been explored in breast milk. It was decided that functional SNPs in the promoter regions of Il4, IL6, and IL10 were another reasonable way to examine variable interleukin concentrations in breast milk.

Genotype data was collected at the University of Pittsburgh School of Nursing using TaqMan allele discrimination assays to genotype the seven functional promoter polymorphisms of IL4 (rs2070874, rs2243250), IL6 (rs1800795, rs1800796), and IL10 (rs1800871, rs1800872, rs1800896). We performed TaqMan allelic discrimination with the ABI Prism 7000 Sequence Detection System and SDS software v1.2.3 (Applied Biosystems Inc., Carlsbad, CA). Negative controls were included and a portion of the samples were repeated to confirm that they repeatedly discriminated into the same genotype. The following cycling conditions were used: 1) Activation at 95°C for 10 minutes, 2) denaturation at 95°C for 15 seconds, 3) anneal/extend at 58°C for 1:30 minutes, 4) go to step 2 50 times, and 5) hold at 10°C forever. We were able to successfully amplify and genotype 100% of the breast milk samples (n=63).

2.1.4 Specific Aims

Due to the tissue available (whey milk), and the scientific reasoning behind examining IL SNPs and their relationship to milk concentrations, it was decided that we would explore how maternal SNPs impact milk IL concentration, and subsequent infant outcomes. Trajectory analyses would remain, but at a protein level. The BMI aim was excluded because genotypes are not impacted by lifestyle factors; however, given the impact of BMI on milk composition, this was included as a covariate. A new set of Specific Aims were constructed to reflect a genotype approach rather than a methylation analysis:

Primary Aim 1: Examine the relationship between maternal IL SNPs and cross-sectional breast milk concentrations. Functional SNPs in the promoter region of IL4, IL6, and IL10 will be genotyped and we will determine if they predict milk interleukin concentration during the first three weeks postpartum.

Specific Aim 2: Describe the trajectories of breast milk IL concentration change over time. Trajectory analysis will be used to examine IL4, IL6, and IL10 milk concentration changes over the first three weeks postpartum.

Exploratory Specific Aim 3: Explore whether maternal IL genotypes predict milk IL trajectory groups. Data from Aim 2 will be used in a univariate analysis to explore whether maternal IL genotype is associated with IL changes.

Specific Aim 4: Examine if cross-sectional IL levels and/or IL trajectories predict infant outcomes. IL levels from significant SNP/IL associations will be examined for relationships to infant outcomes, including: SNAPPEII scores, length of stay, weight at 6 weeks, days on oxygen, fecal calprotectin levels, necrotizing enterocolitis, sepsis, retinopathy of prematurity, bronchopulmonary dysplasia, feeding intolerance, and blood transfusions.

<u>Exploratory Aim 5:</u> Explore if there is a relationship between maternal IL SNPs and infant outcomes. Controlling for gestational age at delivery and ratio of human milk to total milk, the association between maternal IL SNPs and infant outcomes (described above) will be analyzed.

2.1.5 Discussion

At the beginning of the study, we did not anticipate poor DNA quality. We assumed that DNA from lysed cells would be available for pyrosequencing. It seems that the available DNA in breast milk whey is not appropriate for examining methylation using pyrosequencing. Despite this, there is enough high quality DNA to successfully genotype and this was done in all of the available breast milk samples. The specific aims of this dissertation were changed to reflect the examination of IL genotypes rather than IL methylation.

2.2 STUDY STRENGTHS AND LIMITATIONS

There were several limitations associated with this dissertation project. The sample size was small, and this was further decreased when we performed race-specific analyses. Trajectory analysis includes, ideally, 100 subjects with at least three time points, and we did not reach this recommended sample size. Hardy-Weinburg Equilibrium was violated for three of the SNPs evaluated; however, many of the SNPs investigated are associated with a variety of obstetrical outcomes so we likely enriched for the alleles of interest because our sample included a group of high-risk women who delivered preterm infants. Additionally, we did not examine donor milk

interleukin levels which could influence the amount of ILs each infant was exposed to. Lastly, we did not measure infant serum IL levels which would have allowed us to also consider endogenous IL exposure.

There were also strengths associated with this study. This is the first study to examine the influence of maternal SNPs on milk IL levels, which is highly innovative. This was a heterogeneous sample of obstetrically high risk mothers, and the collection of preterm breast milk in this population is difficult. While the sample size was small, there were consistent findings that inform current knowledge related to mechanisms for variability, and how milk immunological profiles might influence NICU outcomes.

2.3 FUTURE STUDIES AND IMPLICATIONS FOR NURSING PRACTICE

Future studies should include a larger sample size, with at least 100 mother-infant dyads so as to perform an adequately powered trajectory analysis. More studies are needed to confirm the relationships observed, particularly among different ethnicities. Because breast milk composition is heavily influenced by lifestyle factors, future studies should examine how epigenetic influences impact milk composition. These studies should collect whole, fresh breast milk that would allow for pyrosequencing.

The administration of breast milk, including donor milk, is done by the bedside nurse. Unfortunately, breast milk is treated as unchanging in the NICU, despite evidence that it is variable between women. This study further informs the knowledge related to the dynamic nature of breast milk, and this is important to nursing because milk differences influence infant outcomes. Nearly 70% of mothers who deliver preterm are unable to provide MOM to their

infants. As a result, donor breast milk has become the standard of care in NICUs. This study has potential clinical application to personalized medicine, where donor breast milk can be screened for SNPs that are associated with high or low interleukin levels, and breast milk can be matched based on infant needs and/or risks related to NICU complications.

3.0 DATA-BASED MANUSCRIPT: THE IMPACT OF PROMOTER POLYMORPHISMS ON CYTOKINE CONCENTRATION IN PRETERM BREAST MILK AND SUBSEQUENT INFANT OUTCOMES

3.1 ABSTRACT

Background: Preterm infants are at risk for complications, and breast milk protects against many complications. Breast milk is variable between women and interleukin (IL) differences are associated with infant outcomes. The molecular mechanism for milk variability remains unknown.

Objective: The aims of this ancillary study were to: 1) examine the relationship between maternal IL genotypes and weekly milk concentrations of IL4, IL6, and IL10, 2) describe the trajectories of milk IL change over the first three weeks postpartum, 3) examine whether maternal IL genotypes predict milk IL trajectories, 4) examine if weekly IL levels and/or IL trajectories predict infant outcomes, and 5) explore a relationship between maternal IL genotypes and infant outcomes.

Methods: Preterm breast milk aliquots (0.5 mL) were collected from each feeding of mom's own milk and pooled weekly for three weeks. DNA was extracted from the whey portion of breast milk using QIAmp DNA Extraction MiniKit and genotyped with TaqMan. Milk IL concentrations were measured using MagPix and Millipore kits. Trajectory analysis examined milk change over time.

Results: Multivariate analysis resulted in associations between IL6 and IL10 SNPs and subsequent IL6 and IL10 milk levels. Infant outcomes associated with varying IL milk levels included calprotectin and IVH. Trajectory analysis resulted in linear group shapes, with two distinct subgroups in IL6, and three subgroups in both IL4 and IL10. Trajectory groups were associated with calprotectin, IVH, and blood transfusions. There were also significant relationships between maternal IL genotypes and NICU outcomes.

Conclusions: Maternal IL SNPs are associated with IL breast milk levels and IL milk levels are associated with infant outcomes.

3.2 BACKGROUND

The immune protection offered through breast milk is perhaps the original function of the mammary gland (Vorbach, Capecchi, & Penninger, 2006). The impact of gestational age of delivery on breast milk composition (Gidrewicz & Fenton, 2014) and the protection provided to preterm infants who receive Mother's Own Milk (MOM) (Corpeleijn et al., 2012; Vohr et al., 2007) suggest that preterm infants may benefit from breast milk that is immunologically appropriate for their needs. MOM provides preterm infants with immunofactors that are at levels appropriate for their development. Preterm birth complications are the leading cause of death among children under the age of five (World Health Organization, 2014). Preterm infants face increased risks of: pneumonia, retinopathy, necrotizing enterocolitis (NEC), and sepsis. Breast milk provides protection against these complications (Barsam et al., 2013; Corpeleijn et al., 2012; Cristofalo et al., 2013; Maayan-Metzger, Avivi, Schushan-Eisen, & Kuint, 2012; Manzoni et al., 2013; Schanler, 2005), and for this reason, the American Academy of Pediatrics (AAP) recommends that all infants, particularly those weighing less than 1500 grams at birth, receive human milk (American Academy of Pediatrics, 2012). The robust immunological profile of breast milk, which contains white blood cells, cytokines, and immunoglobulins, may contribute to this protective influence on vulnerable infants.

Preterm infants are no longer receiving maternal immunological protection *in utero*; therefore, it is essential that they receive passive immunity through breast milk. ILs are an

integral part of the inflammatory response, and the preterm infant is vulnerable to complications due, in part, to an underdeveloped immunological system. ILs are present in breast milk, though their concentration is highly variable between women (Lawrence & Lawrence, 2005). Interleukin-4 (IL4) is involved in adaptive immunity and acts as both a pro- and anti-inflammatory cytokine. Breast milk from mothers with allergies has a higher IL4 concentration when compared to breast milk from mothers who do not report allergies (Bottcher, Jenmalm, Garofalo, & Bjorksten, 2000; Marek et al., 2009). IL4 variability is relevant to the preterm population because it induces both antibody and IgE production. IL4 also contributes to macrophage activation, which results in microbial phagocytosis. Varying levels of IL4 in breast milk may contribute to disparate outcomes. To date, no studies have examined the relationship between maternal SNPs, subsequent milk composition, and resulting infant outcomes; however, there is evidence that varying IL4 in breast milk may contribute to the development of allergic dermatitis in healthy term infants (Ochiai et al., 2013).

Interleukin-6 (IL6), a pleiotropic cytokine, is involved in both adaptive and innate immune responses. IL6 stimulates both neutrophil production and proteins by hepatocytes to help with acute-phase responses. IL6 also promotes the growth of monoclonal antibodies. There are a number of factors associated with varying IL6 concentration in breast milk, including: mastitis (Mizuno et al., 2012), preeclampsia (Erbagci et al., 2005), cesarean section delivery (Mehta & Petrova, 2011), and maternal smoking (Etem Piskin et al., 2012). IL6 concentration is relevant to the preterm population because this interleukin is poorly regulated in preterm infants (Currie et al., 2011); therefore it is essential that vulnerable infants receive adequate amounts of IL6 through breast milk. Variable IL6 in breast milk has been related to subsequent outcomes,

with higher levels associated with decreased infant weight gain, percent fat, and fat mass among healthy breastfed infants (Fields & Demerath, 2012).

Interleukin-10 (IL10) is produced by many cells of both the adaptive and innate immune systems. The role of IL10 in the prevention of inflammatory pathologies, as well as its function as a feedback regulator (Saraiva & O'Garra, 2010), illustrates its potential impact on preterm infants who are vulnerable to infection. IL10 concentration in breast milk is higher at three months postpartum among mothers who report allergies (Prokesova et al., 2006). Variable levels of IL10 in breast milk have been implicated in disparate neonatal outcomes. Milk fed to infants who developed NEC had immeasurable levels of IL10 when compared with milk fed to infants who did not develop NEC (Fituch et al., 2004). In another study, higher IL10 milk levels were associated with neonatal jaundice (Zanardo et al., 2007).

Potential mechanisms for IL breast milk differences include variability in the DNA coding for these ILs, for example, single nucleotide polymorphisms (SNPs), particularly in regulatory regions of the genes. Serum concentrations of IL4 have been associated with two SNPs at the gene's promoter region in pregnant women (rs2244350 -589 T/C) (Nguyen et al., 2004), and children with malaria or anemia with the C allele (rs2070874 -33 T/C) produce higher levels of IL4 (Cabantous et al., 2009). Patients in septic shock experience higher IL6 if they have the G allele (rs1800795 -174 C/G), and Malarstig demonstrated that subjects at risk for a cardiovascular event had lower IL6 levels if they had the G allele for another SNP for the same IL6 gene (rs1800796 -572 C/G) (Malarstig, Lindahl, Wallentin, & Siegbahn, 2006). Women with idiopathic recurrent miscarriage had reduced IL10 production with the minor alleles at two SNPs in the promoter region of IL10 (rs1800871 -819 C/T; rs1800872 -592 A/C) (Qaddourah et al., 2014). Lowe had similar results in one of these SNPs (rs1800872), with the A allele resulting in

lower IL10 production (Lowe, Galley, Abdel-Fattah, & Webster, 2003). In a third SNP in the IL10 promoter region (rs1800896 -1082 G/A), the A allele is also associated with low production in healthy individuals (Yilmaz, Yentur, & Saruhan-Direskeneli, 2005), and preterm infants with RDS (Capasso et al., 2007).

A relationship between IL concentration and SNPs has been established in serum. Additionally, amniotic fluid shows variable IL6 concentration associated with a haplotype containing rs1800795 and rs1800796 (Velez et al., 2008). IL10 concentration is also variable in the amniotic fluid of women who experienced term and preterm birth, and rs1800896 is associated with this relationship. The relationship between these SNPs and breast milk IL concentration has never been explored, despite evidence that IL concentration is variable between women and linked to infant outcomes. If maternal SNPs do impact breast milk composition, which influences NICU outcomes, it is reasonable to explore a direct relationship between maternal IL SNPs and subsequent infant outcomes. The aims of this study were to, over the first three weeks postpartum and in a population who delivered preterm: 1) examine the relationship between maternal IL genotypes and cross-sectional (weekly) breast milk concentrations of IL4 (rs2243250 -589 T/C, rs2070874 -33 T/C), IL6 (rs1800795 -174 C/G, rs1800796 -572 G/C), and IL10 (rs1800871 -819 C/T, rs1800872 -592 A/C, rs1800896 -1082 G/A), 2) describe the trajectories of breast milk IL concentration change over time, 3) examine whether maternal IL genotypes are associated with breast milk IL trajectories, 4) examine if weekly IL levels and/or IL trajectories predict infant outcomes, and 5) explore a relationship between maternal IL genotypes and infant outcomes.

3.3 MATERIALS AND METHODS

3.3.1 STUDY POPULATION

This ancillary study included women (n=64) who delivered infants (n=73, including multiples) with a birth weight <1500 grams and delivered at Tampa General Hospital (Tampa, FL). We were able to collect genotype data from DNA in prospectively collected breast milk samples (n=192) over the first three weeks postpartum. The parent study, *The Association between Preterm Milk Immunobiology and Infant Health Study* (NINR, R21 NR01309401A1), was conducted at the University of South Florida, and investigated the relationship between milk immunity and milk volume with clinical outcomes in preterm infants. Mothers with HIV, infants with major congenital anomalies, and moribund infants were excluded from enrollment. All aspects of the parent study were approved by the University of South Florida's Institutional Review Board (IRB). Separate IRB approval from the University of Pittsburgh was also obtained for this study, which added genomic data collection to the parent project.

The following variables were collected and available for analyses: maternal age, parity, income, education, ethnicity, race, marital status, working status, and pregnancy history. Medical record data provided information about the labor and delivery of the infant(s). Maternal BMI was self-reported prepregnancy weight (pounds), and the height was obtained from the subjects' medical chart (inches). The following CDC recommended equation was used to obtain a BMI for each subject: weight (pounds) / [height (inches)] ^{2} x 703. Infant data was obtained from the NICU medical record, and included: gender, ethnicity, gestational age at birth, birth weight, APGAR scores, ratio of human milk to total milk, SNAPPE-II scores (severity of illness), length of stay, weight gain at 6 weeks, days on oxygen, sepsis, retinopathy of prematurity (ROP),

necrotizing enterocolitis (NEC), intraventricular hemorrhage (IVH), blood transfusions, and feeding intolerance.

3.3.2 Breast milk collection and whey separation

Any volume and source of milk the infant received (including MOM, donor breast milk, and formula) was recorded from the infant's medical record. Breast milk aliquots from each feeding were collected for up to three weeks and these aliquots were pooled weekly (n=192). All milk was collected and stored frozen until brought to the laboratory twice weekly. The pooled MOM was centrifuged, defatted, and filtered, and the whey was frozen at -80 C. Although infants received both donor milk and MOM, only MOM was examined and genotyped for this study.

3.3.3 Interleukin measurement

IL4, IL6, and IL10 concentration was measured using a bead based assay on a MagPix instrument (Luminex, Austin, TX), and Millipore kits (Emd Millipore, Billerica, MA), and is detailed in previous work (M. Groer, Kane, B., Williams, N. , 2013). IL concentrations were measured weekly in all pooled samples of MOM for the first three weeks. Each assay included a standard curve and quality controls, and all samples were done in duplicate.

3.3.4 Fecal calprotectin

Fecal calprotectin has been used as a biomarker of inflammation within the preterm population, as calprotectin is an accurate indicator of neutrophil migration toward the GI tract (Kapel et al.,

2005). Calprotectin levels were measured in a weekly stool sample using the PhiCal™ Fecal Calprotectin Immunoassay (Geneva Diagnostics, 2006).

3.3.5 Genotyping

Genotype data was collected using TaqMan allele discrimination assays to genotype seven functional promoter polymorphisms of IL4 (rs2070874, rs2243250), IL6 (rs1800795, rs1800796), and IL10 (rs1800871, rs1800872, rs1800896). We performed TaqMan allelic discrimination with the ABI Prism 7000 Sequence Detection System and SDS software v1.2.3 (Applied Biosystems Inc., Carlsbad, CA). Negative controls were included and a portion of the samples were repeated to confirm that they repeatedly discriminated into the same genotype. We also included duplicates, performed independent blinded double calls, and discrepancies were regenotyped. The following cycling conditions were used: 1) Activation at 95°C for 10 minutes, 2) denaturation at 95°C for 15 seconds, 3) anneal/extend at 58°C for 1:30 minutes, 4) go to step 2-3 50 times, and 5) hold at 10°C forever. Blinded raw data was reexamined of the SNPs for which Hardy-Weinburg Equilibrium (HWE) was violated to rule out genotyping error.

3.4 STATISTICAL ANALYSIS

3.4.1 Preliminary analysis

All statistical analyses were performed using SAS (v. 9.4). Univariate outliers were assessed using frequency tables and graphical methods including histograms and normal probability plots.

Multivariate outliers were assessed using scatterplots. Missing data was assessed for both amount (percentage), and pattern (random versus nonrandom). In addition to the Shapiro-Wilks test, normality was also evaluated graphically and at each time point with: frequency histograms and normal probability plots. Linearity, independence, and homoscedasticity were assessed by evaluating bivariate scatterplots. Data transformations were performed when a regression assumption (normality, linearity, homoscedasticity) was compromised. Transformations were checked after each application to ensure that the appropriate transformation was applied. There were eight sets of multiples, including one set of triplets. Milk-specific analyses excluded one twin, or two triplets, removed randomly, to ensure that each mother was represented only once (n=64). For infant-specific aims, all infants were included in the analyses (n=73).

3.4.2 Trajectory Modeling

Trajectory modeling with the censored normal model was used to examine changes in breast milk IL levels over the first three weeks postpartum. All modeling was done using the PROC TRAJ procedure in SAS (v 9.4). This approach is based on a semiparametric, group-based modeling strategy that is a mixture of probability distributions specified to describe the data (Jones, 2001). This procedure allows for estimation of multiple, distinct groups, within a population, as opposed to a traditional regression or growth curve that models only one mean within the population (Arrandale, 2006). Ultimately, PROC TRAJ allows for identification of distinct subgroup memberships within a population and estimates a regression model for each identified subgroup.

When determining the number of trajectory groups, all group orders were set to second order when fitting the maximum number of groups (Nagin, 2005). After the number of groups

was determined, trajectory shapes were evaluated in a step-wise manner, up to a second order polynomial, as only three time points are represented. Subjects were assigned to a trajectory group based on their highest posterior group probability, which measures the probability that a subject with a specific milk IL concentration profile belongs to a definite trajectory group. We compared Bayesian Information Criteria (BIC) between the models to determine the appropriate number of groups and trajectory weights. The best fit model was the one with the lowest BIC score.

3.4.3 Univariate analysis

Univariate analyses were first performed for each association and any relationship with a p-value ≤ 0.20 was subjected to multivariate regression models, where $p \leq 0.10$ was considered significant due to: 1) the exploratory nature of a pilot study, 2) small sample size, and 3) race-specific subgroup analyses, which further decreased our sample size. We also considered relationships with $p \leq 0.150$ as trending toward significance. SNP-specific analyses included both maternal genotype and minor allele presence (yes/no). The Fisher's Exact test was used to examine the relationships between: 1) SNP and categorical infant outcomes, 2) SNP and IL trajectory group, and 3) IL trajectory group and categorical infant outcomes.

3.4.4 Multivariate analysis

Multivariate breast milk IL concentration analyses controlled for both gestational age and prepregnancy BMI. Multivariate infant outcomes analyses controlled for gestational age and the ratio of human milk to total milk administered. Infant outcomes were examined using a

multivariate approach only when there was a significant relationship ($p \leq 0.10$) between maternal SNP and IL milk concentration in the multivariate analysis. The association between maternal SNPs and IL trajectory grouping was examined without covariates because this relationship has never been explored and our sample size is less than ideal for trajectory modeling.

Due to a small sample size and the need for ethnic subset analyses which further decreased our sample size, SNP-specific analyses included minor allele absence, with the exception of rs2243250. The minor alleles for rs2243250 are inconsistent between the represented ethnicities in this sample; therefore, genotypes were included in the rs2243250 analysis for each ethnicity and total population. Multivariate models that included the total population were examined using both minor allele absence and genotype as an independent variable. This was done because including the total population increased our sample size, giving more power to examine genotype-specific relationships.

Continuous outcomes were examined using multiple linear regression and binary outcomes were examined with multiple logistical regression. The relationship between trajectory group and continuous infant outcomes was assessed by generating multiple contingency tables using IL trajectory group and outcomes. We also performed general linear regression models to measure these associations, since a samples size of at least 100 is ideal to perform trajectory analyses (Nagin, 2005) and our sample size was smaller than this recommendation. We also used a contingency table and chi-square test of independence to evaluate if the independent variable was associated with the dependent variable for each aim.

3.5 RESULTS

3.5.1 Demographics, Genotype Frequencies, and Univariate Analysis

The average maternal age was 28.27 years old, the sample was mostly African American (39.68%), and the average gestational age at delivery was 28 weeks. Additional maternal and infant demographics are presented in Tables 5 and 6. There were three SNPs not in HWE (Table 7). Relationships in the univariate analyses (Table 8) that met our criteria for inclusion in the multivariate model that examined *SNP/milk IL* (Table 9) were: A) rs2070874: Caucasian IL4 weeks 1, 2; Hispanic IL4 weeks 2, 3, B) rs2243250: Caucasian IL4 week 1; Hispanic IL4 weeks 1, 2, 3, C) rs1800795: Caucasian IL6 week 1; African American IL6 weeks 1, 2, average; Hispanic IL6 weeks 1, 2, 3, D) rs1800796: Caucasian IL6 week 1; African American IL6 week 2; E) rs1800871: African American IL10 weeks 1, 2, 3, average; F) rs1800872: African American IL10 weeks 1, 2, 3, average, G) rs1800896: Total population IL10 week 1, Caucasians IL10 weeks 1, 3, average; African Americans IL10 weeks 3, average; Hispanic IL10 weeks 1, 3. There were no significant univariate findings between SNP and milk IL trajectory group (Table 10). Relationships in the univariate analysis (Tables 11 and 12) that met our criteria for inclusion in the multivariate model that examined *milk IL trajectory groups and infant outcomes* (Tables 13 and 14) were: A) IL4 and sepsis, B) IL4 and blood transfusions, C) IL6 and IVH, D) IL6 and fecal calprotectin, and E) IL10 and fecal calprotectin. Relationships between SNP and interleukin concentration that met our criteria for inclusion in the multivariate model that examined *cross-sectional milk IL concentration and subsequent infant outcomes* (Tables 15 and 16) were: A) Caucasians: IL6 and sepsis, ROP, transfusions, feeding intolerance, calprotectin, SNAPPEII, length of stay, weight at six weeks, and days on oxygen, B) African Americans: IL6 and IVH,

calprotectin, SNAPPEII, weight at six weeks, and days on oxygen, and C) African Americans: IL10 and sepsis, IVH, feeding intolerance, calprotectin, SNAPPEII, length of stay, weight at six weeks, and days on oxygen. Relationships in the univariate analysis (Tables 17 and 18) that met our criteria for inclusion in the multivariate model that examined *maternal SNP and infant outcomes* (Tables 19 and 20) were: A) rs2070874: length of stay, days on oxygen, weight at six weeks (Caucasian); weight at six weeks, calprotectin, and SNAPPEII (African American); calprotectin (Hispanic); ROP (total population); B) rs2243250: length of stay, weight at six weeks, days on oxygen, calprotectin (Caucasian); length of stay, weight at six weeks, oxygen, calprotectin (African American); calprotectin (Hispanic); blood transfusions and calprotectin (total population); C) rs1800795: days on oxygen, calprotectin (Caucasian); IVH and SNAPPEII (African American); length of stay, calprotoctin (Hispanic); ROP, blood transfusions, SNAPPEII, days on oxygen, calprotectin (total population); D) rs1800796: calprotectin (Caucasian); sepsis, length of stay, SNAPPEII, calprotectin (African American); feeding intolerance, weight at six weeks, calprotectin (Hispanic); ROP, BPD, NEC, blood transfusions, feeding intolerance, length of stay, weight at six weeks, calprotectin (Total population); E) rs1800871: IVH, length of stay, calprotectin (Caucasian); sepsis, length of stay (African American); calprotectin (Hispanic); calprotectin (total population); F) rs1800872: length of stay, weight at six weeks, and calprotectin (Caucasian); length of stay (African American); calprotectin (Hispanic); length of stay and calprotectin (Total population); G) rs1800896: calprotectin (Caucasian); days on oxygen (African American); length of stay and calprotectin (Hisapnic); sepsis, length of stay, and calprotectin (total population).

3.5.2 Interleukin trajectory modeling

All interleukins levels were natural log transformed to fulfill the normality distribution assumption. The resulting model from the IL4 trajectory model included three groups with order 1) low linear (34.7%), 2) middle linear (46.5%) and high linear (18/8%) (Figure 2). The IL6 trajectory resulted in two groups, ordered: 1) low linear (49.7%), and 3) high linear (50.3%) (Figure 3). IL10 resulted in three groups, ordered: 1) low linear (33.2%), 2) middle linear (46.5%) and 3) high linear (20.3%) (Figure 4).

3.5.3 Maternal Interleukin SNPs and Breast Milk Interleukin Concentration/Interleukin Trajectory

Controlling for gestational age at delivery and maternal prepregnancy BMI, there were no significant associations between IL4 genotypes and subsequent IL4 breast milk concentrations. However, there were trends ($p \leq 0.150$) toward significance between rs2070874 minor allele absence and IL4 milk concentration among Caucasians at week one (p=0.1417), and for Hispanics at weeks two (p=0.1406). Caucasians have a similar relationship between rs2243250 genotype and IL4 milk concentration at week one (p=0.1492), however there were no TT genotype subjects represented during this week (Table 9).

There was a significant inverse relationship between absence of minor allele rs1800795 and IL6 milk concentration among Caucasians at week three (p=0.0966). At week one, Caucasians experience a trend toward significance between rs1800796 minor allele absence and IL6 milk concentrations (p=0.1173). African Americans have a significant inverse relationship between rs1800796 minor allele absence and IL6 milk levels at week two (p=0.0772).

Additionally, prepregnancy BMI was inversely associated with IL6 milk concentrations in only the African American population in the rs1800795 model for week one (p=0.0411) and average IL6 (p=0.0126). A similar association between prepregnancy BMI and IL6 milk concentration was also seen among African Americans in the rs1800796 model for week two (p=0.0288).

While there were no significant relationships between rs1800871 minor allele absence and subsequent IL10 milk concentration, prepregnancy BMI was inversely associated with milk IL10 levels at week one (p=0.061), week two (p=0.0350), and average IL10 (p=0.0263), though this relationships was seen only among African Americans. Prepregnancy BMI was also significantly related to IL10 milk concentration when examining minor allele absence of rs1800872 at weeks one (p=0.0260), two (p=0.0250), and average IL10 (p=0.015), though there was only a trend toward significance between minor allele absence and IL10 at week three (p=0.1662). Absence of minor allele rs1800896 was significantly associated with IL10 at week three among African Americans (p=0.0705).

3.5.4 Interleukin SNP and infant outcomes

When controlling for gestational age at delivery and ratio of human milk to total milk received, there was a significant association between rs2070874 genotype TT and ROP in the total population (p=0.0706) (Tables 19 and 20). There was also a significant association, among African American infants, between minor allele absence of rs2070874 and fecal calprotectin at week two (p=0.0589). When examining the relationship between rs2243250 genotype and outcomes, Caucasian infants with a TT genotype was associated with a longer length of stay (p=0.0720). Among African Americans, it was the CC genotype that significantly associated with length of stay (p=0.0518), and days on oxygen (p=0.0809).

There was a significant association between rs1800795 genotype CC and SNAPPEII scores among the total population (p=0.0625). There was also an association between genotype GG and SNAPPEII (p=0.1075), though there was no significance between rs1800795 CG genotype and SNAPPEII scores. Rs1800795 genotype GG was also significantly associated with number of days on oxygen (p=0.0316), as was minor allele absence (p=0.0316) in the total infant population. Among Caucasians, rs1800795 minor allele absence was significantly associated with calprotectin at week two (p=0.0222). When examining the relationship between rs1800796 and outcomes, there was a significant association between ROP and genotype AA (p=0.0687) and minor allele absence (p=0.0573) in the total population. There was also an association between NEC and minor allele absence rs1800796 (p=0.0833), though there were only three NEC cases in this sample. Additionally, among the total population, there was a significant inverse relationship between minor allele absence rs1800796 and length of stay (p=0.0688), genotype GG and length of stay (p=0.0034), and genotype GG and calprotectin at week three (p=0.0213). Among Caucasians, there was a significant association between rs1800796 minor allele absence and calprotectin at week 3 (p=0.0429). African Americans exhibit a significant relationship between rs1800796 minor allele absence and length of stay (p=0.0158), SNAPPEII (p=0.0497), and calprotectin at week three (p=0.0868). There is a significant relationship among Hispanics between rs1800796 minor allele absence and weight at six weeks (p=0.0272).

When examining the relationship between rs1800871 and infant outcomes, Caucasians had a significant association between absence of minor allele and length of stay (p=0.0989) and calprotectin week one (p=0.0890). African Americans also had a relationship between minor allele absence rs1800871 and length of stay (p=0.0989). There was a significant relationship between rs1800871 genotype TT and calprotectin week three in the total population (p=0.0270).

There were significant associations between rs1800872 minor allele absence among Caucasians and calprotectin at week one (p=0.0196) and this relationship was also observed in the total population but with genotype AA at week three (p=0.0158). There was a significant association between rs1800896 minor allele absence and days on oxygen (p=0.0320) in African Americans. rs1800896 minor allele absence was also significant associated with calprotectin week two (p=0.0737) among Caucasians. Among the total population, there was a significant association between rs1800896 genotype AA calprotectin at weeks two (p=0.0045) and three (p=0.0744), and minor allele absence and calprotectin weeks two p=0.0057) and three (p=0.0825).

3.5.5 Interleukin Trajectory, Interleukin Concentration, and Infant Outcomes

Multivariate analysis of trajectory group and infant outcomes are presented in Tables 13 and 14. When controlling for gestational age at delivery and ratio of human milk to total milk received, infants who received breast milk from group 2 were 4.16 times more likely to receive a blood transfusion when compared with infants who received breast milk from trajectory group 3 (OR=4.162, CI 0.778, 22.277, p=0.0712). When controlling for gestational age at delivery and ratio of human milk to total milk, there was a significant association between IL6 group 1 membership and IVH (OR=6.275, CI 1.076, 36.584, p=0.0412). There was also a significant relationship between IL6 group 1 and fecal calprotectin level at week 3 (p=0.0822). There was no significant association between IL SNPs and interleukin trajectory group. There was a significant and positive association between IL6 milk levels at weeks one and two and calprotectin at week three (p=0.0794, p=0.0978) among African Americans. Caucasians also experience a significant relationship between IL6 at three weeks and infant calprotectin at week two (p=0.0290). There is a positive and significant relationship between IL6

at weeks one (p=0.1059) and a trend toward significance between IL6 at week two (p=0.1362) and subsequent IVH among African Americans (p=0.1059, p=0.1362) (Tables 15 and 16).

3.6 DISCUSSION

This study suggests that maternal SNPs influence IL milk concentration, and resulting IL levels impact neonatal outcomes. Further, our trajectory analysis of IL change over time illustrates the dynamic nature of breast milk and that IL patterns also contribute to infant outcomes. This study is consistent with previous work that demonstrates a functional impact of SNPs on subsequent IL concentration for rs2243250 (Nguyen et al., 2004), rs1800795 (Tischendorf et al., 2007), rs1800796 (Malarstig et al., 2006), rs1800871 (Qaddourah et al., 2014), rs1800872 (Lowe et al., 2003), rs1800896 (Capasso et al., 2007; Yilmaz et al., 2005); however, this is the first study to reveal this relationship in breast milk. There are race and/or ethnic specific associations, as noted in the Caucasian (1800795) and African American (1800796, 1800896) subgroups.

Interestingly, IL6 was implicated for both 1) SNP/IL level and 2) IL level/outcome in the African American subgroup. While rs1800796 does not impact week one IL6 milk levels, there is a significant positive relationship in African Americans between IL6 week one levels and subsequent infant IVH. There is a trend toward significance at week two between milk IL6 levels and subsequent IVH, and this is a time when there is a significant relationship between rs180076 and IL6 milk levels. Additionally, the trajectory analysis revealed that infants who received breast milk from the low-linear IL6 trajectory group one were more likely to develop IVH that infants who received milk from IL6 trajectory group two. Together, these findings reveal a

relationship between maternal genetics' influence on milk IL6 levels and subsequent IVH. Given the time-sensitive associations among subgroups, it appears that its influence may occur over time.

IL6 has been proposed as a strong candidate to modify the risk of perinatal brain injury (Baier, 2006). Preterm infants face deficient cerebral structural support, and are vulnerable to brain injury from compromised blood flow. IL6 crosses the blood brain barrier and enters the cerebrospinal fluid and interstitial spaces of the brain and spinal cord (Banks, Kastin, & Broadwell, 1995). IL6 helps to induce coagulation (Laine et al., 2014; Singh, Vennila, Snijesh, George, & Sunny, 2015), and enhances the expression of tissue factor which activates the coagulation cascade (Poralla, Hertfelder, et al., 2012; Poralla, Traut, et al., 2012). Additionally, IL6 activation decreases Vitamin K dependent coagulation factors and subsequent IVH development (Poralla, Hertfelder, et al., 2012). Furthermore, IL6 serves a neuroprotective role in hypoxic-ischemic injury, with adult brain injury patients experiencing improved long term outcomes after administration of exogenous IL6 (Winter, Pringle, Clough, & Church, 2004). Neurodevelopmental complications among infants who experience IVH are lessened if they receive breast milk (Gibertoni et al., 2015). When adequate amounts of IL6 are present in breast milk, infants may receive exogenous coagulative protection against IVH, and neuroprotection to infants who had IVH may last into childhood. IL6 has been observed in high amounts in umbilical vein blood among infants who develop IVH (Kassal et al., 2005), though other studies have found no such relationship (Bhandari et al., 2011; Sorokin et al., 2014). Conflicting infant IL6 levels and how they relate to IVH risk, coupled with our results, indicate that exogenous milk IL6 should be examined closely as potentially mediating IVH development.

Our sample also included very few NEC cases (n=3, 4.1%). While this study was not efficiently powered to detect a significant relationship between SNPs, ILs, and subsequent NEC diagnoses, we were able to detect significant relationships between SNPs, ILs, and infant fecal calprotectin levels. Calprotectin levels are derived mostly from neutrophils and monocytes (Selmi et al., 2015), when white blood cell migration to the intestines activates neutrophils to release this protein (Yoon et al., 2014). Fecal calprotectin has been directly associated with inflammation severity in the small intestine, including NEC (Aydemir et al., 2012; Dabritz, Jenke, Wirth, & Foell, 2012; Josefsson, Bunn, & Domellof, 2007; Yoon et al., 2014). Fecal calprotectin is also involved in microbiota establishment in preterm infants, including *Clostridium sp* and *Staphylococcus sp* (Rouge et al., 2010), establishing its potential influence on long-term outcomes.

Our study observed that maternal IL6 SNPs influence IL6 milk levels, and IL6 milk levels are associated with subsequent calprotectin levels. There was a significant positive association between IL6 milk levels at weeks one and two and week three infant fecal calprotectin levels among African Americans, which suggests that milk IL6 is influencing the infant's intestinal inflammatory hemostasis. The risk of NEC among preterm infants peaks between 13 and 21 days (Llanos et al., 2002; Snyder et al., 1997), and our findings suggest that this window of risk is reflected in infants' calprotectin levels. There was also a positive association between higher IL6 levels at week three and calprotectin levels at week two, which, given its nonsequential relationship, appears to be a coincidental finding.

In vitro IL6 expression is increased in ileum mucosal tissue within the NEC population (Lu et al., 2013), and increased serum IL6 levels reflect the clinical severity of NEC (Goepfert et al., 2004; Morecroft, Spitz, Hamilton, & Holmes, 1994). Likewise, and as confirmed with the

trajectory analysis, lower IL6 milk levels is associated with lower calprotectin levels. This is consistent with the pro-inflammatory role of IL6 in the gut, as illustrated by Hegazy (2010), who ameliorated colitis in vitro by downregulating IL6 (Hegazy & El-Bedewy, 2010). The proinflammatory properties of IL6 include a role in neutrophil transition to monocyte infiltration during inflammation, suggesting a magnified effect when IL6 levels are introduced at higher rates, like via breast milk. Higher IL6 levels may predispose the infant to a hyperinflammatory intestinal environment, thereby increasing calprotectin. Cury (2013) reports that increased levels of inflammatory mediators, including IL6, result in a loss of bowel homeostasis, which may lead to disease development (Cury et al., 2013).

The trajectory of IL4 change over time was significantly associated with blood transfusions, as indicated by our finding that infants who received milk from IL4 group 2 were 4.162 times more likely to receive a blood transfusion when compared with infants who received milk from group 3. While all groups are linear in shape, group three is much higher, suggesting that those infants receive much more IL4 via the breast milk. IL4 has a suppressive role on erythropoiesis (Sawada, Sato, & Koike, 1995; Thawani, Tam, & Stevenson, 2009). This is not consistent with our findings, which suggest that infants who receive less IL4 via breast milk are more likely to require a blood transfusion. It is possible that exogenous IL4 signals the infant to make less IL4, thereby balancing IL4 levels and subsequent erythopoietic activity; however, IL4 group 1 was not associated with blood transfusions and this group had lower milk IL4 than group 2. There may be a threshold for endogenous IL4 production that group 2 reaches, but to our knowledge this has never been studied in the preterm population. This finding should be interpreted cautiously, since the overall impact of trajectory grouping on transfusions is not

significant, and it is only when comparing groups two and three that we uncover a potential relationship.

The relationship between maternal IL SNPs and subsequent neonatal outcomes suggests that milk composition may mediate the relationship between breast milk and neonatal outcomes. The three ILs examined in this study have numerous downstream immunological roles, many of which have been implicated in neonatal outcomes. For example, IL4 promotes Th2 cellular response and T-helper transcription factors are positively correlated with calprotectin levels (Garcia-Rodriguez et al., 2012). Likewise, IL6 plays a major role in transitioning neutrophils to monocytes during infiltration after injury, and monocyte levels, similar to calprotectin levels, are positively associated with NEC (Christensen, Jensen, Maheshwari, & Henry, 2010; Maheshwari et al., 2014). IL10 inhibits TNFα, which is involved in systemic inflammation, and this protein is positively correlated with fecal calprotectin in infants (Kapel et al., 2005). It is likely that our small sample size prevented us from identifying more shared significant relationships between maternal SNPs, milk IL levels, and subsequent infant outcomes; however, we believe the results from this exploratory aim can be used to design future studies that examine appropriate pathways of bioactive milk components and their influence on outcomes.

While there was only one time period (week three) in which maternal SNP (rs1800896) impacted IL10 milk levels, the impact of prepregnancy BMI on milk IL10 levels remained consistent over time, though this was observed only among African Americans. There are racial/ethnic differences with respect to BMI, and visceral adipose fat is lower in African American women when compared with both Hispanic and White women (Carroll, Franks, Smith, & Phelps, 2009). Despite lower visceral adipose fat, African American women have higher IL6 serum concentrations when compared with white women (Carroll et al., 2009). Abdominal

obesity is associated with low-grade inflammation and this increases plasma IL6 (Yudkin, Stehouwer, Emeis, & Coppack, 1999). Maternal adiposity was positively correlated to cord blood IL6 levels (Catalano et al., 2009), but a negative correlation was found in a sample of Mexican mothers (Vega-Sanchez et al., 2010), suggesting a race-specific influence of BMI on IL production. In all of the models for which BMI was significantly associated with milk IL levels, it was an inverse association, suggesting that higher BMIs result in lower IL6 milk concentration in African Americans.

The impact of maternal BMI on interleukin levels among African American mothers suggests that, in addition to maternal SNPs, environmental factors impact milk composition. Other immunofactors present in milk are greatly influenced by maternal weight, including lower levels of TGF-β2 and sCD14 levels in the breast milk of overweight mothers when compared with normal weight mothers (Collado et al., 2012). Other lifestyle factors, including exercise (M. W. Groer & Shelton, 2009) and smoking (Szlagatys-Sidorkiewicz, Wos, et al., 2013), influence the immunolgoical profile of breast milk. It is estimated that between 50-75% of IL10 variability can be exlained by polymorphisms (Riiskjaer et al., 2011). The remaining contributors that influence IL10 production are unknown, though several studies suggest an epigenetic influence (Saraiva & O'Garra, 2010). Epigenetic influences on IL levels may explain why we did not uncover more SNP/IL relationships. For example, women who received probiotics/dietary counselling have higher IL10 levels in their breast milk compared with women who did not receive this intervention (Hoppu, Isolauri, Laakso, Matomaki, & Laitinen, 2012). Elevated IL10 is also present in breast milk expressed by mothers with allergies (Prokesova et al., 2006), though it is unclear if that represents an endogenous or exogenous exposure. These findings implore an epigenetic approach to understand the molecular mechanism for breast milk variation.

3.7 LIMITATIONS AND CONSIDERATIONS TO CONCLUSIONS

There were several limitations to this study, including a small sample size. Due to differences in allele frequencies across races, we further decreased our power by doing subgroup analyses of Caucasians, Hispanics, and African Americans. Our p-value cutoff was more liberal ($p \leq 0.10$) because of the exploratory nature of the study and the small sample size. This is especially relevant to the trajectory analysis, which ideally includes 100 subjects (Nagin, 2005). We did obtain good group membership percentages, and far surpassed the minimum of 10% for each trajectory group; however, we did not see any relationship between SNP and trajectory group and this may be due to inadequate power. Specifically, our analysis that examined IVH as an infant outcome was only significant in the African American population, which included only two infants in the minor allele absence group. The trend toward significance in the Caucasian population suggests a true relationship, especially since there are more infants who belonged to both minor allele absence groups. Additionally, our calprotectin analysis should be interpreted cautiously, since calprotectin levels follow their own trajectory patterns with a decrease during the first week and a subsequent steady increase into the eighth week of life (Josefsson et al., 2007). This is relevant since we examined cross-sectional calprotectin levels and not patterns over time.

Much of this study is based on self-reported variables, including ethnicity and pre-pregnancy BMI. These self-reported variables are less than ideal, since study participants tend to under-report weight (Connor Gorber, Tremblay, Moher, & Gorber, 2007). Additionally, self-reported ethnicity does not adequately capture inherent biological differences, and ancestral markers are a more reliable way of obtaining biologically relevant information that accounts for admixture (Yaeger et al., 2008). The Hispanic subgroup included both “white Hispanic” and

"black Hispanic", and it could be argued that they should have been assigned Hispanics to either the Caucasian or African American groups. Hispanics were examined separately because some of the SNPs in this study have markedly different minor allele frequencies among Hispanics when compared with Caucasians and/or African Americans. Additionally, HapMap reports a separate Hispanic minor allele frequency, further supporting our decision to do a separate Hispanic-specific analysis. However, given the small sample size, it may have been advantageous to have categorized Hispanics into the Caucasian or African American subgroups. HWE was violated for three SNPs (rs2070874, rs2243250, rs1800796), though we only report significant findings for one of these (rs1800796). We were able to eliminate genotyping error; therefore, we believe HWE violation was due to a biased sample of women who delivered preterm infants which enriched for the alleles under investigation. The SNPs included in this study have been implicated in a variety of obstetrical complications, including SGA (rs2070874 and rs2243250) (Engel et al., 2005), spontaneous preterm birth (rs1800795) (Wu et al., 2013), and pregnancy loss (rs1800871, rs1800872) (Cochery-Nouvellon et al., 2009).

Infants in this study also received donor breast milk, and this milk was not included in the analysis of this study. According to Molinari, infants who receive donor milk are exposed to variable amounts of protein and bioactive components (Molinari, Casadio, Hartmann, Arthur, & Hartmann, 2013). While we acknowledge that this could certainly contaminate our IL/outcomes analysis, it is important to note that the Holder pasteurization method which donor milk is subjected to eliminates many interleukins, including IL6 and IL10 (Ewaschuk et al., 2011; Reeves, Johnson, Vasquez, Maheshwari, & Blanco, 2013; Untalan, Keeney, Palkowetz, Rivera, & Goldman, 2009). Ewaschuk found that Holder pasteurization did not significantly alter IL4 milk concentration (Ewaschuk et al., 2011) which means that infants in our study who received

donor breast milk likely received IL4 from both their MOM and donor breast milk. This may explain our findings that infants from the MOM middle-linear IL4 trajectory group were more likely to receive blood transfusions than infants who received milk from the high-linear IL4 trajectory group. It is possible that the infants in Group 2 received more donor breast milk, and therefore extra IL4 than present in MOM, suppressing erythopoesis; however, we did not control for donor milk administration. It is not possible to measure the bioavailable IL exposure of even the infants who received exclusive MOM because it is unclear to what extent IL4, IL6, and IL10 are bioavailable once they reach the stomach. Calhoun (1999) has demonstrated a sequestration that may protect cytokines until they reach the intestine to be absorbed, but this is an area that has not been adequately studied (Calhoun, Lunoe, Du, Staba, & Christensen, 1999).

Although infants in this study received one maternal allele from all the examined SNPs, we do not know which allele they received, and since we do not have infant genotypes or serum IL6 levels, we cannot examine the relationship between infant genotype and subsequent IVH; however, Baier reports no association between IL6 SNPs and IVH in a race-specific subgroup analysis (Baier, 2006). This finding further highlights the potential relevance of breast milk as it relates to the predisposition to or protection from IVH.

3.8 IMPLICATIONS FOR NURSING PRACTICE

Breast milk is currently treated as unchanging between women in most NICUs, despite a growing body of evidence that suggests great variability between women that may impact infant outcomes. Protein biomarkers have clinical relevance, including milk lactose and glucose as a marker of mastitis (Fetherston, Wells, & Hartmann, 2006), and metabolites to identify diabetic

mothers (Arthur, Kent, & Hartmann, 1994). Our study shows that maternal SNPs also contribute to the immunological profile of breast milk, specifically ILs, which also impact neonatal outcomes. This has relevance to donor breast milk administration practices. Donor breast milk is becoming the norm in most NICUs, where the best clinical practice is to provide human milk when MOM is unavailable. Most donor breast milk is expressed by women who have delivered healthy term infants (Dempsey & Miletin, 2010), so the immunological properties delivered to preterm infants via donor breast milk may be suboptimal. Future donor milk administration practices may include screening the DNA in breast milk for clinically relevant SNPs that are specific to the infants' needs.

APPENDIX A

DATA-BASED MANUSCRIPT: TABLES AND FIGURES

Table 5: Maternal Demographics (n=64)

Age	28.27 (± 6.79)
Total pregnancies	3.09 (± 2.36)
Prepregnancy Body Mass Index (BMI)	27.83 (± 7.25)
Ethnicity	
Caucasian	21 (33.33%)
African American	25 (39.68%)
Hispanic	13 (20.63%)
Asian	2 (3.17%)
Other	1 (1.59%)
Education	
Grammar/elementary school	4 (6.25%)
Middle School	6 (9.38%)
High School	36 (56.25%)
College graduate	14 (21.88%)
Post graduate degree	4 (6.25%)
Delivery method	
Vaginal	15 (23.44%)
Cesarean Section	49 (76.56%)

Table 6: Infant demographics (n=73)

Gender	
Male	38 (52.05%)
Female	35 (47.95%)
Gestational Age at Delivery	28.29 (± 2.39)
Birthweight (grams)	1069.61 (± 216.82)
Apgar 1 minute	5.97 (± 1.93)
Apgar 5 minutes	7.44 (± 1.53)
SNAPPE II Score	19.51 (± 16.82)
Time to enteral feeding	12.6 (± 5.09)
Days on oxygen	15.19 (± 21.32)
Length of stay (days)	70.52 (± 37.04)
ROP (yes)	13 (19.12%)
BPD (yes)	4 (5.56%)
Sepsis (yes)	10 (14.08%)
NEC (yes)	3 (4.17%)
IVH (yes)	9 (12.86%)
Blood transfusions (yes)	33 (45.21%)
Feeding Intolerance (yes)	15 (21.13%)

Table 7: Genotype frequency and Hardy-Weinburg Equilibrgium, total population (n=64)

SNP	Frequency (%)	Study MAF	HWE
rs2070874		T=0.148	p=0.001*
Genotype CC	45 (70.31%)		
Genotype TT	8 (12.5%)		
Genotype CT	11 (17.19%)		
rs2243250		n/a	p=0.00001*
Genotype CC	28 (43.75%)		
Genotype TT	20 (31.25%)		
Genotype CT	16 (25%)		
rs1800795		C=0.195	p=0.3011
Genotype CC	5 (7.81%)		
Genotype GG	39 (60.94%)		
Genotype CG	20 (31.25%)		
rs1800796		C=0.109	p=0.0423*
Genotype CC	3 (4.69%)		
Genotype GG	50 (78.13%)		
Genotype CG	11 (17.19%)		
rs1800871		T=0.227	p=0.252
Genotype TT	7 (11.11%)		
Genotype CC	34 (53.97%)		
Genotype CT	22 (34.92%)		
rs1800872		A=0.242	p=0.1232
Genotype CC	32 (50.79%)		
Genotype AA	9 (14.29%)		
Genotype AC	22 (34.92%)		
rs1800896		G=0.313	p=0.7227
Genotype GG	11 (17.46%)		
Genotype AA	23 (36.51%)		
Genotype AG	29 (46.03%)		

MAF=Minor Allele Frequency, HWE=Hardy-Weinburg Equilibrium $\chi 2$ goodness-of-fit test

Table 8: Univariate analysis of maternal SNP and milk interleukin concentrations

Rs2070874	**CC** (n=43)	**TT** (n=7)	**CT** (n=11)	**p**
Total Population				
lnIL4 week 1	1.033 (±2.005)	0.047 (±1.543)	1.126 (±2.499)	0.7186
lnIL4 week 2	0.679 (±1.777)	0.971 (±0.782)	1.246 (±2.46)	0.659
lnIL4 week 3	1.042 (±1.798)	0.307 (±1.386)	1.019 (±2.237)	0.6214
lnIL4 first three weeks	1.091 (±1.768)	0.524 (±1.152)	1.157 (±2.36)	0.7297
Caucasian	**CC** (n=19)	**TT** (n=1)	**CT** (n=1)	**p**
lnIL4 week 1	1.387 (±1.892)	n/a	4.675 (n/a)	0.113*
lnIL4 week 2	1.616 (±1.637)	2.228 (n/a)	4.261 (n/a)	0.3075
lnIL4 week 3	1.766 (±1.542)	0.476 (n/a)	3.559 (n/a)	0.3806
lnIL4 first three weeks	1.701 (±1.447)	1.695 (n/a)	4.263 (n/a)	0.2512
African American	**CC** (n=16)	**TT** (n=3)	**CT** (n=4)	**P**
lnIL4 week 1	0.882 (±2.05)	0.326 (±2.073)	0.253 (±2.413)	0.861
lnIL4 week 2	0.387 (±1.706)	1.063 (±0.830)	0.115 (±1.946)	0.8159
lnIL4 week 3	0.899 (±1.751)	0.614 (±2.122)	-0.004 (±1.649)	0.674
lnIL4 first three weeks	0.819 (±1.785)	0.389 (±1.769)	0.098 (±1.879)	0.674
Hispanic	**CC** (n=7)	**TT** (n=3)	**CT** (n=3)	**P**
lnIl4 week 1	0.3199 (±2.425)	-0.511 (n/a)	1.455 (±2.989)	0.7676
lnIL4 week 2	-0.946 (0.437)	0.49 (0.195)	1.486 (3.014)	0.1032*
lnIL4 week 3	-1.139 (0)	-0.056 (0.955)	1.353 (3.031)	0.1135*
lnIL4 first three weeks	-0.131 (2.054)	0.267 (0.196)	1.435 (3.009)	0.5706

Rs2070874 MAP	**No** (n=43)	**Yes** (n=18)	**p**
Total Population			
lnIL4 week 1	1.033 (±2.005)	0.856 (±2.283)	0.801
lnIL4 week 2	0.679 (±1.778)	1.149 (±1.998)	0.385
lnIL4 week 3	1.042 (±1.798)	0.742 (±1.936)	0.5697
lnIL4 first three weeks	1.091 (±1.768)	0.91 (±1.962)	0.7253
Caucasian	**No** (n=19)	**Yes** (n=2)	**p**
lnIL4 week 1	1.387 (±1.892)	4.675 (n/a)	0.1125*
lnIL4 week 2	1.616 (±1.637)	3.244 (n/a)	0.198*
lnIL4 week 3	1.766 (±1.542)	2.018 (n/a)	0.8339
lnIL4 first three weeks	1.701 (±1.447)	2.979 (n/a)	0.2563
African American	**No** (n=16)	**Yes** (n=7)	**p**
lnIL4 week 1	0.882 (±2.05)	0.283 (±1.997)	0.579
lnIL4 week 2	0.388 (±1.706)	0.431 (±1.682)	0.9581
lnIL4 week 3	0.899 (±1.751)	0.261 (±1.723)	0.4386
lnIL4 first three weeks	0.819 (±1.785)	0.223 (±1.683)	0.4622
Hispanic	**No** (n=7)	**Yes** (n=6)	**p**
lnIL4 week 1	0.319 (±2.425)	0.963 (±2.631)	0.7144
lnIL4 week 2	-0.946 (±0.473)	0.988 (±1.986)	0.0427*
lnIL4 week 3	-1.139 (n/a)	0.648 (±2.153)	0.0694*
lnIL4 first three weeks	-0.131 (±2.053)	0.851 (±2.011)	0.4039

MAP=Minor Allele Presence

rs2243250	**CC** (n=27)	**TT** (n=19)	**CT** (n=15)	**p**
Total Population				
lnIL4 week 1	0.865 (±1.884)	1.307 (±2.379)	0.866 (±2.157)	0.8115
lnIL4 week 2	0.75 (±1.723)	0.865 (±1.892)	0.887 (± 2.098)	0.9697
lnIL4 week 3	1.044 (±1.723)	0.827 (±1.967)	0.946 (±1.947)	0.9308
lnIL4 first three weeks	1.054 (±1.719)	1.118 (±1.879)	0.908 (±2.004)	0.9454
Caucasian	**CC** (n=16)	**TT** (n=2)	**CT** (n=3)	**p**
lnIL4 week 1	1.24 (±1.887)	n/a	3.167 (± 2.046)	0.1336*
lnIL4 week 2	1.596 (±1.665)	1.299 (±1.313)	3.006 (±1.803)	0.3951
lnIL4 week 3	1.886 (±1.429)	0.127 (±0.495)	2.427 (±2.187)	0.2459
lnIL4 first three weeks	1.729 (±1.399)	0.906 (±1.115)	2.932 (±1.981)	0.2994
African American	**CC** (n=5)	**TT** (n=13)	**CT** (n=5)	**p**
lnIL4 week 1	-0.043 (± 1.084)	1.219 (±2.239)	0.188 (±2.299)	0.4688
lnIL4 week 2	0.04 (±1.105)	0.702 (±1.877)	-0.056 (±1.614)	0.6441
lnIL4 week 3	0.613 (±1.222)	0.95 (±1.941)	-0.096 (±1.474)	0.5917
lnIL4 first three weeks	0.115 (±1.173)	1.172 (±1.883)	-0.228 (±1.548)	0.2409
Hispanic	**CC** (n=6)	**TT** (n=3)	**CT** (n=4)	**p**
lnIL4 week 1	0.685 (2.637)	4.724 n/a	-0.503(0.905)	0.1372*
lnIL4 week 2	-0.908 (0.518)	1.896 (2.502)	-0.225 (1.061)	0.0570
lnIL4 week 3	-1.139 (n/a)	1.417 (±3.005)	-0.375 (±0.885)	0.1189*
lnIL4 first three weeks	0.037 (2.196)	1.778 (2.575)	-0.341 (0.943)	0.3779

rs1800795	**CC** (n=5)	**GG** (n=36)	**CG** (n=20)	**p**
Total Population				
lnIL6 week 1	2.07 (±0.911)	2.81 (±1.26)	2.492 (±1.498)	0.5529
lnIL6 week 2	1.617 (±1.991)	2.169 (±1.019)	1.912 (±1.378)	0.5927
lnIL6 week 3	1.735 (±1.301)	1.547 (±1.291)	1.709 (±1.293)	0.8917
lnIL6 first three weeks	2.117 (±1.021)	2.351 (±1.097)	2.166 (±1.343)	0.8149
Caucasian	**CC** (n=5)	**GG** (n=6)	**CG** (n=10)	**p**
lnIL6 week 1	2.07 (±0.911)	2.267 (±1.623)	2.973 (±1.498)	0.5545
lnIL6 week 2	1.617 (±1.991)	1.918 (±1.072)	2.389 (±1.099)	0.5883
lnIL6 week 3	1.735 (±1.301)	0.733 (±0.981)	1.994 (±1.324)	0.163
lnIL6 first three weeks	2.117 (±1.021)	1.859 (±1.057)	2.579 (±1.271)	0.4767
African American	**CC** (n=0)	**GG** (n=13)	**CG** (n=6)	**p**
lnIL6 week 1	n/a	2.905 (±1.279)	1.705 (±1.653)	0.1002*
lnIL6 week 2	n/a	2.388 (±1.083)	1.294 (±1.509)	0.0766*
lnIL6 week 3	n/a	1.796 (±1.422)	1.242 (±1.38)	0.4539
lnIL6 first three weeks	n/a	2.587 (±1.24)	1.481 (±1.498)	0.0891*
Hispanic	**CC** (n=0)	**GG** (n=10)	**CG** (n=3)	**p**
lnIL6 week 1	n/a	2.766 (±0.711)	2.12 (±0.369)	0.1915*
lnIL6 week 2	n/a	2.002 (±0.943)	1.019 (±0.992)	0.1528*
lnIL6 week 3	n/a	1.614 (±1.234)	1.063 (±0.873)	0.4967
lnIL6 first three weeks	n/a	2.175 (±0.829)	1.617 (±0.455)	0.2969

rs1800795 MAP	**No** (n=36)	**Yes** (n=25)	**p**
Total Population			
lnIL6 week 1	2.819 (±1.26)	2.435 (±1.423)	0.3314
lnIL6 week 2	2.169 (±1.019)	1.861 (±1.452)	0.3532
lnIL6 week 3	1.547 (±1.291)	1.714 (±1.265)	0.6315
lnIL6 first three weeks	2.351 (±1.097)	2.156 (±1.265)	0.5241
Caucasian	**No** (n=6)	**Yes** (n=15)	**p**
lnIL6 week 1	2.267 (±1.632)	2.748 (±1.396)	0.5458
lnIL6 week 2	1.918 (±1.072)	2.151 (±1.391)	0.7217
lnIL6 week 3	0.733 (±0.981)	1.919 (±1.273)	0.0575*
lnIL6 first three weeks	1.859 (±1.057)	2.425 (±1.178)	0.3197
African American	**No** (n=13)	**Yes** (n=6)	**p**
lnIL6 week 1	2.905 (±1.279)	1.705 (±1.653)	0.1002*
lnIL6 week 2	2.387 (±1.083)	1.294 (±1.509)	0.0766*
lnIL6 week 3	1.796 (±1.422)	1.242 (±1.38)	0.4539
lnIL6 first three weeks	2.587 (±1.24)	1.481 (±1.497)	0.0891*
Hispanic	**No** (n=10)	**TT** (n=3)	**p**
lnIL6 week 1	2.766 (±0.711)	2.12 (±0.369)	0.1915*
lnIL6 week 2	2.002 (±0.943)	1.019 (±0.992)	0.1528*
lnIL6 week 3	1.614 (±1.234)	1.063 (±0.873)	0.4967
lnIL6 first three weeks	2.175 (±0.829)	1.617 (±0.455)	0.2969

rs1800796				
Total Population	***CC** (n=3)*	**GG** (n=49)	**CG** (n=9)	**p**
lnIL6 week 1	3.041 (n/a)	2.584 (±1.395)	2.94 (±1.063)	0.8008
lnIL6 week 2	1.819 (±0.299)	2.042 (±1.271)	2.097 (±1.095)	0.9592
lnIL6 week 3	2.226 (±0.6989)	1.571 (±1.352)	1.701 (±0.93)	0.7629
lnIL6 first three weeks	2.414(±0.58)	2.275 (±1.204)	2.205 (±1.157)	0.9644
Caucasian	***CC** (n=0)*	**GG** (n=19)	**CG** (n=2)	**p**
lnIL6 week 1	n/a	2.456 (±1.333)	5.006 (n/a)	0.0834*
lnIL6 week 2	n/a	2.164 (±1.261)	1.338 (±1.588)	0.4007
lnIL6 week 3	n/a	1.633 (±1.309)	0.941 (±1.331)	0.4878
lnIL6 first three weeks	n/a	2.284 (±1.018)	2.068 (±2.764)	0.8076
African American	**CC** (n=0)	**GG** (n=21)	**CG** (n=2)	**p**
lnIL6 week 1	n/a	2.549 (±1.562)	2.335 (±0.291)	0.853
lnIL6 week 2	n/a	1.953 (±1.27)	3.239 (±0.942)	0.1839*
lnIL6 week 3	n/a	1.629 (±1.433)	1.995 (±1.385)	0.7348
lnIL6 first three weeks	n/a	2.255 (±1.419)	2.754 (±0.707)	0.6340
Hispanic	**CC** (n=2)	**GG** (n=7)	**CG** (n=4)	**P**
lnIL6 week 1	3.041 (n/a)	2.391 (±0.854)	2.655 (±0.398)	0.7045
lnIL6 week 2	2.031 (n/a)	1.513 (±1.192)	2.113 (±0.76)	0.6584
lnIL6 week 3	1.739 (n/a)	1.189 (±1.371)	1.912 (±0.757)	0.6299
lnIL6 first three weeks	2.460 (±0.809)	1.861 (±0.867)	2.159(±0.706)	0.6286

rs1800796 MAP			
Total Population	**No** (n=49)	**Yes** (n=12)	**p**
lnIL6 week 1	2.584 (±1.395)	2.954 (±0.971)	0.3249
lnIL6 week 2	2.04 (±1.271)	2.047 (±0.99)	0.9907
lnIL6 week 3	1.57 (±1.352)	1.796 (±0.886)	0.6013
lnIL6 first three weeks	2.275 (±1.204)	2.257 (±1.021)	0.9642
Caucasian	**No** (n=19)	**Yes** (n=2)	**p**
lnIL6 week 1	2.456 (±1.333)	5.005 (n/a)	0.0834*
lnIL6 week 2	2.164 (±1.261)	1.388 (±1.588)	0.4007
lnIL6 week 3	1.633 (±1.309)	0.941 (±1.331)	0.4878
lnIL6 first three weeks	2.284 (±1.018)	2.068 (±2.764)	0.8076
African American	**No** (n=21)	**Yes (n=2)**	**P**
lnIL6 week 1	2.549 (±1.562)	2.335 (±0.291)	0.8530
lnIL6 week 2	1.953 (±1.27)	3.239 (±0.942)	0.1839*
lnIL6 week 3	1.629 (±1.433)	1.995 (±1.385)	0.7348
lnIL6 first three weeks	2.255 (±1.419)	2.754 (±0.707)	0.6340
Hispanic	**No** (n=7)	**Yes (n=6)**	**P**
lnIL6 week 1	2.391 (±0.854)	2.751 (±0.378)	0.4624
lnIL6 week 2	1.513 (±1.192)	2.096 (±0.659)	0.3485
lnIL6 week 3	1.1896 (±1.371)	1.878 (±0.659)	0.3272
lnIL6 first three weeks	1.861 (±0.867)	2.263(±0.675)	0.3776

rs1800871				
Total Population	**TT** (n=7)	**CC** (n=31)	**CT** (n=22)	**p**
lnIL10 week 1	2.364 (±2.175)	1.41 (±1.875)	1.581 (±1.386)	0.5423
lnIL10 week 2	1.026 (±1.578)	1.012 (±1.869)	0.96 (±1.369)	0.9931
lnIL10 week 3	0.812 (±1.663)	0.669 (±1.769)	0.464 (±1.484)	0.8652
lnIl10 first three weeks	1.614(±1.839)	1.149 (±1.757)	1.113 (±1.352)	0.7639
Caucasian	**TT** (n=1)	**CC** (n=13)	**CT** (n=6)	**p**
lnIL10 week 1	2.715 (n/a)	1.659 (±2.115)	1.938 (±2.059)	0.8807
lnIL10 week 2	1.785 (n/a)	1.137 (±1.808)	1.389 (±2.032)	0.9250
lnIL10 week 3	0.285 (n/a)	0.949 (±1.949)	1.116 (±1.727)	0.9193
lnIL10 first three weeks	2.01 (n/a)	1.461 (±1.801)	1.414 (±1.893)	0.9546
African American	**TT** (n=1)	**CC** (n=11)	**CT** (n=11)	**p**
lnIL10 week 1	4.474 (n/a)	1.255 (±1.614)	1.635 (±1.058)	0.1128*
lnIL10 week 2	3.616 (n/a)	0.876 (±1.518)	0.973 (±1.245)	0.1933*
lnIL10 week 3	3.904 (n/a)	0.464 (±1.276)	0.256 (±1.214)	0.0385*
lnIL10 first three weeks	4.063 (n/a)	0.942 (±1.396)	1.091 (±1.174)	0.0911*
Hispanic	**TT** (n=4)	**CC** (n=6)	**CT** (n=3)	**p**
lnIL10 week1	1.545 (±2.483)	1.088 (±2.481)	0.261 (±0.739)	0.8319
lnIL10 week 2	0.06 (±1.039)	0.373 (±2.272)	0.371 (±0.751)	0.9668
lnIL10 week 3	-0.036 (±0.956)	0.215 (±2.307)	-0.489 (±0.596)	0.8604
lnIL10 first three weeks	1.133 (±2.031)	0.451 (±2.171)	0.042 (±0.664)	0.7494

rs1800871 MAP			
Total Population	**No** (n=31)	**Yes** (n=29)	**p**
lnIL10 week 1	1.41 (±1.875)	1.759 (±1.574)	0.5003
lnIL10 week 2	1.012 (±1.869)	0.975 (±1.387)	0.9348
lnIL10 week 3	0.668 (±1.769)	0.541 (±1.499)	0.7711
lnIL10 first three weeks	1.149 (±1.757)	1.234 (±1.464)	0.8422
Caucasian	**No** (n=13)	**Yes** (n=7)	**p**
lnIL10 week 1	1.659 (±2.115)	2.067 (±1.869)	0.7029
lnIL10 week 2	1.137 (±1.808)	1.455 (±1.825)	0.7266
lnIL10 week 3	0.949 (±1.949)	0.997 (±1.607)	0.9560
lnIL10 first three weeks	1.461 (±1.801)	1.499 (±1.743)	0.9640
African American	**No** (n=11)	**Yes** (n=12)	**p**
lnIL10 week 1	1.255 (±1.614)	1.919 (1.342)	0.3140
lnIL10 week 2	0.876 (±1.518)	1.214 (1.424)	0.6051
lnIL10 week 3	0.588 (±1.593)	0.464 (1.276)	0.8469
lnIL10 first three weeks	0.942 (±1.396)	1.339 (1.41)	0.5055
Hispanic	**No** (n=6)	**Yes** (n=7)	**p**
lnIL10 week1	1.088 (±2.481)	1.031 (±1.927)	0.970
lnIL10 week 2	0.373 (±2.272)	0.216 (±0.829)	0.8770
lnIL10 week 3	0.215 (±2.307)	-0.263 (±0.754)	0.6405
lnIL10 first three weeks	0.451 (±2.171)	0.665 (±1.597)	0.8414

rs1800872				
Total Population	**CC** (n=29)	**AA** (n=9)	**AC** (n=22)	**p**
lnIL10 week 1	1.472 (±1.931)	2.041 (±2.309)	1.425 (±1.369)	0.7242
lnIL10 week 2	0.879 (±1.883)	1.365 (±1.697)	0.996 (±1.352)	0.7926
lnIL10 week 3	0.567 (±1.771)	1.163 (±1.779)	0.517 (±1.443)	0.6554
lnIL10 first three weeks	1.075 (±1.789)	1.509 (±1.972)	1.18 (±1.308)	0.7833
Caucasian	**CC** (n=14)	**AA** (n=3)	**AC** (n=4)	**p**
lnIL10 week 1	1.739 (±2.07)	1.727 (±2.522)	1.319 (±2.214)	0.9451
lnIL10 week 2	0.997 (±1.775)	2.592 (±1.141)	1.458 (±2.169)	0.5036
lnIL10 week 3	0.909 (±1.901)	1.775 (±2.106)	0.757 (±1.648)	0.8066
lnIL10 first three weeks	1.347 (±1.793)	1.435 (±2.339)	1.295 (±1.906)	0.9952
African American	**CC** (n=8)	**AA** (n=1)	**AC** (n=13)	**p**
lnIL10 week 1	1.259 (±1.785)	4.474 (n/a)	1.628 (±1.146)	0.1351*
lnIL10 week 2	0.549 (±1.459)	3.616 n/a	1.085 (±1.354)	0.1468*
lnIL10 week 3	0.077 (±1.094)	3.904 n/a	0.621 (±1.341)	0.0325*
lnIL10 first three weeks	0.752 (±1.402)	4.063 n/a	1.185 (±1.248)	0.0821*
Hispanic	**CC** (n=6)	**AA** (n=4)	**AC** (n=3)	**P**
lnIL10 week 1	1.088 (±2.481)	1.545 (±2.483)	0.261 (±0.739)	0.8319
lnIL10 week 2	0.373 (±2.272)	0.06 (±1.039)	0.371 (±0.751)	0.9668
lnIL10 week 3	0.215 (±2.307)	-0.036 (±0.956)	-0.489 (±0.596)	0.8604
lnIL10 first three weeks	0.451 (±2.171)	1.133 (±2.031)	0.042 (±0.665)	0.7494

rs1800872 MAP			
Total Population	**No** (n=29)	**Yes** (n=31)	**p**
lnIL10 week 1	1.472 (±1.931)	1.598 (±1.656)	0.8137
lnIL10 week 2	0.879 (±1.883)	1.092 (±1.424)	0.6399
lnIL10 week 3	0.567 (±1.771)	0.684 (±1.528)	0.7923
lnIL10 first three weeks	1.075 (±1.786)	1.232 (±1.506)	0.7147
Caucasian	**No** (n=14)	**Yes** (n=7)	**p**
lnIL10 week 1	1.793 (±2.07)	1.494 (±2.149)	0.8158
lnIL10 week 2	0.997 (±1.775)	1.912 (±1.751)	0.3350
lnIL10 week 3	0.909 (±1.901)	1.096 (±1.671)	0.8376
lnIL10 first three weeks	1.347 (±1.793)	1.355 (±1.909)	0.9931
African American	**No** (n=8)	**Yes** (n=14)	**P**
lnIL10 week 1	1.259 (±1.785)	1.865 (±1.367)	0.4341
lnIL10 week 2	0.549 (±1.459)	1.28 (±1.474)	0.3028
lnIL10 week 3	0.077 (±1.094)	0.895 (±1.592)	0.2228
lnIL10 first three weeks	0.752 (±1.402)	1.391 (±1.391)	0.3212
Hispanic	**No** (n=6)	**Yes** (n=7)	**P**
lnIL10 week1	1.088 (±2.481)	1.031 (±1.927)	0.970
lnIL10 week 2	0.373 (±2.272)	0.216 (±0.829)	0.8770
lnIL10 week 3	0.215 (±2.307)	-0.263(±0.754)	0.6405
lnIL10 first three weeks	0.451 (±2.171)	0.665 (±1.597)	0.8414

rs1800896				
Total Population	**GG** (n=10)	**AA** (n=22)	**AG** (n=28)	**p**
lnIL10 week 1	1.537 (±2.373)	1.983 (±1.878)	1.136 (±1.291)	0.3420
lnIL10 week 2	1.469 (±2.191)	0.981 (±1.658)	0.907 (±1.421)	0.6503
lnIL10 week 3	1.161(±2.378)	0.809 (±1.626)	0.327 (±1.271)	0.3363
lnIL10 first three weeks	1.563(±2.216)	1.404 (±1.669)	0.889 (±1.292)	0.3933
Caucasian	**GG** (n=5)	**AA** (n=7)	**AG** (n=9)	**p**
lnIL10 week 1	1.777 (±2.19)	2.572 (±2.081)	0.759 (±1.816)	0.2925
lnIL10 week 2	1.674 (±2.077)	1.925 (±2.005)	0.614 (±1.421)	0.3562
lnIL10 week 3	1.804 (±2.354)	1.641 (±1.708)	0.05 (±1.148)	0.1136*
lnIL10 first three weeks	1.832 (±2.098)	2.083 (±1.761)	0.511 (±1.423)	0.1737*
African American	**CC** (n=3)	**TT** (n=6)	**CT** (n=14)	**p**
lnIL10 week 1	0.922 (±2.618)	2.247 (±1.342)	1.424 (±1.019)	0.4036
lnIL10 week 2	0.637 (±2.042)	1.726 (±1.641)	0.89 (±1.271)	0.4948
lnIL10 week 3	-0.618 (±0.903)	1.461 (±1.829)	0.435 (±1.179)	0.1204*
lnIL10 first three weeks	0.725 (±2.156)	2.004 (±1.448)	0.872 (±1.126)	0.2161
Hispanic	**CC** (n=2)	**TT** (n=8)	**CT** (n=2)	**P**
lnIL10 week 1	1.979 (±3.897)	0.959 (±1.943)	0.377 (±0.576)	0.7781
lnIL10 week 2	2.208 (±3.776)	-0.167 (±0.74)	0.71 (±0.747)	0.1999*
lnIL10 week 3	2.221 (±3.645)	-0.367 (±0.779)	-0.510 (±0.763)	0.1372*
lnIL10 first three weeks	2.145 (±3.763)	0.448 (±1.524)	0.312 (±0.688)	0.5104

rs1800896 MAP			
Total Population	**No** (n=22)	**Yes** (n=38)	**p**
lnIL10 week 1	1.983 (±1.878)	1.257 (±1.654)	0.1752*
lnIL10 week 2	0.981 (±1.658)	1.059 (±1.65)	0.8697
lnIL10 week 3	0.809 (±1.626)	0.552 (±1.649)	0.5820
lnIL10 first three weeks	1.404 (±1.669)	1.067 (±1.582)	0.4389
Caucasian	**No** (n=7)	**Yes** (n=14)	**p**
lnIL10 week 1	2.572 (±2.081)	1.129 (±1.918)	0.1704*
lnIL10 week 2	1.925 (±2.005)	0.992 (±1.687)	0.3255
lnIL10 week 3	1.641 (±1.708)	0.676 (±1.81)	0.2820
lnIL10 first three weeks	2.083 (±1.761)	0.983 (±1.741)	0.1899*
African American	**No** (n=6)	**Yes** (n=17)	**p**
lnIL10 week 1	2.247 (±1.542)	1.308 (±1.403)	0.2058
lnIL10 week 2	1.726 (±1.641)	0.843 (±1.363)	0.2415
lnIL10 week 3	1.461 (±1.829)	0.237 (±1.183)	0.0919*
lnIL10 first three weeks	2.005 (±1.448)	0.847 (±1.27)	0.0777*
Hispanic	**No** (n=8)	**Yes** (n=4)	**p**
lnIL10 week 1	0.959 (±1.945)	1.178 (±2.455)	0.8854
lnIL10 week 2	-0.166 (±0.740)	1.459 (±2.384)	0.1186*
lnIL10 week 3	-0.367 (±0.779)	0.855 (±2.666)	0.2717
lnIL10 first three weeks	0.448 (±1.524)	1.228 (±2.449)	0.5069

Table 9: Multivariate analysis of maternal SNP and milk interleukin concentrations

Multivariate regression model for Interleukin 4 concentration with minor allele presence rs2070874

Subgroup	Outcome	Predictor	Estimate	p-value
Caucasian	lnIL4 week 1	Minor allele presence – no	-3.271	0.1417
		Minor allele presence – yes (reference)		
		Gestational Age	0.0126	0.9675
		Prepregnancy BMI	0.018	0.8156
Caucasian	lnIL4 week 2	Minor allele presence – no	-1.619	0.2640
		Minor allele presence – yes (reference)		
		Gestational Age	0.072	0.8357
		Prepregnancy BMI	0.016	0.8049
Hispanic	lnIL4 week 2	Minor allele presence – no	-2.059	0.1406
		Minor allele presence – yes (reference)		
		Gestational Age	-0.064	0.8689
		Prepregnancy BMI	-0.075	0.6565
Hispanic	lnIL4 week 3	Minor allele presence – no	-1.988	0.1920
		Minor allele presence – yes (reference)		
		Gestational Age	-0.234	0.5979
		Prepregnancy BMI	-0.013	0.9463

Multivariate regression model for Interleukin 4 concentration with genotype rs2243250

Subgroup	Outcome	Predictor	Estimate	p-value
Caucasian	lnIL4 week 1	Genotype CC	-2.019	0.1492
		Genotype TT	n/a	n/a
		Genotype CT (reference)		
		Gestational Age	-0.066	0.8324
		Prepregnancy BMI	0.025	0.7531
Hispanic	lnIL4 week 1	Genotype CC	-0.13	0.9665
		Genotype TT	4.309	0.2868
		Genotype CT (reference)		
		Gestational Age	-1.159	0.4261
		Prepregnancy BMI	-0.078	0.8620
Hispanic	lnIL4 week 2	Genotype CC	-0.636	0.7221
		Genotype TT	1.966	0.2601
		Genotype CT (reference)		
		Gestational Age	-0.266	0.5566
		Prepregnancy BMI	-0.033	0.8738
Hispanic	lnIL4 week 3	Genotype CC	-1.15	0.5565
		Genotype TT	1.737	0.3425
		Genotype CT (reference)		
		Gestational Age	-0.455	0.3664
		Prepregnancy BMI	0.058	0.7947

Multivariate regression model for Interleukin 6 concentration with minor allele presence rs1800795

Subgroup	Outcome	Predictor	Estimate	p-value
Caucasian	lnIL6 week 3	Minor allele presence – no	-1.134	0.0966*
		Minor allele presence – yes (reference)		
		Gestational Age	-0.139	0.5439
		Prepregnancy BMI	-0.013	0.7854
African American	lnIL6 week 1	Minor allele presence – no	0.510	0.4503
		Minor allele presence – yes (reference)		
		Gestational Age	-0.111	0.2429
		Prepregnancy BMI	-0.086	0.0411*
African American	lnIL6 week 2	Minor allele presence – no	0.661	0.2753
		Minor allele presence – yes (reference)		
		Gestational Age	-0.038	0.6580
		Prepregnancy BMI	-0.069	0.0640*
African American	lnIL6 average	Minor allele presence – no	0.498	0.4054
		Minor allele presence – yes (reference)		
		Gestational Age	-0.041	0.6244
		Prepregnancy BMI	-0.096	0.0126*
Hispanic	lnIL6 week 1	Minor allele presence – no	0.444	0.5059
		Minor allele presence – yes (reference)		
		Gestational Age	0.205	0.4975
		Prepregnancy BMI	0.001	0.9933

Hispanic	lnIL6 week 2	Minor allele presence – no	0.701	0.4613
		Minor allele presence – yes (reference)		
		Gestational Age	-0.055	0.8479
		Prepregnancy BMI	0.018	0.8695

Multivariate regression model for Interleukin 6 concentration with minor allele presence rs1800796

Subgroup	Outcome	Predictor	Estimate	p-value
Caucasian	lnIL6 week 1	Minor allele presence – no	-2.725	0.1173
		Minor allele presence – yes (reference)		
		Gestational Age	-0.069	0.7501
		Prepregnancy BMI	0.005	0.9300
African American	lnIL6 weeks 2	Minor allele presence – no	-1.571	0.0772*
		Minor allele presence – yes (reference)		
		Gestational Age	-0.102	0.2425
		Prepregnancy BMI	-0.076	0.0288

Multivariate regression model for Interleukin 10 concentration with minor allele presence rs1800871

Subgroup	Outcome	Predictor	Estimate	p-value
African American	lnIL10 week 1	Minor allele presence – no	-0.57	0.3882
		Minor allele presence – yes (reference)		
		Gestational Age	-0.011	0.9131
		Prepregnancy BMI	-0.082	0.0601*
African American	lnIL10 week 2	Minor allele presence – no	-0.221	0.7116
		Minor allele presence – yes (reference)		
		Gestational Age	0.122	0.2368
		Prepregnancy BMI	-0.091	0.0350*
African American	lnIL10 week 3	Minor allele presence – no	-0.374	0.5796
		Minor allele presence – yes (reference)		
		Gestational Age	-0.139	0.3483
		Prepregnancy BMI	-0.086	0.2012
African American	lnIL10 average	Minor allele presence – no	-0.231	0.6754
		Minor allele presence – yes (reference)		
		Gestational Age	0.063	0.5073
		Prepregnancy BMI	-0.09	0.0263*

Multivariate regression model for Interleukin 10 concentration with minor allele presence rs1800872

Subgroup	Outcome	Predictor	Estimate	p-value
African American	lnIL10 week 1	Minor allele presence – no	-0.883	0.2415
		Minor allele presence – yes (reference)		
		Gestational Age	0.006	0.9612
		Prepregnancy BMI	-0.099	0.0260*
African American	lnIL10 week 2	Minor allele presence – no	-0.769	0.2468
		Minor allele presence – yes (reference)		
		Gestational Age	0.116	0.3209
		Prepregnancy BMI	-0.099	0.0254*
African American	lnIL10 week 3	Minor allele presence – no	-0.962	0.1662
		Minor allele presence – yes (reference)		
		Gestational Age	-0.148	0.3909
		Prepregnancy BMI	-0.085	0.1885
African American	lnIL10 average	Minor allele presence – no	-0.68	0.2594
		Minor allele presence – yes (reference)		
		Gestational Age	0.069	0.5129
		Prepregnancy BMI	-0.099	0.0150*

Multivariate regression model for Interleukin 10 concentration with minor allele presence rs1800896

Subgroup	Outcome	Predictor	Estimate	p-value
Total Population	lnIL10 week 1	Minor allele presence – no	0.598	0.2935
		Minor allele presence – yes (reference)		
		Gestational Age	-0.024	0.8168
		Prepregnancy BMI	-0.049	0.2163
Total Population	lnIL10 week 1	Genotype GG	0.227	0.7595
		Genotype AA	0.673	0.2829
		Genotype AG (reference)		
		Gestational Age	-0.023	0.8233
		Prepregnancy BMI	-0.047	0.2494
Caucasian	lnIL10 week 1	Minor allele presence – no	1.493	0.1950
		Minor allele presence – yes (reference)		
		Gestational Age	-0.236	0.4608
		Prepregnancy BMI	-0.025	0.7670
Caucasian	lnIL10 week 3	Minor allele presence – no	1.154	0.2303
		Minor allele presence – yes (reference)		
		Gestational Age	0.065	0.8475
		Prepregnancy BMI	0.061	0.3967
Caucasian	lnIL10 average	Minor allele presence – no	1.185	0.1863
		Minor allele presence – yes (reference)		
		Gestational Age	-0.194	0.4526
		Prepregnancy BMI	-0.015	0.8189

African American	lnIL10 week 3	Minor allele presence – no	1.325	0.0705*
		Minor allele presence – yes (reference)		
		Gestational Age	-0.153	0.2522
		Prepregnancy BMI	-0.082	0.1746
African American	lnIL10 average	Minor allele presence – no	1.185	0.1863
		Minor allele presence – yes (reference)		
		Gestational Age	-0.194	0.4526
		Prepregnancy BMI	-0.015	0.8189
Hispanic	lnIL10 week 2	Minor allele presence – no	-2.093	0.2417
		Minor allele presence – yes (reference)		
		Gestational Age	0.057	0.9015
		Prepregnancy BMI	0.048	0.8099
Hispanic	lnIL10 week 3	Minor allele presence – no	-1.888	0.3229
		Minor allele presence – yes (reference)		
		Gestational Age	0.012	0.9807
		Prepregnancy BMI	0.063	0.7747

Table 10: Univariate analysis of maternal SNP and milk interleukin trajectory group

rs2070874	**Trajectory Group 1 (n=22)**	**Trajectory Group 2 (n=27)**	**Trajectory Group 3 (n=12)**	
CC	14 (22.95%)	21 (34.43%)	8 (13.11%)	
TT	3 (4.92%)	4 (6.56%)	0	p = 0.2245
CT	5 (8.20%)	2 (3.28%)	4 (6.56%)	

rs2070874 MAP	**Trajectory Group 1 (n=22)**	**Trajectory Group 2 (n=27)**	**Trajectory Group 3 (n=12)**	
No	14 (22.95%)	21 (34.43%)	8 (13.11%)	p = 0.5467
Yes	8 (13.11%)	6 (9.84%)	4 (6.56%)	

rs2243250	**Trajectory Group 1 (n=22)**	**Trajectory Group 2 (n=27)**	**Trajectory Group 3 (n=12)**	
CC	9 (14.75%)	14 (22.95%)	4 (6.56%)	
TT	7 (11.48%)	8 (13.11%)	4 (6.56%)	p = 0.7962
CT	6 (9.84%)	5 (8.2%)	4 (6.56%)	

rs1800795	**Trajectory Group 1**	**Trajectory Group 2**	
CC	3 (4.92%)	2 (3.28%)	p = 0.5807
GG	14 (22.95%)	22 (36.07%)	
CG	10 (16.39%)	10 (16.39%)	

rs1800795 MAP	**Trajectory Group 1**	**Trajectory Group 2**	
No	14 (22.95%)	22 (36.07%)	p = 0.4323
Yes	13 (21.31%)	12 (19.67%)	

rs1800796	**Trajectory Group 1**	**Trajectory Group 2**	
CC	0 (0%)	3 (4.92%)	p = 0.4539
GG	23 (37.7%)	26 (42.62%)	
CG	4 (6.56%)	5 (8.2%)	

rs1800796 MAP	**Trajectory Group 1**	**Trajectory Group 2**	
No	23 (37.7%)	26 (42.62%)	p = 0.522
Yes	4 (6.56%)	8 (13.11%)	

rs1800871	**Trajectory Group 1**	**Trajectory Group 2**	**Trajectory Group 3**
TT	1 (1.67%)	4 (6.67%)	2 (3.33%)
CC	11 (18.33%)	13 (21.67%)	7 (11.67%)
CT	6 (10%)	12 (20%)	4 (6.67%)

rs1800871 MAP	**Trajectory Group 1**	**Trajectory Group 2**	**Trajectory Group 3**	
No	11 (18.33%)	13 (21.67%)	7 (11.67%)	p = 0.5293
Yes	7 (11.67%)	16 (26.67%)	6 (10%)	

rs1800872	**Trajectory Group 1**	**Trajectory Group 2**	**Trajectory Group 3**	
CC	11 (18.33%)	12 (20%)	6 (10%)	
AA	2 (3.33%)	4 (6.67%)	3 (5%)	p = 0.7560
AC	6 (10%)	12 (20%)	4 (6.67%)	

rs1800872 MAP	**Trajectory Group 1**	**Trajectory Group 2**	**Trajectory Group 3**	
No	11 (18.33%)	12 (20%)	6 (10%)	p =0.6301
Yes	8 (13.33%)	16 (26.67%)	7 (11.67%)	

rs1800896	**Trajectory Group 1**	**Trajectory Group 2**	**Trajectory Group 3**	
TT	4 (6.67%)	2 (3.33%)	4 (6.67%)	p = 0.2913
TT	6 (10%)	11 (18.33%)	5 (8.33%)	
GT	8 (13.33%)	16 (26.67%)	4 (6.67%)	

rs1800896 MAP	**Trajectory Group 1**	**Trajectory Group 2**	**Trajectory Group 3**	
No	6 (10%)	11 (18.33%)	5 (8.33%)	p = 1.0
Yes	12 (20%)	18 (30%)	8 (13.33%)	

Table 11: Univariate Analysis of Interleukin Trajectory Group and Continuous Infant Outcomes

Interleukin 4				
Total Population	**Trajectory Group 1** (n=25)	**Trajectory Group 2** (n=24)	**Trajectory Group 3** (n=13)	**p**
LnSNAPPEII	2.867 (±0.714)	2.903 (±0.589)	2.758 (±0.696)	0.8124
LOS	63.8 (±25.762)	74.161 (40.670)	59.455 (±19.386)	0.3343
Weight at 6 weeks	7.52 (±0.19)	7.501 (±0.156)	7.532 (±0.217)	0.7376
Days on oxygen	12.208 (±25.108)	16.419 (±20.508)	14.636 (±16.244)	0.7765
lnCalprotectin Week 1	5.239 (±0.909)	4.836 (±0.856)	5.137 (±0.871)	0.3007
Ln Calprotectin Week 2	5.475 (±0.747)	5.133 (±0.609)	5.052 (±0.836)	0.2203
Ln Calprotectin Week 3	5.224 (±0.688)	5.196 (±0.621)	5.471 (±0.666)	0.4966

Interleukin 6			
Total Population	**Trajectory Group 1** (n=33)	**Trajectory Group 2** (n=34)	**p**
LnSNAPPEII	2.853 (±0.694)	2.861 (±0.625)	0.9657
LOS	63.727 (±31.881)	71.912 (±34.049)	0.3139
Weight at 6 weeks	7.521 (±0.164)	7.506 (±0.189)	0.7376
Days on oxygen	13.781 (±24.247)	15.353 (±18.853)	0.7690
lnCalprotectin Week 1	4.917 (±0.892)	5.114 (±0.873)	0.4120
Ln Calprotectin Week 2	5.256 (±0.766)	5.213 (±0.680)	0.8390
Ln Calprotectin Week 3	5.111 (±0.682)	5.421 (±0.586)	0.0825*

Interleukin 10				
Total Population	**Trajectory Group 1** (n=23)	**Trajectory Group 2** (n=32)	**Trajectory Group 3** (n=13)	**p**
LnSNAPPEII	2.911 (±0.594)	2.842 (±0.669)	2.801 (±0.747)	0.8846
LOS	69.869 (±34.492)	68.825 (±36.376)	62.083 (±19.538)	0.7956
Weight at 6 weeks	7.489 (±0.172)	7.538 (±0.169)	7.494 (±0.211)	0.5915
Days on oxygen	15.5 (±20.43)	12.25 (±23.923)	19.167 (±16.634)	0.6246
lnCalprotectin Week 1	5.021 (±0.999)	4.989 (±0.847)	5.088 (±0.840)	0.9517
Ln Calprotectin Week 2	5.249 (±0.731)	5.267 (±0.646)	5.125 (±0.887)	0.8702
Ln Calprotectin Week 3	4.976 (±0.577)	5.375 (±0.644)	5.445 (±0.671)	0.0810*

Table 12: Univariate Analysis of Interleukin Trajectory Group and Categorical Infant Outcomes

Interleukin 4 **Total Population**	**Trajectory Group 1** (n=25)	**Trajectory Group 2** (n=24)	**Trajectory Group 3** (n=13)	**p**
Sepsis				
Yes	1 (1.67%)	6 (10%)	0 (%)	0.0785*
No	21 (35%)	21 (35%)	11 (18.33%)	
ROP				
Yes	3 (5.26%)	4 (7.02%)	1 (1.75%)	1.00
No	18 (31.58%)	22 (38.6%)	9 (15.79%)	
BPD				
Yes	1 (1.67%)	0 (%)	1 (1.64%)	0.3066
No	21 (35%)	27 (44.26)	11 (18.03%)	
NEC				
Yes	0 (0%)	2 (3.28%)	0 (%)	0.6754
No	22 (36.07%)	25 (40.98%)	12 (19.67%)	
IVH				
Yes	3 (5.08%)	5 (8.47%)	1 (1.69%)	0.8035
No	19 (32.3%)	21 (35.59%)	10 (16.95%)	
Blood transfusion				
Yes	9 (14.75%)	16 (26.23%)	3 (4.92%)	0.1273*
No	13 (21.31%)	11 (18.03%)	9 (14.75%)	
Feeding intolerance				
Yes	3 (5%)	7 (11.67%)	1 (1.67%)	0.4985
No	19 (31.67%)	20 (33.33%)	10 (16.67%)	

Interleukin 6			
Total Population	**Trajectory Group 1** (n=27)	**Trajectory Group 2** (n=33)	**p**
Sepsis			
Yes	2 (3.33%)	5 (8.33%)	0.4422
No	25 (41.67%)	28 (46.67%)	
ROP			
Yes	3 (5.26%)	5 (8.77%)	1.00
No	22 (38.6%)	27 (47.37%)	
BPD			
Yes	1 (1.64%)	1 (1.64%)	1.00
No	26 (42.62%)	33 (54.1%)	
NEC			
Yes	0 (0%)	2 (3.28%)	0.4984
No	29 (44.26%)	32 (52.46%)	
IVH			
Yes	7 (11.86%)	2 (3.39%)	0.0351*
No	19 (32.2%)	31 (52.54%)	
Blood transfusion			
Yes	14 (22.95%)	14 (22.95%)	0.4473
No	13 (21.31%)	20 (32.79%)	
Feeding intolerance			
Yes	4 (6.67%)	7 (11.67%)	0.7391
No	23 (38.22%)	26 (43.44%)	

Interleukin 10 **Total Population**	**Trajectory Group 1** (n=19)	**Trajectory Group 2** (n=29)	**Trajectory Group 3** (n=12)	**p**
Sepsis				
Yes	2 (3.33%)	5 (8.33%)	0 (0%)	0.3830
No	17 (28.33%)	24 (40%)	12 (20%)	
ROP				
Yes	3 (5.26%)	4 (7.02%)	1 (1.75%)	1.000
No	15 (26.32%)	24 (42.11%)	10 (17.54%)	
BPD				
Yes	0 (0%)	1 (1.64%)	1 (1.64%)	0.4770
No	19 (31.15%)	28 (45.90%)	12 (19.67%)	
NEC				
Yes	0 (%)	2 3.28%)	0 (%)	0.6989
No	19 (31.15%)	27 (44.26%)	13 (21.31%)	
IVH				
Yes	5 (8.47%)	3 (5.08%)	1 (1.69%)	0.3581
No	14 (23.73%)	25 (42.37%)	11 (18.64%)	
Blood transfusion				
Yes	12 (19.67%)	11 (18.03%)	5 (8.20%)	0.2211
No	7 (11.48%)	18 (29.51%)	8 (13.11%)	
Feeding intolerance				
Yes	4 (6.67%)	5 (8.33%)	2 (3.33%)	1.000
No	15 (25%)	24 (40%)	10 (16.67%)	

Table 13: Multivariate analysis of interleukin trajectory group and continuous infant outcomes

Outcome	Predictor	Estimate	p-value
lnCalprotectin Week 3	Interleukin 6 trajectory group 1	-0.312	0.0822*
	Interleukin 6 trajectory group 2 (reference)		
	Gestational Age	0.081	0.0521*
	Ratio of mom's own milk to all milk	0.333	0.395
lnCalprotectin Week 3	Interleukin 10 trajectory group 1	-0.391	0.1161
	Interleukin 10 trajectory group 2	-0.006	0.9778
	Interleukin 10 trajectory group 3 (reference)		
	Gestational age	0.072	0.0877*
	Ratio of mom's own milk to all milk	0.359	0.372

Table 14: Multivariate analysis of interleukin trajectory group and categorical infant outcomes

Outcome	Odds Ratio	Point estimate, 95% Confidence Interval	p
Sepsis	IL4 Group Membership		0.2731
	IL4 group 1 versus 3	0.132 (0.004, 4.028)	0.1416
	IL4 group 2 versus 3	1.050 (0.086, 12.818)	0.2594
	Gestational Age	0.348 (0.155, 0.781)	0.0105*
	Ratio of mom's own milk to all milk	0.075 (0.001, 4.81)	0.2228
Blood transfusion	IL4 Group Membership		0.1917
	IL4 group 1 versus 3	1.712 (0.298, 9.837)	0.7956
	IL4 group 2 versus 3	4.162 (0.778, 22.277)	0.0712*
	Gestational Age	0.454 (0.298, 0.693)	0.0002*
	Ratio of mom's own milk to all milk	1.838 (0.148, 22.815)	0.6357
Intraventricular Hemorrhage	IL6 Group Membership		0.0412*
	IL6 group 1 versus 2	6.275 (1.076, 36.584)	0.0412*
	Gestational Age	0.632 (0.381, 1.050)	0.0762
	Ratio of mom's own milk to all milk	5.285 (0.077, 364.919)	0.4410

Table 15: Multivariate analysis of interleukin concentrations and continuous infant outcomes

Subgroup	Outcome	Predictor	Estimate	p-value
Caucasian	Lncalprotectin week 2	lnIL6 week 3	0.303	0.0290*
		Gestational Age	-0.009	0.9180
		Ratio of mom's own milk to all milk	-0.829	0.1984
African American	Lncalprotectin week 3	lnIL6 week 1	0.169	0.0794*
		Gestational Age	0.056	0.2941
		Ratio of mom's own milk to all milk	-0.129	0.7946
African American	Lncalprotectin week 3	lnIL6 week 2	0.137	0.0978*
		Gestational Age	0.031	0.5153
		Ratio of mom's own milk to all milk	-0.329	0.4946

*Note: only significant SNP/IL models were included

Table 16: Multivariate analysis of interleukin concentrations and continuous infant outcomes

Subgroup	Outcome	Variable	Odds Ratio	95% Confidence Interval	p
African American	IVH	lnIL6 week 1	0.329	(0.085, 1.266)	0.1059*
		Gestational Age	0.336	(0.085, 1.330)	0.1202
		Ratio of mom's own milk to all milk	0.033	(<0.001, 55.895)	0.3689
African American	IVH	lnIL6 week 2	0.278	(0.052, 1.496)	0.1362
		Gestational Age	0.246	(0.034, 1.754)	0.1616
		Ratio of mom's own milk to all milk	0.026	(<0.001, 77.223)	0.3724

*Note: only significant SNP/IL models were included

Table 17: Univariate analysis of maternal SNP and continuous infant outcomes

2070874				
Total Population	**CC** (n=48)	**TT** (n=9)	**CT** (n=14)	**p**
LnSNAPPEII	2.958 (±0.734)	2.659 (±0.610)	2.815 (±0.454)	0.4436
LOS	71.542 (±38.069)	67.667 (±28.579)	68.857 (±40.405)	0.9442
Weight at 6 weeks	7.496 (±0.177)	7.586 (±0.168)	7.515 (±0.192)	0.4002
Days on oxygen	15.083 (±21.534)	16.111 (±23.645)	14.923 (±20.601)	0.9903
lnCalprotectin Week 1	4.986 (±0.884)	5.081 (±0.599)	4.914 (±0.983)	0.9234
Ln Calprotectin Week 2	5.239 (±0.636)	5.467 (±0.877)	4.946 (±0.819)	0.2649
Ln Calprotectin Week 3	5.287 (±0.659)	5.168 (±0.308)	5.329 (±0.778)	0.8882
Caucasian	**CC** (n=20)	**TT** (n=1)	**CT (**n=1)	
LnSNAPPEII	2.833 (±0.762)	3.258 (n/a)	2.197 (n/a)	0.6168
LOS	61.15 (±14.865	1.23 (n/a)	4.7 (n/a)	0.0019*
Weight at 6 weeks	7.555 (±0.133)	7.269 (n/a)	1.138 (n/a)	0.0631*
Days on oxygen	11.555 (±15.856)	6.9 (n/a)	0 (n/a)	0.0065*
lnCalprotectin Week 1	5.07 (0.759)	5.09 (n/a)	6.19 (n/a)	0.3805
lnCalprotectin Week 2	5.334 (±0.574)	4.656 (n/a)	4.968 (n/a)	0.4627
lnCalprotectin Week 3	5.289 (±0.495)	4.905 (n/a)	5.803 (n/a)	0.4517
African American	**CC** (n=20)	**TT** (n=4)	**CT** (n=7)	
LnSNAPPEII	3.142 (±0.559)	2.61 (±0.826)	2.829 (±0.524)	0.2152
LOS	71.5 (±42.745)	60.25 (±24.047)	54.286 (±16.70)	0.5435
Weight at 6 weeks	7.372 (±0.158)	7.604 (±0.064)	7.446 (±0.246)	0.0091*
Days on oxygen	17.70 (±24.989)	9.5 (±13.82)	14.167 (±21.302)	0.7989
lnCalprotectin Week 1	4.776 (±0.925)	5.373 (±0.433)	5.257 (±0.862)	0.4165
lnCalprotectin Week 2	5.001 (±0.607)	5.489 (±0.726)	4.293 (±0.598)	0.0243*
lnCalprotectin Week 3	4.974 (±0.574)	5.024 (±0.079)	4.848 (±0.624)	0.8955
Hispanic	**CC** (n=7)	**TT** (n=3)	**CT** (n=4)	**p**
LnSNAPPEII	2.846 (±1.055)	2.677 (±0.417)	2.867 (±0.245)	0.9450
LOS	80.857 (±31.243)	68.0 (±20.075)	114.667 (±70.358)	0.3721
Weight at 6 weeks	7.532 (±0.231)	7.616 (±0.205)	7.469 (±0.102)	0.6370
Days on oxygen	19.714 (±27.134)	12.667 (±17.786)	32.0 (±25.339)	0.6435
lnCalprotectin Week 1	5.493 (±1.298)	4.878 (±1.101)	4.295 (±1.158)	0.6486
lnCalprotectin Week 2	5.929 (±0.669)	5.960 (±1.536)	5.429 (±0.446)	0.6486

lnCalprotectin Week 3	5.885 (±0.669)	5.204 (±0.347)	4.986 (±0.704)	0.2196

2070874 MAP

Total Population	**No** (n=48)	**Yes** (n=23)	**P**
LnSNAPPEII	2.958 (±0.734)	2.751 (±0.516)	0.2454
LOS	71.542 (±38.069)	68.391 (±35.526)	0.740
Weight at 6 weeks	7.496 (±0.177)	7.542 (±0.183)	0.3220
Days on oxygen	15.083 (±21.534)	15.409 (±21.351)	0.9532
lnCalprotectin Week 1	4.986 (±0.884)	4.976 (±0.848)	0.9679
Ln Calprotectin Week 2	5.239 (±0.636)	5.128 (±0.855)	0.5870
Ln Calprotectin Week 3	5.287 (±0.659)	5.273 (±0.634)	0.9391

Caucasian	**No** (n=20)	**Yes** (n=2)	**P**
LnSNAPPEII	2.8333 (±0.762)	2.728 (±0.75)	0.8569
LOS	61.15 (±14.865)	85.0 (±53.74)	0.103*
Weight at 6 weeks	7.555 (±0.133)	7.504 (±0.332)	0.651
Days on oxygen	11.55 (±15.856)	34.5 (±48.79)	0.1175*
lnCalprotectin Week 1	5.076 (±0.759)	5.642 (±0.778)	0.3319
lnCalprotectin Week 2	5.334 (±0.574)	4.812 (±0.220)	0.2298
lnCalprotectin Week 3	5.289 (±0.495)	5.354 (±0.635)	0.8558

African American	**No** (n=20)	**Yes** (n=11)	**p**
LnSNAPPEII	3.142 (±0.559)	2.742 (±0.627)	0.0941*
LOS	71.5 (±42.745)	56.455 (±18.704)	0.2786
Weight at 6 weeks	7.37 (±0.157)	7.509 (±0.204)	0.0902*
Days on oxygen	17.7 (±24.989)	12.3 (±17.932)	0.5485
lnCalprotectin Week 1	4.776 (±0.925)	5.307 (±0.662)	0.1842*
lnCalprotectin Week 2	5.001 (±0.607)	4.691 (±0.844)	0.3079
lnCalprotectin Week 3	4.974 (±0.574)	4.898 (±0.518)	0.767

Hispanic	**No** (n=7)	**Yes** (n=7)	**p**
LnSNAPPEII	2.846 (±1.055)	2.786 (±0.313)	0.8883
LOS	80.857 (±31.243)	91.333 (±52.864)	0.6660
Weight at 6 weeks	7.532 (±0.232)	7.532 (±0.159)	0.9979

Days on oxygen	19.714 (±27.134)	22.333 (±22.214)	0.8542	
lnCalprotectin Week 1	5.493 (±1.298)	4.489 (±1.067)	0.2166	
lnCalprotectin Week 2	5.929 (±0.638)	5.606 (±0.816)	0.5715	
lnCalprotectin Week 3	5.885 (±0.669)	5.095 (±0.470)	0.0764*	

2243250

Total Population	**CC** (n=31)	**TT** (n=21)	**CT** (n=19)	**p**
LnSNAPPEII	2.95 (±0.811)	2.848 (±0.546)	2.816 (±0.551)	0.7922
LOS	74.581 (±35.029)	60.381 (±25.593)	75.105 (±48.908)	0.3314
Weight at 6 weeks	7.495 (±0.196)	7.497 (±0.187)	7.559 (±0.138)	0.4693
Days on oxygen	18.258 (±25.048)	14.19 (±19.372)	11.056 (±16.126)	0.5121
lnCalprotectin Week 1	5.053 (±0.916)	5.116 (±0.768)	4.765 (±0.880)	0.4379
Ln Calprotectin Week 2	5.258 (±0.699)	5.162 (±0.842)	5.136 (±0.619)	0.8620
Ln Calprotectin Week 3	5.403 (±0.655)	50.35 (±0.497)	5.347 (±0.743)	0.190*
Caucasian	**CC** (n=17)	**TT** (n=2)	**CT** (n=3)	**p**
LnSNAPPEII	2.882 (±0.769)	3.258 (n/a)	2.197 (n/a)	0.4189
LOS	61.765 (±15.335)	92.5 (±43.134)	52.667 (±14.364)	0.0597*
Weight at 6 weeks	7.56 (±0.142)	7.373 (±0.147)	7.616 (±0.126)	0.171*
Days on oxygen	12.645 (±16.985)	39.0 (±42.426)	2.333 (±2.517)	0.1093*
lnCalprotectin Week 1	5.107 (±0.818)	4.858 (±0.33)	5.447 (±0.724)	0.6971
lnCalprotectin Week 2	5.44 (±0.531)	4.607 (±0.069)	4.797 (±0.242)	0.0573*
lnCalprotectin Week 3	5.319 (±0.485)	4.756 (±0.209)	5.558 (±0.461)	0.1974*
African American	**CC** (n=8)	**TT** (n=15)	**CT** (n=8)	**p**
LnSNAPPEII	3.186 (±0.633)	2.822 (±0.598)	3.163 (±0.557)	0.3112
LOS	96.625 (±53.774)	57.333 (±24.127)	52.25 (±13.562)	0.0163*
Weight at 6 weeks	7.274 (±0.151)	7.463 (±0.188)	7.515 (±0.162)	0.078*
Days on oxygen	29.125 (±34.428)	14.333 (±16.443)	4.143 (±8.821)	0.0939*
lnCalprotectin Week 1	4.55 (±0.823)	5.407 (±0.799)	4.632 (±0.817)	0.0730*
lnCalprotectin Week 2	4.738 (±0.652)	5.02 (±0.851)	4.854 (±0.557)	0.7134
lnCalprotectin Week 3	5.15 (±0.446)	4.946 (±0.429)	4.727 (±0.455)	0.3976
Hispanic	**CC** (n=6)	**TT** (n=3)	**CT** (n=5)	**p**

LnSNAPPEII	2.784 (±1.141)	2.848 (±0.124)	2.836 (±0.425)	0.9915
LOS	81.5 (±34.175)	61.0 (±22.627)	100.6 (±53.374)	0.5257
Weight at 6 weeks	7.504 (±0.241)	7.561 (±0.192)	7.547 (±0.159)	0.9069
Days on oxygen	19.667 (±29.723)	2.5 (±3.536)	29.8 (±18.674)	0.4265
lnCalprotectin Week 1	5.961 (±1.102)	4.594 (±0.699)	4.367 (±1.147)	0.1891*
lnCalprotectin Week 2	6.059 (±0.843)	5.841 (±1.105)	5.445 (±0.456)	0.6533
lnCalprotectin Week 3	5.976 (±0.705)	4.969 (±0.679)	5.292 (±0.289)	0.163*

1800795

Total Population	**CC** (n=5)	**GG** (n=45)	**CG** (n=21)	**p**
LnSNAPPEII	3.369 (±0.305)	2.969 (±0.652)	2.649 (±0.679)	0.0944*
LOS	67 (±36.083)	70.911 (±39.183)	70.524 (±34.064)	0.9759
Weight at 6 weeks	7.438 (±0.135)	7.508 (±0.188)	7.541 (±0.167)	0.5591
Days on oxygen	21.4 (±39.905)	11.886 (±17.362)	20.619 (±26.012)	0.2444
lnCalprotectin Week 1	5.044 (±0.282)	5.005 (±0.797)	4.922 (±1.091)	0.9381
Ln Calprotectin Week 2	5.534 (±0.635)	5.094 (±0.81)	5.337 (±0.476)	0.3424
Ln Calprotectin Week 3	5.2631 (±0.349)	5.295 (±0.701)	5.258 (±0.589)	0.9809
Caucasian	**CC** (n=5)	**GG** (n=7)	**CG** (n=10)	**p**
LnSNAPPEII	3.369 (±0.305)	2.446 (±0.431)	2.768 (±0.831)	0.3068
LOS	67 (±36.083)	63 (±17.776)	61.7 (±10.12)	0.8946
Weight at 6 weeks	7.438 (±0.135)	7.57 (±0.103)	7.581 (±0.168)	0.2486
Days on oxygen	21.4 (±29.9)	4.429 (±4.276)	16.2 (±19.629)	0.3022
lnCalprotectin Week 1	5.044 (±0.282)	5.012 (±0.597)	5.266 (±1.019)	0.7941
lnCalprotectin Week 2	5.543 (±0.635)	4.87 (±0.495)	5.452 (±0.467)	0.0901*
lnCalprotectin Week 3	5.263 (±0.349)	5.291 (±0.546)	5.313 (±0.540)	0.9896
African American	**CC** (n=0)	**GG** (n=25)	**CG** (n=6)	**p**
LnSNAPPEII	0	3.078 (±0.614)	2.637 (±0.449)	0.1437*
LOS	0	68.76 (±38.285)	55.333 (±27.457)	0.4269
Weight at 6 weeks	0	7.442 (±0.196)	7.383 (±0.164)	0.6238
Days on oxygen	0	14.452 (±17.088)	21.333 (±40.038)	0.5218
lnCalprotectin Week 1	0	4.948 (±0.818)	4.848 (±1.289)	0.8402
lnCalprotectin Week 2	0	4.783 (±0.749)	5.191 (±0.467)	0.2259

lnCalprotectin Week 3	0	4.914 (±0.513)	5.082 (±0.707)	0.5557
Hispanic	**CC** (n=0)	**GG** (n=10)	**CG** (n=4)	**p**
LnSNAPPEII	**0**	2.979 (±0.746)	2.406 (±0.665)	0.2066
LOS	0	74.889 (±29.817)	110.0 (±56.739)	0.1627*
Weight at 6 weeks	0	7.534 (±0.216)	7.526 (±0.164)	0.9434
Days on oxygen	0	15.778 (±24.748)	32.5 (±20.273)	0.2634
lnCalprotectin Week 1	0	5.316 (±1.132)	4.254 (±1.162)	0.1884*
lnCalprotectin Week 2	0	5.829 (±0.753)	5.311 (±0.732)	0.4175
lnCalprotectin Week 3	0	5.708 (±0.719)	5.013 (±0.077)	0.2278

1800795 MAP

Total Population	**No** (n=45)	**Yes** (n=26)	**p**
LnSNAPPEII	2.969 (±0.653)	2.743 (±0.684)	0.2005*
LOS	70.911 (±39.183)	69.846 (±33.743)	0.9081
Weight at 6 weeks	7.508 (±0.188)	7.523 (±0.164)	0.7541
Days on oxygen	11.886 (±17.362)	20.769 (±26.166)	0.0923*
lnCalprotectin Week 1	5.005 (±0.797)	4.944 (±0.989)	0.7969
Ln Calprotectin Week 2	5.094 (±0.81)	5.376 (±0.499)	0.1657*
Ln Calprotectin Week 3	5.295 (±0.701)	5.259 (±0.550)	0.8440
Caucasian	**No** (n=7)	**Yes** (n=15)	**p**
LnSNAPPEII	2.446 (±0.431)	2.906 (±0.776)	0.3465
LOS	63 (±17.776)	63.467 (±21.084)	0.9601
Weight at 6 weeks	7.57 (±0.103)	7.54 (±0.168)	0.6716
Days on oxygen	4.429 (±4.276)	17.933 (±22.575)	0.1368*
lnCalprotectin Week 1	5.012 (±0.597)	5.198 (±0.851)	0.616
lnCalprotectin Week 2	4.87 (±0.495)	5.479 (±0.501)	0.0267*
lnCalprotectin Week 3	5.291 (±0.546)	5.299 (±0.478)	0.9732
African American	**No** (n=25)	**Yes** (n=6)	**p**
LnSNAPPEII	3.078 (±0.614)	2.637 (±0.449)	0.1437*
LOS	68.76 (±38.285)	55.333 (±27.457)	0.4269
Weight at 6 weeks	7.442 (±0.196)	7.383 (±0.164)	0.6238

Days on oxygen	14.452 (±17.088)	21.333 (±40.038)	0.5218
lnCalprotectin Week 1	4.948 (±0.818)	4.848 (±1.289)	0.8402
lnCalprotectin Week 2	4.783 (±0.749)	5.191 (±0.467)	0.2259
lnCalprotectin Week 3	4.914 (±0.513)	5.082 (±0.707)	0.5557
Hispanic	**No** (n=10)	**Yes** (n=4)	**p**
LnSNAPPEII	2.979 (±0.746)	2.406 (±0.665)	0.2066
LOS	74.889 (±29.817)	110.0 (±56.739)	0.1627*
Weight at 6 weeks	7.534 (±0.216)	7.526 (±0.164)	0.9434
Days on oxygen	15.778 (±24.748)	32.5 (±20.273)	0.2634
lnCalprotectin Week 1	5.316 (±1.132)	4.254 (±1.162)	0.1884*
lnCalprotectin Week 2	5.829 (±0.753)	5.311 (±0.732)	0.4175
lnCalprotectin Week 3	5.708 (±0.719)	5.013 (±0.077)	0.2278

1800796

Total Population	**GG** (n=3)	**AA** (n=58)	**AG** (n=11)	**p**
LnSNAPPEII	3.105 (±0.998)	2.845 (±0.659)	3.028 (±0.665)	0.6419
LOS	146 (±97.581)	67.569 (±28.194)	72.364 (±54.099)	0.0110*
Weight at 6 weeks	7.491 (±0.152)	7.498 (±0.183)	7.615 (±0.128)	0.2260
Days on oxygen	3.0 (±2.8285)	16.105 (±22.457)	12.636 (±16.439)	0.6389
lnCalprotectin Week 1	5.287 (±1.437)	4.979 (±0.897)	4.907 (±0.545)	0.8046
Ln Calprotectin Week 2	4.874 (n/a)	5.158 (±0.684)	5.382 (±0.876)	0.5996
Ln Calprotectin Week 3	6.194 (±0.656)	5.211 (±0.615)	5.319 (±0.639)	0.036*
Caucasian	**GG** (n=0)	**AA** (n=20)	**AG** (n=2)	**p**
LnSNAPPEII	0	2.802 (±0.758)	3.09 (n/a)	0.7176
LOS	0	64.4 (±20.109)	52.5 (±13.435)	0.4279
Weight at 6 weeks	0	7.559 (±0.153)	7.469 (±0.01)	0.4273
Days on oxygen	0	14.55 (±20.379)	4.5 (±6.364)	0.504
lnCalprotectin Week 1	0	5.139 (±0.789)	5.077 (±0.6396)	0.916
lnCalprotectin Week 2	0	5.326 (±0.572)	4.875 (±0.449)	0.303
lnCalprotectin Week 3	0	5.352 (±0.484)	4.851 (±0.343)	0.1805*
African American	**GG** (n=0)	**AA** (n=27)	**AG** (n=4)	**p**

LnSNAPPEII	0	2.921 (±0.564)	3.649 (±0.645)	0.0467*
LOS	0	60.963 (±21.517)	101.25 (±85.885)	0.0367*
Weight at 6 weeks	0	7.434 (± 0.189)	0 (n/a)	n/a
Days on oxygen	0	15.384 (±23.233)	19.25 (±21.654)	0.7574
lnCalprotectin Week 1	0	4.969 (±0.926)	4.663 (±0.400)	0.5840
lnCalprotectin Week 2	0	4.879 (±0.690)	4.910 (±0.886)	0.9384
lnCalprotectin Week 3	0	4.867 (±0.494)	5.511 (±0.650)	0.0547*
Hispanic	**GG** (n=2)	**AA** (n=9)	**AG** (n=4)	**p**
LnSNAPPEII	3.559 (±0.869)	2.661 (±0.816)	2.754 (±0.427)	0.3345
LOS	77.0 (n/a)	92.556 (±47.355)	68 (±20.075)	0.6910
Weight at 6 weeks	7.383 (n/a)	7.485 (±0.200P)	7.675 (±0.097)	0.1893*
Days on oxygen	5.0 (n/a)	23.778 (±27.376)	17.667 (±16.623)	0.7659
lnCalprotectin Week 1	5.493 (±1.969)	4.618 (±1.315)	4.945 (±0.794)	0.7356
lnCalprotectin Week 2	4.874 (n/a)	5.647 (±0.678)	6.105 (±0.816)	0.3848
lnCalprotectin Week 3	5.949 (±0.708)	5.783 (±0.675)	4.959 (±0.472)	0.1976*

1800796 MAP

Total Population	**No** (n=58)	**Yes** (n=13)	**p**
LnSNAPPEII	2.846 (±1.609)	3.047 (±0.709)	0.3526
LOS	67.569 (±28.195)	83.692 (±63.222)	0.1575*
Weight at 6 weeks	7.498 (±0.184)	7.59 (±0.131)	0.1370*
Days on oxygen	16.105 (±22.457)	11.154 (±15.459)	0.4539
lnCalprotectin Week 1	4.979 (±0.898)	4.995 (±0.771)	0.9527
Ln Calprotectin Week 2	5.158 (±0.684)	5.339 (±0.848)	0.4457
Ln Calprotectin Week 3	5.211 (±0.615)	5.521 (±0.725)	0.1319*
Caucasian	**No** (n=20)	**Yes** (n=2)	**p**
LnSNAPPEII	2.802 (±0.758)	3.091 (n/a)	0.7176
LOS	64.40 (±20.109)	52.5 (±13.43)	0.4279
Weight at 6 weeks	7.559 (±0.153)	7.469 (±0.015)	0.4273
Days on oxygen	14.55 (±20.379)	1.5 (±6.364)	0.504
lnCalprotectin Week 1	5.139 (±0.789)	5.077 (±0.639)	0.916
lnCalprotectin Week 2	5.326 (±0.572)	4.875 (±0.449)	0.303

lnCalprotectin Week 3	5.352 (±0.484)	4.851 (±0.343)	0.1805*
African American	**No** (n=27)	**Yes** (n=4)	**p**
LnSNAPPEII	2.921 (±0.564)	3.649 (±0.645)	0.0467*
LOS	60.963 (±21.517)	101.25 (±85.885)	0.0367*
Weight at 6 weeks	7.434 (± 0.189)	0 (n/a)	n/a
Days on oxygen	15.384 (±23.233)	19.25 (±21.654)	0.7574
lnCalprotectin Week 1	4.969 (±0.926)	4.663 (±0.400)	0.5840
lnCalprotectin Week 2	4.879 (±0.690)	4.910 (±0.886)	0.9384
lnCalprotectin Week 3	4.867 (±0.494)	5.511 (±0.650)	0.0547*
Hispanic	**No** (n=9)	**Yes** (n=6)	**p**
LnSNAPPEII	2.661 (±0.816)	3.022 (±0.659)	0.3925
LOS	92.556 (±47.355)	70.25 (±16.998)	0.3886
Weight at 6 weeks	7.485 (±0.200)	7.616 (±0.155)	0.231
Days on oxygen	23.778 (±27.376)	14.5 (±14.9778)	0.5434
lnCalprotectin Week 1	4.618 (±1.315)	5.164 (±1.173)	0.5079
lnCalprotectin Week 2	5.647 (±0.678)	5.797 (±0.906)	0.7827
lnCalprotectin Week 3	5.783 (±0.675)	5.356 (±0.729)	0.3645

1800871

Total Population	**TT** (n=7)	**CC** (n=40)	**CT** (n=24)	**p**
LnSNAPPEII	3.062 (±0.643)	2.894 (±0.691)	2.804 (±0.672)	0.7082
LOS	58.667 (±17.489)	75.075 (±38.711)	67.792 (±37.383)	0.5189
Weight at 6 weeks	7.534 (±0.178)	7.501 (±0.181)	7.528 (±0.182)	0.8358
Days on oxygen	8.333 (±7.005)	18.0 (±24.515)	12.917 (±17.90)	0.4668
lnCalprotectin Week 1	5.199 (±0.969)	4.927 (±0.783)	5.204 (±4.046)	0.7453
Ln Calprotectin Week 2	5.507 (±0.560)	5.228 (±0.665)	4.976 (±0.892)	0.3050
Ln Calprotectin Week 3	5.849 (±0.539)	5.274 (±0.631)	5.149 (±0.654)	0.0980*
Caucasian	**TT** (n=1)	**CC** (n=14)	**CT** (n=6)	**p**
LnSNAPPEII	2.89 (n/a)	2.824 (±0.729)	2.714 (±1.033)	0.9676
LOS	57 (n/a)	64.74 (±20.447)	67.5 (±14.46)	0.8686
Weight at 6 weeks	7.719 (n/a)	7.533 (±0.146)	7.563 (±0.158)	0.4838

Days on oxygen	10 (n/a)	12.429 (±19.782)	5.625 (±1.004)	0.7863
lnCalprotectin Week 1	5.473 (n/a)	4.948 (±0.647)	5.625 (±1.004)	0.2411
lnCalprotectin Week 2	5.881 (n/a)	5.363 (±0.538)	4.671 (±0.259)	0.079*
lnCalprotectin Week 3	n/a	5.333 (±0.466)	5.198 (±0.594)	0.6133
African American	**TT** (n=1)	**CC** (n=17)	**CT** (n=13)	**p**
LnSNAPPEII	n/a	3.094 (±0.603)	2.934 (±0.60)	0.3324
LOS	36.0 (n/a)	77.118 (±44.084)	54.154 (±18.032)	0.1640*
Weight at 6 weeks	n/a	7.435 (±0.20)	7.433 (±0.186)	0.9743
Days on oxygen	4.0 (n/a)	19.562 (±27.043)	12.307 (±16.977)	0.6178
lnCalprotectin Week 1	4.874 (n/a)	5.068 (±0.769)	4.681 (±1.111)	0.6211
lnCalprotectin Week 2	4.857 (n/a)	4.956 (±0.712)	4.754 (±0.764)	0.8206
lnCalprotectin Week 3	n/a	4.972 (±0.596)	4.982 (±0.514)	0.8493
Hispanic	**TT** (n=4)	**CC** (n=8)	**CT** (n=3)	**p**
LnSNAPPEII	3.321 (±0.643)	2.894 (±0.691)	2.804 (±0.672)	0.7082
LOS	58.667 (±17.489)	75.075 (±38.711)	670792 (±37.383)	0.5189
Weight at 6 weeks	7.534 (±0.178)	7.501 (±0.181)	7.528 (±0.182)	0.8358
Days on oxygen	8.333 (±7.005)	18.0 (±24.515)	12.917 (±17.9)	0.4668
lnCalprotectin Week 1	5.199 (±0.969)	4.927 (±0.783)	5.024 (±1.046)	0.7453
lnCalprotectin Week 2	5.507 (±0.560)	5.228 (±0.665)	4.976 (±0.892)	0.3050
lnCalprotectin Week 3	5.849 (±0.539)	5.274 (±0.631)	5.149 (±0.654)	0.0980*

1800871 MAP

Total Population	**No** (n=40)	**Yes** (n=31)	**p**
LnSNAPPEII	2.894 (±0.691)	2.863 (±0.662)	0.8621
LOS	75.075 (±38.71)	65.967 (±34.276)	0.3102
Weight at 6 weeks	7.501 (±0.181)	7.529 (±0.178)	0.5497
Days on oxygen	18 (±24.515)	12 (±16.312)	0.2514
lnCalprotectin Week 1	4.927 (±0.783)	5.077 (±1.004)	0.5263
Ln Calprotectin Week 2	5.228 (±0.665)	5.143 (±0.826)	0.6830
Ln Calprotectin Week 3	5.274 (±0.631)	5.295 (±0.686)	0.9065
Caucasian	**No** (n=14)	**Yes** (n=7)	**p**

LnSNAPPEII	2.824 (±0.729)	2.749 (±0.897)	0.8637
LOS	64.714 (±20.447)	66 (±13.784)	0.8829
Weight at 6 weeks	7.533 (±0.146)	7.585 (±0.156)	0.4536
Days on oxygen	12.429 (±19.782)	17.857 (±21.302)	0.5698
lnCalprotectin Week 1	4.948 (±0.647)	5.599 (±0.90)	0.0885*
lnCalprotectin Week 2	5.363 (±0.538)	4.973 (±0.641)	0.2368
lnCalprotectin Week 3	5.334 (±0.466)	5.198 (±0.594)0.	0.6133
African American	**No** (n=17)	**Yes** (n=14)	**p**
LnSNAPPEII	3.094 (±0.603)	2.872 (±0.61)	0.3469
LOS	77.118 (±44.084)	52.857 (±17.991)	0.0639
Weight at 6 weeks	7.435 (±0.20)	7.433 (±0.186)	0.9743
Days on oxygen	19.562 (±27.043)	11.714 (±16.462)	0.3539
lnCalprotectin Week 1	5.068 (±0.769)	4.703 (±1.042)	0.339
lnCalprotectin Week 2	4.956 (±0.712)	4.766 (±0.716)	0.5338
lnCalprotectin Week 3	4.972 (±0.597)	4.928 (±0.514)	0.8493
Hispanic	**No** (n=8)	**Yes** (n=7)	**p**
LnSNAPPEII	2.693 (±0.852)	2.938 (±0.612)	0.5618
LOS	93.286 (±52.554)	76.833 (±23.404)	0.495
Weight at 6 weeks	7.492 (±0.198)	7.584 (±0.183)	0.3939
Days on oxygen	28.857 (±29.249)	11.667 (±12.801)	0.211
lnCalprotectin Week 1	4.541 (±1.263)	5.241 (±1.179)	0.3919
lnCalprotectin Week 2	5.668 (±0.673)	5.769 (±0.916)	0.853
lnCalprotectin Week 3	5.584 (±0.896)	5.555 (±0.545)	0.9530

1800872

Total Population	**CC** (n=39)	**AA** (n=9)	**AC** (n=24)	**p**
LnSNAPPEII	2.883 (±0.713)	2.963 (±0.626)	2.878 (±0.653)	0.9505
LOS	77.684 (±37.757)	55.75 (±19.64)	66.042 (±38.589)	0.2201
Weight at 6 weeks	7.507 (±0.172)	7.562 (±0.160)	7.521 (±0.202)	0.9474
Days on oxygen	16.703 (±20.513)	7 (±6.503)	16.208 (±25.578)	0.5023
lnCalprotectin Week 1	4.903 (±0.787)	5.169 (±0.858)	5.091 (±1.071)	0.6389
Ln Calprotectin Week 2	5.173 (±0.689)	5.507 (±0.560)	5.131 (±0.844)	0.5367

Ln Calprotectin Week 3	5.272 (±0.652)	5.848 (±0.482)	5.117 (±0.634)	0.0552*
Caucasian	**CC** (n=15)	**AA** (n=3)	**AC** (n=4)	**p**
LnSNAPPEII	2.824 (±0.789)	2.741 (±0.487)	2.886 (±1.192)	0.9748
LOS	67.73 (±19.692)	50.333 (±22.745)	56.5 (±14.978)	0.2931
Weight at 6 weeks	7.523 (±0.139)	7.602 (±0.164)	7.627 (±0.18)	0.4163
Days on oxygen	12.8 (±18.88)	5.333 (±4.509)	23 (±28.959)	0.5019
lnCalprotectin Week 1	4.933 (±0.626)	5.2 (±0.412)	5.99 (±1.169)	0.0798*
lnCalprotectin Week 2	5.299 (±0.572)	5.881 (n/a)	4.797 (±0.242)	0.2891
lnCalprotectin Week 3	5.258 (±0.509)	5.845 (n/a)	5.289 (±0.445)	0.5415
African American	**CC** (n=14)	**AA** (n=1)	**AC** (n=5)	**p**
LnSNAPPEII	3.113 (±0.646)	2.197 (n/a)	2.968 (±0.563)	0.3512
LOS	82.143 (±44.487)	36.0 (n/a)	56.067 (±21.215)	0.1001*
Weight at 6 weeks	7.458 (±0.193)	(n/a)	7.411 (±0.191)	0.5767
Days on oxygen	16.231 (±16.996)	4.0 (n/a)	17.467 (±28.099)	0.8587
lnCalprotectin Week 1	5.042 (±0.813)	4.874 (n/a)	4.839 (±1.075)	0.8767
lnCalprotectin Week 2	4.804 (±0.718)	4.857 (n/a)	4.976 (±0.746)	0.8534
lnCalprotectin Week 3	4.978 (±0.661)	(n/a)	4.901 (±0.49)	0.7586
Hispanic	**CC** (n=8)	**AA** (n=4)	**AC** (n=3)	**p**
LnSNAPPEII	2.693 (±0.852)	3.321 (±0.584)	2.428 (±0.400)	0.2614
LOS	93.286 (±52.54)	72.667 (±7.505)	81.0 (±35.511)	0.780
Weight at 6 weeks	7.492 (±0.198)	7.468 (±0.199)	7.7 (±0.065)	0.2350
Days on oxygen	28.857 (±29.249)	12.0 (±7.549)	11.333 (±18.771)	0.4750
lnCalprotectin Week 1	4.541 (±1.263)	5.137 (±1.335)	5.656 n/a	0.667
lnCalprotectin Week 2	5.668 (±0.673)	5.345 (±0.417)	7.046 n/a	0.1220*
lnCalprotectin Week 3	5.584 (±0.896)	5.704 (±0.498)	4.956 n/a	0.6875

1800872 MAP

Total Population	**No** (n=39)	**Yes** (n=33)	**p**
LnSNAPPEII	2.883 (±0.713)	2.901 (±0.636)	0.9157
LOS	77.684 (±37.757)	63.469 (±34.82)	0.1087*
Weight at 6 weeks	7.507 (±0.172)	7.522 (±0.19)	0.7454

Days on oxygen	16.703 (±20.513)	13.906 (±22.614)	0.5920
lnCalprotectin Week 1	4.903 (±0.787)	5.119 (±0.977)	0.3540
Ln Calprotectin Week 2	5.173 (±0.689)	5.238 (±0.779)	0.7484
Ln Calprotectin Week 3	5.272 (±0.652)	5.293 (±0.672)	0.9105
Caucasian	**No** (n=15)	**Yes** (n=7)	**p**
LnSNAPPEII	2.824 (±0.0729)	2.813 (±0.818)	0.9785
LOS	67.733 (±19.692)	53.857 (±17.189)	0.1258*
Weight at 6 weeks	7.523 (±0.138)	7.619 (±0.158)	0.1841*
Days on oxygen	12.8 (±18.88)	15.429 (±22.699)	0.7781
lnCalprotectin Week 1	4.933 (±0.625)	5.599 (±0.897)	0.0707
lnCalprotectin Week 2	5.299 (±0.572)	5.158 (±0.649)	0.7051
lnCalprotectin Week 3	5.258 (±0.509)	5.428 (±0.457)	0.5583
African American	**No** (n=14)	**Yes** (n=16)	**p**
LnSNAPPEII	3.113 (±0.646)	2.93 (±0.58)	0.4464
LOS	82.143 (±44.487)	54.813 (±21.1)	0.0366*
Weight at 6 weeks	7.458 (±0.194)	7.41 (±0.191)	0.5767
Days on oxygen	16.231 (±16.996)	16.625 (±27.354)	0.9642
lnCalprotectin Week 1	5.042 (±0.813)	4.842 (±1.014)	0.6043
lnCalprotectin Week 2	4.804 (±0.718)	4.965 (±0.712)	0.5839
lnCalprotectin Week 3	4.978 (±0.661)	4.901 (±0.49)	0.7586
Hispanic	**No** (n=8)	**Yes** (n=7)	**p**
LnSNAPPEII	2.693 (±0.852)	2.938 (±0.672)	0.5618
LOS	93.286 (±52.554)	76.833 (±23.404)	0.4950
Weight at 6 weeks	7.492 (±0.198)	7.584 (±0.183)	0.3939
Days on oxygen	28.857 (±29.249)	11.667 (±12.801)	0.2110
lnCalprotectin Week 1	4.541 (±1.263)	5.241 (±1.179)	0.3919
lnCalprotectin Week 2	5.668 (±0.673)	5.769 (±0.916)	0.8530
lnCalprotectin Week 3	5.584 (±0.896)	5.555 (±0.545)	0.9530

1800896				
Total Population	**GG** (n=13)	**AA** (n=32)	**AG** (n=34)	**p**
LnSNAPPEII	2.803 (±0.472)	2.868 (±0.766)	2.914 (±0.678)	0.9004
LOS	76.75 (±38.581)	59.125 (±23.184)	75.735 (±43.461)	0.1987*

Weight at 6 weeks	7.529 (±0.163)	7.542 (±0.179)	7.496 (±0.181)	0.6594
Days on oxygen	15.333 (±18.593)	17.833 (±27.142)	13.0 (±17.897)	0.7084
lnCalprotectin Week 1	4.741 (±1.031)	5.115 (±0.766)	4.992 (±0.867)	0.5031
Ln Calprotectin Week 2	5.106 (±0.689)	5.631 (±0.709)	4.981 (±0.661)	0.0164*
Ln Calprotectin Week 3	5.227 (±0.584)	5.499 (±0.669)	5.145 (±0.648)	0.1896*
Caucasian	**GG** (n=5)	**AA** (n=8)	**AG** (n=9)	**p**
LnSNAPPEII	2.563 (±0.634)	2.769 (±0.864)	3.008 (±0.694)	0.7036
LOS	62.2 (±10.257)	56.75 (±16.59)	69.77 (±25.114)	0.4103
Weight at 6 weeks	7.62 (±0.103)	7.554 (±0.119)	7.508 (±0.183)	0.4152
Days on oxygen	12 (±15.443)	13.75 (±21.393)	14.444 (±22.187)	0.9776
lnCalprotectin Week 1	5.034 (±1.014)	4.896 (±0.557)	5.361 (±0.802)	0.484
lnCalprotectin Week 2	5.415 (±0.369)	5.667 (±0.54)	4.946 (±0.537)	0.0573*
lnCalprotectin Week 3	5.632 (±0.451)	5.479 (±0.379)	4.99 (±0.441)	0.0436*
African American	**GG** (n=5)	**AA** (n=6)	**AG** (n=20)	**p**
LnSNAPPEII	2.953 (±0.505)	2.947 (±0.649)	3.022 (±0.641)	0.9598
LOS	66.0 (±8.155)	55.83 (±28.673)	69.3 (±42.678)	0.7427
Weight at 6 weeks	7.447 (±0.222)	7.329 (±0.142)	7.452 (±0.191)	0.6048
Days on oxygen	7.8 (±10.257)	27.667 (±38.79)	14.316 (±17.954)	0.3212
lnCalprotectin Week 1	5.009 (±0.951)	5.093 (±0.631)	4.845 (±0.971)	0.8539
lnCalprotectin Week 2	4.493 (±0.782)	4.969 (±0.406)	4.962 (±0.729)	0.4962
lnCalprotectin Week 3	5.006 (±0.491)	4.697 (±0.462)	5.004 (±0.591)	0.6142
Hispanic	**GG** (n=3)	**AA** (n=9)	**AG** (n=2)	**p**
LnSNAPPEII	2.842 (±0.293)	2.989 (±0.836)	1.903(±0.416)	0.2131
LOS	140.0 (±77.782)	68.5 (±26.333)	97.0 (±26.87)	0.0843*
Weight at 6 weeks	7.516 (±0.045)	7.586 (±0.204)	7.512 (±0.174)	0.7861
Days on oxygen	42.5 (±24.749)	19.0 (±26.338)	6.5 (±0.7.778)	0.3677
lnCalprotectin Week 1	3.903 (±1.043)	5.409 (±1.133)	4.742 n/a	0.2250
lnCalprotectin Week 2	5.408 (±0.544)	5.947 (±0.886)	n/a	0.3855
lnCalprotectin Week 3	4.488 n/a	5.823 (±0.644)	5.067 n/a	0.1869*

1800896 MAP			
Total Population	**No** (n=25)	**Yes** (n=47)	**p**
LnSNAPPEII	2.868 (±0.766)	2.886 (±0.629)	0.920

LOS	59.125 (±23.84)	76.0 (±41.823)	0.0715*
Weight at 6 weeks	7.542 (±0.179)	7.507 (±0.174)	0.4676
Days on oxygen	17.833 (±27.142)	13.622 (±17.902)	0.4417
lnCalprotectin Week 1	5.115 (±0.766)	4.915 (±0.914)	0.4051
Ln Calprotectin Week 2	5.631 (±0.709)	5.021 (±0.663)	0.0046*
Ln Calprotectin Week 3	5.49 (±0.668)	5.166 (±0.626)	0.0715*
Caucasian	**No** (n=8)	**Yes** (n=14)	**p**
LnSNAPPEII	2.769 (±0.864)	2.859 (±0.672)	0.8162
LOS	56.75 (±16.593)	67.071 (±20.849)	0.2455
Weight at 6 weeks	7.554 (±0.119)	7.548 (±0.163)	0.9376
Days on oxygen	13.75 (±21.393)	13.571 (±19.437)	0.9842
lnCalprotectin Week 1	4.896 (±0.557)	5.26 (±0.843)	0.3203
lnCalprotectin Week 2	5.667 (±0.54)	5.126 (±0.519)	0.0681*
lnCalprotectin Week 3	5.479 (±0.379)	5.205 (±0.528)	0.2766
African American	**No** (n=6)	**Yes** (n=25)	**p**
LnSNAPPEII	2.947 (±0.649)	3.01 (±0.61)	0.8367
LOS	55.833 (±28.673)	68.64 (±38.142)	0.4487
Weight at 6 weeks	7.329 (±0.142)	7.451 (±0.194)	0.3103
Days on oxygen	27.667 (±38.79)	0.12.958 (±16.669)	0.1594*
lnCalprotectin Week 1	5.093 (±0.631)	4.889 (±0.942)	0.6533
lnCalprotectin Week 2	4.969 (±0.406)	4.873 (±0.743)	0.8309
lnCalprotectin Week 3	4.697 (±0.462)	5.004 (±0.559)	0.3174
Hispanic	**No** (n=9)	**Yes** (n=6)	**p**
LnSNAPPEII	2.989 (±0.836)	2.466 (±0.592)	0.2510
LOS	68.5 (±26.333)	118.5 (±53.607)	0.0503*
Weight at 6 weeks	7.586 (±0.204)	7.515 (±0.093)	0.4771
Days on oxygen	19.0 (±26.338)	24.5 (±25.619)	0.7381
lnCalprotectin Week 1	5.049 (±1.133)	4.113 (±0.949)	0.0966*
lnCalprotectin Week 2	5.947 (±0.887)	5.408 (±0.543)	0.3855
lnCalprotectin Week 3	5.822 (±0.644)	4.778 (±0.409)	0.0723*

Table 18: Univariate analysis of maternal SNP and categorical infant outcomes

2070874				
Total Population	**CC** (n=49)	**TT** (n=9)	**CT** (n=15)	**p**
Sepsis				
Yes	5 (10.2%)	3 (33.3%)	2 (13.3%)	0.218
No	42 (85.7%)	6 (66.7%)	13 (86.7%)	
ROP				
Yes	7 (14.3%)	4 (44.4%)	2 (13.3%)	0.132*
No	37 (75.6%)	5 (55.6%)	13 (86.7%)	
BPD				
Yes	4 (8.26%)	0 (0%)	0 (0%)	0.748
No	44 (89.8%)	9 (100%)	15 (100%)	
NEC				
Yes	1 (2%)	1 (11.1%)	1 (6.7%)	0.256
No	47 (95.9%)	8 (88.9%)	14 (93.3%)	
IVH				
Yes	9 (18.4%)	0 (0%)	0 (0%)	0.082*
No	37 (75.5%)	9 (100%)	15 (100%)	
Blood transfusion				
Yes	24 (48.9%)	4 (44.4%)	5 (33.3%)	0.593
No	25 (51%)	5 (55.6%)	10 (66.7%)	
Feeding intolerance				
Yes	10 (10.4%)	1 (11.1%)	4 (26.7%)	0.606
No	39 (49.6%)	8 (88.9%)	9 (60%)	
Caucasian	**CC** (n=20)	**TT** (n=1)	**CT** (n=1)	
Sepsis				
Yes	2 (10%)	1 (100%)	0 (0%)	0.260
No	18 (90%)	0 (0%)	1 (100%)	
ROP				
Yes	2 (10%)	1 (100%)	0 (0%)	0.260
No	18 (90%)	0 (0%)	1 (100%)	
BPD				

Yes	1 (5%)	0 (0%)	0 (0%)	1.0
No	19 (95%)	1 (100%)	1 (100%)	
NEC				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	20 (100%)	1 (100%)	1 (100%)	
IVH				
Yes	4 (20%)	0 (0%)	0 (0%)	1.0
No	16 (80%)	1 (100%)	1 (100%)	
Blood transfusion				
Yes	8 (40%)	1 (100%)	0 (0%)	0.662
No	12 (60%)	0 (0%)	1 (100%)	
Feeding intolerance				
Yes	4 (20%)	0 (0%)	0 (0%)	1.0
No	16 (80%)	1 (100%)	1 (100%)	
African American	**CC** (n=20)	**TT** (n=4)	**CT** (n=7)	**P**
Sepsis				
Yes	3 (15%)	1 (25%)	1 (14.3%)	0.80
No	17 (85%)	3 (75%)	6 (85.7%)	
ROP				
Yes	3 (21.4%)	0 (0%)	0 (0%)	0.709
No	14 (70%)	4 (100%)	7 (100%)	
BPD				
Yes	3 (15%)	0 (0%)	0 (0%)	0.704
No	17 (85%)	4 (100%)	7 (100%)	
NEC				
Yes	0 (0%)	0 (0%)	1 (14.3%)	0.355
No	20 (100%)	4 (100%)	6 (85.7%)	
IVH				
Yes	3 (15%)	0 (0%)	0 (0%)	0.707
No	15 (75%)	4 (100%)	7 (100%)	
Blood transfusion				
Yes	11 (55%)	1 (25%)	3 (42.9%)	0.667
No	9 (45%)	3 (75%)	4 (57.1%)	
Feeding intolerance				
Yes	2 (10%)	0 (0%)	1 (14.3%)	1.0
No	18 (90%)	4 (100%)	5 (71.4%)	

Hispanic	**CC** (n=8)	**TT** (n=3)	**CT** (n=4)	**P**
Sepsis				
Yes	0 (0%)	1 (33.3%)	1 (25%)	0.269
No	6 (75%)	2 (66.7%)	3 (75%)	
ROP				
Yes	2 (25%)	2 (66.7%)	2 (50%)	0.790
No	4 (50%)	1 (33.3%)	2 (50%)	
BPD				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	7 (87.5%)	3 (100%)	4 (100%)	
NEC				
Yes	0 (0%)	1 (33.3%)	0 (0%)	0.214
No	7 (87.5%)	2 (66.7%)	4 (100%)	
IVH				
Yes	2 (25%)	0 (0%)	0 (0%)	0.692
No	5 (62.5%)	3 (100%)	4 (100%)	
Blood transfusion				
Yes	4 (50%)	2 (66.7%)	2 (50%)	1.0
No	4 (50%)	1 (33.3%)	2 (50%)	
Feeding intolerance				
Yes	3 (37.5%)	1 (33.3%)	3 (75%)	0.339
No	5 (62.5%)	2 (66.7%)	0 (0%)	

2070874 MAP			
Total Population	**No** (n=49)	**Yes** (n=24)	**p**
Sepsis			

Yes	5 (10.2%)	5 (20.8%)	0.289
No	42 (85.7%)	19 (79.2%)	
ROP			
Yes	7 (14.3%)	6 (25%)	0.520
No	37 (75.5%)	18 (75%)	
BPD			
Yes	4 (8.2%)	0 (0%)	0.294
No	44 (89.8%)	24 (100%)	
NEC			
Yes	1 (2%)	2 (8.3%)	0.256
No	47 (95.9%)	22 (91.7%)	
IVH			
Yes	9 (18.4%)	0 (0%)	0.023*
No	37 (75.5%)	24 (100%)	
Blood transfusion			
Yes	24 (48.9%)	9 (37.5%)	0.455
No	25 (51%)	15 (62.5%)	
Feeding intolerance			
Yes	10 (20.4%)	5 (20.8%)	1.0
No	39 (79.6%)	17 (70.8%)	
Caucasian	**No** (n=20)	**Yes** (n=2)	**P**
Sepsis			
Yes	2 (10%)	1 (50%)	0.260
No	18 (90%)	1 (50%)	
ROP			
Yes	2 (10%)	1 (50%)	0.260
No	18 (90%)	1 (50%)	
BPD			
Yes	2 (10%)	0 (0%)	1.0
No	19 (90%)	1 (50%)	
NEC			
Yes	0 (0%)	0 (0%)	n/a
No	20 (100%)	2 (100%)	
IVH			
Yes	2 (10%)	0 (0%)	1.0

No	16 (80%)	2 (100%)	
Blood transfusion			
Yes	1 (5%)	1 (50%)	1.0
No	12 (60%)	1 (50%)	
Feeding intolerance			
Yes	2 (10%)	0 (0%)	1.0
No	16 (80%)	2 (100%)	
African American	**No** (n=20)	**Yes** (n=11)	**p**
Sepsis			
Yes	3 (15%)	2 (18.2%)	1.0
No	17 (85%)	9 (81.8%)	
ROP			
Yes	3 (21.4%)	0 (0%)	0.258
No	14 (70%)	11 (100%)	
BPD			
Yes	3 (21.4%)	0 (0%)	0.535
No	17 (85%)	11 (100%)	
NEC			
Yes	0 (0%)	1 (9.1%)	0.355
No	20 (100%)	10 (90.9%)	
IVH			
Yes	3 (15%)	0 (0%)	0.268
No	15 (75%)	11 (100%)	
Blood transfusion			
Yes	11 (55%)	4 (36.3%)	0.458
No	9 (45%)	7 (63.4%)	
Feeding intolerance			
Yes	2 (10%)	1 (9.1%)	1.0
No	18 (90%)	9 (81.8%)	
Hispanic	**No** (n=8)	**Yes** (n=7)	**P**
Sepsis			
Yes	0 (0%)	2 (28.6%)	0.462
No	6 (75%)	5 (71.3%)	
ROP			

Yes	2 (25%)	4 (57.1%)	0.592
No	4 (50%)	3 (48.9%)	
BPD			
Yes	0 (0%)	0 (0%)	n/a
No	7 (87.5%)	7 (100%)	
NEC			
Yes	0 (0%)	1 (14.3%)	1.0
No	7 (87.5%)	6 (85.7%)	
IVH			
Yes	2 (25%)	0 (0%)	0.462
No	5 (62.5%)	7 (100%)	
Blood transfusion			
Yes	4 (50%)	4 (57.1%)	1.0
No	4 (50%)	3 (42.8%)	
Feeding intolerance			
Yes	3 (37.5%)	4 (57.1%)	0.592
No	5 (62.5%)	2 (28.6%)	

2243250				
Total Population	**CC** (n=32)	**TT** (n=22)	**CT** (n=19)	**P**
Sepsis				
Yes	5 (15.6	4 (18.2%)	1 (5.6%)	0.468

No	25 (78.1	18 (81.8%)	18 (94.7%)	
ROP				
Yes	6 (18.8	4 (18.2%)	3 (15.8%)	1.0
No	23 (71.9	16 (72.7%)	16 (84.2%)	
BPD				
Yes	4 (12.5	0 (0%)	0 (0%)	0.120*
No	27 (84.4	22 (100%)	19 (100%)	
NEC				
Yes	0 (0	2 (9.1%)	1 (5.3%)	0.179
No	31 (96.9	20 (90.9%)	18 (94.7%)	
IVH				
Yes	6 (18.8	3 (13.6%)	0 (0%)	0.127*
No	25 (78.1	17 (77.3%)	19 (100%)	
Blood transfusion				
Yes	18 (56.3	10 (45.5%)	5 (26.3%)	0.128*
No	14 (43.8	12 (54.5%)	14 (76.7%)	
Feeding intolerance				
Yes	8 (25	3 (13.6%)	4 (21.1%)	0.703
No	24 (75	18 (18.8%)	14 (73.7%)	
Caucasian	**CC** (n=17)	**TT** (n=2)	**CT** (n=3)	
Sepsis				
Yes	2 (13.3%)	1 (50%)	0 (0%)	0.294
No	15 (88.2%)	1 (50%)	3 (100%)	
ROP				
Yes	1 (5.9%)	1 (50%)	1 (33.3%)	0.117*
No	16 (94.1%)	1 (50%)	2 (66.7%)	
BPD				
Yes	1 (5.9%)	0 (0%)	0 (0%)	1.0
No	16 (94.1%)	2 (100%)	3 (100%)	
NEC				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	17 (100%)	2 (100%)	3 (100%)	
IVH				
Yes	3 (76.7%)	1 (50%)	0 (0%)	0.396
No	14 (82.3%)	1 (50%)	3 (100%)	

Blood transfusion				
Yes	7 (41.2%)	2 (100%)	0 (0%)	0.131*
No	10 (58.8%)	0 (0%)	3 (100%)	
Feeding intolerance				
Yes	3 (17.7%)	1 (50%)	0 (0%)	0.396
No	14 (82.3%)	1 (50%)	3 (100%)	
African American	**CC** (n=8)	**TT** (n=15)	**CT** (n=8)	**P**
Sepsis				
Yes	3 (37.5%)	2 (13.3%)	0 (0%)	0.125*
No	5 (62.5%)	13 (86.7%)	8 (100%)	
ROP				
Yes	3 (37.5%)	0 (0%)	0 (0%)	0.011*
No	4 (50%)	13 (86.7%)	8 (100%)	
BPD				
Yes	3 (37.5%)	0 (0%)	0 (0%)	0.025*
No	5 (62.5%)	15 (100%)	8 (100%)	
NEC				
Yes	0 (0%)	1 (6.7%)	0 (0%)	1.0
No	8 (100%)	14 (93.3%)	8 (100%)	
IVH				
Yes	1 (12.5%)	2 (13.3%)	0 (0%)	0.772
No	7 (87.5%)	11 (73.3%)	8 (100%)	
Blood transfusion				
Yes	7 (12.5%)	7 (46.7%)	1 (12.5%)	0.011*
No	1 (87.5%)	8 (53.3%)	7 (87.5%)	
Feeding intolerance				
Yes	2 (25%)	1 (6.7%)	0 (0%)	0.0293*
No	6 (75%)	14 (93.3%)	7 (87.5%)	
Hispanic	**CC** (n=7)	**TT** (n=3)	**CT** (n=5)	**p**
Sepsis				
Yes	0 (0%)	1 (33.3%)	1 (20%)	0.679
No	5 (71.4%)	2 (66.7%)	4 (80%)	
ROP				
Yes	2 (28.6%)	2 (66.7%)	2 (40%)	1.0

No	3 (42.9%)	1 (33.3%)	3 (60%)	
BPD				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	6 (85.7%)	3 (100%)	5 (100%)	
NEC				
Yes	0 (0%)	1 (33.3%)	0 (0%)	0.214
No	6 (85.7%)	2 (66.7%)	5 (100%)	
IVH				
Yes	2 (28.6%)	0 (0%)	0 (0%)	0.473
No	4 (57.1%)	3 (100%)	5 (100%)	
Blood transfusion				
Yes	4 (57.1%)	1 (33.3%)	3 (60%)	1.0
No	3 (42.9%)	2 (66.7%)	2 (40%)	
Feeding intolerance				
Yes	3 (42.9%)	1 (33.3%)	3 (60%)	1.0
No	4 (57.1%)	1 (33.3%)	2 (40%)	

1800795				
Total Population	**CC** (n=5)	**GG** (n=47)	**CG** (n=21)	**P**
Sepsis				
Yes	1 (20%)	6 (12.8%)	3 (14.3%)	0.869
No	4 (80%)	39 (82.9%)	18 (85.7%)	
ROP				
Yes	1 (20%)	10 (21.3%)	2 (9.5%)	0.546
No	4 (80%)	34 (72.3%)	17 (80.1%)	
BPD				
Yes	0 (0%)	2 (4.3%)	2 (9.5%)	0.690
No	5 (100%)	44 (93.6%)	19 (90.5%)	
NEC				
Yes	0 (0%)	3 (6.4%)	0 (0%)	0.636

No	5 (100%)	43 (91.5%)	21 (100%)	
IVH				
Yes	1 (20%)	5 (10.6%)	3 (14.3%)	0.606
No	4 (80%)	40 (85.1%)	17 (80.1%)	
Blood transfusion				
Yes	4 (80%)	19 (41.3%)	10 (47.6%)	0.235
No	1 (20%)	28 (60.9%)	11 (52.4%)	
Feeding intolerance				
Yes	1 (20%)	10 (21.3%)	4 (19.1%)	1.0
No	4 (80%)	35 (74.5%)	17 (80.9%)	
Caucasian	**CC** (n=5)	**GG** (n=7)	**CG** (n=10)	
Sepsis				
Yes	1 (20%)	0 (0%)	2 (20%)	0.568
No	4 (80%)	7 (100%)	8 (80%)	
ROP				
Yes	1 (20%)	2 (28.6%)	0 (0%)	0.208*
No	4 (80%)	5 (71.4%)	10 (100%)	
BPD				
Yes	0 (0%)	0 (0%)	1 (10%)	1.0
No	5 (100%)	7 (100%)	9 (90%)	
NEC				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	5 (100%)	7 (100%)	10 (100%)	
IVH				
Yes	1 (20%)	2 (28.6%)	1 (10%)	0.785
No	4 (80%)	5 (71.4%)	9 (90%)	
Blood transfusion				
Yes	4 (80%)	1 (14.3%)	4 (40%)	0.082*
No	1 (20%)	6 (85.7%)	6 (60%)	
Feeding intolerance				
Yes	1 (20%)	2 (28.6%)	1 (10%)	0.785
No	4 (80%)	5 (71.4%)	9 (90%)	
African American	**CC** (n=0)	**GG** (n=25)	**CG** (n=6)	**P**
Sepsis				

Yes	0 (0%)	5 (20%)	0 (0%)	0.553
No	0 (0%)	20 (80%)	6 (100%)	
ROP				
Yes	0 (0%)	3 (12%)	0 (0%)	1.0
No	0 (0%)	21 (84%)	4 (66.7%)	
BPD				
Yes	0 (0%)	2 (8%)	1 (16.7%)	0.488
No	0 (0%)	23 (92%)	5 (83.3%)	
NEC				
Yes	0 (0%)	1 (4%)	0 (0%)	1.0
No	0 (0%)	24 (96%)	6 (100%)	
IVH				
Yes	0 (0%)	1 (4%)	2 (33.3%)	0.068*
No	0 (0%)	23 (92%)	3 (50%)	
Blood transfusion				
Yes	0 (0%)	12 (48%)	3 (50%)	1.0
No	0 (0%)	13 (52%)	3 (50%)	
Feeding intolerance				
Yes	0 (0%)	2 (8%)	1 (16.7%)	0.501
No	0 (0%)	22 (88%)	5 (83.3%)	
Hispanic	**CC** (n=0)	**GG** (n=11)	**CG** (n=4)	**P**
Sepsis				
Yes	0 (0%)	1 (9.1%)	1 (25%)	1.0
No	0 (0%)	8 (72.7%)	3 (75%)	
ROP				
Yes	0 (0%)	4 (36.4%)	2 (50%)	1.0
No	0 (0%)	5 (45.5%)	2 (50%)	
BPD				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	0 (0%)	10 (90.9%)	4 (100%)	
NEC				
Yes	0 (0%)	1 (9.1%)	0 (0%)	1.0
No	0 (0%)	9 (81.8%)	4 (100%)	
IVH				
Yes	0 (0%)	2 (18.2%)	0 (0%)	1.0

No	0 (0%)	8 (72.7%)	4 (100%)	
Blood transfusion				
Yes	0 (0%)	5 (45.5%)	3 (75%)	0.569
No	0 (0%)	6 (54.5%)	1 (25%)	
Feeding intolerance				
Yes	0 (0%)	5 (45.5%)	2 (50%)	1.0
No	0 (0%)	5 (45.5%)	2 (50%)	

1800795 MAP			
Total Population	**No** (n=47)	**Yes** (n=26)	**p**
Sepsis			
Yes	6 (12.8%)	4 (15.4%)	1.0
No	39 (82.9%)	22 (84.6%)	
ROP			
Yes	10 (21.3%)	3 (11.5%)	0.355
No	34 (72.3%)	21 (80.8%)	
BPD			
Yes	2 (4.3%)	2 (7.7%)	0.616
No	44 (93.6%)	24 (92.3%)	
NEC			
Yes	3 (6.4%)	0 (0%)	0.549
No	43 (91.5%)	26 (100%)	
IVH			
Yes	5 (10.6%)	4 (15.3%)	0.712
No	40 (86.1%)	21 (80.7%)	
Blood transfusion			

Yes	19 (41.3%)	14 (53.9%)	0.330
No	28 (60.9%)	12 (46.2%)	
Feeding intolerance			
Yes	10 (21.3%)	5 (19.2%)	1.0
No	35 (74.5%)	21 (80.8%)	
Caucasian	**No** (n=7)	**Yes** (n=15)	**P**
Sepsis			
Yes	0 (0%)	3 (20%)	0.523
No	7 (100%)	12 (80%)	
ROP			
Yes	2 (28.6%)	1 (6.7%)	0.227
No	5 (71.4%)	14 (93.3%)	
BPD			
Yes	0 (0%)	1 (6.7%)	1.0
No	7 (100%)	14 (93.3%)	
NEC			
Yes	0 (0%)	0 (0%)	n/a
No	7 (100%)	15 (100%)	
IVH			
Yes	2 (28.6%)	2 (20%)	0.565
No	5 (71.4%)	13 (80%)	
Blood transfusion			
Yes	1 (14.3%)	8 (53.3%)	0.165*
No	6 (85.7%)	7 (46.7%)	
Feeding intolerance			
Yes	2 (28.6%)	2 (20%)	0.565
No	5 (71.4%)	13 (80%)	
African American	**No** (n=25)	**Yes** (n=6)	**p**
Sepsis			
Yes	5 (20%)	0 (0%)	0.553
No	20 (80%)	6 (100%)	
ROP			
Yes	3 (12%)	0 (0%)	1.0
No	21 (84%)	4 (66.7%)	

BPD			
Yes	2 (8%)	1 (16.7%)	0.488
No	23 (92%)	5 (83.3%)	
NEC			
Yes	1 (4%)	0 (0%)	1.0
No	24 (96%)	6 (100%)	
IVH			
Yes	1 (4%)	2 (33.3%)	0.068*
No	23 (92%)	3 (50%)	
Blood transfusion			
Yes	12 (48%)	3 (50%)	1.0
No	13 (52%)	3 (50%)	
Feeding intolerance			
Yes	2 (8%)	1 (16.7%)	0.501
No	22 (88%)	5 (83.3%)	
Hispanic	**No** (n=12)	**Yes** (n=4)	**P**
Sepsis			
Yes	1 (8.3%)	1 (25%)	1.0
No	8 (66.7%)	3 (75%)	
ROP			
Yes	4 (33.3%)	2 (50%)	1.0
No	5 (41.7%)	2 (50%)	
BPD			
Yes	0 (0%)	0 (0%)	n/a
No	10 (83.3%)	4 (100%)	
NEC			
Yes	1 (8.3%)	0 (0%)	1.0
No	9 (75%)	4 (100%)	
IVH			
Yes	4 (33.3%)	0 (0%)	1.0
No	8 (66.7%)	2 (50%)	
Blood transfusion			
Yes	5 (41.7%)	3 (75%)	0.569
No	6 (50%)	1 (25%)	
Feeding intolerance			

Yes	5 (41.7%)	2 (50%)	n/a	
No	5 (41.7%)	2 (50%)		

1800796				
Total Population	**GG** (n=3)	**AA** (n=58)	**AG** (n=12)	**p**
Sepsis				
Yes	1 (33.3	7 (12.1%)	2 (16.7%)	0.225
No	1 (33.3	50 (86.2%)	10 (83.3%)	
ROP				
Yes	1 (33.3	8 (13.8%)	4 (33.3%)	0.132*
No	1 (33.3	46 (79.3%)	8 (66.7%)	
BPD				
Yes	0 (0	2 (3.5%)	2 (16.7%)	0.275
No	3 (100	55 (94.8%)	10 (83.3%)	
NEC				
Yes	2 (66.7	1 (1.7%)	0 (0%)	0.003*
No	1 (33.3	56 (96.6%)	12 (100%)	
IVH				
Yes	1 (33.3	7 (12.1%)	1 (8.3%)	0.569
No	2 (66.7	48 (82.8%)	11 (91.7%)	
Blood transfusion				
Yes	3 (100	26 (44.8%)	4 (33.3%)	0.120*
No	0 (0	32 (55.2%)	8 (66.7%)	
Feeding intolerance				
Yes	3 (100	10 (17.2%)	2 (16.7%)	0.014*
No	0 (0	47 (81%)	9 (75%)	

Caucasian	**GG** (n=0)	**AA** (n=20)	**AG** (n=2)	**p**
Sepsis				
Yes	0 (0	3 (15%)	0 (0%)	1.0
No	0 (0	17 (85%)	2 (100%)	
ROP				
Yes	0 (0	3 (15%)	0 (0%)	1.0
No	0 (0	17 (85%)	2 (100%)	
BPD				
Yes	0 (0	1 (5%)	0 (0%)	1.0
No	0 (0	19 (95%)	2 (100%)	
NEC				
Yes	0 (0	0 (0%)	0 (0%)	n/a
No	0 (0	20 (100%)	2 (100%)	
IVH				
Yes	0 (0	3 (15%)	1 (50%)	0.338
No	0 (0	17 (85%)	1 (50%)	
Blood transfusion				
Yes	0 (0	8 (40%)	1 (50%)	1.0
No	0 (0	12 (60%)	1 (50%)	
Feeding intolerance				
Yes	0 (0	3 (15%)	1 (50%)	0.338
No	0 (0	17 (85%)	1 (50%)	
African American	**GG** (n=0)	**AA** (n=27)	**AG** (n=4)	**p**
Sepsis				
Yes	0 (0%)	3 (11.1%)	2 (50%)	0.112*
No	0 (0%)	24 (88.9%)	2 (50%)	
ROP				
Yes	0 (0%)	1 (3.7%)	2 (50%)	0.045*
No	0 (0%)	23 (85.2%)	2 (50%)	
BPD				
Yes	0 (0%)	1 (3.7%)	2 (50%)	0.037*
No	0 (0%)	26 (96.3%)	2 (50%)	
NEC				
Yes	0 (0%)	1 (3.7%)	0 (0%)	1.0

No	0 (0%)	26 (96.3%)	4 (100%)	
IVH				
Yes	0 (0%)	3 (11.1%)	0 (0%)	1.0
No	0 (0%)	22 (81.5%)	4 (100%)	
Blood transfusion				
Yes	0 (0%)	13 (48.2%)	2 (50%)	1.0
No	0 (0%)	14 (51.9%)	2 (50%)	
Feeding intolerance				
Yes	0 (0%)	2 (7.4%)	1 (25%)	0.360
No	0 (0%)	24 (88.9%)	3 (75%)	
Hispanic	**GG** (n=2)	**AA** (n=9)	**AG** (n=4)	**p**
Sepsis				
Yes	1 (50%)	1 (11.1%)	0 (0%)	0.231
No	0 (0%)	7 (77.8%)	4 (100%)	
ROP				
Yes	1 (50%)	4 (44.4%)	1 (25%)	0.559
No	0 (0%)	4 (44.4%)	3 (75%)	
BPD				
Yes	0 (50%)	0 (0%)	0 (0%)	n/a
No	2 (100%)	8 (88.9%)	4 (100%)	
NEC				
Yes	1 (50%)	0 (0%)	0 (0%)	0.143*
No	1 (50%)	8 (88.9%)	4 (100%)	
IVH				
Yes	1 (50%)	1 (11.1%)	0 (0%)	0.341
No	1 (50%)	7 (77.8%)	4 (100%)	
Blood transfusion				
Yes	2 (100%)	5 (55.6%)	1 (25%)	0.386
No	0 (0%)	4 (44.4%)	3 (75%)	
Feeding intolerance				
Yes	2 (100%)	5 (55.6%)	0 (0%)	0.094*
No	0 (0%)	4 (44.4%)	3 (75%)	

1800796 MAP			
Total Population	**No** (n=57)	**Yes** (n=15)	**P**
Sepsis			
Yes	7 (12.3%)	3 (20%)	0.402
No	50 (87.7%)	11 (73.3%)	
ROP			
Yes	8 (14%)	5 (33.3%)	0.122*
No	46 (80.7%)	9 (60%)	
BPD			
Yes	2 (3.5%)	2 (13.3%)	0.189*
No	55 (96.5%)	13 (86.7%)	
NEC			
Yes	1 (1.8%)	2 (13.3%)	0.108*
No	56 (98.2%)	13 (86.7%)	
IVH			
Yes	7 (12.3%)	2 (13.3%)	1.0
No	48 (84.2%)	13 (86.7%)	
Blood transfusion			
Yes	26 (45.6%)	7 (46.7%)	1.0
No	32 (56.1%)	8 (53.3%)	
Feeding intolerance			
Yes	10 (17.5%)	5 (33.3%)	0.155*
No	47 (82.5%)	9 (60%)	
Caucasian	**No** (n=20)	**Yes** (n=2)	**p**
Sepsis			
Yes	3 (15%)	0 (0%)	1.0
No	17 (85%)	2 (100%)	
ROP			

Yes	3 (15%)	0 (0%)	1.0
No	17 (85%)	2 (100%)	
BPD			
Yes	1 (5%)	0 (0%)	1.0
No	19 (95%)	2 (100%)	
NEC			
Yes	0 (0%)	0 (0%)	n/a
No	20 (100%)	2 (100%)	
IVH			
Yes	3 (15%)	1 (50%)	0.338
No	17 (85%)	1 (50%)	
Blood transfusion			
Yes	8 (40%)	1 (50%)	1.0
No	12 (60%)	1 (50%)	
Feeding intolerance			
Yes	3 (15%)	1 (50%)	0.338
No	17 (85%)	1 (50%)	

African American	**No** (n=27)	**Yes** (n=4)	**p**
Sepsis			
Yes	3 (11.1%)	2 (50%)	0.112*
No	24 (88.9%)	2 (50%)	
ROP			
Yes	1 (3.7%)	2 (50%)	0.045*
No	23 (85.2%)	2 (50%)	
BPD			
Yes	1 (3.7%)	2 (50%)	0.037*
No	26 (96.3%)	2 (50%)	
NEC			
Yes	1 (3.7%)	0 (0%)	1.0
No	26 (96.3%)	4 (100%)	
IVH			
Yes	3 (11.1%)	0 (0%)	1.0
No	22 (81.5%)	4 (100%)	
Blood transfusion			
Yes	12 (44.4%)	2 (50%)	1.0

No	14 (51.9%)	2 (50%)	
Feeding intolerance			
Yes	2 (7.4%)	1 (25%)	0.360
No	24 (88.9%)	3 (75%)	

Hispanic	**No** (n=9)	**Yes** (n=6)	**p**
Sepsis			
Yes	1 (11.1%)	1 (16.7%)	1.0
No	7 (77.8%)	4 (66.7%)	
ROP			
Yes	4 (44.4%)	2 (33.3%)	1.0
No	4 (44.4%)	3 (50%)	
BPD			
Yes	0 (0%)	0 (0%)	n/a
No	8 (88.9%)	6 (100%)	
NEC			
Yes	0 (0%)	1 (16.7%)	0.429
No	8 (88.9%)	5 (83.3%)	
IVH			
Yes	1 (11.1%)	1 (16.7%)	1.0
No	7 (77.8%)	5 (83.3%)	
Blood transfusion			
Yes	5 (55.6%)	3 (50%)	1.0
No	4 (44.4%)	3 (50%)	
Feeding intolerance			
Yes	5 (55.6%)	2 (33.3%)	1.0
No	4 (44.4%)	3 (50%)	

1800871				
Total Population	**TT** (n=7)	**CC** (n=40)	**CT** (n=24)	**p**
Sepsis				
Yes	1 (14.3	8 (20%)	1 (4.4%)	0.160*
No	5 (71.4	32 (80%)	23 (95.7%)	
ROP				
Yes	1 (14.3	9 (22.5%)	3 (12.5%)	0.709
No	4 (57.1	30 (75%)	20 (83.3%)	
BPD				
Yes	0 (0	4 (10%)	0 (0%)	0.40
No	7 (100	36 (90%)	24 (100%)	
NEC				
Yes	1 (14.3	0 (0%)	2 (8.3%)	0.058*
No	6 (85.7	40 (100%)	22 (91.7%)	
IVH				
Yes	1 (14.3	4 (10%)	4 (16.7%)	0.551
No	5 (71.4	36 (90%)	19 (79.2%)	
Blood transfusion				
Yes	2 (28.6	18 (45%)	12 (50%)	0.646
No	5 (74.4	23 (57.5%)	12 (50%)	
Feeding intolerance				
Yes	3 (42.9	7 (17.5%)	5 (20.8%)	0.333
No	4 (57.1	32 (80%)	19 (79.2%)	

Caucasian	**TT** (n=1)	**CC** (n=14)	**CT** (n=6)	**p**
Sepsis				
Yes	0 (0%)	3 (21.4%)	0 (0%)	0.589
No	1 (100%)	11 (78.6%)	6 (100%)	
ROP				
Yes	0 (0%)	2 (14.3%)	1 (16.7%)	1.0
No	1 (100%)	12 (85.7%)	5 (83.3%)	
BPD				
Yes	0 (0%)	1 (7.1%)	0 (0%)	1.0

No	1 (100%)	13 (92.9%)	6 (100%)	
NEC				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	1 (100%)	14 (100%)	6 (100%)	
IVH				
Yes	0 (0%)	1 (7.1%)	3 (50%)	0.088*
No	1 (100%)	13 (92.9%)	3 (50%)	
Blood transfusion				
Yes	0 (0%)	5 (35.7%)	3 (50%)	0.779
No	1 (100%)	9 (64.3%)	3 (50%)	
Feeding intolerance				
Yes	0 (0%)	2 (14.3%)	2 (33.3%)	0.635
No	1 (100%)	12 (85.7%)	4 (66.7%)	
African American	**TT** (n=1)	**CC** (n=17)	**CT** (n=13)	**p**
Sepsis				
Yes	0 (0%)	4 (23.5%)	1 (7.7%)	0.459
No	1 (100%)	13 (76.5%)	12 (97.3%)	
ROP				
Yes	0 (0%)	3 (17.7%)	0 (0%)	0.238
No	0 (0%)	13 (76.5%)	12 (92.3%)	
BPD				
Yes	0 (0%)	3 (17.7%)	0 (0%)	0.312
No	1 (100%)	14 (82.3%)	13 (100%)	
NEC				
Yes	0 (0%)	0 (0%)	1 (7.7%)	0.452
No	1 (100%)	17 (100%)	12 (92.3%)	
IVH				
Yes	0 (0%)	2 (11.8%)	1 (7.7%)	1.0
No	0 (0%)	15 (88.2%)	11 (84.6%)	
Blood transfusion				
Yes	0 (0%)	9 (52.9%)	6 (46.2%)	1.0
No	1 (100%)	8 (47.1%)	7 (53.8%)	
Feeding intolerance				
Yes	0 (0%)	1 (5.9%)	2 (15.4%)	0.616
No	1 (100%)	15 (88.2%)	11 (84.6%)	

Hispanic	**TT** (n=4)	**CC** (n=8)	**CT** (n=3)	**p**
Sepsis				
Yes	1 (25%)	1 (12.5%)	0 (0%)	1.0
No	2 (50%)	6 (75%)	3 (100%)	
ROP				
Yes	1 (25%)	3 (37.5%)	2 (66.6%)	1.0
No	2 (50%)	4 (50%)	1 (33.3%)	
BPD				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	4 (100%)	7 (87.5%)	3 (100%)	
NEC				
Yes	1 (25%)	0 (0%)	0 (0%)	0.500
No	3 (75%)	7 (87.5%)	3 (100%)	
IVH				
Yes	1 (25%)	1 (12.5%)	0 (0%)	1.0
No	3 (75%)	6 (75%)	3 (100%)	
Blood transfusion				
Yes	2 (50%)	4 (50%)	2 (66.7%)	1.0
No	2 (50%)	4 (50%)	1 (33.3%)	
Feeding intolerance				
Yes	3 (75%)	4 (50%)	0 (0%)	0.241
No	1 (25%)	3 (37.5%)	3 (100%)	

1800871 MAP			
Total Population	**No** (n=40)	**Yes** (n=31)	**P**

Sepsis			
Yes	8 (20%)	2 (6.5%)	0.171*
No	32 (80%)	28 (90.3%)	
ROP			
Yes	9 (22.5%)	4 (12.9%)	0.533
No	30 (75%)	24 (77.4%)	
BPD			
Yes	4 (10%)	0 (0%)	0.126*
No	36 (90%)	31 (100%)	
NEC			
Yes	0 (0%)	3 (9.7%)	0.079*
No	40 (100%)	28 (90.3%)	
IVH			
Yes	4 (10%)	5 (16.1%)	0.477
No	36 (90%)	24 (77.4%)	
Blood transfusion			
Yes	18 (45%)	14 (45.2%)	1.0
No	23 (57.5%)	17 (54.8%)	
Feeding intolerance			
Yes	7 (17.5%)	8 (25.8%)	0.560
No	32 (80%)	23 (74.2%)	

Caucasian	**No** (n=14)	**Yes** (n=7)	**P**
Sepsis			
Yes	3 (21.4%)	0 (0%)	0.521
No	11 (78.6%)	7 (100%)	
ROP			
Yes	2 (14.3%)	1 (14.3%)	1.0
No	12 (85.7%)	6 (85.7%)	
BPD			
Yes	1 (7.1%)	0 (0%)	1.0
No	13 (92.9%)	7 (100%)	
NEC			
Yes	0 (0%)	0 (0%)	n/a
No	14 (100%)	7 (100%)	
IVH			

Yes	1 (7.1%)	3 (42.9%)	0.088*
No	13 (92.9%)	4 (57.1%)	
Blood transfusion			
Yes	5 (35.7%)	3 (42.9%)	1.0
No	9 (64.3%)	4 (57.1%)	
Feeding intolerance			
Yes	2 (14.3%)	2 (28.6%)	0.574
No	12 (85.7%)	5 (71.4%)	
African American	**No** (n=17)	**Yes** (n=14)	**P**
Sepsis			
Yes	4 (2.4%)	1 (7.1%)	0.344
No	13 (76.5%)	13 (92.9%)	
ROP			
Yes	3 (17.6%)	0 (0%)	0.238
No	13 (76.5%)	12 (85.7%)	
BPD			
Yes	3 (17.6%)	0 (0%)	0.232
No	14 (82.4%)	14 (100%)	
NEC			
Yes	0 (0%)	1 (7.1%)	0.452
No	17 (100%)	13 (92.9%)	
IVH			
Yes	2 (11.8%)	1 (7.1%)	1.0
No	15 (88.2%)	11 (78.6%)	
Blood transfusion			
Yes	9 (52.9%)	6 (42.9%)	0.722
No	8 (47.1%)	8 (57.1%)	
Feeding intolerance			
Yes	1 (5.9%)	2 (14.3%)	0.586
No	15 (88.2%)	12 (85.7%)	
Hispanic	**No** (n=8)	**Yes** (n=7)	**P**
Sepsis			
Yes	1 (12.5%)	1 (14.3%)	1.0
No	6 (75%)	5 (71.4%)	

ROP			
Yes	3 (37.5%)	3 (42.8%)	1.0
No	4 (50%)	3 (42.8%)	
BPD			
Yes	0 (0%)	0 (0%)	n/a
No	7 (87.5%)	7 (100%)	
NEC			
Yes	0 (0%)	1 (14.3%)	1.0
No	7 (87.5%)	6 (85.7%)	
IVH			
Yes	1 (12.5%)	1 (14.3%)	1.0
No	6 (75%)	6 (85.7%)	
Blood transfusion			
Yes	4 (50%)	4 (57.1%)	1.0
No	4 (50%)	3 (42.9%)	
Feeding intolerance			
Yes	4 (50%)	3 (42.9%)	1.0
No	3 (37.5%)	4 (57.1%)	

1800872				
Total Population	**CC** (n=39)	**AA** (n=9)	**AC** (n=24)	**p**
Sepsis				
Yes	8 (20.5%)	1 (11.1	1 (4.2%)	0.189*
No	30 (76.9%)	7 (77.8	23 (95.8%)	
ROP				
Yes	9 (23.1%)	2 (22.2	2 (8.3%)	0.314

No	29 (74.4%)	5 (55.6	21 (87.5%)	
BPD				
Yes	3 (7.7%)	0 (0	1 (4.2%)	1.0
No	35 (89.7%)	9 (100	23 (95.8%)	
NEC				
Yes	0 (0%)	1 (11.1	2 (8.3%)	0.119*
No	38 (97.4%)	8 (88.9	22 (92.7%)	
IVH				
Yes	5 (12.8%)	1 (11.1	3 (12.5%)	1.0
No	33 (84.6%)	7 (77.8	21 (87.5%)	
Blood transfusion				
Yes	18 (46.2%)	3 (33.3	12 (50%)	0.711
No	21 (53.9%)	6 (66.7	12 (50%)	
Feeding intolerance				
Yes	8 (20.5%)	3 (33.3	4 (16.7%)	0.601
No	29 (74.4%)	6 (66.7	20 (83.3%)	
Caucasian	**CC** (n=15)	**AA** (n=3)	**AC** (n=4)	**p**
Sepsis				
Yes	3 (20%)	0 (0%)	0 (0%)	1.0
No	12 (80%)	3 (100%)	4 (100%)	
ROP				
Yes	2 (13.3%)	1 (33.3%)	0 (0%)	0.432
No	13 (86.7%)	2 (66.7%)	4 (100%)	
BPD				
Yes	1 (6.7%)	0 (0%)	0 (0%)	1.0
No	14 (93.3%)	3 (100%)	4 (100%)	
NEC				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	15 (100%)	3 (100%)	4 (100%)	
IVH				
Yes	3 (20%)	0 (0%)	1 (25%)	1.0
No	12 (80%)	3 (100%)	3 (75%)	
Blood transfusion				
Yes	6 (40%)	1 (33.3%)	2 (50%)	1.0
No	9 (60%)	2 (66.7%)	2 (50%)	

Feeding intolerance				
Yes	3 (20%)	0 (0%)	1 (25%)	1.0
No	12 (80%)	3 (100%)	3 (75%)	
African American	**CC** (n=14)	**AA** (n=1)	**AC** (n=15)	**p**
Sepsis				
Yes	4 (28.6%)	0 (0%)	1 (6.7%)	0.307
No	10 (71.4%)	1 (100%)	14 (93.3%)	
ROP				
Yes	4 (28.6%)	0 (0%)	0 (0%)	0.222
No	11 (78.6%)	0 (0%)	14 (93.3%)	
BPD				
Yes	2 (14.3%)	0 (0%)	1 (6.7%)	0.638
No	12 (85.7%)	1 (100%)	14 (93.3%)	
NEC				
Yes	0 (0%)	0 (0%)	1 (6.7%)	1.0
No	14 (100%)	1 (100%)	14 (93.3%)	
IVH				
Yes	1 (7.4%)	0 (0%)	2 (13.3%)	1.0
No	13 (92.9%)	0 (0%)	13 (86.7%)	
Blood transfusion				
Yes	8 (57.1%)	0 (0%)	7 (46.7%)	0.715
No	6 (42.9%)	1 (100%)	8 (53.3%)	
Feeding intolerance				
Yes	1 (7.4%)	0 (0%)	2 (2%)	1.0
No	12 (85.7%)	1 (100%)	13 (86.7%)	
Hispanic	**CC** (n=8)	**AA** (n=4)	**AC** (n=3)	**p**
Sepsis				
Yes	1 (12.5%)	1 (25%)	0 (0%)	1.0
No	6 (75%)	2 (50%)	3 (100%)	
ROP				
Yes	3 (37.5%)	1 (25%)	2 (66.7%)	1.0
No	4 (50%)	2 (50%)	1 (33.3%)	
BPD				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a

No	7 (87.5%)	4 (100%)	3 (100%)	
NEC				
Yes	0 (0%)	1 (25%)	0 (0%)	0.500
No	7 (87.5%)	3 (75%)	3 (100%)	
IVH				
Yes	1 (12.5%)	1 (25%)	0 (0%)	1.0
No	6 (75%)	3 (75%)	3 (100%)	
Blood transfusion				
Yes	4 (50%)	2 (50%)	2 (66.7%)	1.0
No	4 (50%)	2 (50%)	1 (33.3%)	
Feeding intolerance				
Yes	4 (50%)	3 (75%)	0 (0%)	0.241
No	3 (37.5%)	1 (25%)	3 (100%)	

1800872			
Total Population	**No** (n=39)	**Yes** (n=33)	**P**
Sepsis			
Yes	8 (20.5%)	2 (6.1%)	0.097*
No	30 (76.9%)	30 (90.9%)	
ROP			
Yes	9 (23.1%)	4 (12.1%)	0.360
No	29 (74.4%)	26 (78.8%)	
BPD			
Yes	3 (7.7%)	1 (3%)	0.618
No	35 (89.7%)	32 (97%)	
NEC			

Yes	0 (0%)	3 (9.1%)	0.095*
No	38 (97.4%)	30 (90.9%)	
IVH			
Yes	5 (12.8%)	4 (12.1%)	1.0
No	33 (84.6%)	28 (84.9%)	
Blood transfusion			
Yes	18 (46.1%)	15 (45.5%)	1.0
No	21 (53.9%)	18 (54.5%)	
Feeding intolerance			
Yes	8 (20.5%)	7 (21.2%)	1.0
No	29 (74.4%)	26 (78.8%)	
Caucasian	**No** (n=15)	**Yes** (n=7)	**P**
Sepsis			
Yes	3 (20%)	0 (0%)	0.523
No	12 (80%)	7 (100%)	
ROP			
Yes	2 (13.3%)	1 (14.3%)	1.0
No	13 (86.7%)	6 (85.7%)	
BPD			
Yes	1 (6.7%)	0 (0%)	1.0
No	14 (93.3%)	7 (100%)	
NEC			
Yes	0 (0%)	0 (0%)	n/a
No	15 (100%)	7 (100%)	
IVH			
Yes	3 (20%)	1 (14.3%)	1.0
No	12 (80%)	6 (85.7%)	
Blood transfusion			
Yes	6 (40%)	3 (42.9%)	1.0
No	9 (60%)	4 (57.1%)	
Feeding intolerance			
Yes	3 (20%)	1 (14.3%)	1.0
No	12 (80%)	6 (85.7%)	
African American	**No** (n=14)	**Yes** (n=16)	**P**

Sepsis			
Yes	4 (28.6%)	1 (6.3%)	0.157*
No	10 (71.4%)	15 (93.8%)	
ROP			
Yes	3 (21.4%)	0 (0%)	0.222
No	11 (78.6%)	14 (87.5%)	
BPD			
Yes	2 (14.3%)	1 (6.3%)	0.586
No	12 (85.7%)	15 (93.7%)	
NEC			
Yes	0 (0%)	1 (6.3%)	1.0
No	14 (100%)	15 (93.7%)	
IVH			
Yes	1 (7.1%)	2 (12.5%)	1.0
No	13 (92.9%)	13 (81.3%)	
Blood transfusion			
Yes	8 (57.1%)	7 (43.8%)	0.715
No	6 (42.9%)	9 (56.2%)	
Feeding intolerance			
Yes	1 (7.1%)	2 (12.5%)	1.0
No	12 (85.7%)	14 (87.5%)	
Hispanic	**No** (n=8)	**Yes** (n=7)	**P**
Sepsis			
Yes	1 (12.5%)	1 (20%)	1.0
No	6 (75%)	5 (71.3%)	
ROP			
Yes	3 (37.5%)	3 (42.9%)	1.0
No	4 (50%)	3 (42.9%)	
BPD			
Yes	0 (0%)	0 (0%)	n/a
No	7 (87.5%)	7 (100%)	
NEC			
Yes	0 (0%)	1 (20%)	1.0
No	7 (87.5%)	6 (85.7%)	
IVH			

Yes	1 (12.5%)	1 (20%)	1.0
No	6 (75%)	6 (85.7%)	
Blood transfusion			
Yes	4 (50%)	4 (57.1%)	1.0
No	4 (50%)	3 (42.9%)	
Feeding intolerance			
Yes	4 (50%)	3 (42.9%)	1.0
No	3 (37.5%)	4 (57.1%)	

1800896				
Total Population	**GG** (n=13)	**AA** (n=32)	**AG** (n=34)	**p**
Sepsis				
Yes	2 (15.4%)	1 (3.1%)	7 (20.6%)	0.283
No	11 (84.6%)	22 (68.8%)	27 (79.4%)	
ROP				
Yes	3 (23.1%)	5 (15.6%)	5 (14.7%)	0.657
No	10 (76.9%)	16 (50%)	28 (82.4%)	
BPD				
Yes	1 (7.7%)	1 (3.1%)	2 (5.9%)	1.0
No	12 (92.3%)	23 (71.9%)	32 (94.1%)	
NEC				
Yes	0 (0%)	1 (12.5%)	2 (5.9%)	1.0
No	13 (100%)	23 (87.5%)	32 (94.1%)	
IVH				
Yes	1 (7.7%)	4 (12.5%)	4 (11.8%)	0.715
No	12 (92.3%)	28 (87.5%)	30 (88.2%)	

Blood transfusion				
Yes	7 (53.9%)	9 (28.1%)	16 (47.1%)	0.519
No	6 (46.1%)	16 (50%)	18 (52.9%)	
Feeding intolerance				
Yes	3 (23.1%)	6 (18.8%)	5 (14.7%)	0.667
No	9 (69.2%)	19 (50%)	28 (82.4%)	
Caucasian	**GG** (n=5)	**AA** (n=8)	**AG** (n=9)	**p**
Sepsis				
Yes	1 (20%)	0 (0%)	2 (22.2%)	0.416
No	4 (80%)	8 (100%)	7 (77.8%)	
ROP				
Yes	0 (0%)	1 (12.5%)	2 (22.2%)	0.766
No	5 (100%)	7 (87.5%)	7 (77.8%)	
BPD				
Yes	1 (20%)	0 (0%)	0 (0%)	0.227
No	4 (80%)	8 (100%)	9 (100%)	
NEC				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	5 (100%)	8 (100%)	9 (100%)	
IVH				
Yes	1 (20%)	0 (0%)	3 (33.3%)	0.234
No	4 (80%)	8 (100%)	6 (66.7%)	
Blood transfusion				
Yes	2 (40%)	2 (25%)	5 (55.6%)	0.557
No	3 (60%)	6 (75%)	4 (44.4%)	
Feeding intolerance				
Yes	1 (20%)	2 (25%)	1 (11.1%)	0.803
No	4 (80%)	6 (75%)	8 (88.9%)	
African American	**GG** (n=5)	**AA** (n=6)	**AG** (n=20)	**p**
Sepsis				
Yes	0 (0%)	0 (0%)	5 (25%)	0.384
No	5 (100%)	6 (100%)	15 (75%)	
ROP				
Yes	1 (20%)	0 (0%)	2 (10%)	0.704

No	4 (80%)	4 (66.7%)	17 (85%)	
BPD				
Yes	0 (0%)	1 (2.8%)	2 (10%)	1.0
No	5 (100%)	5 (83.3%)	18 (90%)	
NEC				
Yes	0 (0%)	0 (0%)	1 (5%)	1.0
No	5 (100%)	6 (100%)	19 (95%)	
IVH				
Yes	0 (0%)	2 (33.3%)	1 (5%)	0.056*
No	5 (100%)	2 (33.3%)	19 (95%)	
Blood transfusion				
Yes	3 (60%)	3 (50%)	9 (45%)	0.877
No	2 (40%)	3 (50%)	11 (55%)	
Feeding intolerance				
Yes	0 (0%)	0 (0%)	3 (15%)	0.747
No	5 (100%)	6 (100%)	16 (80%)	
Hispanic	**GG** (n=3)	**AA** (n=9)	**AG** (n=2)	**p**
Sepsis				
Yes	1 (33.3%)	1 (11.1%)	0 (0%)	1.0
No	2 (66.7%)	6 (66.7%)	2 (100%)	
ROP				
Yes	2 (66.7%)	3 (33.3%)	1 (50%)	1.0
No	1 (33.3%)	4 (44.4%)	1 (50%)	
BPD				
Yes	0 (0%)	0 (0%)	0 (0%)	n/a
No	3 (100%)	8 (88.9%)	2 (100%)	
NEC				
Yes	0 (0%)	1 (11.1%)	0 (0%)	1.0
No	3 (100%)	7 (77.8%)	2 (100%)	
IVH				
Yes	0 (0%)	2 (22.2%)	0 (0%)	1.0
No	3 (100%)	6 (66.7%)	2 (100%)	
Blood transfusion				
Yes	2 (66.7%)	4 (44.4%)	1 (50%)	1.0
No	1 (33.3%)	5 (55.6%)	1 (50%)	

Feeding intolerance				
Yes	2 (66.7%)	4 (44.4%)	0 (0%)	0.217
No	0 (0%)	5 (55.6%)	2 (100%)	

1800896 MAP			
Total Population	**No** (n=25)	**Yes** (n=47)	**p**
Sepsis			
Yes	1 (4%)	9 (19.1%)	0.149*
No	22 (88%)	38 (80.9%)	
ROP			
Yes	5 (20%)	8 (17%)	0.526
No	16 (64%)	38 (81%)	
BPD			
Yes	1 (4%)	3 (6.4%)	1.0
No	23 (92%)	44 (93.6%)	
NEC			
Yes	1 (4%)	2 (4.3%)	1.0
No	23 (92%)	45 (95.7%)	
IVH			
Yes	4 (16%)	5 (10.6%)	0.452
No	18 (72%)	42 (89.4%)	
Blood transfusion			
Yes	9 (36%)	23 (48.9%)	0.329

No	16 (64%)	24 (51.1%)	
Feeding intolerance			
Yes	6 (24%)	8 (17%)	0.548
No	19 (76%)	37 (78.7%)	
Caucasian	**No** (n=8)	**Yes** (n=14)	**p**
Sepsis			
Yes	0 (0%)	3 (21.4%)	0.273
No	8 (100%)	11 (78.6%)	
ROP			
Yes	1 (12.5%)	2 (14.3%)	1.0
No	7 (87.5%)	12 (85.7%)	
BPD			
Yes	0 (0%)	1 (7.1%)	1.0
No	8 (100%)	13 (92.9%)	
NEC			
Yes	0 (0%)	0 (0%)	n/a
No	8 (100%)	14 (100%)	
IVH			
Yes	0 (0%)	4 (28.6%)	0.254
No	8 (100%)	10 (71.4%)	
Blood transfusion			
Yes	2 (25%)	7 (50%)	0.380
No	6 (75%)	7 (50%)	
Feeding intolerance			
Yes	2 (25%)	2 (14.3%)	0.602
No	6 (75%)	12 (85.7%)	
African American	**No** (n=6)	**Yes** (n=25)	**p**
Sepsis			
Yes	0 (0%)	5 (20%)	0.553
No	6 (100%)	20 (80%)	
ROP			
Yes	0 (0%)	3 (12%)	1.0
No	4 (66.7%)	21 (84%)	
BPD			

Yes	1 (16.7%)	2 (8%)	0.488
No	5 (83.3%)	23 (92%)	
NEC			
Yes	0 (0%)	1 (4%)	1.0
No	6 (100%)	24 (96%)	
IVH			
Yes	2 (33.3%)	1 (4%)	0.042*
No	2 (33.3%)	24 (96%)	
Blood transfusion			
Yes	3 (50%)	12 (48%)	1.0
No	3 (50%)	13 (52%)	
Feeding intolerance			
Yes	0 (0%)	3 (12%)	1.0
No	6 (100%)	21 (84%)	

Hispanic	**No** (n=9)	**Yes** (n=6)	**p**
Sepsis			
Yes	1 (11.1%)	1 (16.7%)	1.0
No	6 (66.7%)	4 (66.7%)	
ROP			
Yes	3 (33.3%)	3 (50%)	1.0
No	4 (44.4%)	2 (33.3%)	
BPD			
Yes	0 (0%)	0 (0%)	n/a
No	8 (88.9%)	5 (83.3%)	
NEC			
Yes	1 (11.1%)	0 (0%)	1.0
No	7 (77.8%)	5 (83.3%)	
IVH			
Yes	2 (22.2%)	0 (0%)	0.487
No	6 (66.7%)	5 (83.3%)	
Blood transfusion			
Yes	3 (33.3%)	4 (66.7%)	1.0
No	5 (55.6%)	2 (33.3%)	
Feeding intolerance			
Yes	4 (44.4%)	2 (33.3%)	1.0

No	5 (55.6%)	2 (33.3%)

ROP=retinopathy of prematurity; BPD=bronchopulmonary dysplasia; NEC=necrotizing enterocolitis; IVH=intraventricular hemorrhage

Table 19: Multivariate analysis of maternal IL SNP and continuous infant outcomes

African Americans: multivariate model for continuous infant outcomes with minor allele presence rs2070874

Outcome	Predictor	Estimate	p-value
lnWeight at 6 weeks	Minor allele presence – no	0.105	0.2627
	Minor allele presence – yes (reference)		
	Gestational Age	0.621	0.0107
	Ratio of mom's own milk to all milk	-0.344	0.026*
lnCalprotectin Week 1	Minor allele presence – no	-0.242	0.5666
	Minor allele presence – yes (reference)		
	Gestational Age	-0.021	0.7258
	Ratio of mom's own milk to all milk	-1.197	0.1025
lnCalprotectin Week 2	Minor allele presence – no	0.5807	0.0589*
	Minor allele presence – yes (reference)		
	Gestational Age	0.030	0.5764
	Ratio of mom's own milk to all milk	-1.111	0.0730*
lnSNAPPEII	Minor allele presence – no	0.284	0.2424
	Minor allele presence – yes (reference)		
	Gestational Age	-0.085	0.0426*
	Ratio of mom's own milk to all milk	0.207	0.6327

Caucasians: multivariate model for continuous infant outcomes with minor allele presence rs2070874

Outcome	Predictor	Estimate	p-value
Length of Stay	Minor allele presence – no	-10.701	0.3684
	Minor allele presence – yes (reference)		
	Gestational Age	-6.836	0.0025*
	Ratio of mom's own milk to all milk	11.978	0.3908
Days on oxygen	Minor allele presence – no	-10.364	0.3968
	Minor allele presence – yes (reference)		
	Gestational Age	-6.761	0.0034*
	Ratio of mom's own milk to all milk	9.584	0.5031
Weight at 6 weeks	Minor allele presence – no	-0.0169	0.8754
	Minor allele presence – yes (reference)		
	Gestational Age	0.055	0.0190*
	Ratio of mom's own milk to all milk	0.042	0.7419

Hispanics: multivariate model for continuous infant outcomes with minor allele presence rs2070874

Outcome	Predictor	Estimate	p-value
lnCalprotectin Week 3	Minor allele presence – no	0.548	0.238
	Minor allele presence – yes (reference)		
	Gestational Age	0.113	0.447
	Ratio of mom's own milk to all milk	-8.510	0.1836

Caucasians: multivariate model for continuous infant outcomes with genotype rs2243250

Outcome	Predictor	Estimate	p-value
Length of stay	Genotype CC	2.978	0.7390
	Genotype TT	25.261	0.0720*
	Genotype CT (reference)		
	Gestational Age	-6.403	0.0027*
	Ratio of mom's own milk to all milk	11.728	0.3629
lnWeight at 6 weeks	Genotype CC	-0.005	0.9550
	Genotype TT	-0.0157	0.2326
	Genotype CT (reference)		
	Gestational Age	0.047	0.0372*
	Ratio of mom's own milk to all milk	0.063	0.6038
Days on oxygen	Genotype CC	4.283	0.6555
	Genotype TT	22.394	0.1312
	Genotype CT (reference)		
	Gestational Age	-6.402	0.0046*
	Ratio of mom's own milk to all milk	9.629	0.4839
lnCalprotectin Week 2	Genotype CC	0.611	0.1651
	Genotype TT	-0.266	0.6481
	Genotype CT (reference)		
	Gestational Age	-0.042	0.6741
	Ratio of mom's own milk to all milk	0.002	0.9966
lnCalprotectin Week 3	Genotype CC	-0.155	0.6540
	Genotype TT	-0.694	0.1744
	Genotype CT (reference)		
	Gestational Age	0.063	0.4862
	Ratio of mom's own milk to all milk	0.136	0.7811

African Americans: multivariate model for continuous infant outcomes with genotype rs2243250

Outcome	Predictor	Estimate	p-value
Length of stay	Genotype CC	24.923	0.0518*
	Genotype TT	-2.289	0.8295
	Genotype CT (reference)		
	Gestational Age	-8.657	<0.001*
	Ratio of mom's own milk to all milk	-17.174	0.3158
lnWeight at 6 weeks	Genotype CC	-0.114	0.2894
	Genotype TT	-0.059	0.5111
	Genotype CT (reference)		
	Gestational Age	0.0369	0.1258
	Ratio of mom's own milk to all milk	-0.262	0.0851*
Days on oxygen	Genotype CC	19.173	0.0809*
	Genotype TT	10.610	0.2608
	Genotype CT (reference)		
	Gestational Age	-2.952	0.0424*
	Ratio of mom's own milk to all milk	12.779	0.3733
lnCalprotectin Week 1	Genotype CC	-47.301	0.6484
	Genotype TT	138.096	0.1520
	Genotype CT (reference)		
	Gestational Age	-12.937	0.3433
	Ratio of mom's own milk to all milk	-197.385	0.1856

Hispanics: multivariate model for continuous infant outcomes with genotype rs2243250

Outcome	Predictor	Estimate	p-value
lnCalprotectin week 1	Genotype CC	360.22	0.1605
	Genotype TT	-100.711	0.7002
	Genotype CT (reference)		
	Gestational Age	-48.106	0.4751
	Ratio of mom's own milk to all milk	495.52	0.4717
lnCalprotectin Week 3	Genotype CC	0.468	0.3610
	Genotype TT	-0.339	0.5638
	Genotype CT (reference)		
	Gestational Age	0.142	0.3377
	Ratio of mom's own milk to all milk	-8.309	0.2048

Total population: multivariate model for continuous infant outcomes with genotype rs2243250

Outcome	Predictor	Estimate	p-value
lnCalprotectin week 3	Genotype CC	29.27	0.6261
	Genotype TT	-89.557	0.1615
	Genotype CT (reference)		
	Gestational Age	16.294	0.1533
	Ratio of mom's own milk to all milk	111.78	0.2880

Caucasians: multivariate model for continuous infant outcomes with minor allele presence rs1800795

Outcome	Predictor	Estimate	p-value
Days on oxygen	Minor allele presence – no	-8.332	0.2698
	Minor allele presence – yes (reference)		
	Gestational Age	-6.95	0.0021*
	Ratio of mom's own milk to all milk	6.801	0.6396
lnCalprotectin Week 2	Minor allele presence – no	-0.743	0.0222*
	Minor allele presence – yes (reference)		
	Gestational Age	0.011	0.9038
	Ratio of mom's own milk to all milk	-0.535	0.3383

African Americans: multivariate model for continuous infant outcomes with minor allele presence rs1800795

Outcome	Predictor	Estimate	p-value
lnSNAPPEII	Minor allele presence – no	0.344	0.2235
	Minor allele presence – yes (reference)		
	Gestational Age	-0.066	0.1187
	Ratio of mom's own milk to all milk	0.513	0.1984

Hispanics: multivariate model for continuous infant outcomes with minor allele presence rs1800795

Outcome	Predictor	Estimate	p-value
Length of stay	Minor allele presence – no	9.701	0.6945
	Minor allele presence – yes (reference)		
	Gestational Age	-7.122	0.2932
	Ratio of mom's own milk to all milk	-184.046	0.0350*
lnCalprotectin Week 1	Minor allele presence – no	0.241	0.7971
	Minor allele presence – yes (reference)		
	Gestational Age	-0.063	0.8116
	Ratio of mom's own milk to all milk	4.662	0.1360

Total Population: multivariate model for continuous infant outcomes with minor allele presence rs1800795

Outcome	Predictor	Estimate	p-value
lnSNAPPEII	Minor allele presence – no	0.179	0.2790
	Minor allele presence – yes (reference)		
	Gestational Age	-0.11	0.0023*
	Ratio of mom's own milk to all milk	-0.187	0.5745
lnSNAPPEII	Genotype CC	0.721	0.0652*
	Genotype GG	0.276	0.1075*
	Genotype CG (reference)		
	Gestational Age	-0.108	0.0023*
	Ratio of mom's own milk to all milk	-0.079	0.8117
Days on oxygen	Minor allele presence – no	-9.588	0.0316*
	Minor allele presence – yes (reference)		
	Gestational Age	-5.048	<0.001*
	Ratio of mom's own milk to all milk	1.02	0.9076
Days on oxygen	Genotype CC	1.529	0.8622
	Genotype GG	-9.29	0.0533*
	Genotype CG (reference)		
	Gestational Age	-5.048	<0.001*
	Ratio of mom's own milk to all milk	1.107	0.9008
lnCalprotectin week 2	Minor allele presence – no	-0.253	0.2117
	Minor allele presence – yes (reference)		
	Gestational Age	0.079	0.0940*
	Ratio of mom's own milk to all milk	0.209	0.6038
lnCalprotectin week 2	Genotype CC	0.279	0.4870
	Genotype GG	-0.163	0.3716
	Genotype CG (reference)		

	Gestational Age	0.083	0.0833*
	Ratio of mom's own milk to all milk	0.217	0.5932

Caucasians: multivariate model for continuous infant outcomes with minor allele presence rs1800796

Outcome	Predictor	Estimate	p-value
lnCalprotectin week 3	Minor allele presence – no	0.815	0.0429*
	Minor allele presence – yes (reference)		
	Gestational Age	0.167	0.0530*
	Ratio of mom's own milk to all milk	0.252	0.5708

African Americans: multivariate model for continuous infant outcomes with minor allele presence rs1800796

Outcome	Predictor	Estimate	p-value
Length of stay	Minor allele presence – no	-32.318	0.0158*
	Minor allele presence – yes (reference)		
	Gestational Age	-9.142	<0.001
	Ratio of mom's own milk to all milk	-10.227	0.5371
lnSNAPPEII	Minor allele presence – no	-0.668	0.0497*
	Minor allele presence – yes (reference)		
	Gestational Age	-0.059	0.1350
	Ratio of mom's own milk to all milk	0.579	0.1310
lnCalprotectin Week 3	Minor allele presence – no	-0.602	0.0868*
	Minor allele presence – yes (reference)		
	Gestational Age	0.016	0.7278
	Ratio of mom's own milk to all milk	-0.214	0.6610

Hispanics: multivariate model for continuous infant outcomes with minor allele presence rs1800796

Outcome	Predictor	Estimate	p-value
lnWeight at 6 weeks	Minor allele presence – no	-0.195	0.0272*
	Minor allele presence – yes (reference)		
	Gestational Age	0.089	0.0027*
	Ratio of mom's own milk to all milk	-0.519	0.0822*
lnCalprotectin Week 3	Minor allele presence – no	0.125	0.7941
	Minor allele presence – yes (reference)		
	Gestational Age	0.159	0.3460
	Ratio of mom's own milk to all milk	-10.593	0.1481

Total population: multivariate model for continuous infant outcomes with minor allele presence rs1800796

Outcome	Predictor	Estimate	p-value
Length of stay	Minor allele presence – no	-17.986	0.0688*
	Minor allele presence – yes (reference)		
	Gestational Age	-8.573	<0.0001
	Ratio of mom's own milk to all milk	-12.514	0.4299
Length of stay	Genotype GG	69.376	0.0034*
	Genotype AA	-7.257	0.4626
	Genotype AG (reference)		
	Gestational Age	-8.411	<0.0001
	Ratio of mom's own milk to all milk	-12.07	0.4194
lnWeight at 6 weeks	Minor allele presence – no	-0.066	0.2108
	Minor allele presence – yes (reference)		
	Gestational Age	0.055	<0.0001
	Ratio of mom's own milk to all milk	-0.091	0.2689
lnWeight at 6 weeks	Genotype GG	-0.100	0.3925
	Genotype AA	-0.088	0.1368
	Genotype AG (reference)		
	Gestational Age	0.055	<0.001
	Ratio of mom's own milk to all milk	-0.097	0.2414
lnCalprotectin week 3	Minor allele presence – no	-0.291	0.1607
	Minor allele presence – yes (reference)		
	Gestational Age	0.052	0.1958
	Ratio of mom's own milk to all milk	0.229	0.5473
lnCalprotectin Week 3	Genotype GG	0.974	0.0213*
	Genotype AA	-0.063	0.7730
	Genotype AG (reference)		
	Gestational Age	0.065	0.0968*

	Ratio of mom's own milk to all milk	0.257	0.4826

Caucasians: multivariate model for continuous infant outcomes with minor allele presence rs1800871

Outcome	Predictor	Estimate	p-value
Length of stay	Minor allele presence – no	15.709	0.0989*
	Minor allele presence – yes (reference)		
	Gestational Age	-9.041	<0.0001*
	Ratio of mom's own milk to all milk	-9.104	0.6104
lnCalprotectin week 1	Minor allele presence – no	-0.763	0.0890*
	Minor allele presence – yes (reference)		
	Gestational Age	-0.103	0.4835
	Ratio of mom's own milk to all milk	-0.357	0.6594
lnCalprotectin Week 2	Minor allele presence – no	0.428	0.2402
	Minor allele presence – yes (reference)		
	Gestational Age	-0.013	0.8998
	Ratio of mom's own milk to all milk	0.231	0.6998

African Americans: multivariate model for continuous infant outcomes with minor allele presence rs1800871

Outcome	Predictor	Estimate	p-value
Length of stay	Minor allele presence – no	15.709	0.0989*
	Minor allele presence – yes (reference)		
	Gestational Age	-9.041	<0.0001*
	Ratio of mom's own milk to all milk	-9.104	0.6104

Hispanics: multivariate model for continuous infant outcomes with minor allele presence rs1800871

Outcome	Predictor	Estimate	p-value
lnCalprotectin week 2	Minor allele presence – no	0.048	0.9084
	Minor allele presence – yes (reference)		
	Gestational Age	0.402	0.0282*
	Ratio of mom's own milk to all milk	-0.302	0.8316

Total Population: multivariate model for continuous infant outcomes with minor allele presence rs1800871

Outcome	Predictor	Estimate	p-value
lnCalprotectin Week 3	Minor allele presence - no	-0.001	0.9949
	Minor allele presence – yes (reference)		
	Gestational Age	0.055	0.1737
	Ratio of mom's own milk to all milk	0.286	0.4647
lnCalprotectin Week 3	Genotype TT	0.732	0.0270*
	Genotype CC	0.143	0.4381
	Genotype CT (reference)		
	Gestational Age	0.059	0.1299
	Ratio of mom's own milk to all milk	0.174	0.6452

Caucasians: multivariate model for continuous infant outcomes with minor allele presence rs1800872

Outcome	Predictor	Estimate	p-value
Length of stay	Minor allele presence – no	7.761	0.2763
	Minor allele presence – yes (reference)		
	Gestational Age	-6.747	0.0026*
	Ratio of mom's own milk to all milk	14.478	0.2895
lnWeight at 6 weeks	Minor allele presence – no	-0.082	0.2041
	Minor allele presence – yes (reference)		
	Gestational Age	0.052	0.0168*
	Ratio of mom's own milk to all milk	0.039	0.7411
lnCalprotectin week 1	Minor allele presence – no	-0.997	0.0196*
	Minor allele presence – yes (reference)	-0.202	0.0987*
	Gestational Age	-0.044	0.9474
	Ratio of mom's own milk to all milk		

African Americans: multivariate model for continuous infant outcomes with minor allele presence rs1800872

Outcome	Predictor	Estimate	p-value
Length of stay	Minor allele presence – no	10.012	0.3255
	Minor allele presence – yes (reference)		
	Gestational Age	-9.762	<0.0001
	Ratio of mom's own milk to all milk	-18.816	0.3375

Hispanics: multivariate model for continuous infant outcomes with minor allele presence rs1800872

Outcome	Predictor	Estimate	p-value
lnCalprotectin week 2	Minor allele presence – no	0.408	0.9084
	Minor allele presence – yes (reference)		
	Gestational Age	0.402	0.0282*
	Ratio of mom's own milk to all milk	-0.302	0.8316

Total Population: multivariate model for continuous infant outcomes with minor allele presence rs1800872

Outcome	Predictor	Estimate	p-value
Length of Stay	Minor allele presence – no	6.078	0.4546
	Minor allele presence – yes (reference)		
	Gestational Age	-8.157	<0.001*
	Ratio of mom's own milk to all milk	-8.832	0.5927
Length of Stay	Genotype CC	4.422	0.6140
	Genotype AA	-7.109	0.6036
	Genotype AC (reference)		
	Gestational Age	-8.057	<0.001*
	Ratio of mom's own milk to all milk	-6.864	0.6868
lnCalprotectin Week 3	Minor allele presence – no	0.006	0.9710
	Minor allele presence – yes (reference)		
	Gestational Age	0.065	0.1499
	Ratio of mom's own milk to all milk	0.341	0.3990
lnCalprotectin Week 3	Genotype CC	0.203	0.2892
	Genotype AA	0.768	0.0158*
	Genotype AC (reference)		
	Gestational Age	0.063	0.1466
	Ratio of mom's own milk to all milk	0.074	0.8523

Caucasians: multivariate model for continuous infant outcomes with minor allele presence rs1800896

Outcome	Predictor	Estimate	p-value
lnCalprotectin Week 2	Minor allele presence – no	0.582	0.0737*
	Minor allele presence – yes (reference)		
	Gestational Age	-0.058	0.5695
	Ratio of mom's own milk to all milk	-0.136	0.8034
lnCalprotectin Week 3	Minor allele presence – no	0.244	0.3576
	Minor allele presence – yes (reference)		
	Gestational Age	0.083	0.3316
	Ratio of mom's own milk to all milk	-0.026	0.9582

African Americans: multivariate model for continuous infant outcomes with minor allele presence rs1800896

Outcome	Predictor	Estimate	p-value
Days on oxygen	Minor allele presence – no	21.589	0.0320*
	Minor allele presence – yes (reference)		
	Gestational Age	-4.699	0.0016*
	Ratio of mom's own milk to all milk	-5.597	0.7021

Hispanics: multivariate model for continuous infant outcomes with minor allele presence rs1800896

Outcome	Predictor	Estimate	p-value
Length of stay	Minor allele presence – no	2.249	0.8656
	Minor allele presence – yes (reference)		
	Gestational Age	-9.808	<0.0001
	Ratio of mom's own milk to all milk	-17.667	0.3866
lnCalprotectin Week 1	Minor allele presence – no	0.614	0.5005
	Minor allele presence – yes (reference)		
	Gestational Age	-0.084	0.7390
	Ratio of mom's own milk to all milk	4.011	0.1763
lnCalprotectin Week 3	Minor allele presence – no	0.771	0.2073
	Minor allele presence – yes (reference)		
	Gestational Age	0.119	0.4411
	Ratio of mom's own milk to all milk	-8.200	0.2258

Total Population: multivariate model for continuous infant outcomes with minor allele presence rs1800896

Outcome	Predictor	Estimate	p-value
Length of stay	Minor allele presence – no	-12.461	0.1594
	Minor allele presence – yes (reference)		
	Gestational Age	-7.938	<0.0001
	Ratio of mom's own milk to all milk	-1.298	0.9396
Length of stay	Genotype GG	-1.128	0.9199
	Genotype AA	-12.832	0.1841
	Genotype AG (reference)		
	Gestational Age	-7.932	<0.0001
	Ratio of mom's own milk to all milk	-0.854	0.9618
lnCalprotectin Week 2	Minor allele presence – no	0.644	0.0057*
	Minor allele presence – yes (reference)		
	Gestational Age	0.058	0.2023
	Ratio of mom's own milk to all milk	-0.251	0.5478
lnCalprotectin week 2	Genotype GG	0.220	0.3795
	Genotype AA	0.741	0.0045*
	Genotype AG (reference)		
	Gestational Age	0.055	0.2297
	Ratio of mom's own milk to all milk	-0.379	0.3928
lnCalprotectin Week 3	Minor allele presence – no	0.342	0.0825*
	Minor allele presence – yes (reference)		
	Gestational Age	0.050	0.2098

	Ratio of mom's own milk to all milk	0.050	0.9006
lnCalprotectin week 3	Genotype GG	0.139	0.5982
	Genotype AA	0.387	0.0744*
	Genotype AG (reference)		
	Gestational Age	0.051	0.2076
	Ratio of mom's own milk to all milk	-0.013	0.9760

Table 20: Multivariate analysis of maternal SNP and categorical infant outcomes

Total Population: multivariate model for categorical infant outcomes with rs2070874

Outcome	Predictor	OR, 95% CI	p
ROP	MAP – no		0.2250
	MAP – no versus yes	0.437 (0.115, 0.1665)	0.2250
	Gestational Age	0.633 (0.433, 0.927)	0.0188*
	Ratio of mom's own milk to all milk	0.877 (0.0.54, 14.340)	0.9268
ROP	Genotype		
	Genotype CC versus CT	0.885 (0.149, 5.240)	0.1636
	Genotype TT versus CT	4.558 (0.552, 37.640)	0.2171
	Gestational Age	0.642 (0.433, 0.952)	0.0706*
	Ratio of mom's own milk to all milk	1.290 (0.079, 21.171)	0.0277*

Total Population: multivariate model for categorical infant outcomes with rs2243250

Outcome	Predictor	OR, 95% CI	p
Transfusion	Genotype		
	Genotype CC versus CT	2.752 (0.618, 12.261)	0.404
	Genotype TT versus CT	2.224 (0.462, 10.719)	0.3098
	Gestational Age	0.464 (0.311, 0.693)	0.6472
	Ratio of mom's own milk to all milk	1.048 (0.090, 12.181)	0.0002*

Caucasians: multivariate model for categorical infant outcomes with minor allele presence rs1800795

Outcome	Predictor	OR, 95% CI	p
ROP	MAP – no		0.1560
	MAP – no versus yes	10.605 (0.406, 276.879)	0.1560
	Gestational Age	0.580 (0.190, 1.774)	0.3399
	Ratio of mom's own milk to all milk	0.910 (0.004, 197.685)	0.9727
Transfusion	MAP – no		0.1145
	MAP – no versus yes	0.102 (0.006, 1.738)	0.1145
	Gestational Age	0.675 (0.374, 1.220)	0.1933
	Ratio of mom's own milk to all milk	0.089 (<0.001, 13.987)	0.3479

African Americans: multivariate model for categorical infant outcomes minor allele presence rs1800795

Outcome	Predictor	OR, 95% CI	p
IVH	MAP - no		0.1116
	MAP – no versus yes	<0.001 (<0.0001, 6.715)	0.1116
	Gestational Age	0.219 (0.029, 1.649)	0.1403
	Ratio of mom's own milk to all milk	<0.001 (<0.001, 510.404)	0.2910

African Americans: multivariate model for categorical infant outcomes minor allele presence rs1800796

Outcome	Predictor	OR, 95% CI	p
Sepsis	MAP - no		0.5109
	MAP – no versus yes	0.076 (<0.001, 165.882)	0.5109
	Gestational Age	0.334 (0.119, 0.938)	0.0375*
	Ratio of mom's own milk to all milk	0.149 (0.001, 20.377)	0.4485

Hispanics: multivariate model for categorical infant outcomes minor allele presence rs1800796

Outcome	Predictor	OR, 95% CI	p
Feeding Intolerance	MAP - no		0.5982
	MAP – no versus yes	2.626 (0.072, 95.151)	0.5982
	Gestational Age	0.187 (0.018, 1.947)	0.1609
	Ratio of mom’s own milk to all milk	<0.001 (<0.001, >999.999)	0.6131

Total Population: multivariate model for categorical infant outcomes with rs1800796

Outcome	Predictor	OR, 95% CI	p
ROP	MAP – no		0.0573*
	MAP – no versus yes	0.223 (0.048, 1.048)	0.0573*
	Gestational Age	0.630 (0.429, 0.925)	0.0185*
	Ratio of mom's own milk to all milk	0.440 (0.023, 8.441)	0.5859
ROP	Genotype		0.1492
	Genotype GG versus AG	1.955 (0.078, 49.084)	0.3782
	Genotype AA versus AG	0.253 (0.048, 1.325)	0.0687*
	Gestational Age	0.630 (0.429, 0.925)	0.0185*
	Ratio of mom's own milk to all milk	0.447 (0.023, 8.597)	0.5938
BPD	MAP – no		0.3860
	MAP – no versus yes	0.337 (0.029, 3.941)	0.3860
	Gestational Age	0.438 (0.199, 0.963)	0.0400*
	Ratio of mom's own milk to all milk	0.320 (0.001, 78.279)	0.6845
NEC	MAP – no		0.0883*
	MAP – no versus yes	0.106 (0.008, 1.401)	0.0883*
	Gestational Age	0.910 (0.557, 1.488)	0.7075

	Ratio of mom's own milk to all milk	0.326 (0.001, 93.334)	0.6975
Transfusion	MAP – no		0.7995
	MAP – no versus yes	0.817 (0.172, 3.875)	0.7995
	Gestational Age	0.454 (0.306, 0.675)	<0.001*
	Ratio of mom's own milk to all milk	0.941 (0.080, 11.132)	0.9617
Feeding intolerance	MAP – no		0.1282
	MAP – no versus yes	0.313 (0.070, 1.398)	0.1282
	Gestational Age	0.632 (0.444, 0.899)	0.0108*
	Ratio of mom's own milk to all milk	0.353 (0.024, 5.232)	0.4491

Caucasians: multivariate model for categorical infant outcomes minor allele presence rs1800871

Outcome	Predictor	OR, 95% CI	p
IVH	MAP - no		0.1189
	MAP – no versus yes	0.116 (0.008, 1.737)	0.1189
	Gestational Age	0.481 (0.559, 3.923)	0.4299
	Ratio of mom's own milk to all milk	151.263 (<0.001, >999.999)	0.4524

Total Population: multivariate model for categorical infant outcomes with rs1800871

Outcome	Predictor	OR, 95% CI	p
Sepsis	MAP – no		0.6683
	MAP – no versus yes	1.52 (0.224, 10.318)	0.6683
	Gestational Age	0.363 (0.186, 0.712)	0.0032*
	Ratio of mom's own milk to all milk	0.199 (0.006, 60587)	0.3662
Sepsis	Genotype		
	Genotype TT versus CT	7.386 (0.283, 193.11)	0.4778
	Genotype CC versus CT	2.693 (0.249, 29.173)	0.2914
	Gestational Age	0.352 (0.176, 0.703)	0.9930
	Ratio of mom's own milk to all milk	0.152 (0.004, 5.585)	0.0031*

African Americans: multivariate model for categorical infant outcomes minor allele presence rs1800872

Outcome	Predictor	OR, 95% CI	p
Sepsis	MAP - no		0.8579
	MAP – no versus yes	0.722 (0.021, 25.407)	0.8579
	Gestational Age	0.269 (0.075, 0.964)	0.0438*

	Ratio of mom's own milk to all milk	0.090 (<0.001, 24.030)	0.3977

Total Population: multivariate model for categorical infant outcomes with rs1800872

Outcome	Predictor	OR, 95% CI	p
Sepsis	MAP – no		0.6620
	MAP – no versus yes	1.531 (0.227, 10.322)	0.6620
	Gestational Age	0.364 (0.186, 0.714)	0.0033*
	Ratio of mom's own milk to all milk	0.197 (0.006, 6.412)	0.3607
Sepsis	Genotype		
	Genotype CC versus AC	2.696 (0.250, 29.095)	0.4861
	Genotype AA versus AC	7.162 (0.276, 186.138)	0.9943
	Gestational Age	0.350 (0.175, 0.700)	0.3004
	Ratio of mom's own milk to all milk	0.147 (0.004, 5.392)	0.0030*

Total Population: multivariate model for categorical infant outcomes with rs1800896

Outcome	Predictor	OR, 95% CI	p
Sepsis	MAP – no		0.1819
	MAP – no versus yes	0.194 (0.017, 2.155)	0.1819
	Gestational Age	0.337 (0.165, 0.687)	0.0028*
	Ratio of mom's own milk to all milk	0.315 (0.010, 10.394)	0.5170
Sepsis	Genotype		0.3975
	Genotype GG versus AG	0.780 (0.107, 5.677)	0.5548
	Genotype AA versus AG	0.178 (0.015, 2.171)	0.1963
	Gestational Age	0.339 (0.166, 0.692)	0.0030*
	Ratio of mom's own milk to all milk	0.341 (0.010, 11.721)	0.5507

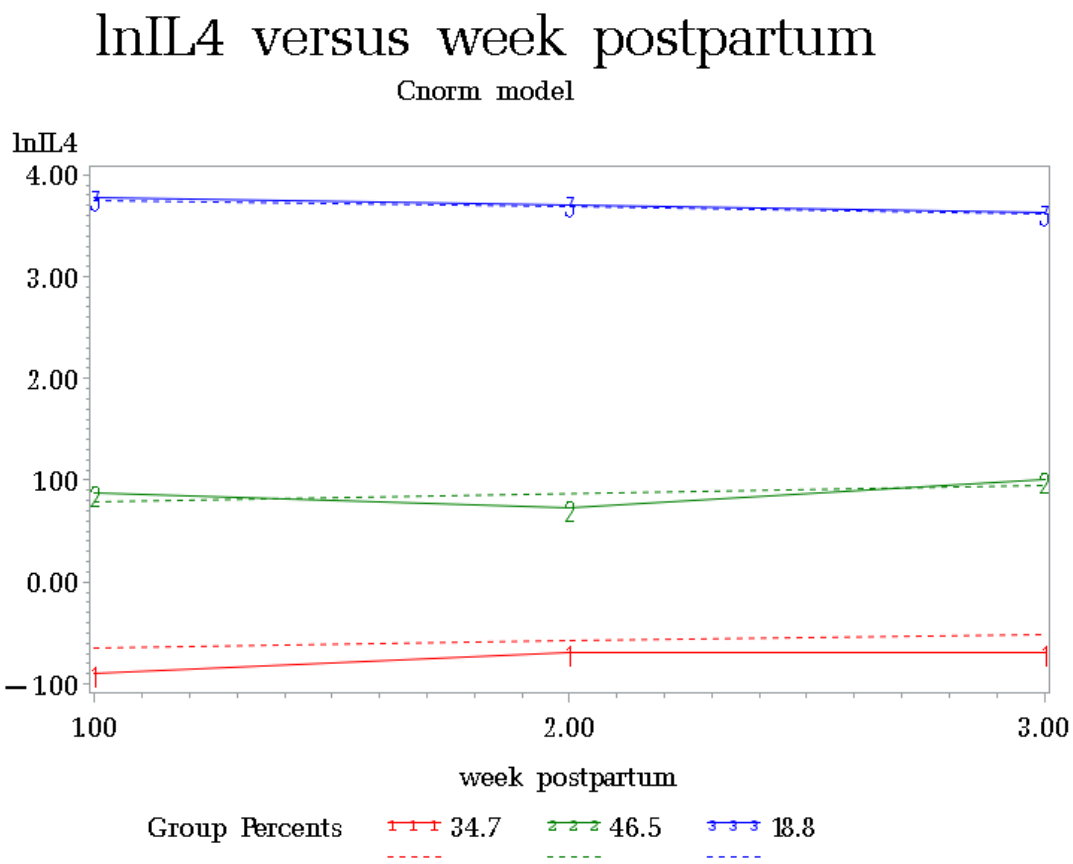

Figure 2: Interleukin 4 Trajectory Groups

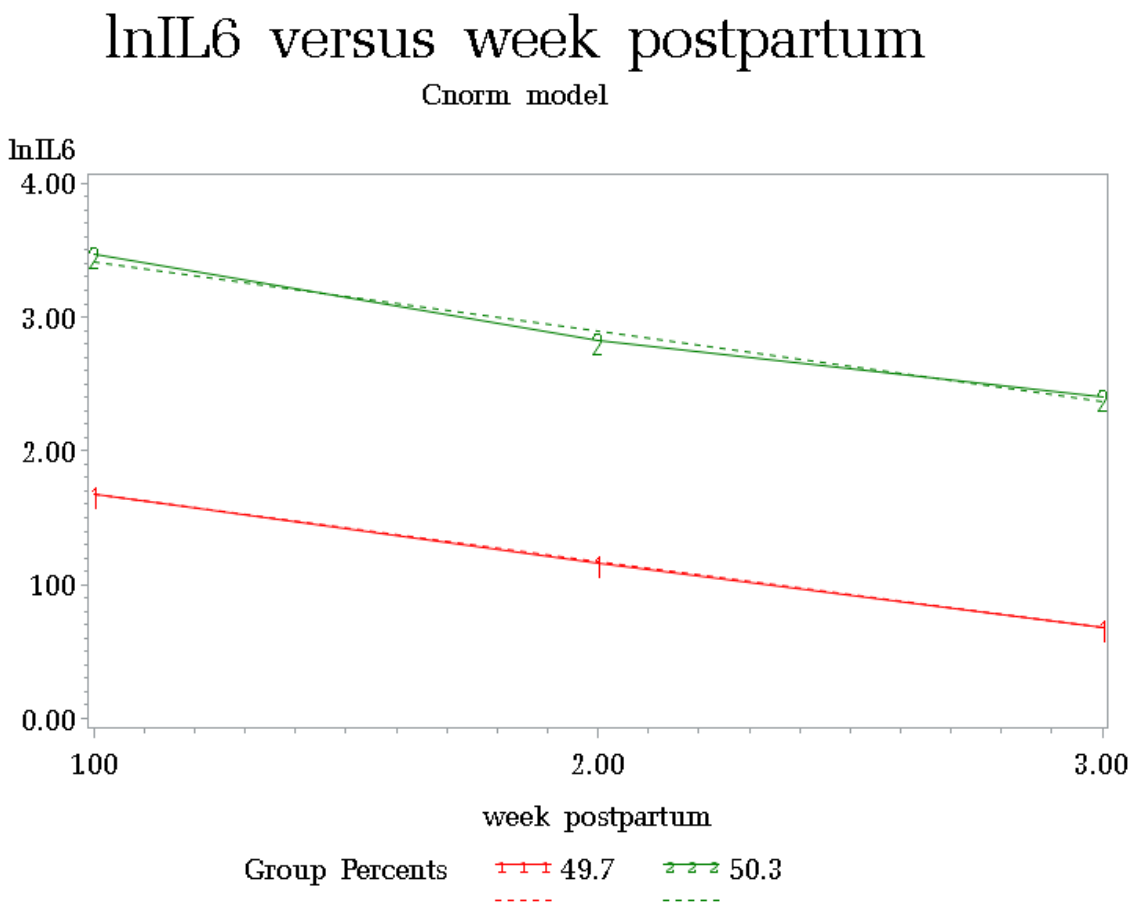

Figure 3: Interleukin 6 Trajectory Groups

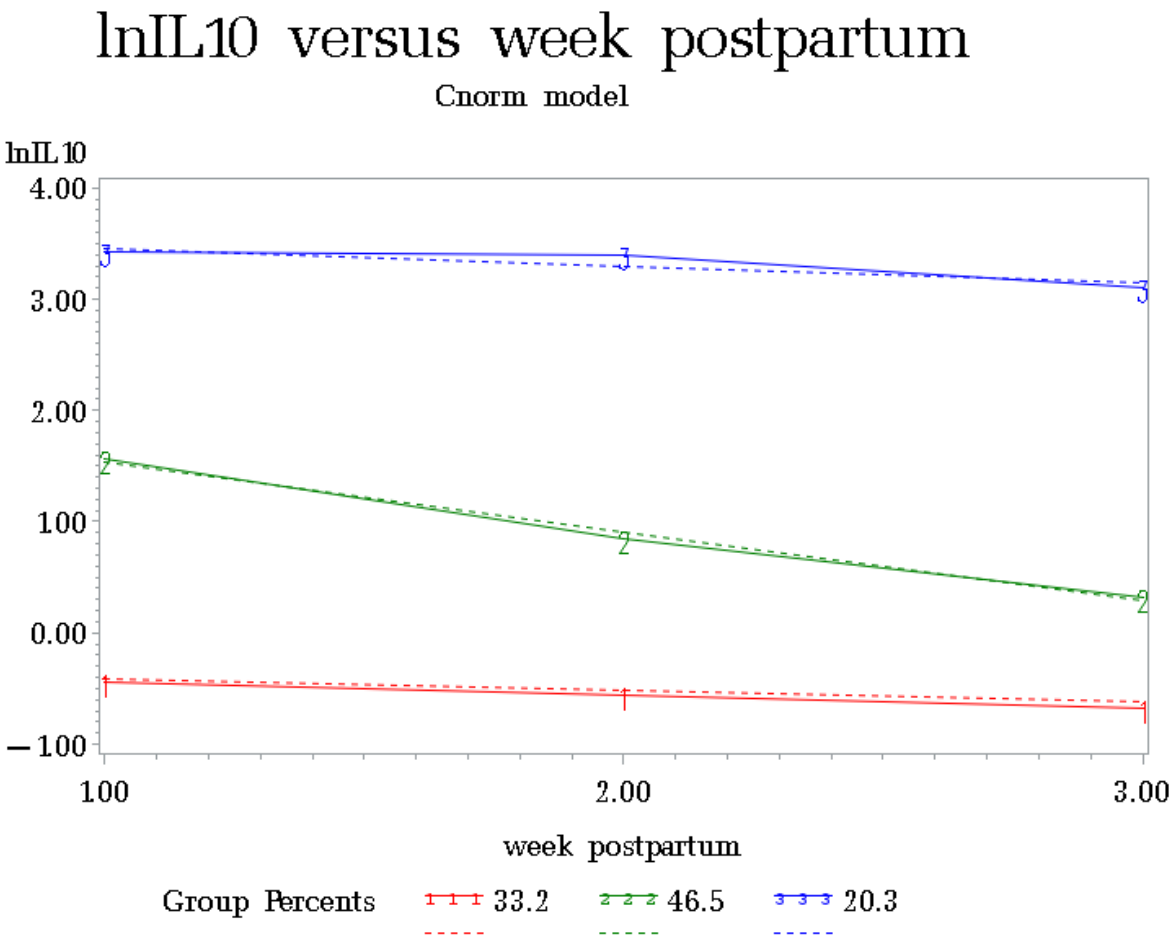

Figure 4: Interleukin 10 Trajectory Groups

APPENDIX B

UNIVERSITY OF PITTSBURGH INSTITUTIONAL REVIEW BOARD'S APPROVAL LETTER

University of Pittsburgh
Institutional Review Board

3500 Fifth Avenue
Pittsburgh, PA 15213
(412) 383-1480
(412) 383-1508 (fax)
http://www.irb.pitt.edu

Memorandum

To: Kelley Baumgartel
From: Christopher Ryan , Ph.D., Vice Chair
Date: 4/22/2013
IRB#: PRO13040181
Subject: Breastmilk is not a uniform substance: epigenetic mechanisms

---

The above-referenced protocol has been reviewed by the University of Pittsburgh Institutional Review Board. Based on the information provided to the IRB, this project includes no involvement of human subjects, according to the federal regulations [§45 CFR 46.102(f)]. That is, the investigator conducting research will not obtain information about research subjects via an interaction with them, nor will the investigator obtain identifiable private information. Should that situation change, the investigator must notify the IRB immediately.

Given this determination, you may now begin your project.

Please note the following information:

- If any modifications are made to this project, use the "**Send Comments to IRB Staff**" process from the project workspace to request a review to ensure it continues to meet the determination.
- Upon completion of your project, be sure to finalize the project by submitting a "**Study Completed**" report from the project workspace.

Please be advised that your research study may be audited periodically by the University of Pittsburgh Research Conduct and Compliance Office.

APPENDIX C

MANUSCRIPT #1: MOLECULAR GENOMIC RESEARCH DESIGNS

Published in final edited form as:
Annu Rev Nurs Res. 2011 ; 29: 1–26.

Molecular Genomic Research Designs

Kelley Baumgartel[1], **Jamie Zelazny**[1], **Theresa Timcheck**[1], **Chantel Snyder**[1], **Mandy Bell**[1], and **Yvette Conley**[1,2,*]

[1]University of Pittsburgh School of Nursing

[2]University of Pittsburgh Department of Human Genetics

Abstract

Genetic and genomic research approaches have the capability to expand our understanding of the complex pathophysiology of disease susceptibility, susceptibility to complications related to disease, trajectory of recovery from acquired injuries and infections, patient response to interventions and therapeutics, as well as informing diagnoses and prognoses. Nurse scientists are actively involved in all of these fields of inquiry and the goal of this manuscript is to assist with incorporation of genetic and genomic trajectories into their research and facilitate the design and execution of these studies. New studies that are going to embark on recruitment, phenotyping, and sample collection will benefit from forethought about research design to ensure that it addresses the research questions or hypotheses being tested. Studies that will utilize existing data or samples will also benefit from forethought about research design for the same reason but to also address the fact that some designs may not be feasible with the available data or samples. This manuscript discusses candidate gene association, genome wide association, candidate gene expression, global gene expression, and epigenetic/epigenomic study designs. Information provided includes rationale for selecting an appropriate study design, important methodology considerations for each design, key technologies available to accomplish each type of study, and online resources available to assist in executing each type of study design.

In the last decade we have progressed from a rough draft of the human genome sequence to availability of an abundance of publicly available databases and high throughput data collection technologies to facilitate genetic and genomic study design. Genetic (focus on one gene at a time) and genomic (focus on entire genome as well as gene-gene interactions) research continues to hold great promise for understanding a wealth of human conditions, providing objective data for diagnosis and prognosis, informing therapeutics, and providing the cornerstone for evidence based practice for genomic health care (Green, Guyer, & National Human Genome Research Institute [NHGRI], 2011; Lander, 2011). The research programs of many nurse scientists are ripe for incorporating a genetic/genomic research component or movement of existing genetic or genomic research in a new direction.

The goal of his paper is to bring together key information about designing studies with a molecular genetic or genomic focus coupled with dynamic resources offered to the reader to

*Corresponding Author: Yvette P. Conley, University of Pittsburgh, 3500 Victoria Street, 440 Victoria Building, Pittsburgh, PA 15261, yconley@pitt.edu, 412-383-7641.

Contributor's list:
Kelley Baumgartel, RN, Doctoral Student, University of Pittsburgh, Pittsburgh, PA
Jamie Zelazny, RN, MPH, Doctoral Student, University of Pittsburgh, Pittsburgh, PA
Theresa Timcheck, RN, Doctoral Student, University of Pittsburgh, Pittsburgh, PA
Chantel Snyder, RN, Doctoral Student, University of Pittsburgh, Pittsburgh, PA
Mandy Bell, RN, Doctoral Student, University of Pittsburgh, Pittsburgh, PA
Yvette P. Conley, PhD, Associate Professor of Nursing and Human Genetics, University of Pittsburgh, Pittsburgh, PA

expand their understanding and ensure access to state of the science information. It is not meant to be an exhaustive resource, but one that sets the stage for contemplation of embarking on such research designs and key issues to ponder during study design phase. This paper is written for the researcher who has a basic understanding of genetics and is contemplating adding a genetic or genomic component to their research or designing the next step in their genetic or genomic program of research. Readers are encouraged to visit an extremely useful resource, the National Human Genome Research Institute's talking glossary at http://www.genome.gov/glossary, for clarification of unfamiliar terms and expansion of knowledge about genetic terminology. Technology to collect genetic and genomic data changes rapidly, therefore proper study design, and selection of appropriate methodology to accomplish a study also change rapidly. This paper incorporates a large number of online resources that are continuously updated in an attempt to keep the paper as up to data as possible. Readers are encouraged to visit these online resources when designing their study to ensure that their study design is state of the science.

DNA POLYMORPHISM BASED ASSOCIATION STUDIES

The overall objective of a polymorphism based association study is to examine the relationship between DNA variation and a phenotype (e.g., diabetes, fatigue). A polymorphism is defined as a DNA variation that is present in at least one percent of the population (NHGRI, n.d.). One advantage of this approach compared to other genetic/genomic approaches is the use of DNA. DNA is a very stable template for experiments, allowing for use of previously collected samples. Such a retrospective approach could save time and money that would be needed to prospectively recruit participants and collect samples; however, attention must be given to subject consent to assure that informed consent was obtained for future genetic/genomic evaluation related to the phenotype of interest. Another advantage is that this approach does not require that subjects be related, which is a requirement for linkage analysis, an approach not discussed in this manuscript. It should be noted that while related individuals are not required, newer software has been developed to allow for the analyses of related individuals within the context of an association study. Two very appealing additional advantages of polymorphism based studies are the fact that polymorphisms do not change over time and the DNA template that is utilized can be extracted from any tissue. The sample for DNA extraction and collection of polymorphism data only need to be collected once, yet that polymorphism data can be evaluated within the context of a phenotype that changes over time. While blood and saliva are the most frequently used cell/tissue type for DNA extraction, any cells/tissues that have a nucleus can serve as samples for polymorphism based studies. Because DNA polymorphisms do not change and are not tissue specific, investigators need not worry about collection of DNA samples over time or from what tissue DNA extraction occurs. These advantages are not carried over to other genomic approaches detailed in this manuscript.

Candidate Gene Association Studies

Rationale for taking a candidate gene association approach—Candidate gene association studies investigate polymorphisms representing a specific gene(s) to determine if it is associated with a phenotype of interest. With this hypothesis-driven approach, the investigator pre-selects the candidate gene(s) to be evaluated. This approach is only appropriate when *a priori* assumptions about the gene(s) that may be involved in the phenotype of interest can be justified.

Genome wide association studies (GWAS); discussed in the next section, have large sample size requirements (e.g., 1000 cases/1000 controls), and one relative advantage of the candidate gene approach is that it often requires half that number or less. This reduced sample size requirement compared to a GWAS is due to the focused evaluation of a

candidate gene(s), which reduces multiple testing concerns. The candidate gene association approach is also ideal when studying rarer phenotypes since attainment of a large sample may not be feasible for a condition with a low population frequency.

Subject and sample considerations—Clearly defined inclusion/exclusion criteria, which include a detailed definition of the phenotype, are essential to the candidate gene association approach. Structured inclusion/exclusion criteria help to ensure that individuals with/without the phenotype of interest are similar in all aspects except for the condition being investigated. Moreover, phenotypic assessment of controls should be as comprehensive as the phenotypic assessment of cases. Ultimately, carefully crafted criteria, and thorough phenotypic assessments help reduce the impact of confounding variables.

Population stratification represents another potential source of confounding in candidate gene association studies utilizing a case-control design. The case-control design compares allele, genotype, or haplotype frequencies between the groups. Because these frequencies can be extremely disparate for different ancestries, it is important to control for ancestry to avoid spurious results/conclusions (e.g., concluding that there is an association between a phenotype/allele when in reality the association is fueled by ancestral differences in allelic frequencies). The risk for population stratification can be mitigated. Subgroup analysis represents one option, but it relies on self report to categorically measure race/ethnicity. An option that controls for population stratification statistically is the use of ancestral informative markers (AIM), which are polymorphisms in the DNA that allow one to calculate an admixture proportion for an individual. The application of these proportions are used for analysis rather than the traditionally used, though unreliable, method of self-reported race/ethnicity. In a recent study, only 30 AIMs were needed to estimate European admixture in a group of African American women (Ruiz-Narváez, Rosenberg, Wise, Reich, & Palmer, 2011). Although different AIMs may be needed to estimate other admixture proportions, this example demonstrates that population stratification can be successfully controlled through the analysis of genetic markers.

Another aspect of the candidate gene association study that should be considered is sample size requirements. Quanto (http://hydra.usc.edu/gxe/) is a freely downloadable computer program that can assist with sample size and/or power calculations for candidate gene association studies. User defined criteria can be manipulated according to the polymorphisms that have been selected for evaluation and according to study design specifications.

Candidate gene selection—Candidate gene selection is often based on biologic plausibility. This plausibility can be based on biological pathways implicated in the condition, biomarker data implicating a gene/gene product in the phenotype of interest, pharmacologic treatments for the condition that may indicate a target gene(s), or data from animal models (Hattersley & McCarthy, 2005). Bio-informatics databases, such as the Gene Ontology (http://www.geneontology.org/), may also aid in the identification of genes whose products may impact the phenotype of interest (The Gene Ontology, 1999–2011). Moreover, consideration should be given to number of genes on which to focus, ranging from a single gene to genes within a candidate biological pathway. Because more biologically global conclusions can be drawn, the study of a biologic pathway has the advantage of being more informative than the singular gene approach in most situations (Jorgenson, Ruczinski, Kessing, Smith, Shugart, & Alberg, 2009).

Polymorphism selection—Once selection of the candidate gene(s) is finalized, polymorphisms must be selected to evaluate candidate gene variability, and these are the genetic data used for analyses. The candidate gene association approach includes the

evaluation of single nucleotide polymorphisms (SNPs), repeat polymorphisms, insertion/deletion polymorphisms (INDEL), and copy number variants (CNV).

Resources for polymorphism selection: The SNP is the most common type of polymorphism and is a nucleotide (also known as a base) in the DNA where the nucleotide present (e.g., A, T, C, G) varies in the population (Genetics Home Reference, 2011). The scientific literature and a variety of online databases provide excellent resources for SNP identification and selection. A simple literature search combining the candidate gene(s) with the keyword "functional polymorphism" will help to identify SNPs known to alter the function of the candidate gene(s). Because functional polymorphisms modify the function of a gene regardless of phenotype, the literature search should not be limited to just the phenotype of interest. In addition to the literature, investigators also commonly use the Database of Single Nucleotide Polymorphisms (dbSNP) (http://www.ncbi.nlm.nih.gov/projects/SNP/) and the International HapMap Project (http://hapmap.ncbi.nlm.nih.gov/) to identify/select SNPs and tagging SNPs, respectively.

HapMap is accessed for the selection of tagging SNPs (tSNP), which represent the current gold standard for the evaluation of genetic variation in the candidate gene association study. The goal of HapMap is to develop a haplotype map of the human genome and to describe common patterns of genetic variation in humans (International HapMap Project, 2006). Essentially, HapMap is based on the premise that DNA is inherited in chunks/blocks (haploblock). Within these haploblocks, certain variants are inherited together. If the genotype of one variant within that block of DNA is known the genotype of a second variant within the same block can be determined since they are inherited together. Thus, HapMap assists the user in selecting SNPs that tag a certain haploblock of DNA (tagging SNPs or tSNPs). Ultimately, utilization of tSNPs allows one to fully evaluate the genetic variability of the candidate genes with the least number of SNPs (International HapMap Project, n.d.).

Repeat polymorphisms are characterized by repeating units of DNA bases. The number of times these DNA units repeat is variable in the population (Passarge, 2007). While repeat polymorphisms are less frequent in the genome than SNPs, they are often more informative as they usually have more alleles in the population than SNPs, which typically only have 2. The short tandem repeat (STR) is typically comprised of a repeating unit of two to four DNA bases (e.g., CAG CAG CAG) while the variable number tandem repeat (VNTR) is comprised of a larger repeating unit (Passarge), usually greater than 5 bases. For the evaluation of STRs and VNTRs, the literature continues to be the best source for identification and characterization.

An INDEL polymorphism occurs when a base(s) is added or subtracted from a place in the DNA. It is the presence or absence of the INDEL that is variable in the population (Nussbaum, McInnes, & Willard, 2007). Like SNPs, the dbSNP can be freely accessed to identify small-scale INDELs.

The CNV occurs when the number of copies of a particular genomic sequence/segment is variable in the population (NHGRI, n.d.). CNVs can be identified through scientific literature and online databases. The Database of Genomic Structural Variation (dbVar) (http://www.ncbi.nlm.nih.gov/dbvar) and The Copy Number Variation Project by the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/humgen/cnv/) are two online resources that may assist in CNV identification.

Genotype data collection technologies—Multiple options are available for SNP genotyping, including the polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) technique, real-time PCR allelic discrimination (e.g.

TaqMan®), multiplexing via mass spectrometry, and bead chip technology. Selection of the genotyping technique is guided by the number of samples and polymorphisms to be genotyped and available resources. PCR-RFLP, which is used to genotype SNPs based on differences in fragment lengths, is suitable when the number of SNPs and samples to be genotyped is relatively small. Real-time PCR allelic discrimination (http://www.appliedbiosystems.com; http://www.roche-applied-science.com), which genotypes SNPs based on allele-specific fluorescence intensity signals, is suitable for a medium number of SNPs and sample size. Because PCR-RFLP and real-time PCR allelic discrimination can only genotype one SNP at a time, the use of high throughput technologies have become the gold standard for SNP genotype collection when the number of SNPs to be evaluated approaches 24. The iPLEX® Gold-SNP Genotyping assay (http://www.sequenom.com), which genotypes SNPs based on differences in molecular mass, allows for the analysis of up to 36 SNPs per assay (Sequenom, 2010) in larger sample sizes. Not only can an investigator analyze multiple SNPs simultaneously, but time, assay to assay variability, and costs are reduced. The GoldenGate Genotyping Assay (http://www.illumina.com) is another high throughput bead based technology that can be utilized when the number of SNPs and samples to be analyzed is too large for other technologies.

There are several genotyping technologies also available for repeat polymorphisms, INDELs, and CNVs. PCR amplification followed by fragment sizing can be used for genotyping repeat polymorphisms. As with SNPs, real-time PCR allelic discrimination can be used to genotype small INDELs. Finally, TaqMan® Copy Number Assays (http://appliedbiosystems.com) or cytogenetic techniques (e.g., Fluorescence In Situ Hybridization) can be utilized for genotyping candidate CNVs.

Genome Wide Association Studies (GWAS)

Rationale for taking a GWAS approach—A GWAS gentoypes thousands to millions of polymorphisms across the genome for individuals who are phenotypically well-characterized (DiStefano & Taverna, 2011). If genetic variability is significantly different between cases and controls, those variations may be associated with susceptibility to or protection from the phenotype of interest and can provide direction as to which region of the genome these differences might be located. Ongoing efforts of the Human Genome Project and the International HapMap Project have made this approach possible through the generation of large databases that reference and map both sequence and variability.

The major advantage of a GWAS approach is that the biology of the phenotype of interest does not need to be completely understood prior to implementing this approach and the SNPs or genes of interest do not need to be defined *a priori*. Instead of selecting genes and polymorphisms *a priori*, polymorphisms that cover haploblocks across the entire genome are used for genotype data collection and non-parametric based analyses determine what genes/regions of the genome are relevant to the phenotype of interest (Hakonarson & Grant, 2011). The data derived from GWAS will provide direction regarding which areas of the genome warrant additional study.

There are several limitations to GWAS. The variant identified may not be what's accounting for the association, but is rather "tagging along" with the actual causal variant(s). This obstacle is also present for candidate gene association studies, particularly those that utilize a tSNP approach. Therefore, it may be necessary to follow up with more focused genotype data collection, including denser polymorphism evaluations and/or sequencing of that specific region of the genome to identify the exact allele accounting for the association (NHGRI, 2010). A major limitation for the GWAS approach, and perhaps a reason why many investigators are unable to pursue this approach, is the need for thousands of subjects

who are phenotypically well characterized and for which DNA is available. The need for large samples sizes for GWAS is due to the inherent issue of multiple testing that accompanies the evaluation of thousands to millions of different genetic variables. Additionally, the need for very large sample sizes, coupled with the cost of commercial genome-wide scanning techniques makes this approach very costly. GWAS approaches are also not optimal to assess rare polymorphisms as the data collection approaches for the GWAS are more focused on optimizing informativeness of the data (Ku, Loy, Pawitan, & Chia, 2010).

Subject and sample considerations—The cross-sectional case-control study design is the most frequently used approach for a GWAS. Study subjects should be selected based on a well-defined and heritable phenotype. Cases are defined as individuals who meet criteria for a phenotype of interest. Controls are individuals who have never met criteria for the phenotype and ideally have passed through the age or period of risk for the phenotype (Hakonarson & Grant, 2011). Like candidate gene associations studies, ancestry must be considered to avoid issues related to population substructure and this is why some investigators have conducted these types of studies with homogeneous populations (Psychiatric GWAS Consortium Coordinating Committee, 2009). Case and control groups should be matched on ancestry as much as possible to avoid false-positives. Despite this consideration, an advantage of GWAS is that whole genome data can provide adequate data to identify stratification and inflation of test statistics due to population substructure can be addressed (Hakonarson & Grant).

Obtaining a sufficiently large sample size is essential to ensure sufficient statistical power for a GWAS approach. Approximately 1,000 cases and a similar number of controls are required to detect 1–5 variants associated with a given trait. A larger sample is needed to uncover additional variants that may have diminishing contributions to the disease (Hakonarson & Grant, 2011).

Informed consent issues: While informed consent is of paramount importance with any research study, researchers who are considering a GWAS should be cognizant of issues related to conducting such as study and the National Institutes of Health (NIH) policy on data sharing for GWAS. In January 2008, the NIH adjusted its policy mandating the sharing of GWAS data obtained in NIH-funded or conducted studies. The details of this policy can be found at http://gwas.nih.gov/. Most NIH-funded GWAS are required to include language in the consent document that addresses public sharing of de-identified genotype and phenotype data. Researchers who are planning to study existing samples must ensure that the original consent signed by the subjects is consistent with conducting a GWAS.

Genotype data collection technologies—There are currently two commonly used vendors that provide technology for collection of GWAS data, Affymetrix and Illumina. The companies use different technological approaches, which are both widely used in the research community. The Affymetrix® Genome Wide SNP Array 6.0 features 1.8 million genetic markers, including 906,600 SNPs and more than 946,000 probes for the detection of CNVs. This platform also includes a high resolution reference map and a copy number polymorphism (CNP) algorithm (see http://www.affymetrix.com for additional information). The Illumina Omni Microarrays provide a multiple bead chip option which will soon include nearly 5 million markers per sample, including both common and rare variants identified by the 1000 Genomes project. Omni microarrays assess structural variation, including CNVs and copy neutral variants (inversions and translocations) which may also be significant contributors to disease (see http://www.illumina.com for additional information).

Resources of interest for GWAS: The Center for Inherited Disease Research (CIDR) at Johns Hopkins University (http://www.cidr.jhmi.edu/requirements/applications.html) is funded by NIH Institutes and provides genotyping and statistical genetic services to investigators who have received access after a competitive peer review process. Interested investigators are required to submit an application for projects supported by the NIH. In order to maximize access to resources, the application process to CIDR should ideally take place before or at the time of grant application, though this is not a requirement.

The repository for GWAS data is currently the Database of Genotypes and Phenotypes (dbGaP; http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gap). This database was developed to archive the results of studies that have investigated the genotype-phenotype interaction and serves as a useful resource in reviewing the work that has already been completed and aids in planning future research. The dbGaP database provides the opportunity for *in silico* research. Researchers have the option of two levels of access (open and closed) to dbGaP: Open-access data are aggregate data that are publicly available while closed level access requires an application and approval process that includes de-identified subject specific data. The genotype data and their linked phenotype data are invaluable resources and researchers are encouraged to investigate this database as it pertains to their phenotypes of interest prior to designing a study.

GENE EXPRESSION STUDIES

Gene expression studies evaluate the activity of a gene using the level of messenger RNA (mRNA) from a gene(s) and determine if that level is associated with the phenotype of interest. DNA contains a code to generate mRNA through a process called transcription. The amount of mRNA produced from a gene, if at all, depends on many factors including tissue type, local cell environment, and point in the cell cycle.

A gene expression study is different from a polymorphism based study because an expression study evaluates mRNA levels that can change over time, uses less stable mRNA instead of DNA, and mRNA levels can be dramatically different based on what tissue is used for analysis, since gene expression is tissue-specific. Gene expression studies therefore should address whether multiple samples over time are needed for evaluation (similar to other types of biomarkers that change over time), RNA stabilization, and what cell/tissue type is most appropriate to evaluate for the phenotype of interest. For these reasons, many stored samples may not be appropriate for this approach.

Candidate Gene Expression Studies

Rationale for taking a candidate gene expression approach—Candidate gene expression studies investigate mRNA levels for a specific gene(s) to determine if it is associated with a phenotype of interest. Similar to a candidate gene association approach, this is a hypothesis-driven approach where the investigator *a priori* selects the candidate gene(s) to be evaluated. This approach is only appropriate if the investigator has ample justification for investigating a specific gene(s).

Subject and sample considerations—Gene expression studies often involve relative comparisons of mRNA levels between two groups (these groups can be different types of tissues, groups that vary for a particular exposure or groups that vary by the presence or absence of a phenotype of interest). Clearly defined inclusion/exclusion criteria are necessary, due to the relative comparison nature of this approach.

RNA stabilization: Stabilization of RNA is essential to obtain accurate gene expression profiles of biological samples. Immediately after sample collection, RNA degradation and

other transcriptional changes begin to occur. These alterations may result in false up or down regulation of gene expression levels. RNA stabilization preserves a representative gene expression profile for later analysis (e.g. quantitative RT-PCR and microarray analysis). RNA stabilization methods vary based on the type of biological sample. Five of the most common RNA stabilization methods are: (a) PAXgene Blood RNA System (http://www.preanalytix.com), which utilizes a single tube (pre-filled with RNA stabilization reagent) for blood collection, RNA stabilization, sample transport and storage, and purification of total RNA (PreAnalytiX, 2010); (b) LeukoLOCK System (http:www.lifetechnologies.com), which filters and isolates leukocytes from whole blood. RNAlater solution is then used to stabilize the RNA of the leukocytes. A notable advantage to the LeukoLOCK System is the ability to remove a large proportion of reticulocyte-derived globin mRNA. Depletion of the globin mRNA allows for the detection of thousands of additional genes on microarray (Life Technologies Corporation LeukoLOCK, 2010); (c) RNAlater (http://www.ambion.com; http:www.qiagen.com) stabilizes RNA in a variety of fresh samples including animal tissue, tissue culture cells, leukocytes, yeast, and bacteria. After collection, the sample is submerged in the RNAlater stabilization solution. This solution permeates and stabilizes the sample eliminating the need for immediate processing or freezing of samples (Life Technologies Corporation RNAlater, 2010; Qiagen, 2006); (d) Oragene RNA for Expression Analysis Self Collection Kit (http://www.dnagenotek.com): allows for the non-invasive collection of RNA from saliva. Donors are instructed to expectorate into a vial, cap the container, and shake vigorously to release a stabilization solution from the cap. Oragene RNA samples can remain stable for months at room temperature (DNA Genotek, 2011); and (e) Snap Freeze quick freezes solid tissues with liquid nitrogen and dry ice can be used to preserve RNA; however, disruptions during freezing and thawing can lead to RNA degradation. Due to potential RNA degradation and difficulty of obtaining and working with liquid nitrogen and dry ice, RNAlater described above may be a more viable option for solid tissue RNA stabilization.

Candidate gene selection—Candidate gene selection must be justified and rationale for selection is similar to selection of candidate genes in the candidate gene association section. The same bio-informatics databases mentioned in that section are also applicable to aiding in the selection of candidate genes for an expression study and as with a candidate gene association study, a candidate gene expression study should consider focusing on a group of genes in a biological pathway versus the value of focusing on a single gene.

Expression data collection technologies—Selection of a data collection technology for a candidate gene expression study should take into account the number of genes/loci and the number of samples to be evaluated. The most frequently used technologies for a candidate gene approach include Northern blotting, quantitative real-time PCR (qRT-PCR), and multiplex platforms that support 3–36 genes/loci per reaction.

Northern blotting requires electrophoresis of RNA, transfer to a membrane and hybridizing the membrane with a probe specific for detection of the mRNA of interest. The advantages of blotting are that most laboratories will have the equipment to conduct this type of data collection and assessment of RNA size is possible. Disadvantages of blotting are that RNA degradation is common, it requires more RNA as a template for the experiment compared to other methods, it is laborious, and is not optimal for quantification of mRNA levels.

Currently, one of the most popular techniques for assessing the level of mRNA for a gene/locus is qRT-PCR. qRT-PCR requires conversion of RNA to a more stable template called cDNA (complementary DNA), PCR amplification and probe hybridization for the gene/locus of interest. The probe is fluorescently labeled and liberation of this fluorescent label is quantified, reflecting the amount of starting mRNA template in the sample. One crucial step

in conducting qRT-PCR is normalization of the data generated. Normalization of the data allows for sample to sample comparisons that have been corrected for noise such as what's introduced when sample dispensing between samples isn't uniform. This is often done using qRT-PCR data collected simultaneously for an endogenous control, which usually represents a stably expressed gene (often referred to as a "housekeeping gene") and allows for normalization of data across samples (Guenin et al., 2009). Thought needs to be given to selection of an appropriate endogenous control given that different tissues will have different stably expressed genes (Guenin et al.). If in doubt, endogenous control panels are available for assessment prior to conducting qRT-PCR. Advantages of qRT-PCR include high sensitivity and reduced RNA template requirements, high throughput capabilities, quantification of starting mRNA template is possible with use of proper exogenous reference controls, and for many genes/loci/pathways off the shelf optimized assays are available (http://www.appliedbiosystems.com; http;//www.roche-applied-science.com).

Multiplex gene expression assays are available when the number of genes/loci to be evaluated is in the range of ~3–36. One example is the QuantiGene® Plex 2.0 assay (for more information see http://www.panomics.com/index.php?id=product_6) that uses Luminex technology to collect the data and the assay can be customized.

Global (Genome Wide) Gene Expression Studies

Rationale for taking a global gene expression approach—Whole genome expression (also known as global gene expression or gene expression profiling) offers a comprehensive view of gene activity within a biological sample by examining mRNA levels for all known genes across the genome. In this way, whole genome expression provides functional information regarding "when and where a protein is expressed, when it is degraded, and with which other proteins it may interact" (Altman & Raychaudhuri, 2001, p. 340). Due to the dynamic nature of expression, gene expression profiles are often relatively compared under multiple conditions (such as comparing different tissue types, comparing normal versus abnormal tissues, comparing tissues before and after an exposure) or over a period of time (Altman & Raychaudhuri, 2001; Arcellana-Panlilio & Robbins, 2002). The use of global gene expression profiling is extremely advantageous when little to nothing is known about the genes influencing a condition, a similar advantage held by the GWAS approach. Thus, whole genome expression can identify novel candidate hypotheses through a non-parametric analysis of genome wide expression data.

Subject and sample considerations

Sample selection: Although this is an approach similar to GWAS, with evaluation of thousands of genes in a nonparametric manner, sample size requirements for global gene expression are usually smaller, requiring approximately 10 subjects per variable. Matching of subjects for key variables known to influence the phenotype under investigation can reduce the number of variables that need to be accounted for in the analyses. A sample size calculator for global gene expression experiments can be found at http://bioinformatics.mdanderson.org/MicroarraySampleSize/. Additionally, as with candidate gene expression studies, mRNA stabilization of the collected samples is crucial.

Gene expression data collection technologies

Microarrays: Microarrays are used to examine the expression profile of a single sample (often referred to as single dye array) or to compare expression levels between two different samples/conditions (often referred to as two dye array). The microarray itself is a solid surface covered with an "ordered arrangement of unique nucleic acid fragments derived from individual genes" (Arcellana-Panilio & Robbins, 2002, p. G397). Fluorescently labeled template hybridizes to these nucleic acid fragments (referred to as probes) on the solid

surface through complementary pairing. The intensity of the florescence at each spot on the microarray corresponds to the amount of sample binding to a particular nucleic acid fragment and thus, the gene expression level. If the microarray reveals any interesting findings, q-RT-PCR should be carried out for validation purposes. For a visual representation of microarray methodology visit this web address: http://www.bio.davidson.edu/courses/genomics/chip/chip.html.

Microarrays have revolutionized gene expression analyses, as this technology is able to simultaneously survey thousands of genes in a short period of time. However, the ability to detect novel genes is limited to the hybridization probes represented on the microarray. Off-the-shelf probe sets that contain reference sequences can be used, or custom probe sets are designed based on specific genes of interest or pathways. Additionally, microarrays require specialized lab equipment and are very useful when analyzing a small sample size but become costly as sample size increases. Two popular microarrays platforms include Affymetrix's GeneChip and Illumina's BeadChip.

Affymetrix's GeneChip platform (for more information see http://www.affymetrix.com) utilizes traditional solid support microarray technology. Affymetrix's latest product, the GeneChip Human Gene 1.0 ST Array, is able to interrogate 28,869 genes and covers over 700,000 distinct probes. A greater number of samples can be processed simultaneously (with this same probe set) using the Human Gene 1.1 Array Strip (4 samples/strip) and the Human Gene 1.1 Array Plate (16, 24, or 96 samples/plate). Affymetrix also provides whole transcript expression analysis technology for mice and rats.

Instead of using a solid support platform, the Illumina BeadChip platform (for more information see http://www.illumina.com) employs silica beads (each covered with thousands of copies of a specific oligonucleotide) self-assembled in microwells of fiber optic bundles or planar silica slides. Illumina's most recent whole genome expression array, the HumanHT-12 v4 BeadChip, provides high throughput processing of twelve samples and covers over 47,000 probes. Illumina also offers whole genome expression BeadChip technology for mice and rats.

Normalization of gene expression data is also important with microarray data collection. Unlike qRT-PCR where an appropriate endogenous control needs to be selected and included in the data collection, microarrays already include a range of endogenous controls for which data is simultaneously collected and from which the investigator can select to use for normalization of the data.

Sequence based technologies that utilize next-generation sequencing (NGS; high throughput sequencing) are also available for collection of genome-wide gene expression data. An example of such a technology is the RNA-Seq method (for more information see http://www.illumina.com). This method requires conversion to cDNA, ligation of the cDNA fragments, creation of a library, sequencing of the template, and collection of frequency data for a transcript. An advantage of this approach over microarrays is that it does not require primers or probes therefore novel transcripts that would not be detectable with a microarray can be identified.

Serial analysis of gene expression: Serial analysis of gene expression (SAGE) provides comprehensive quantitative gene expression data. SAGE technology is based on three main principles: (1) a short sequence tag (9–17 bases) contains sufficient information to distinctively identify a transcript, (2) sequence tags can be linked together to form one long molecule that can be cloned and sequenced to allow efficient analysis of transcripts, and (3) the number of times a particular tag is observed corresponds to the expression level of the

transcript (Sagenet, 2005; Velculescu, Zhang, Vogelstein, & Kinzler, 1995). One of the main advantages of SAGE, similar to RNA-Seq, is the ability to detect novel genes as it does not require prior sequence information or hybridization probes for each transcript like microarrays (Velculescu et. al, 1997). Another advantage of SAGE is that is utilizes common laboratory equipment and techniques. Any laboratory that performs PCR and manual sequencing could also execute SAGE. Nonetheless, due to cloning and sequencing, SAGE can be expensive, time consuming, and labor intensive.

EPIGENETIC STUDIES

An epigenetic mechanism is a biochemical alteration to the DNA molecule that does not change the sequence of the DNA but does influence gene expression. Epigenetics is often defined as the "study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Russo, Martienssen, & Riggs, 1996, p. 1).

The epigenetic/epigenomic approach shares many advantages and disadvantages with DNA polymorphism based approaches and gene expression based approaches. Like DNA polymorphism based approaches, the epigenetic/epigenomic approach uses DNA as its template for data collection. Since both DNA sequence and its chemical modifications are stable, stored samples are more likely to be appropriate for this approach than gene expression approaches. Similar to a gene expression based approach; epigenetic/epigenomic alterations can change over time and can differ dramatically between cell/tissue types. Although template stability is not an issue, the investigator should give great consideration to whether multiple samples over time are needed for evaluation and what cell/tissue type is most appropriate to evaluate for the given phenotype of interest. For these reasons, similar to gene expression studies, many stored samples may not be appropriate for this approach.

Chromatin remodeling, non-coding RNAs, histone modifications, and DNA methylation are all epigenetic/epigenomic alterations that impact gene expression. Chromatin remodeling is an enzymatic process that results in altered chromatin and nucleosome composition. This transformed structure provides regulatory proteins access to the DNA molecule. Non-coding RNAs are not translated into protein but have considerable involvement in gene expression through interactions with DNA/mRNA. While chromatin remodeling and non-coding RNAs are important to gene regulation, this paper will focus primarily on the commonly examined epigenetic mechanisms for which the most technology for data collection is available: histone modifications and methylation.

Rationale for Taking an Epigenetic/Epigenomic Approach

The decision to take an epigenetic (candidate gene) or an epigenomic (genome wide) approach is based upon wanting to evaluate the mechanism for gene regulation. There are many environmental factors that impact the severity and frequency of epigenetic/epigenomic alterations and subsequent gene expression; therefore, this approach is often used to examine multifactorial diseases that have an environmental component associated with it. Epigenetic approaches to examine transcriptional regulation have contributed to a more comprehensive understanding of complex conditions that demonstrate aberrant gene expression, including: cancer (Wilop et al., 2011), mental health (Read, Bentall, & Fosse, 2009), and cardiovascular disorders (Ordovás & Smith, 2010). Furthermore, the investigation of diseases for which DNA mutations have not been revealed may benefit from an epigenetic approach.

Subject and Sample Considerations

The epigenome is subject to frequent alterations; therefore, longitudinal sample collection is recommended if evaluating time-sensitive trends. Subject size recommendations for an epigenetic study follow similar guidelines to a gene expression study, and vary on whether the investigator will examine the entire genome (hypothesis generating/larger sample size) or a candidate gene profile (hypothesis driven/smaller sample size). Like the other approaches described, an epigenetic study does not require that subjects be related. The advantages and disadvantages of conducting a genome-wide versus candidate gene epigenetic study are similar to those described in previous sections.

The epigenome is largely determined by cell type, and this is especially true for methylation patterns; therefore, tissue source is extremely important to consider for this type of approach. For example, the methylation profile of a skin cell is very different than the methylation profile of a liver cell, since different genes are expressed in each cell type, and methylation is a driving force behind tissue specific gene expression. Similar to a gene expression study, an epigenetic design requires the samples for epigenetic analyses be from a tissue that appropriately addresses the phenotype of interest. Tissue specific sample collection will capture epigenetic patterns that impact gene expression which are potentially contributing to the disease. Unlike a gene expression study which examines RNA, this design requires DNA, which is advantageous for the investigator who has access to previously collected samples, assuming they were collected from an appropriate tissue for the phenotype under investigation.

Epigenetic and Epigenomic Data Collection Technologies

This section will focus on the two epigenetic mechanisms most frequently studied: (a) histone modification and (b) methylation. Post-translational histone modifications include alteration of the histone tail through biochemical changes that ultimately impact gene activity. Genome-wide histone modifications can be captured with chromatin immunoprecipitation technology (ChIP), and quantified with a microarray (ChIP-chip). Methylation refers to the addition of a methyl group to a cytosine, often at CpG islands, which are regions of the genome that are rich in CG base sequences. Hypermethylation of a gene typically leads to gene suppression, while hypomethylation results in gene expression. Genome-wide methylation intensities can also be measured with affinity-based immunoprecipitation (MeDIP), and quantified with a microarray (MeDIP-chip, Infinium platform). Methylation of candidate genes can also be measured with restriction enzymes that recognize only demethylated CpG regions (HELP assay), or pyrosequencing. Next Generation Sequencing approaches are also becoming increasingly popular, more cost effective, and provide global sequencing for histone modification (ChIP-seq) and methylation (MeDIP-seq), often integrating these with other epigenetic mechanisms. This section will describe each method and provide the reader with technologies and recommendations to aide in the design and implementation of an epigenetic study.

Histone modification analysis

Histone modification signals can be captured with chromatin immunoprecipitation (ChIP), which provides modification position approximation on the genome (Collas, 2010). The ChIP-chip technique combines this ChIP technology with a microarray (chip) to quantify the sum of binding sites on the genome (Aparacio, Geisberg, & Struhl, 2004). The ChIP-seq technique (see Next Generation Sequencing) has become a popular technique compared to ChIP-chip. Unlike ChIP-seq, ChIP-chip requires more amplification, multiplexing is not possible (Park, 2009), and the results have a lower resolution that are limited to the coverage provided by the selected microarray (Evertts, Zee, & Garcia, 2010). Nimble Gen offers a whole-genome ChIP-chip tiling array that allows the investigator to choose between

ordering the entire genome set or individual arrays within a set (http://www.nimblegen.com/products/chip/wgt/index.html). Single gene ChIP technologies are available that target antibodies against specific histone modifications. Mass spectrometry also allows the measurement of mass-to-charge ratio of peptides (Evertts et al.) and allows for changes in modification to be quantified during chromatin assembly (Deal & Henikoff, 2010).

When performing any microarray experiment, it is important to address concerns that may compromise the integrity of the experiment, including: image acquisition, background subtraction, standard normalization and the need to control for biases from dye (Buck & Lieb, 2004). Additionally, the reproducibility of the histone-modification results depends on the quality and specificity of antibodies used. Antibodies may exhibit appropriate specificity, but are ineffective when subjected to ChIP reagents (Egelhoffer et al., 2010). The Center for Biomedical Informatics at Harvard Medical School has developed an online repository that allows investigators to search for antibodies subjected to validation tests (http://compbio.med.harvard.edu/antibodies/about). It is important to note that this validation data should be used as a guide and investigators are encouraged to validate their own findings.

Bisulfite-conversion based methylation analyses

Bisulfite-conversion of unmethylated cytosines to uracils remains the gold-standard to evaluate methylation (Huang, Huang, & Feng, 2010). Bisulfite-conversion based microarrays use probes that hybridize targets to methylated and unmethylated regions, and release a fluorescent intensity that denotes methylation status (Huang et al.). Recent research indicates that tissue-specific methylation occurs in CpG island shores rather than previously targeted CpG islands (Irizarry et al., 2009); therefore, CpG islands alone are not sufficient to reveal differentially methylated regions and methylome evaluation should also include CpG shores (Gupta, Nagarajan, & Wajapeyee, 2010). Like other non-sequencing-based methods, the results of this platform are "susceptible to certain polymorphisms that were not known or considered at the time the array was designed" (Rakyan, Down, & Balding, 2011, p. 532). Illumina offers the Infinium HumanMethylation450K which provides a whole-genome analysis of methylation intensities of more than 450,000 sites, including CpG islands, shores and other CpG sites outside of islands (for more information see http://www.illumina.com/products/methylation_450_beadchip_kits.ilmn). Candidate gene methylation assessment can be accomplished through technologies such as the EpiTYPER (for more information see http://www.sequenom.com) that uses bisulfite converted DNA as a template for PCR and after modification and cleavage of the PCR product, mass spectrometry is performed to quantify methylated and non-methylated DNA.

Bisulphite-based sequencing (BS-seq) uses bisulphite converted DNA as a template, PCR amplification occurs, and sequencing of the resulting fragments provide a global view of methylation with minimal bias toward CpG dense regions. This approach provides the highest level of coverage and resolution, but is not capable of distinguishing between methylated and hydroxymethylated cytosine bases. BS-seq can be used for both a genome-wide or candidate gene approach. Pyrosequencing examines the methylation intensity of specific sites or genes of interest. Illumina offers a single-site resolution methylation assay that uses bisulfite conversion and pyrosequencing to produce high resolution results (http://www.illumina.com/technology/veracode_methylation_assay.ilmn).

Affinity-based methylation analyses

Genome-wide affinity-based microarrays use enzyme recognition sites within CpG sites that enrich the methylated fraction of the genome. The MeDIP-chip technique (Methylated DNA

Immunprecipitation-chromatin Immunoprecipitation) immunoprecipitates the methylated portion of genomic DNA with an antibody, and is followed by quantification of methylation with a microarray. This technique yields a restricted resolution that is limited by the type of array used. MeDIP-chip should be validated with quantitative PCR, though referencing is not required since bisulfite conversion does not occur. ArrayStar offers MeDIP-chip services that include quality assessments for both methods (http://www.arraystar.com/Microarray/service_main.asp?id=181).

Restriction endonuclease-based methylation analysis

Restriction endonucleases have been adapted to discriminate methylated from unmethylated regions in the DNA (Edwards et al., 2010). This approach uses restriction enzymes that recognize only unmethylated sites, and are therefore unable to cut methylated portions of DNA. This method, combined with high throughput sequencing is limited by the availability of restriction enzyme sites in the target DNA (Gupta et al., 2010). Additionally, this technique requires large amounts of DNA (Biotage, 2007). Advantages for this approach include: a simplified data analysis, straightforward protocol, and it does not require bisulfite-conversion. The use of restriction enzymes to analyze methylation can be used for either candidate-gene or genome-wide studies (Gupta et al.) and has been used as a method of methylation mapping analysis (Edwards et al., 2010).

Data Quality assessments are important to incorporate into an epigenetic study. Quantile and LOESS normalization is recommended, which assumes a similar total strength (source). Additionally, bisulfite-based experiments, especially pyrosequencing since PCR is highly variable, should include verification in independent samples to distinguish methylation from incomplete bisulfite conversion (Laird, 2010). Incomplete conversion of methylated cytosines remains a major weakness of bisulfite-conversion based analysis techniques. Fully methylated and fully unmethylated controls should be provided by commercial vendors which allow the investigator to evaluate bisulphite-conversion efficiency.

Next generation sequencing (NGS) for histone modification analysis

DNA sequencing from epigenetic events may provide a first step toward quantification of epigenetic mechanisms. Similar to ChIP-chip, ChIP-seq uses antibodies to enrich for histone modifications, but is instead followed by high-throughput sequencing that measures gene expression levels (Evertts et al., 2010). This technique determines the genome-wide patterns of modified chromatin, including: histone methylation, acetylation status and binding regions for proteins (Werner, 2010). Unlike ChIP-chip, ChIP-Seq offers higher resolution with fewer artifacts, greater coverage, and requires less DNA. Illumina offers a ChIP-seq assay that provides a wide range of binding sites with varying strength (http://www.illumina.com/technology/chip_seq_assay.ilmn).

Next generation sequencing (NGS) for methylation analysis

MeDIP-seq (Methylated DNA Immunoprecipitation-Sequencing) is a high throughput sequencing technique of methylated DNA fragments that is aligned to a referenced genome. This technique is comparatively easier to analyze and interpret (Gupta et al., 2010); however, this method is best used to study hypermethylation of CpG-rich areas, since methylated CpG-rich sequences are more efficiently enriched than methylated CpG-poor sequences (Bibkova & Fan, 2009).

CONCLUSIONS

Nurse scientists should give much thought to how a genetic or genomic study could positively impact and move forward their program of research. When designing a genetic or

genomic research study it is paramount that one decides if they will take a polymorphism based, gene expression based or epigenetic based approach and then within the context of that study whether they will take a genetic or a genomic approach. This paper, while not providing an exhaustive review of available technologies, demonstrates the variety of technologies available for commonly used approaches, each with advantages and disadvantages. Availability of databases housing information to facilitate study design, data collection, interpretation of findings, and dissemination of data have greatly improved over the past decade. Investigators are encouraged to visit and utilize *in silico* resources when designing a research study to ensure they are conducting novel investigations and using up to date information.

Acknowledgments

The authors would like to acknowledge support available through a National Institutes of Health, National Institute of Nursing Research award "Targeted Research and Academic Training Program for Nurses in Genomics" (T32 NR009759).

References

Altman RB, Raychaudhuri S. Whole-genome expression analysis: challenges beyond clustering. Current Opinion in Structural Biology. 2001; 11(3):340–347.10.1016/S0959-440X(00)00212-8 [PubMed: 11406385]

Aparicio, O.; Geisberg, J.; Struhl, K. Current Protocols in Cell Biology. Vol. Chapter 17. Los Angeles, CA, USA: John Wiley & Sons, Inc; 2004. Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo.

Arcellana-Panlilio M, Robbins SM. Global gene expression profiling using DNA microarrays. American Journal of Physiology – Gastrointestinal and Liver Physiology. 2002; 282(3):G397–G402.10.1152/ajpgi.00519.2001 [PubMed: 11841989]

Bibkova M, Fan J. Genome-wide DNA methylation profiling. WIRE Systems Biology and Medicine. 2009; 2(2):210–223.10.1002/wsbm.35

Biotage. CpG methylation analysis by pyrosequencing: benchmarks and application. 2007 Mar. Retrieved July 11, 2011, from http://www.pyrosequencing.com/graphics/7424.pdf

Buck MJ, Lieb JD. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics. 2004; 83(3):349–360.10.1016/j.ygeno.2003.11.004 [PubMed: 14986705]

Campbell, AM. Molecular movies: DNA microarray methodology. 2001. Retrieved July 13 2011 , from http://www.bio.davidson.edu/courses/genomics/chip/chip.html

Collas P. The current state of chromatin immunoprecipitation. Molecular Biotechnology. 2010; 45(1):87–100.10.1007/s12033-009-9239-8 [PubMed: 20077036]

Corvin A, Craddock N, Sullivan PF. Genome-wide association studies: A primer. Psychological Medicine. 2010; 40(7):1063–1077.10.1017/S0033291709991723 [PubMed: 19895722]

Deal R, Henikoff S. Capturing the dynamic epigenome. Genome Biology. 2010; 11(10):218.10.1186/gb-2010-11-10-218 [PubMed: 20959022]

DiStefana, JK.; Taverna, DM. Technical issues and experimental design of gene association studies. In: DiStefano, Joanna K., editor. Disease Gene Identification: Methods and Protocols, Methods in Molecular Biology. 2011. p. 3-16.

DNA Genotek. Oragene RNA. 2011. Retrieved July 13, 2001, from http://www.dnagenotek.com/DNA_Genotek_Product_RNA_Overview.html

Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, Milekic MH, Bestor TH. Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Research. 2010; 20(7):972–980.10.1101/gr.101535.109 [PubMed: 20488932]

Egelhoffer TA, Minoda A, Klugman S, Lee K, Kolasinska-Zwierz P, Alekseyenko AA, Lieb JD. An assessment of histone-modification antibody quality. Nature Structural & Molecular Biology. 2010; 18(1):91–93.10.1038/nsmb.1972

Everts A, Zee B, Garcia B. Modern approaches for investigating epigenetic signaling pathways. Journal of Applied Physiology. 2010; 109(3):927–933.10.1152/japplphysiol.00007.2010 [PubMed: 20110548]

Genetics Home Reference. What are single nucleotide polymorphisms (SNPs)?. 2011. Retrieved July 19, 2011, from http://ghr.nlm.nih.gov/handbook/genomicresearch/snp

Grant S, Hakonarson H. Microarray technology and applications in the arena of genome-wide association. Clinical Chemistry. 2008; 54(7):1116–1124.10.1373/clinchem.2008.105395 [PubMed: 18499899]

Green ED, Guyer MS. National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. Nature. 2011; 470(7333):204–213.10.1038/nature09764 [PubMed: 21307933]

Guenin S, Mauriat M, Pelloux J, Van Wuytswinkel O, Bellini C, Gutierrez L. Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. Journal of Experimental Botany. 2009; 60(2):487–493.10.1093/jxb/ern305 [PubMed: 19264760]

Gupta R, Nagarajan A, Wajapeyee N. Advances in genome-wide DNA methylation analysis. Biotechniques. 2010; 49(4):iii–xi.10.2144/000113493 [PubMed: 20964631]

Hakonarson H, Grant S. Planning a genome-wide association study: Points to consider. Annals of Medicine. 2011; 43(6):451–460.10.3109/07853890.2011.573803 [PubMed: 21595511]

Hattersly AT, McCarthy MI. Genetic Epidemiology 5: What makes a good genetic association study? The Lancet. 2005; 366(9493):1315–1323.10.1016/S0140-6736(05)67531-9

Huang Y, Huang T, Wang L. Profiling DNA methylomes from microarray to genome-scale sequencing. Technology in Cancer Research and Treatment. 2010; 9(2):139–147. Retrieved from PubMed database. [PubMed: 20218736]

International HapMap Project. About the International HapMap Project. 2006. Retrieved July 18, 2011, from http://hapmap.ncbi.nlm.nih.gov/abouthapmap.html

International HapMap Project. What is the HapMap?. n.d. Retrieved July 18, 2011, from http://hapmap.ncbi.nlm.nih.gov/whatishapmap.html.en

Irizarry RA, Ladd-Acosta B, Wen Z, Wu C, Montano P, Onyango H, Feinberg AP. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genetics. 2009; 41(2):178–186.10.1038/ng.298 [PubMed: 19151715]

Jorgensen TJ, Ruczinski I, Kessing B, Smith MW, Shugart YY, Alberg AJ. Hypothesis-drive candidate gene association studies: Practical design and analytical considerations. American Journal of Epidemiology. 2009; 170(8):986–993.10.1093/aje/kwp242 [PubMed: 19762372]

Ku CK, Loy EY, Pawitan Y, Chia KS. The pursuit of genome-wide association studies: Where are we now? Journal of Human Genetics. 2010; 55(4):195–206.10.1038/jhg.2010.19 [PubMed: 20300123]

Laird P. Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics. 2010; 11:191–203.10.1038/nrg2732

Lander ES. Initial impact of the sequencing of the human genome. Nature. 2011; 470(7333):187–197.10.1038/nature09792 [PubMed: 21307931]

Life Technologies Corporation. LeukoLOCK total RNA isolation system. 2010. Retrieved July 13, 2011, from http://www.ambion.com/techlib/prot/fm_1923.pdf

Life Technologies Corporation. RNAlater tissue collection: RNA stabilization solution. 2010. Retrieved July 13, 2011, from http://www.ambion.com/techlib/prot/bp_7020.pdf

National Center for Biotechnology Information. Microarrays: Chipping away at the mysteries of science and medicine. 2007. Retrieved June 6, 2011, from http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html

National Human Genome Research Institute [NHGRI]. Talking Glossary of Genetic Terms: Copy Number Variation (CNV). n.d. Retrieved July 19, 2011, from http://www.genome.gov/glossary/index.cfm?id=40

National Human Genome Research Institute [NHGRI]. Talking Glossary of Genetic Terms: Polymorphism. n.d. Retrieved July 26, 2011, from http://www.genome.gov/glossary/index.cfm?id=160

Nussbaum, RL.; McInnes, RR.; Huntington, FW. Thompson & Thompson genetics in medicine. 7. Philadelphia, PA: Saunders Elsevier; 2007.

Ordovás JM, Smith C. Epigenetics and cardiovascular disease. Nature Reviews Cardiology. 2010; 7(9):510–519.10.1038/nrcardio.2010.104

Park P. ChIP-seq: advantages and challenges of a maturing technology. Nature Reviews. 2009; 10(10): 669–680.10.1038/nrg2641

Passarge, E. Color Atlas of Genetics. New York: Thieme; 2007.

PreAnalytiX. PAXgene blood RNA: The better the source, the more to explore. 2010. Retrieved July 13, 2011, from http://www.qiagen.com/literature/render.aspx?id=200337

Psychiatric GWAS Consortium Coordinating Committee. Genomewide Association Studies: History, rationale, and prospects for psychiatric disorders. American Journal of Psychiatry. 2009; 166(5): 540–556.10.1176/appi.ajp.2008.08091354 [PubMed: 19339359]

Qiagen. RNAlater handbook. 2006. Retrieved July 13, 2011, from www.qiagen.com/literature/render.aspx?id=403

Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nature Reviews Genetics. 2011; 12(8):529–541.10.1038/nrg3000

Read J, Bentall R, Fosse R. Time to abandon the bio-bio-bio model of psychosis: Exploring the epigenetic and psychological mechanisms by which adverse life events lead to psychotic symptoms. Epidemiologia e Psichiatria Sociale. 2009; 18(4):299–310.10.1017/S1121189X00000257 [PubMed: 20170043]

Ruiz-Naváez EA, Rosenberg L, Wise LA, Reich D, Palmer JR. Validation of a small set of ancestral informative markers for control of population admixture in African Americans. American Journal of Epidemiology. 2011; 173(5):587–592.10.1093/aje/kwq401 [PubMed: 21262910]

Russo, VA.; Martienssen, RA.; Riggs, AD. Epigenetic Mechanisms of Gene Regulation. Vol. 32. Plainview, NY: Cold Spring Harbor Laboratory Press; 1996.

Sagenet. Description of SAGE. 2003–2005. Retrieved June 6, 2011, from http://www.sagenet.org/findings/index.html

Sequenom. MassARRAY® iPLEX® gold-SNP genotyping: From target discovery to HTP validating (version 2) [Brochure]. 2010.

The Gene Ontology. An Introduction to the Gene Ontology. 1999–2011. Retrieved July 19, 2011, from http://www.geneontology.org/GO.doc.shtml

The University of Texas MD Anderson Cancer Center: Department of Bioinformatics and Computational Biology. Sample size for microarray experiments. 2003–2010. Retrieved July 13, 2011, from http://bioinformatics.mdanderson.org/MicroarraySampleSize/

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science. 1995; 270(5235):484–487.10.1017/S1121189X00000257 [PubMed: 7570003]

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Kinzler KW. Characterization of the yeast transcriptome. Cell. 1997; 88:243–251.10.1016/S0092-8674(00)81845-0 [PubMed: 9008165]

Werner T. Next generation sequencing in functional genomics. Briefings in Bioinformatics. 2010; 2(5):499–511.10.1093/bib/bbq018 [PubMed: 20501549]

Wilop S, Fernandez AF, Jost E, Herman JG, Brummendorf TH, Esteller M, Galm O. Array-based DNA methylation profiling in acute myeloid leukaemia. British Journal of Haematology. 2011; 155(1):65–72.10.1111/j.1365-2141.2011.08801.x [PubMed: 21790528]

Table 1

Online Genome Databases and Resources

Name and Address	Description
Database of Short Genetic Variations (aka SNP database) http://www.ncbi.nlm.nih.gov/snp/?term=	This database houses documented SNPs, microsatellites, and small-scale INDELs. It provides population specific allele frequencies; genotype data, genome location, and information on function (e.g., change in an amino acid).
International HapMap Project http://hapmap.ncbi.nlm.nih.gov/	This database is used to identify and select tagging SNPs. User defined criteria under the configure tab include population selection, R^2 cutoff values, and mean allele frequency cutoff. SNPs identified in the literature or dbSNP can also be included in the tagger SNP configuration.
Database of Genomic Structural Variation http://www.ncbi.nlm.nih.gov/dbvar	This database houses information on documented structural variants, including CNVs. User defined limits include criteria such as study design, method type (e.g., SNP genotyping, FISH), project ID, and variant type
Copy Number Variation (CNV) Project http://www.sanger.ac.uk/humgen/cnv/	This database provides CNV data from two projects (Global CNV assessment; High-resolution CNV discovery)
Genetics Home Reference http://ghr.nlm.nih.gov/	This website by the National Library of Medicine contains information concerning genetic conditions, genes, and chromosomes.
Talking Glossary of Genetic Terms http://www.genome.gov/glossary/index.cfm	This glossary provides definitions, illustrations, and animations of commonly used genetic/genomic terms.
The Gene Ontology Project http://www.geneontology.org/	This database can be used to identify genes whose products may impact a phenotype of interest. The domains covered include cellular component, molecular function, and biological process.
Catalog of Published Genome-Wide Association Studies http://www.genome.gov/gwastudies/	Database containing all published GWA studies attempting to genotype at least 100,000 SNPs in the initial stage
Genome-Wide Association Studies Data Repository http://was.nih.gov/	Website for the NIH Genome Wide Association Study Portal
The Genes, Environment, and Health Initiative http://www.genesandenvironment.nih.gov	Website for Genes, Environment and Health Initiative (GEI)
Database of Genotypes and Phenotypes http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gap	Database containing results of studies investigating genotype-phenotype interaction. Currently houses NIH GWAS repository.
Center for Inherited Disease Research http://www.cidr.jhmi.edu	Provides genotyping and statistical genetic services to investigators approved for access through competitive peer review process
Understanding the Basics of Microarrays http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html	This publication from the National Center for Biotechnology Information (NCBI) provides an overview of DNA microarrays explaining gene expression, the technology underlying microarrays, the purpose and importance of microarrays, and the basics of microarray experiments.
Gene Expression Omnibus http://www.ncbi.nlm.nih.gov/geo	GEO: the Gene Expression Omnibus. GEO serves as public repository and online resource for storage and retrieval of gene expression data. GEO currently maintains microarray and serial analysis of gene expression (SAGE) data on over 100
European Bioinformatics Institute http://www.ebi.ac.uk/	The European Bioinformatics Institute (EBI) is a nonprofit organization that focuses on research and services in bioinformatics. EBI's website enables access to gene expression databases (Array Express Archive and Gene Expression Atlas) and microarray analysis tools (Expression Profiler, Next Generation and Bioconductor).
Serial Analysis of Gene Expression Portal http://www.sagenet.org	Sagenet provides a detailed description of serial analysis of gene expression (SAGE). This website also provides SAGE applications, publications, and resources.
Histone Database http://www.research.nhgri.nih.gov/histones	NHGRI histone database Histone sequence information, including posttranslational modifications
Antibody Validation Database http://compbio.med.harvard.edu/antibodies/about	Collect and to share experimental results on antibodies that would otherwise remain in individual laboratories, thus aiding researchers in selection and validation of antibodies.
Chromatin Structure and Function http://www.chromatin.us	Information on chromatin biology, histones and epigenetics (hosted by Jim Bone)

Name and Address	Description
Database for DNA Methylation and Environmental Epigenetic Effects http://www.methdb.de/	Human DNA methylation Database DNA methylation data readily available to public Future develop includes environmental impact on methylation
CpG Island Searcher http://www.uscnorris.com/cpgislands2/cpg.aspx	CpG island searcher CpG Island sequence search algorithm Allows for selection of % methylation and length of (ISLAND?) and gaps between islands
Catalogue of Parent of Origin Effects http://igc.otago.ac.nz/home.html	Imprinted Gene Catalogue Catalogue of parent of origin effects Can search by taxon, chromosome, gene name or key word
Database of Noncoding RNAs http://www.noncode.org	Knowledge database dedicated to ncRNA Information on: class, name, location, related publications, mechanism through which it exerts its function Includes all traditional ncRNAs, but excludes tRNAs and rRNAs
MicroRNA Database http://www.mirbase.org	MicroRNA data resource Searchable database of >16,000 published miRNA sequences and annotation – includes location and sequence of mature miRNA Can search by name, keyword, reference and/or annotation
Epigenome Network of Excellence http://www.epigenome-noe.net	Epigenome Network of Excellence Web site of European interdisciplinary epigenetics research network Includes protocols, an antibody database and reference information on epigenetics
Human Epigenome Project http://www.epigenome.org	The Human Epigenome Project Research Consortium Collaborative effort to catalogue and interpret genome-wide methylation patterns of all human genes and major tissues

Table 2

Online Commercial Resources Used in Manuscript

Name	Address
Applied Biosystems Incorporated	http://www.appliedbiosystems.com
Roche Applied Science	http://www.roche-applied-science.com
Illumina Incorporated	http://www.illumina.com
Affymetrix Incorporated	http://www.affymetrix.com
Millipore	http://www.millipore.com
Sequenom Incorporated	http://www.sequenom.com
Preanalytix	http://www.preanalytix.com
Life Technologies Corporation	http://www.lifetechnologies.com
Ambion	http://www.ambion.com
Qiagen Incorporated	http://www.qiagen.com
DNAGenoteck Incorporated	http://www.dnagenotek.com
Panomics	http://www.panomics.com
Roche Niblegen Incorporated	http://www.nimblegen.com
Arraystar Incorporated	http://www.arraystar.com

APPENDIX D

LICENSE AGREEMENT FOR MANUSCRIPT #1

Customer Information

Customer: Kelley Baumgartel
Account Number: 3000909913
Organization: Kelley Baumgartel
Email: klb134@pitt.edu
Phone: +1 (412)5238731

Search order details by: Choose One Go

This is not an invoice

Order Details

ANNUAL REVIEW OF NURSING RESEARCH

Billing Status: **N/A**

Order detail ID: 67058924
ISSN: 0739-6686
Publication Type: Monographic Series
Volume:
Issue:
Start page:
Publisher: SPRINGER PUBLISHING COMPANY

Permission Status: **Granted**
Permission type: Republish or display content
Type of use: Thesis/Dissertation
Order License Id: 3627671204045

View details

Note: This item was invoiced separately through our **RightsLink service.** More info **$ 0.00**

APPENDIX E

MANUSCRIPT #2: PSYCHOMETRIC EVALUATION OF THE EPWORTH SLEEPINESS SCALE IN AN OBSTETRIC POPULATION

Sleep Medicine 14 (2013) 116–121

Contents lists available at SciVerse ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

Psychometric evaluation of the Epworth Sleepiness Scale in an obstetric population

Kelley L. Baumgartel [a,*], Lauren Terhorst [a], Yvette P. Conley [a,b], James M. Roberts [c,d,e]

[a] School of Nursing, University of Pittsburgh, United States
[b] Department of Human Genetics, University of Pittsburgh, United States
[c] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, United States
[d] Magee-Women's Research Institute, Department of Obstetrics Gynecology and Reproductive Sciences, School of Medicine, University of Pittsburgh, United States
[e] Clinical and Translational Research Institute, University of Pittsburgh, United States

ARTICLE INFO

Article history:
Received 14 June 2012
Received in revised form 10 October 2012
Accepted 11 October 2012
Available online 4 December 2012

Keywords:
Epworth Sleepiness Scale
Pregnancy
Psychometrics
Sleepiness
Sleep
Confirmatory factor analysis
Exploratory factor analysis
Principal component analysis

ABSTRACT

Background: The Epworth Sleepiness Scale (ESS) was initially developed to measure daytime sleep propensity in patients affected by a variety of sleep disorders. Obstetrical research has measured sleepiness in pregnant women with the ESS, although psychometric analyses and dimensionality evaluations have never been conducted with this population.
Objective: The objective was to perform a psychometric evaluation of the ESS in an obstetric population. The design was a secondary data analysis of the subjects enrolled in the Prenatal Exposures and Preeclampsia Prevention III (PEPP) study. The setting for the subjects who received prenatal care was at Magee-Women's Hospital UPMC in Pittsburgh, Pennsylvania and included 337 pregnant women in their first trimester that completed the ESS.
Methods: Principal components analysis and confirmatory factor analysis were performed using SPSS and M-Plus. Additionally, reliability was assessed and construct validity was measured using the Life Orientation Test (LOT). Lastly, a relationship between daytime sleepiness and snoring was investigated using item 5e from the Pittsburgh Sleep Quality Index (PSQI).
Results: PCA with varimax rotation yielded two factors that explained approximately 50% of the variance and CFA results verified this two-factor solution. An overall Cronbach's alpha (0.751) revealed moderate reliability (Factor 1 α = .754; Factor 2 α = .524). Both convergent and discriminant validity were established.
Conclusion: The ESS is appropriate for use in an obstetric population to measure daytime sleepiness. Future work should include additional evaluations of the ESS in a diverse group of pregnant women.

1. Introduction

Poor sleep is associated with devastating conditions in the obstetric population including: preterm labor [1], hypertensive disorders [2], and gestational diabetes mellitus [3]. When compared to the general population, pregnant women are at a higher risk of developing sleep-related problems [4], likely due to physical and hormonal changes through pregnancy. Sleep disturbances are common among pregnant women and worsen during pregnancy [4]. The mechanism of these changes remains unknown and an accurate measurement of daytime sleepiness is needed to evaluate this phenomenon. The Epworth Sleepiness Scale (ESS) is an eight item Likert-based instrument that has adequately measured daytime sleepiness in diverse populations [5–7]. The scale's ease of use, low expense, and minimal time to complete it, make an attractive option to measure sleep in both clinical and research populations. Although the ESS has been utilized to measure daytime sleepiness in both obstetrical and other clinical populations, the psychometric properties have not been examined in a sample of pregnant women.

2. Background

2.1. Development of the Epworth Sleepiness Scale

The ESS was developed to quickly and conveniently measure daytime sleep propensity in populations suffering from a variety of sleep disorders [8]. The sleep disorders experienced by subjects in the development of this tool include: primary snoring, obstructive sleep apnea syndrome (OSAS), narcolepsy, idiopathic hypersomnia, insomnia and periodic limb movement disorder. Johns [8] admits that, although the ESS does not "distinguish the nature of long-term physiological or pathological processes that produce a

* Corresponding author. Address: University of Pittsburgh, 3500 Victoria Street, 440 Victoria Building, Pittsburgh, PA 15261, United States. Tel.: +1 412 523 8731; fax: +1 412 624 8521.
E-mail address: klb134@pitt.edu (K.L. Baumgartel).

1389-9457/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sleep.2012.10.007

particular level of sleep propensity", it is capable of distinguishing normal subjects from patients suffering from OSAS, narcolepsy and idiopathic hypersomnia [8].

Prior to the development of the ESS, the Multiple Sleep Latency Test (MSLT) and the Maintenance of Wakefulness Test (MWT) were frequently used as objective, physiologic measures of sleepiness [9,10]. Despite the MSLT's high test–retest reliability over a period of months [11], this test is burdensome, time consuming and requires a polysomnographer. The Maintenance of Wakefulness Test, though capable of differentiating daytime wake tendencies from impaired wake tendencies, is costly [10].

The purpose of the development of the ESS was to provide a convenient, standardized and cost-effective way to measure sleep propensity in patients who suffer from sleep disorders [8]. Additionally, variability in daily routines is accounted for this instrument, as frequency is not assessed. For example, not everyone sits and reads every day; however, nearly everyone can assess their likelihood of dozing off when in this situation. The development of the ESS has provided a cost-effective, unobtrusive instrument that quickly and accurately assesses daytime sleepiness.

2.2. Prior psychometric evaluations

Since its development in 1991, the ESS has been administered to a variety of populations to determine its reliability. Overall, when measured in a sleep disturbed population, the reliability of the scale has been established in foreign languages (α = .81–.88) [12–16], and English-based studies (α = .78–.88) [17–19,5]. When administered to populations not recruited from a sleep clinic, the reliability of the ESS is more variable, with Myotonic Dystrophy patients exhibiting the lowest internal consistency (α = .24) [20] and Parkinsons' patients providing the highest reliability score (α = .84) [7].

The factor structure of the ESS has also been examined in a variety of populations. A principal components analysis (PCA) of the ESS was included in Heaton's (2007) evaluation of long-haul truck drivers [5]. PCA of the ESS was also performed using a sample of college students [21], patients with sleep disorders [18], and victims of Parkinson's Disease [7]. Izci et al., [16] examined the dimensionality of a Turkish version of the ESS, and Johns et al. [19,22] performed factor analysis of the ESS in a group of patients with sleep disorders; however, these studies did not describe the factor analytic method. Factor analyses of the ESS have yielded both one [14,16,19,22] and two factors [5,7,18]. Only one study, to our knowledge, has performed a confirmatory factor analysis (CFA) with the ESS [17]. The examination of the ESS in OSAS patients resulted in a one factor solution only when items six and eight were deleted [17].

2.3. Use of ESS in obstetric population

Previous studies focusing on the obstetric population have utilized the ESS to measure daytime sleepiness. Mindell (2000) revealed that despite 67.2% of subjects reporting "moderate" to "very great" daytime sleepiness, no differences were found in ESS total scores throughout pregnancy [4]. Izci (2005) uncovered that among pregnant women who snore, ESS scores were higher, which indicated an increase in daytime sleepiness [23]. Despite the extensive use of the ESS in the obstetric population to examine quality of life [24] and risk of late stillbirth [25], the psychometric properties have not been explored using a sample of pregnant women.

3. Purpose

Psychometric evaluations of the ESS have been performed in a variety of populations with varied results. The wide variability among findings in clinical populations suggest that further exploration of the dimensionality of the ESS is needed to facilitate appropriate scoring. A psychometric evaluation of the ESS is necessary if pregnancy-related research aims to accurately measure sleep propensity in the obstetric population. This is especially vital since poor sleep is associated with pregnancy complications, including preterm labor [26] and gestational diabetes mellitus [27]. The purpose of this study was to assess the psychometric performance of the ESS in women who are in their first trimester of pregnancy (n = 337).

4. Methods

4.1. Sample and parent study (PEPP III)

This study was a secondary analysis of data collected through a larger project, Prenatal Exposures and Preeclampsia Prevention III (PEPP) study, which aims to better understand the role of obesity in the pathogenesis of preeclampsia, though subjects of all BMIs are invited to participate. The PEPP III study has recruited both longitudinal (before preeclampsia) and cross-sectional (preeclamptic or suspected preeclamptic) subjects, and remains an ongoing project through the Magee-Women's Research Institute. Longitudinal subjects were recruited during the first trimester at a prenatal appointment at the WomanCare Clinic at Magee-Womens Hospital. Subjects had a total of five study visits that occur during regularly scheduled appointments: (1) first trimester, (2) second trimester, (3) third trimester, (4) immediately postpartum and (5) 6-week postpartum clinic visit. PEPP III has been approved by the University of Pittsburgh's Institutional Review Board (IRB), and all participating subjects signed an informed consent before any research procedures were performed. Additionally, approval to conduct this secondary analysis was obtained by the University of Pittsburgh's IRB.

The sample for this analysis reflects only the longitudinal subjects who were recruited by the PEPP III study between January 2009 and September 2011 (n = 337). Cross-sectional subjects with suspected preeclampsia were not included in this project, since these women delivered at different gestational ages and completion rates for the ESS were lower in this subgroup. Mindell reports no differences in the ESS scores throughout pregnancy [4]; thus, data from one time point during the first trimester was analyzed in this study. At this point, subjects completed the ESS, Life Orientation Test (LOT) and Pittsburgh Sleep Quality Index (PSQI), among other questionnaires not included in this secondary analysis.

4.2. Instruments

4.2.1. Epworth Sleepiness Scale (ESS)

The ESS is a self-administered, eight-item questionnaire that takes two–three minutes to complete [22]. Items address daily lifestyle activities and the respondent is asked to rate their likelihood of dozing in each situation, from: "would never doze" (0) to "high chance of dozing" (3). The ESS provides a cumulative score between 0 and 24, with higher numbers indicating greater daytime sleepiness.

4.2.2. Life Orientation Test (LOT)

The Life Orientation Test measures general optimism versus pessimism with ten Likert-scaled items that range from "strongly disagree" (0) to "strongly agree" (4) [28]. This scale is designed to assess overall expectations for positive or negative outcomes [29] and is widely used in social and medical sciences, including obstetrics [30]. Reports of the scale's reliability are mostly adequate, but with large variability ranging from α = .35[31] to α = 0.85[32].

4.2.3. Pittsburgh Sleep Quality Index

Item 5e from the Pittsburgh Sleep Quality Index (PSQI) contains ten questions and measures nighttime sleep disturbances. Item 5e asks the subject to identify the frequency of trouble sleeping in the past month due to coughing or snoring (not during the past month, less than once a week, once or twice a week or three or more times a week). The PSQI has been shown to have high internal consistency ($\alpha = .83$) in clinical populations [33].

4.3. Statistical analyses

4.3.1. Preliminary analyses

Descriptive statistics were computed for the demographic variables and the percent of missing data was assessed. Distributions for each ESS item, PSQI (item 5e) and LOT score were examined with histograms and q–q plots. The Shapiro–Wilk test was used as a formal test for univariate normality for all items of each instrument used in the analysis. Potential multivariate outliers were assessed three ways: (1) scatterplots, (2) Mahalanobis Distance and (3) leverage values. The initial screening of data and principal components analysis (PCA) were performed using Statistical Packages for the Social Sciences [34], while the confirmatory analyses were performed using MPLUS [35].

4.3.2. Principal components analysis

The inter-item correlation table was obtained to summarize the interrelationships among the eight items that comprise the ESS. The Kaiser–Meyer–Olkin statistic (KMO) and Bartlett's test of sphericity were obtained to ensure the suitability of factor analysis. KMO values of .6 indicated that a factor analysis was appropriate [36]. Communalities, the scree plot, total variance explained and eigenvalues were assessed along with factor loadings in order to determine factor structure. Factor loadings of greater than .32 were considered sufficient, while items with factor loadings of .32 or greater on more than one factor were considered cross-loading [37]. PCA was performed and both varimax and promax rotations were applied in one half of the sample ($n = 168$). Varimax cross-loadings are reported, since this orthogonal rotation assumes uncorrelated factors and therefore, provides a clearer picture of loadings [38]. Next, the emergent factor structure was tested using confirmatory techniques with the other half of the sample ($n = 169$). In order to assess model fit, the Root Mean Square Error of Approximation (RMSEA) and Comparative Fit Index (CFI) were examined. RMSEA values of less than .05 and values higher than .90 for CFI were indicative of good model fit.

4.3.3. Reliability and validity

The internal consistency of the ESS was assessed with Cronbach's alpha to obtain: (1) an overall reliability, and (2) subscale reliabilities. Cronbach's alpha levels of .80 or greater were considered acceptable since the ESS is an established instrument [39]. Data from a gold standard measure of daytime sleepiness were not available; therefore, hypotheses were tested to further evaluate validity [40]:

1. The ESS subscale score(s) will be weakly and not significantly correlated with LOT scores, since these instruments measure different constructs. A weak r-value and not significant p-value will reinforce that the LOT and ESS measure unrelated (discriminant) concepts. This hypothesis was tested by using appropriate correlational techniques.
2. The ESS subscale score(s) will be positively correlated with item 5e of the PSQI, which measures difficulty sleeping due to coughing or snoring. Previous studies have shown snoring is positively correlated with ESS scores in pregnant women [23]. Hypothesis 2 was tested by computing correlations between the ESS subscale(s) and item 5e of the PSQI.

5. Results

There were no missing cases for the ESS data and missing data was less than 3% for the LOT questionnaire and item 5e of the PSQI. The descriptives and analyses do not include data from eight subjects that were removed as influential multivariate outliers. Descriptives (Table 1) are reported for each analysis group. Sum ESS scores were similar between both groups and were consistent with previous findings that used the ESS in a pregnant population. The examination of the descriptives, box plots, histograms and q–q plots indicated that the items of the ESS, LOT and item 5e (PSQI) were not normally distributed.

5.1. Principal components analysis

Correlations from the inter-item correlation matrix ranged from $r = .07$ to $r = .455$. There were no correlations greater than 0.5, indicating no problems with multicollinearity [41]. The KMO of 0.788 indicated that the dataset was suitable for factor analysis. Additionally, Bartlett's test of sphericity was significant ($p < .001$); therefore, it was appropriate to perform a factor analysis with this sample. Communalities, which quantify the amount of variance between items, ranged from 0.300 to 0.784.

We uncovered a two-factor structure through the examination of the scree plot, eigenvalues, and factor loadings. These findings were similar to a previous factor analysis of the ESS [5]. An examination of the scree plot (Fig. 1) revealed the possibility of two factors. Two factors had eigenvalues greater than one and explained 50.59% of the total variance (Table 2). Three items are loaded onto factor one (items two, four and five) and two items loaded onto factor two (items six and eight). Items one and three, though technically cross-loading (greater than 0.32 on both factors), were most theoretically appropriate under factor one, for which the loading was higher (Table 2). Item seven (sitting quietly after lunch without alcohol) has nearly identical cross-loading values on both factors (0.492 and 0.497). Factor one is comprised of life situations in which dozing off is socially acceptable, and factor two depicts life situations in which dozing off is socially unacceptable.

5.2. Confirmatory factor analysis

In order to determine if the findings of the PCA were appropriate, we compared the two-factor model to a one-factor model of previous findings [14,16,19,22] using confirmatory factor analysis.

Table 1
Demographics.

	PCA ($n = 168$)	CFA ($n = 169$)
Mean age (years)	23.92	24.7
Race		
African American (%)	101 (60.1)	96 (56.8)
Caucasian (%)	61 (36.3)	70 (41.4%)
Other (%)	6 (3.5)	3 (1.8)
Education		
Less than H.S. (%)	26 (15.5)	25 (14.7)
H.S. Diploma/GED (%)	60 (35.7)	57 (33.7)
Some education after H.S. (%)	50 (29.7)	55 (32.5)
College/vocational degree/cert (%)	26 (15.4)	27 (15.9)
Graduate school/degree (%)	4 (2.3)	3 (1.8)
No answer (%)	2 (1.2)	2 (1.1)
Income		
Less than $19,999	66 (39.3)	70 (41.4)
$20,000 to $49,999	25 (14.9)	39 (23)
$50,000 to $75,000	6 (3.6)	7 (4.1)
Greater than $75,000	2 (0.6)	7 (4.1)
Unsure/no answer	56 (33.3)	46 (27.2)
Mean ESS score	7.17	7.24

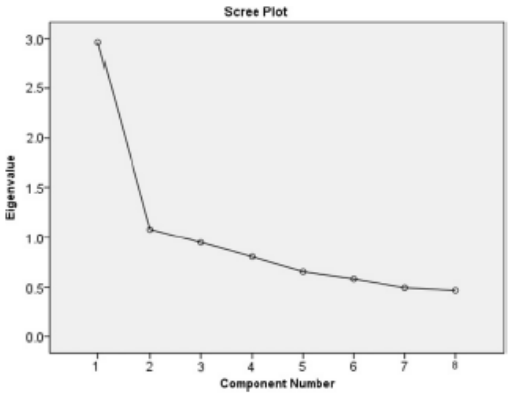

Fig. 1. Scree plot.

Table 2
PCA factor loadings and total variance explained.

ESS item	Factor one	Factor two
1 – Sitting and reading	0.576	
2 – Watching TV	0.705	
3 – Sitting inactive in a public place	0.497	
4 – As a passenger in a car for an hour without a break	0.691	
5 – Lying down to rest in the afternoon	0.747	
6 – Sitting and talking to someone		0.886
7 – Sitting quietly after lunch without alcohol	0.492	0.447
8 – In a car, while stopped for a few minutes in traffic		0.653
Total variance explained (%) = 50.592	37.078	13.514

Table 3
CFA Indices.

Structure	Chi square	RMSEA	CFI
One factor	$\chi^2_{(20)} = 30.115$, $p = .06$	.055	.975
Two factors	$\chi^2_{(19)} = 26.83$, $p = .108$	.049	.981

The indices of the two-factor model were compared to the indices of the one-factor solution. While both models yielded a good model fit, the two-factor solution produced a slightly better solution, with lower RMSEA and higher CFI indices (Table 3).

5.3. Reliability

The overall Cronbach's alpha of the ESS was very close to the recommended value for acceptable reliability of an established instrument [39], ($\alpha = .751$). The reliability of factor one was also acceptable ($\alpha = .743$), while the reliability of factor two was well below the acceptable level ($\alpha = .524$). As previously discussed, rotated factor loadings for item seven were similar for both factors. To determine the appropriate factor for this item, reliability was examined with and without item seven on the overall scale and on each factor. The resulting Cronbach's alpha for the overall score without item seven ($\alpha = .708$) was less than when it was included in the overall score ($\alpha = .751$). Additionally, the reliability of factor one was lower without item seven ($\alpha = 706$). The reliability of factor two with item seven was lower than when item seven was omitted from factor two ($\alpha = .488$ and $\alpha = .524$, respectively). The results of the reliability analysis provided evidence that item seven should remain on factor one.

5.4. Validity

5.4.1. Hypothesis 1

As expected, the correlations between the emergent ESS subscales and the LOT scores were very low and not significant. The correlation between factor one (sleepiness in appropriate situations) and the overall LOT score was very weak ($r = .094$, $p = .088$); while the correlation between factor two (sleepiness in inappropriate situations) and the overall LOT score was even lower ($r = .075$, $p = .169$). Additionally, the magnitude of the relationship between the overall score of the ESS and the overall LOT score was diminutive ($r = .095$, $p = .084$).

5.4.2. Hypothesis 2

A significant and positive correlation was revealed between factor two of the ESS (items six and eight) and item 5e (PSQI) ($r_s = .119$, $p = .031$). The positive direction indicated that as trouble sleeping due to coughing or snoring increased, sleepiness in socially inappropriate situations increased.

6. Discussion

There have been several previous psychometric evaluations of the ESS in various populations, but to our knowledge, none have assessed the validity or reliability in a sample of pregnant women. A previous evaluation of the ESS has revealed a potential gender bias, as women who report unrest are less likely than men to have an elevated ESS score >10 [42]. However, another study uncovered similar ESS scores between men and women [43]. This represents a dearth of information related to the performance of the ESS in specific clinical populations, including pregnant women. According to the Standards of the American Educational Research Association (1999) [40], it is the responsibility of the user of an instrument to justify its use, either by synthesizing prior evidence or by providing new supporting evidence. This study provided an overview of prior psychometric studies, and supplied new evidence obtained with a sample of obstetric patients.

Due to the conflicting results of prior factor analytic studies, it was our goal to determine the appropriate dimensional structure of the ESS when used in samples of pregnant women. Previous studies have reported both one and two factor structures. In order to score the ESS appropriately, the researcher must be familiar with the dimensionality of the instrument. Using an overall (unidimensional) score of a multidimensional instrument may result in a loss of information that could reflect an important characteristic of the population. It could also lead to incorrect inferences with consequential outcomes. Performing a PCA allowed us to uncover a preliminary structure that reflected two factors: sleepiness in appropriate situations and sleepiness in inappropriate situations. In order to justify and substantiate this finding, we performed a CFA and found this two factor structure to be appropriate. These findings imply that future researchers should score the ESS in two parts. High scores on sleepiness in inappropriate situations could be an indication of further complications during pregnancy. Previous work has shown that sleepiness in pregnant women can be related to preterm labor [26] and gestational diabetes mellitus [27]; however, the relationships between these clinical characteristics and the factors of the ESS have yet to be explored.

Once the two-factor structure was confirmed, we continued to substantiate psychometric evidence by examining the reliability of the emergent subscales. The internal consistency of factor one

was adequate; however, the internal consistency of factor two was below the standard. The low Cronbach's alpha of factor two ($\alpha = .524$) could be due to the limited number of items ($n = 2$). The two items of factor two are related to situations in which sleeping is socially unacceptable. We recommend that more content relevant items representative of sleepiness in socially unacceptable situations be added to factor two.

Item seven (sitting quietly after lunch without alcohol) may be unclear to a population to which this scenario may not apply, and its inclusion illustrates the justification for psychometric testing of the ESS in an obstetric population. The wordage of this item within an obstetric population deserves further explanation, since only one out of eight pregnant women reported alcohol consumption [44], a rate that is remarkably lower than that of the general population. The inclusion of item seven does increase the reliability of factor one; therefore, we recommend that when using the ESS with an obstetric population, item seven could be reworded to: "sitting quietly after lunch". This wordage continues to measure postprandial doziness, yet considers the life experience common to this population.

The absence of data obtained from a gold standard measure of daytime sleepiness led us to test a series of hypotheses in order to further establish the validity of the two subscales obtained from the ESS. The emergent subscales of the PCA were assessed in order to further substantiate the use of the two-factor structure. The LOT questionnaire and ESS subscales were very weakly correlated and not significant, which supported hypothesis 1. The LOT questionnaire, which measures orientation to life situations as positive (optimistic) or negative (pessimistic), is not expected to be related to sleep propensity; therefore, this implies that both subscales of the ESS are measuring a different construct than the LOT. The significant relationship between factor two of the ESS and item 5e of the PSQI supports hypothesis 2. Izci (2005) reported a significant positive relationship between snoring and ESS scores among pregnant women in the third trimester [23]. Our findings support this relationship and suggest that this pattern may also exist in the first trimester. The significant correlation with only factor two during the first trimester, supports a two-factor structure rather than a unidimensional scale, as dozing off during socially unacceptable scenarios may be more common among sleepier subjects.

While these findings seem to support the use of a two-factor structure when using the ESS in a sample of pregnant women, some limitations to the current investigation should be addressed. A large proportion of our sample was African American (58.9%), which could skew the findings. Previous work has shown that African Americans score higher on the ESS and generally experience more sleepiness than Caucasians [45]. The relationship between ESS scores and race should be further investigated, as the larger sample of African Americans in our sample could have led to the emergence of factor two.

Other limitations of this study include the unavailability of parity status, as children in the home may impact sleep cycles and subsequent daytime sleep propensity; however, previous work has revealed that children at home did not significantly impact sleepiness [4] or sleep duration [46]. Additionally, depression symptoms including fatigue, low energy, and lack of vigor may confound ESS scores [16], and employment status, which influence sleep patterns [47], were not investigated in this secondary analysis. Additionally, item 5e (PSQI) includes both snoring and coughing; however, only snoring has been associated with ESS scores [23], sleep disturbances [4] and pregnancy complications [23]. We were not able to delineate between the coughing and snoring effects to accurately conclude which is most related to daytime sleepiness in socially appropriate and inappropriate situations.

7. Conclusions

This is the first time the psychometric properties of the ESS have been examined in an obstetric sample. The performance of this instrument is generally consistent with previous evaluations of the ESS [5]. The findings of this investigation supply additional reliability and validity evidence for the ESS and provide evidence of a multidimensional, two-factor structure of the ESS when it is used as a measure of daytime sleep propensity in pregnant women. Further evaluation of this instrument is needed to investigate the relationships between the two factor structure with relevant demographic and clinical characteristics. Also, the structure should be confirmed in pregnant women at various stages of pregnancy.

Conflict of Interest

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: http://dx.doi.org/10.1016/j.sleep.2012.10.007.

References

[1] Micheli K, Komninos I, Bagkeris E, et al. Sleep patterns in late pregnancy and risk of preterm birth and fetal growth restriction. Epidemiology 2011; 22(5):738–44.
[2] Williams MA, Millers RS, Qiu C, Cripe SM, Gelaye B, Enquobahrie D. Associations of early pregnancy sleep duration with trimester-specific blood pressures and hypertensive disorders in pregnancy. Sleep 2010; 33(10):1363–71.
[3] Qiu C, Enguobahrie D, Frederick IO, Abetew D, Williams MA. Glucose intolerance and gestational diabetes risk in relation to sleep duration and snoring during pregnancy: a pilot study. BMC Womens Health 2010;10:17.
[4] Mindell J, Jacobson B. Sleep disturbances during pregnancy. JOGNN 2000; 29(6):590–7.
[5] Heaton K, Anderson D. A psychometric analysis of the Epworth Sleepiness Scale. J Nurs Meas 2007;15(3):177–88.
[6] Pahwa P, Karunanayake CP, Hagel L, et al. Prevalence of high Epworth Sleepiness Scale scores in a rural population. Can Respir J 2012;19(2):e10–14.
[7] Hagell P, Broman J. Measurement properties and hierarchical item structure of the Epworth Sleepiness Scale in Parkinson's disease. J Sleep Res 2007; 16(1):102–9.
[8] Johns M. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. Sleep 1991;14:540–5.
[9] Carskadon MA, Dement WC, Mitler MM, et al. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. Sleep 1986; 9(4):519–24.
[10] Doghramji K, Mitler MM, Sangal RB, et al. A normative study of the maintenance of wakefulness test (MWT). Electroencephalogr Clin Neurophysiol 1997;103(5):554–62.
[11] Zwyghuizen-Doorenbos A, Roehers T, Schaeffer M, Roth T. Test–retest reliability of the MSLT. Sleep 1988;11:562–5.
[12] Bertolazi AN, Fagondes SC, Hoff LS, et al. Portuguese-language version of the Epworth sleepiness scale: validation for use in Brazil. J Bras Pneumol 2009; 35(9):877–83.
[13] Beiske KK, Kjelsberg FN, Ruud EA, Stavem K. Relability and validity of a Norweigian version of the Epworth Sleepiness Scale. Sleep Breath 2009; 13(1):65–72.
[14] Chen N, Johns M, Li H, et al. Validation of a Chinese version of the Epworth Sleepiness Scale. Qual Life Res 2002;11:817–21.
[15] Kaminska M, Jobin V, Mayer P, et al. The Epworth Sleepiness Scale: self-administration versus administration by the physician, and validation of a French version. Can Respir J 2010;17(2):e27–34.
[16] Izci B, Ardic S, Firat H, et al. Reliability and validity studies of the Turkish version of the Epworth Sleepiness Scale. Sleep Breath 2008;12(2):161–8.
[17] Smith SS, Oei TP, Douglas JA, Brown I, Jorgensen G, Andrews J. Confirmatory factor analysis of the Epworth Sleepiness Scale in patients with obstructive sleep apnoea. Sleep Med 2008;9(7):739–44.
[18] Violani C, Lucidi F, Robusto E, Devoto A, Zucconi M, Ferini-Strambi L. The assessment of daytime sleep propensity: a comparison between the Epworth Sleepiness Scale and a newly developed Resistance to Sleepiness Scale. Clin Neurophysiol 2003;114:1027–33.
[19] Johns M. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 1992;15:376–81.
[20] Laberge L, Gagnon C, Jean S, Mathieu J. Fatigue and daytime sleepiness rating scales in myotonic dystrophy: a study of reliability. J Neurol Neurosurg Psychiatry 2005;76(10):1403–5.
[21] Pilcher JJ, Pury CL, Muth ER. Assessing subjective daytime sleepiness: an internal state versus behavior approach. Behav Med 2003;29(20):60–7.

[22] Johns M. Sleepiness in different situations measured by the Epworth Sleepiness Scale. Sleep 1994;17:703–10.
[23] Izci B, Martin SD, Dundas KC, Wang LA, Calder AA, Douglas N. Sleep complaints: snoring and daytime sleepiness in pregnant and pre-eclamptic women. Sleep Med 2005;6:163–9.
[24] Amador-Licona N, Gulzar-Mendoza JM. Daytime sleepiness and quality of life: are they associated in obese pregnant women? Arch Gynecol Obstet 2012;295(1):105–9.
[25] Stacey T, Thompson JM, Mitchell EA, Ekeroma AJ, Zuccollo JM, McCowan LM. Association between maternal sleep practices and risk of late stillbirth: a case-control study. BMJ 2011;342:d3403.
[26] Okun ML, Schetter CD, Glynn LM. Poor sleep quality is associated with preterm birth. Sleep 2011;34(11):1493–8.
[27] Reutrakul S, Zaidi N, Wroblewski K, et al. Sleep disturbances and their relationship to glucose tolerance in pregnancy. Diabetes Care 2011;34(11):2454–7.
[28] Scheier MF, Carver CS. Optimism, coping and health: assessment and implications of generalized outcome expectancies. Health Psychol 1985;4:219–47.
[29] Vassar M, Bradley G. A reliability generalization study of coefficient alpha for the life orientation test. J Pers Assess 2010;92(4):362–70.
[30] Moyer C, Ekpo G, Calhoun C, et al. Quality of life, optimism/pessimism, and knowledge and attitudes toward HIV screening among pregnant women in Ghana. Womens Health Issues 2008;18(4):301–9.
[31] Graci G. Optimism, health locus of control, and quality of life women with recurrent breast cancer; 1998. Available from ProQuest Dissertation and Theses database [UMI No. 1392755].
[32] Eid J, Matthews MD, Meland NT, Johnsen BH. Dispositional optimism and self-assessed situation awareness in a Norweigian military training exercise. Percept Mot Skills 2005;100:649–58.
[33] Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28:129–213.
[34] IBM SPSS Statistics, Version 19.0.0 Copyright 2010.
[35] Muthén LK, Muthén BO. Mplus user's guide. 6th ed. Los Angeles (CA): Muthén & Muthén; 1998.
[36] Kaiser H. An index of factorial simplicity. Psychometrika 1974;39:32–6.
[37] Costello A, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Evalu 2005;10(7).
[38] Cureton EE, Mulaik SA. The weighted varimax rotation and promax rotation. Psychometrika 1975;40(2):183–95.
[39] Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw Hill; 1994.
[40] American Educational Research Association. Standards for educational and psychological testing, Washington (DC), American Educational Research Association, 1999.
[41] Pett M. Making sense of factor analysis: the use of factor analysis of instrument development in health care research. Thousand Oaks: Sage Publications; 2003.
[42] Baldwin CM, Kapur VK, Holberg CJ, et al. Associations between gender and measures of daytime somnolence in the sleep heart health study. Sleep 2004;27(2):305–11.
[43] Mohsenin V. Gender differences in the expression of sleep-disordered breathing: role of upper airway dimensions. Chest 2001;120(5):1442–7.
[44] Centers for Disease Control and Prevention. Alcohol use among women of childbearing age – United States, 1991–2005. Available at: http://www.cdc.gov/ncbddd/fasd/data.html. Accessibility verified May 14, 2012.
[45] Hayes A, Spilsbury J, Patel S. The Epworth score in African American populations. J Clin Sleep Med 2009;5(4):344–8.
[46] Hedman C, Pohjasvaara T, Tolonen U, Suhonen-Malm AS, Myllyla VV. Effects of pregnancy on mother's sleep. Sleep Med 2002;3:37–42.
[47] Akerstedt T, Nordin M, Alfredsson L, Westerholm P, Kecklund G. Predicting changes in sleep complaints from baseline values and changes in work demands, work control, and work preoccupation – the WOLF-project. Sleep Med 2012;13(1):73–80.

APPENDIX F

LICENSE AGREEMENT FOR MANUSCRIPT #4

ELSEVIER LICENSE
TERMS AND CONDITIONS

Apr 16, 2015

This is a License Agreement between Kelley L Baumgartel ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Kelley L Baumgartel
Customer address	1619 Three Degree Rd MARS, PA 16046
License number	3610860184768
License date	Apr 16, 2015
Licensed content publisher	Elsevier
Licensed content publication	Sleep Medicine
Licensed content title	Psychometric evaluation of the Epworth Sleepiness Scale in an obstetric population
Licensed content author	None
Licensed content date	January 2013
Licensed content volume number	14
Licensed content issue number	1
Number of pages	6
Start Page	116
End Page	121
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes

Will you be translating?	No
Title of your thesis/dissertation	The Impact of Promoter Polymorphisms on Cytokine Concentration in Preterm Breast Milk and Subsequent Infant Outcomes
Expected completion date	May 2015
Estimated size (number of pages)	100
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described

in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or

enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
    - via their non-commercial person homepage or blog
    - by updating a preprint in arXiv or RePEc with the accepted manuscript
    - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
    - directly by providing copies to their students or to research collaborators for their personal use
    - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
    - via non-commercial hosting platforms such as their institutional repository
    - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final

record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third

party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access

APPENDIX G

MANUSCRIPT #3: THE UTILITY OF BREAST MILK FOR GENETIC OR GENOMIC STUDIES

BREASTFEEDING MEDICINE
Volume 8, Number 3, 2013

DOI: 10.1089/bfm.2012.0054

Review

The Utility of Breastmilk for Genetic or Genomic Studies: A Systematic Review

Kelley L. Baumgartel[1] and Yvette P. Conley[1,2]

Abstract

This study synthesized scientific literature that applies genetic or genomic approaches to breastmilk. A literature search of PubMed was conducted in March 2012 using the key words "breast milk," "lactation," "genetic," "gene expression," and "epigenetic." Additional articles were identified/selected for evaluation with MeSH term searches, and a review of article reference lists was obtained from the search. The initial 657 abstracts retrieved from the literature search were reviewed, and 16 studies were selected for evaluation. Studies that examined the transmission of viruses/bacteria into breastmilk and/or measured concentration of specific proteins without examination of genetic material from milk were excluded. Data related to subjects, tissue, purpose, setting, gene/protein, approach (candidate versus genome-wide), platform, statistical analysis, and results were extracted. Gene expression and epigenetic/epigenomic study designs have been successfully implemented using breastmilk. A major weakness of both gene expression studies and epigenetic studies that examine breastmilk is the omission of maternal information known to influence milk composition. This review article is the first to synthesize evidence related to the application of breastmilk to evaluate RNA and epigenetic modifications. Additional research is needed that applies epigenetic analyses to human breastmilk samples. Findings from this review can be used for future research designs that use breastmilk for genetic analyses.

Introduction

Breastmilk composition is a constantly changing substance that is influenced by maternal lifestyle factors, including diet,[1] medications,[2] and exercise,[3] among others. Unalterable maternal factors that also impact breastmilk composition include the time of day,[4] number of days postpartum,[5,6] and gestational week at delivery.[7–11] Breastmilk composition varies greatly between women who deliver prematurely and those who deliver at term, although the mechanism for this difference is poorly understood. There is also evidence of variation in breastmilk composition *between* women who deliver at the same gestational age,[12] suggesting that breastmilk may not be a uniform substance. For example, interleukin-10, an anti-inflammatory cytokine present in mature human milk, is influenced by gestational age and is found in lower levels among preterm infants.[13] Additionally, evidence of undetectable interleukin-10 in the milk of women whose infants developed necrotizing enterocolitis[14] suggests that milk variability may also influence neonatal outcomes.

Variable levels of protective components may be driven by genetic or genomic factors, and the procurement of breastmilk is a noninvasive way to examine these factors. Epithelial cells, which contain both RNA and DNA, make up 50–90% of cell types found in human breastmilk.[15] Analyses of RNA and DNA from human milk provide a platform to better understand the mechanism for compositional variability and neonatal outcomes.

A potential link between breastmilk variation and subsequent neonatal outcomes has been minimally explored at both the protein and gene levels; however, recent work has examined breastmilk profiles with particular attention paid to gene expression and, more recently, epigenetics. Exploring the potential genetic mechanism for breastmilk variation between women may lead to the improvement of neonatal diet through breastmilk optimization, including (1) donor breastmilk fortification, (2) maternal dietary supplements, and (3) maternal lifestyle changes. The collection of breastmilk is a noninvasive way to understand gene regulation in mammary epithelial cells that may explain a potential mechanism for outcome disparities among breastfed infants. Despite evidence that has focused on breastmilk variability at the protein level, a critical review and synthesis of the literature are needed that address genetic or genomic approaches to investigate the mechanism for these differences. This systematic review will describe and critique the recent science that has examined breastmilk using genetic or genomic approaches. This review

[1]School of Nursing and [2]Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania.

can be used as a reference by investigators who hope to examine human breastmilk using a genetics or genomics.

Data Collection Method

We used the PubMed database to identify studies that examined human breastmilk using a genetic or genomic approach, including gene expression, candidate gene, and epigenetic analyses. The key words used were "breast," "milk," "lactation," "milk fat globule," "gene," "genetics," "expression," "epigenetic," "epigenomic," "methylation," "gestational age," and "preterm." The following combinations of key words were used: "breast" AND "milk" AND "gene" AND "expression"; "lactation" AND "gene" AND "expression"; "breast milk" AND "genetics" AND "gestational age"; "breast milk" AND "genetics" AND "preterm"; "milk fat globule" AND "genetic"; and "milk fat globule" AND "gene expression." We limited the literature search, which covered literature through May 2012, to the English language and articles involving human subject research and excluded articles related to human immunodeficiency virus, cytomegalovirus, and cancer. We relaxed the inclusion criteria for epigenetic studies to include both human and non-human studies, as few have examined breastmilk from an epigenetic approach.

After completing the literature search, we reviewed abstracts of retrieved articles for relevance, excluding duplicate articles, review articles, and those that did not address the use of genetic approaches to examine breastmilk. We also reviewed PubMed e-mail updates and reference lists of selected articles to identify additional studies. After independent review, we met to discuss findings and synthesize results. We extracted from gene expression studies data related to maternal information, RNA isolation method, type of milk, gene product of interest, data collection platform, and results. From epigenetic studies, we extracted the following data: maternal information, epigenetic modification of interest, gene(s) of interest, data collection platform, and results. This information was summarized in tabular format.

Results

We reviewed 35 articles of the initial 657 results. Of these 35 articles, 16 were included in the final analysis. Of those excluded, 10 articles used milk to examine protein levels, four used human mammary epithelial cells, two were animal gene expression studies, one only evaluated DNA damage, one article used milk only to identify epithelial cells, and one article quantified DNA to assess feasibility of DNA adduct evaluation. There were 16 articles that fulfilled our inclusion criteria and were included in this review. Of these, 13 were gene expression articles, and the remaining three were epigenetic studies. Table 1 includes a summary of the results and characteristics of the articles that examined breastmilk at the gene expression level. Table 2 includes a summary of the results and characteristics of the articles that examined breastmilk from an epigenetic approach. Although we encountered many studies that examined maternal DNA polymorphisms with regard to breastmilk properties and/or milk protein concentration measured with enzyme-linked immunosorbent assay, we did not use these studies in the current review. Only articles that examined epigenetics or gene expression using DNA or RNA isolated from breastmilk were included.

Discussion

We conducted this systematic review in order to critique and synthesize scientific literature that used DNA or RNA found in breastmilk in their methodology. Breastmilk is a unique tissue source; therefore it is ideal to investigate dynamic templates such as epigenetic changes to the DNA or mRNA levels. The resulting articles are overwhelmingly gene expression related. This is appropriate because our focus of this review is on the utility of breastmilk as a biospecimen, and expression studies require the tissue of interest. Few articles in this review are epigenetically focused, likely because of the relatively new technologies available for epigenetic analyses.

Use of breastmilk in gene expression studies

The methodological article by Lindquist et al.[16] documents one of the first approaches of RNA isolation from human breastmilk to examine β-casein mRNA using a northern blot. Most of the studies examined in this review incorporated reverse transcription (RT)-polymerase chain reaction (PCR) to examine gene expression of epithelial cells from breastmilk, including genes for cytokines,[17] defensins,[18] interleukin-18,[19] and interferon-γ.[20] One study incorporated PCR with a western blot to examine M-ficolin expression.[21] Andersson et al.[22] acknowledged that many cell types, including leukocytes and macrophages, exist in breastmilk. To ensure that epithelial cell gene expression data were obtained, they performed a Southern blot on commercially available human mammary epithelial cells and found identical fragments from human milk cells. Obermeier et al.[23] subjected both whole epithelial cells and isolated RNA from milk to RT-PCR in order to determine if RNA isolation is a necessary step. The investigators concluded that the cell fraction from fresh human milk is an appropriate model to examine glucose transporter gene expression, and this was confirmed with Southern and western blotting.

As technology advanced, the available platforms to examine gene expression became more efficient and comprehensive. In addition to RT-PCR, spectrophotometry allowed Alcorn et al.[24] to quantify total RNA concentration extracted from breastmilk to examine transporter genes implicated in drug disposition. Another study[25] used a microarray approach on one breastmilk sample that focused on cytokine-related genes and found the gene for osteopontin was the most highly expressed gene among those tested. A western blot examined osteopontin expression differences among colostrum, early, and mature milk. RT-PCR was performed on remaining milk donor samples to confirm high osteopontin gene expression. Three additional studies, all performed by the same investigators, also used a microarray platform to examine milk fat globule gene expression throughout the day,[26] when different pumping protocols were used,[27] and after administration of recombinant human growth hormone.[28] Maningat et al.[27] used a traditional nonparametric approach to analyze results from one subject's milk that was subjected to the microarray platform, and RT-PCR confirmed these results using milk from 10 other donors. The remaining microarray study[28] used a traditional nonparametric analysis approach using GeneSpring GX 9 (Agilent Technologies, Santa Clara, CA).

Table 1. Gene Expression Studies That Examine mRNA from Breastmilk

Reference (year)	*Maternal information*	*RNA isolation method*	*Fresh vs. frozen milk*	*Gene product*	*Platform*	*Results*
Alcorn et al.[24] (2002)	$n=6$ healthy women, 1–11 months postpartum	RNeasy	Fresh	Drug transporter genes	RT-PCR	Lactating MEC had fourfold higher RNA levels of OCT1, OCTN1, PEPT2, CNT1, CNT3, and ENT3.
Andersson et al.[22] (1996)	$n=1$, 9 days postpartum; 200 mL provided (50 mL used)	Unclear	Fresh	PTHrP	RT-PCR	mRNA encoding three PTHrP variants is present in human milk cells.
Frankenberger et al.[21] (2008)	$n=3$; additional information not provided; 20–50 mL	TRI-Reagent	Fresh	M-Ficolin	RT-PCR Confirmed with western blot	Macrophage M-ficolin expression from breastmilk was lower than in blood monocyte M-ficolin expression.
Lindquist et al.[16] (1994)	Additional information not provided; 50 mL	Unclear	Fresh and frozen	β-Casein	PCR Northern blot	PCR amplification of β-casein gene was successfully performed on mRNA and genomic DNA from breastmilk. Freezing milk degraded β-casein mRNA.
Maningat et al.[27] (2007)	Microarray: $n=1$, additional information not provided. RT-PCR: $n=10$ healthy women 18–35 years old, 6–12 weeks postpartum, term singletons, exclusively breastfeeding, standard diet, pumping protocol; 10 mL	RNeasy; switched to TRizol (less RNA degradation)	Fresh	α-Lactalbumin	Microarray of MFG RNA from one subject; QRT-PCR confirmed array results. Breastmilk from the 10 subjects enrolled in the pumping study was subjected to RT-PCR.	RT-PCR confirmed microarray results: GHR and IGF1 were undetectable in up to 2 μg of MFG RNA. IGF1-R was expressed at low levels using 1–4 μg of RNA. α-Lactalbumin is abundant in the MFG, and its expression may be regulated by pumping.
Maningat et al.[26] (2009)	$n=5$ healthy women, uncomplicated pregnancies, 18–35 years old, exclusively breastfeeding, 6–12 weeks postpartum, standard diet; 10 mL	TRizol	Fresh	Global	Microarray of MFG RNA	1,029 genes were influenced by the time of day milk was expressed. Genes implicated include those involved in cell development, growth, proliferation, and cell morphology.

(*continued*)

Table 1. (Continued)

Reference (year)	*Maternal information*	*RNA isolation method*	*Fresh vs. frozen milk*	*Gene product*	*Platform*	*Results*
Maningat et al.[28] (2011)	$n = 5$ healthy women, 18–35 years old, 6–12 weeks postpartum, singleton, term, uncomplicated birth, exclusively breastfeeding, BMI ≤ 27 kg/m^2, standard diet, administration of rhGH; 10 mL	TRizol	Fresh	Global	Microarray of MFG RNA	681 unique gene probes experienced altered expressions following rhGH administration. These networks are involved in cell cycle, DNA replication, recombination and repair, and cancer. Genes influenced by circadian cycles were not altered by rhGH.
Nagatomo et al.[25] (2004)	n = unknown, healthy mothers, healthy term infants. Microarray analysis performed on pooled cDNA of milk cells from colostrum, early, and mature milk; unknown volume	ISOGEN	Fresh	240 Cytokine-related genes	Microarray to first determine highly expressed genes of cytokines/growth factors. Confirmed with RT-PCR, OPN gene expression quantification.	240 cytokine-related genes are highly expressed. OPN ranked highest. OPN mRNA levels in early or mature milk were more than 3 times higher than in colostrum. Late mature milk had the highest OPN mRNA levels of all periods.
Obermeier et al.[23] (2000)	1.5–8 months postpartum; 30 mL	Guanidine thiocyanate method or milk epithelial cells were directly subjected to RT-PCR without RNA isolation.	Fresh	Glucose transporters	Total RNA was prepared and subjected to RT, and nested-PCR was performed. Milk epithelial cells were directly used for RT-PCR without prior RNA isolation.	Cells that are isolated from fresh breastmilk are an acceptable source for investigating gene expression.
Srivastava et al.[17] (1996)	Healthy women delivering preterm (<37 weeks) or term (37–42 weeks); colostrum, transitional, and mature milk, collected by pump or manual expression; unknown volume	Guanidine thiocyanate method	Fresh	Cytokines	RT-PCR	Maternal cells in breastmilk expressed mRNA for MCP-1, IL-8, TGF-β_1, TGF-β_2, M-CSF, IL-6, and IL-1β. mRNA not detected in maternal cells from breastmilk included that for IL-2, IL-10, IFN-γ, and TNF-α. RANTES was weakly expressed. There was variability between individual women, gestational age at delivery, and breastmilk type.

(*continued*)

Table 1. (Continued)

Reference (year)	*Maternal information*	*RNA isolation method*	*Fresh vs. frozen milk*	*Gene product*	*Platform*	*Results*
Takahata et al.[19] (2001)	$n=116$ milk samples; demographics not provided but complications reported; colostrum; unknown volume	MagExtractor MFX-2000	Fresh	IL-18	Semi-QRT-PCR of 7 colostrum samples	There was no correlation between IL-18 mRNA levels and IL-18 protein concentration in milk. There was also not a clear correlation between IL-18 milk mRNA and IL-18 monocyte mRNA.
Takahata et al.[20] (2003)	$n=127$ (preterm $n=39$, term $n=55$); $n=6$ milk samples; colostrum, early, and mature milk; manual breast pump; unknown volume	ISOGEN	Unclear	MIG and IP-10	Semi-QRT-PCR	IP-10 and MIG are expressed in breastmilk. Unclear if this was conducted in preterm/term or different milk types.
Tunzi et al.[18] (2000)	$n=6$ healthy women, 5 days–7 months postpartum; unknown volume	TRizol	Unclear	HBD-1 and HBD-2	RT-PCR	All milk samples expressed the HBD-1 mRNA transcript.

BMI, body mass index; CNT, concentrative nucleoside transporter; ENT, equilibrative nucleoside transporter; GHR, growth hormone receptor; HBD, hemoglobin δ; IFN-γ, interferon-γ; IGF, insulin-like growth factor; IGF1-R, insulin-like growth factor receptor 1; IL, interleukin; IP-10, interferon-γ-inducible protein of 10 kDa; M-CSF, macrophage colony-stimulating factor; MCP1, monocyte chemoattractant protein 1; MEC, mammary epithelial cell; MFG, milk fat globule; MIG, monokine induced by interferon-γ; OCT, organic cation transporter; OPN, osteopontin; PCR, polymerase chain reaction; PEPT2, peptide transporter 2; PTHrP, parathyroid hormone-related protein; QRT, quantitative reverse transcription; RANTES, regulated upon Activation, normal T-cell expressed, and secreted; rhGH, recombinant human growth hormone; RT-PCR, reverse transcription–polymerase chain reaction; TGF-β, transforming growth factor, β; TNF-α, tumor necrosis factor α.

Table 2. Epigenetic Studies That Examine DNA from Breastmilk

Reference (year)	Maternal information	Epigenetic modification	Gene(s) of interest	Platform	Results
Browne et al.[29] (2011)	$n=134$ lactating women with history of biopsy for a suspicious lump; volume varied	Methylation	RASSF1, SFRP1, GSTP1	Bisulfite conversion and PCR amplification	Mean methylation scores for RASSF1 and GSTP1 were higher in women with breast biopsy history.
Qin et al.[32] (2012)	$n=32$ healthy women who delivered term; 3 milk samples: (1) within 10 days of delivery, (2) 2 months after lactation started, and (3) weaning period; unknown volume	Methylation	KLK6	Bisulfite conversion and PCR amplification	KLK6 methylation varied among women and with breastfeed duration. Methylation intensity was not associated with protein level.
Wong et al.[30] (2010)	$n=102$ healthy, exclusive breastfeeders, 19–45 years old, breastfeeding for 30–820 days, no history of breast cancer; at least 10 mL with varied volumes	Methylation	PYCARD, CDH1, GSTP1, RBP1, SFRP1, RASSF1	Pyrosequencing	There was significant relationship between breast cancer risk and those who had a biopsy and methylation scores for CpG sites in CDH1, GSTP1, SFRP1, and RBP1. Regardless of reproductive past, RASSF1 methylation was positively correlated with age.

CDH1, cadherin 1, type 1, E-cadherin (epithelial); GSTP1, glutathione S-transferase π 1; KLK6, kallikrein-related peptidase 6; PCR, polymerase chain reaction; PYCARD, pyrin domain and caspase recruitment domain containing; RASSF1, Ras association domain family member 1; RBP1, retinol binding protein 1, cellular; SFRP1, secreted frizzled-related protein 1.

Use of breastmilk in epigenetic studies

An epigenetic mechanism is a biochemical alteration to the DNA that does not change the sequence but does influence gene expression. This relatively new field has shown great promise in diseases with multifactorial origins because these epigenetic alterations are greatly influenced by the environment. Despite the influence of environmental factors on breastmilk composition, only three studies were found that used an epigenetic approach to examine breastmilk, and all had an oncology focus.

Breastmilk provides a potentially rich source of maternal genetic information. Breastmilk collection is noninvasive, and there is great potential in this practice for individualized screening of breast cancer risk. Two of the epigenetic studies included in this review used methylation analyses to examine promoter regions of tumor suppressor genes known to influence breast cancer risk.[29,30] The remaining epigenetic study examined the methylation of the promoter region of KLK6, which is down-regulated in breast cancers.[31] Wong et al.[30] were able to attain a sufficient quantity of DNA for methylation analyses and used pyrosequencing to attain mean overall methylation concentrations for six genes of interest. Browne et al.[29] also used pyrosequencing to attain methylation concentrations for three genes of interest in women with a history of a breast biopsy. Similarly, pyrosequencing was used to quantify methylation intensities of KLK6 in healthy women at three time points to examine the extent of epigenetic regulation on protein levels at three stages of lactation.[32]

Limitations of gene expression studies that use breastmilk

RNA isolation techniques varied, and some articles completely omitted this important information. One investigator switched the isolation method technique during the study from RNeasy® (Qiagen, Düsseldorf, Germany) to TRizol® (Life Technologies, Carlsbad, CA) due to RNA degradation.[27] Additionally, the use of fresh versus frozen breastmilk should be explicit. Lindquist et al.[16] subjected both fresh and frozen breastmilk to northern blotting and found that the frozen samples experienced β-casein mRNA degradation. Most articles examined in this review used fresh milk that was immediately processed; however, there were a few articles that omitted this important information.[18,20] When samples are placed on ice, RNA degradation occurs and increases the longer samples remain there.[33] One study[24] placed milk samples on ice, although the authors noted that samples were processed within 1 hour. Three studies reported placing milk samples on ice but failed to report the elapsed time between

this and processing.[26–28] Four studies described "immediately" processing milk samples.[17,19,21,25] Five articles did not report any information on sample preservation before processing.[16,18,20,22,23] It is important that the way in which breastmilk samples were preserved and time to processing are reported, as this may impact the RNA quality and final conclusions.

Breastmilk composition and subsequent gene expression are influenced by many factors, including circadian patterns,[26] diet,[1] frequency of pumping,[27] gestational age at delivery,[7] and number of days since delivery. Most of these studies failed to inform the reader about the time of day milk was expressed, the type of pump used, and how frequently the mothers were pumping. Gestational age also contributes to breastmilk composition, as protective factors in preterm breastmilk include an increase in levels of immunoglobulins,[7] fatty acids,[34] and cytokines.[8] Although most of the articles in this review included subjects who were described as healthy and delivered term babies, some included preterm deliveries. The phenotypic definition of preterm was variable: some defined preterm as less than 37 weeks,[17] others defined preterm as less than 36 weeks,[19] whereas others who included preterm deliveries did not provide exclusion criteria.[20] Other factors that influence gene expression and/or breastmilk composition that were largely overlooked include mixed fore and hind milk, maternal drug/alcohol use, differences in pumps used (or manual expression), maternal diet, and omission of maternal demographics, maternal health, and obstetrical complications.

Limitations of epigenetic studies that use breastmilk

The use of both methylated and unmethylated controls should be used when applying methylation analyses to ensure bisulfite-conversion efficiency. Wong et al.[30] and Qin et al.[32] appropriately used both methylated and unmethylated controls, although it is unclear if Browne et al.[29] did the same. It is unclear if Wong et al.[30] collected milk from subjects with a history of a breast biopsy. Wong et al.[30] reported a minimum breastmilk volume for analysis (10 mL) but did not reveal how many women, if any, were unable to produce this volume. Browne et al.[29] did not reveal a minimum breastmilk volume requirement for analysis but did report a range from 56 mL to 86 mL. Qin et al.[32] reported KLK6 methylation is not associated with protein levels; however, DNA isolation was performed on samples from only 32 women, and a large portion of these were from weaning milk for which the time of collection varied greatly. All three epigenetic studies asked subjects to provide breastmilk samples with breast pumps *or* manual expression. The method of milk collection should be identical among subjects because resulting breastmilk volumes vary between methods,[35] potentially implicating both gene expression and epigenetic regulatory mechanisms.

Conclusions

Breastmilk is an appropriate source of RNA or DNA when conducting a gene expression or epigenetic study. Although techniques for RNA isolation and the volume of breastmilk collected varied, the yield was generally sufficient to conduct gene expression studies. If a gene expression approach is used, frozen breastmilk should be avoided because of RNA degradation. Multiple platforms have been applied to examine mRNA in breastmilk, including RT-PCR, microarrays, and western blots. When the epigenome is examined, regardless of tissue source, it is important to use methylated and unmethylated controls to ensure efficient bisulfite conversion. Many lifestyle factors are known to influence breastmilk composition and are therefore contributors to gene methylation and should be accounted for in the analyses.

Only four of the 16 articles used a nonparametric genome-wide approach to examine gene expression. The mechanisms that contribute to breastmilk variability remain elusive; therefore, genome-wide evaluations are needed to better understand breastmilk genetics. Most of the articles examined a specific pathway or physiological process, and few had a direct clinical application. Although physiological and pathway-related studies are important, more clinically relevant studies are needed.

Investigators can utilize the findings from this review to design future genetic or epigenetic studies that use breastmilk. First, research addressing a more standard approach to breastmilk collection and RNA isolation is needed. Steps to mitigate this include reporting volume of breastmilk collected, RNA isolation technique, and RNA yield. Second, research that explores breastmilk genetics in any capacity should report factors known to impact milk composition and subsequent gene expression. A better understanding of the mechanism that underlies breastmilk variability may lead to approaches that optimize donor breastmilk, a practice that is becoming increasingly popular as an increasing number of premature infants become viable because of neonatal advances. Lastly, epigenetic approaches to explore breastmilk show great promise and may provide a way to capture environmental influences on human milk. A molecular examination of breastmilk is applicable to all researchers in lactation science, including behaviorists and traditional bench scientists, as knowledge gained on the mechanisms for milk variability will move the science forward.

Acknowledgments

We are grateful for the assistance of Mary Lou Klem, Ph.D., MLIS, for her guidance in the literature search process. Additionally, we thank Denise Charron-Prochownik, Ph.D., for her help with editing. The author(s) disclose receipt of the following financial support for the research, authorship, and/or publication of this article: National Institute of Nursing Research: Targeted Research and Academic Training Program for Nurses in Genomics (grant T32NR009759) and the Corrine M. Barnes Award.

Disclosure Statement

No competing financial interests exist.

References

1. Peng Y, Zhou T, Wang Q, et al. Fatty acid composition of diet, cord blood and breast milk in Chinese mothers with different dietary habits. *Prostaglandins Leukot Essent Fatty Acids* 2009;81:325–330.
2. Fríguls B, Joya X, García-Algar O, et al. A comprehensive review of assay methods to determine drugs in breast milk and the safety of breastfeeding when taking drugs. *Anal Bioanal Chem* 2010;397:1157–1179.

3. Groër MW, Shelton M. Exercise is associated with elevated proinflammatory cytokines in human milk. *J Obstet Gynecol Neonatal Nurs* 2009;38:35–41.
4. Mitoulas LR, Kent JC, Cox DB, et al. Variation in fat, lactose and protein in human milk over 24 h and throughout the first year of lactation. *Br J Nutr* 2002;88:29–37.
5. Martin JC, Bougnouz P, Antoine JM, et al. Triacylglycerol structure of human colostrum and mature milk. *Lipids* 1993;28:637–643.
6. Harzer G, Haug M, Dieterich I, et al. Changing patterns of human milk lipids in the course of the lactation and during the day. *Am J Clin Nutr* 1983;37:612–621.
7. Araújo ED, Gonçalves AK, Cornetta Mda C, et al. Evaluation of the secretory immunoglobulin A levels in the colostrum and milk of mothers of term and pre-term newborns. *Braz J Infect Dis* 2005;9:357–362.
8. Ustundag B, Yilmaz E, Dogan Y, et al. Levels of cytokines (IL-1beta, IL-2, IL-6, IL-8, TNF-alpha) and trace elements (Zn, Cu) in breast milk from mothers of preterm and term infants. *Mediat Inflamm* 2005;2005:331–336.
9. Bokor S, Koletzko B, Decsi T. Systematic review of fatty acid composition of human milk from mothers of preterm compared to full-term infants. *Ann Nutr Metab* 2007;51:550–556.
10. Elmlinger MW, Hochhaus F, Loui A, et al. Insulin-like growth factors and binding proteins in early milk from mothers of preterm and term infants. *Horm Res* 2007;68:124–131.
11. Hunt CD, Friel JK, Johnson LK. Boron concentrations in milk from mothers of full-term and premature infants. *Am J Clin Nutr* 2004;80:1327–1333.
12. Ballabio C, Bertino E, Coscia A, et al. Immunoglobulin-A profile in breast milk from mothers delivering full term and preterm infants. *Int J Immunopathol Pharmacol* 2007;20:119–128.
13. Blahnik M, Ramanathan R, Riley C, et al. Lipopolysaccharide-induced tumor necrosis factor-alpha and IL-10 production by lung macrophages from preterm and term neonates. *Pediatr Res* 2001;50:726–731.
14. Fituch CC, Palkowetz KH, Goldman AS, et al. Concentrations of IL-10 in preterm human milk and in milk from mothers of infants with necrotizing enterocolitis. *Acta Paediatr* 2004;93:1496–1500.
15. Boutinaud M, Jammes H. Potential uses of milk epithelial cells: A review. *Reprod Nutr Dev* 2002;42:133–147.
16. Lindquist S, Hansoon L, Hernell O, et al. Isolation of mRNA and genomic DNA from epithelial cells in human milk and amplification by PCR. *Biotechniques* 1994;17:692–696.
17. Srivastava M, Srivastava A, Brouhard B, et al. Cytokines in human milk. *Res Commun Mol Pathol Pharmacol* 1996;93:263–287.
18. Tunzi C, Harper PA, Bar-oz B, et al. β-Defensin expression in human mammary gland epithelia. *Pediatr Res* 2000;48:30–35.
19. Takahata Y, Hidetoshi T, Nomura A, et al. Interleukin-18 in human milk. *Pediatr Res* 2001;50:268–272.
20. Takahata Y, Takada H, Nomura A, et al. Detection of interferon-gamma-inducible chemokines in human milk. *Acta Paediatr* 2003;92:659–665.
21. Frankenberger M, Schwaeble W, Ziegler-Heitbrock L. Expression of M-ficolin in human monocytes and macrophages. *Mol Immunol* 2008;45:1424–1430.
22. Andersson Y, Lindquist S, Bergstrom S, et al. Three variants of parathyroid hormone-related protein messenger RNA are expressed in human mammary gland. *Pediatr Res* 1996;41:380–383.
23. Obermeier S, Huselweh B, Tinel H, et al. Expression of glucose transporters in lactating human mammary gland epithelial cells. *Eur J Nutr* 2000;3:194–200.
24. Alcorn J, Lu X, Moscow A, et al. Transporter gene expression in lactating and nonlactating human mammary epithelial cells using real-time reverse transcription-polymerase chain reaction. *J Pharmacol Exp Ther* 2002;303:487–496.
25. Nagatomo T, Ohga S, Takada H, et al. Microarray analysis of human milk cells: Persistent high expression of osteopontin during the lactation period. *Clin Exp Immunol* 2004;138:47–53.
26. Maningat PD, Sen P, Rijnkels M, et al. Gene expression in the human mammary epithelium during lactation: The milk fat globule transcriptome. *Physiol Genomics* 2009;37:12–22.
27. Maningat PD, Sen P, Sunehag AL, et al. Regulation of gene expression in human mammary epithelium: Effect of breast pumping. *J Endocrinol* 2007;195:503–511.
28. Maningat P, Sen P, Rijnkels M, et al. Short-term administration of rhGH increases markers of cellular proliferation but not milk protein gene expression in normal lactating women. *Physiol Genomics* 2011;43:381–391.
29. Browne EP, Punska EC, Lenington S, et al. Increased promoter methylation in exfoliated breast epithelial cells in women with a previous breast biopsy. *Epigenetics* 2011;6:1425–1435.
30. Wong CM, Anderton DL, Smith-Schneider S, et al. Quantitative analysis of promoter methylation in exfoliated epithelial cells isolated from breast milk of healthy women. *Epigenetics* 2010;5:1–11.
31. Yousef GM, Yacoub GM, Polymeris ME, et al. Kallikrein gene downregulation in breast cancer. *Br J Cancer* 2004;90:167–172.
32. Qin W, Zhang K, Kliethermes B, et al. Differential expression of cancer associated proteins in breast milk based on age at first full term pregnancy. *BMC Cancer* 2012;12:100.
33. Ibberson D, Benes V, Muckenthaler M, et al. RNA degradation compromises the reliability of microRNA expression profiling. *BMC Biotechnol* 2009;9:102.
34. Jackson MB, Lammi-Keefe CJ, Jensen RG, et al. Total lipid and fatty acid composition of milk from women with and without insulin-dependent diabetes mellitus. *Am J Clin Nutr* 1994;60:353–361.
35. Paul VK, Singh M, Deorari AK, et al. Manual and pump methods of expression of breast milk. *Indian J Pediatr* 1996;63:87–92.

Address correspondence to:
Kelley L. Baumgartel, BSN
University of Pittsburgh
3500 Victoria Street
440 Victoria Building
Pittsburgh, PA 15261

E-mail: klb134@pitt.edu

APPENDIX H

LICENSE AGREEMENT FOR MANUSCRIPT #3

Ballen, Karen <KBallen@liebertpu Mark as unread

Fri 4/17/2015 4:42 PM

Inbox

Dear Kelley:

Copyright permission is granted for this request to use your article from BREASTFEEDING MEDICINE in your dissertation.

Kind regards,

Karen Ballen
Manager

APPENDIX I

MANUSCRIPT #4: WIC (THE SUPPLEMENTAL NUTRITION PROGRAM FOR WOMEN, INFANTS, AND CHILDREN): POLICY VERSUS PRACTICE REGARDING BREASTFEEDING

HHS Public Access

Author manuscript

Nurs Outlook. Author manuscript; available in PMC 2015 March 11.

Published in final edited form as:

Nurs Outlook. 2013 ; 61(6): 466–470. doi:10.1016/j.outlook.2013.05.010.

WIC (The Special Supplemental Nutrition Program for Women, Infants, and Children): Policy versus practice regarding breastfeeding

Kelley L. Baumgartel, BSN[a], **Diane L. Spatz, PhD, RN-BC, FAAN**[b,*], and **American Academy of Nursing Expert Breastfeeding Panel**

[a]University of Pittsburgh School of Nursing, Pittsburgh, PA

[b]University of Pennsylvania School of Nursing & The Children's Hospital of Philadelphia, Philadelphia, PA

Abstract

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) provides foods, education, and referrals to participants who are considered to be at nutritional risk. The outreach of the program is impressive, and nearly 9.17 million people participated in the program in 2010. WIC participation is associated with many positive outcomes, including improved birthweights and childhood dietary practices. Despite these benefits, WIC mothers experience lower breastfeeding rates when compared with demographically similar women who do not participate in the WIC program. According to WIC, "A breastfeeding mother and her infant shall be placed in the highest priority level." Despite this statement and others that support breastfeeding, WIC allocates only 0.6% of its budget toward breastfeeding initiatives. Formula expenses accounted for 11.6% ($850 million) of WIC's 2009 expenses. The inconsistency between WIC's policies that encourage breastfeeding vs. practices that favor formula begs further examination. Research shows consistent success with peer counseling programs among WIC participants; however, little money is budgeted for these programs. Rebates included, WIC spends 25 times more on formula than on breastfeeding initiatives. The American Academy of Nursing Expert Panel on Breastfeeding is calling for a re-evaluation of how these taxpayer dollars are spent. Additionally, the American Academy of Nursing recommends a shift from formula bargaining to an investment in structured peer counseling programs. All WIC programs should offer peer counseling support services that encourage breastfeeding and meet the needs of the families they serve.

Keywords

WIC; Breastfeeding; Peer counseling; Formula

Approximately $3.6 billion would be saved if breast-feeding rates were increased to the U.S. Surgeon General's recommendations (Weimer, 2001). Breast milk provides infants with

* Corresponding author: Dr. Diane L. Spatz, 418 Curie Boulevard, Philadelphia, PA 19104. spatz@nursing.upenn.edu (D.L. Spatz).

immunologic, developmental, psychological, and nutritional benefits that prevent illness and optimize the health of our nation's children. The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) serves a large number of people who are at nutritional risk. The policies set forth by the program appear to promote breastfeeding; however, funding points to a practice that favors formula. This article outlines WIC policy vs. WIC practice surrounding infant feeding. Specific attention is made to the lack of peer counseling support offered to WIC participants, despite their consistent performance of improved breastfeeding initiation and duration (Gross, 2009; Kistin, Abramson, & Dublin, 1994; Yun et al., 2009). Lastly, recommendations, presented on behalf of the American Academy of Nursing (AAN) Expert Panel on Breastfeeding include budget re-evaluation that improves funding to peer counseling programs.

WIC: A Program Overview and Its Breastfeeding Policy

In 2010 WIC provided services to 9.17 million people. The population served includes low-income pregnant, postpartum, and breastfeeding women and infants and children up to 5 years who are at nutrition risk. This is a federal program, with funding administered by the United States Department of Agriculture's Food and Nutrition Service. Foods and services provided by WIC to participants at no cost include WIC foods, nutrition education, breastfeeding promotion and support, and administrative costs. Potential participants must be state residents, meet income guidelines (at or below 185% of the U.S. Poverty Income Guidelines), and be determined to be at "nutrition risk" by a health professional (United States Department of Agriculture [USDA], 2012a, 2012b, 2012c). This article focuses on WIC participants who are low-income pregnant, post-partum, and breastfeeding women and children up to age 5 who are at nutrition risk. For purposes of this article, this population is considered vulnerable to health outcomes that can otherwise be prevented with adequate breastfeeding support.

WIC participation among pregnant women is positively associated with gestational age and mean birth weight. Additionally, WIC participation is associated with improved and enriched childhood diets that are higher in iron, contain fewer sugars, and have an overall greater food variety. WIC participation is also associated with greater use of preventative and restorative dental care services. Although breast-feeding rates are now trending upward, WIC participants have not benefited from this improvement at the same rate as nonparticipants (Ryan & Zhou, 2006). WIC appears to be an ideal platform for lactation promotion for a vulnerable population with historically low breastfeeding rates. Statements from WIC appear to support the following agenda: "WIC recognizes and promotes breastfeeding as the optimal source of nutrition for infants" (USDA, 2012b).

Despite this ostensible support for breastfeeding mothers, the breastfeeding rate of WIC participants is at least 20% lower than non-WIC participants (Lawrence, 2006). Mothers not enrolled in WIC are more than twice as likely to breastfeed at 6 months (Ryan & Zhou, 2006). This inequality, although always present, is now becoming more divided. In 1984, a non-WIC mother was 1.41 times more likely to breastfeed than a WIC mother (Ryan, Rush, Krieger, & Lewandowski, 1991). Currently, non-WIC participants are 2.11 times more likely to breastfeed than WIC participants (Ryan & Zhou, 2006). Additionally, families who

were income eligible for WIC but not participants experienced higher breastfeeding rates than their WIC-participating counterparts (Li, Darling, Maurice, Barker, & Grummer-Strawn, 2005).

According to WIC, "A breastfeeding mother and her infant shall be placed in the highest priority level" (USDA, 2012c). It remains unclear then why only $34 million or 0.6% of the total WIC budget is designated for breast-feeding initiatives (Lawrence, 2006). Meanwhile, formula accounted for $850 million (11.6%) of WIC's fiscal year 2009 expenses (Neuberger, 2010). Approximately 1 of 9 WIC participants are pregnant or breastfeeding mothers according to the National WIC Association's breastfeeding strategic plan; yet, 44% of all food items purchased through WIC is infant formula (USDA, 2012a). WIC reports cost savings in the form of rebates from formula manufacturers in exchange for their business and that they are legally required to bid for contracts with formula makers. This exchange allows WIC to serve more women (USDA, 2012b). Unfortunately, formula companies are the primary beneficiary of this practice because WIC purchases account for more than half of all the infant formula sold in the United States (Neuberger, 2010). Formula companies have capitalized on this business exchange and submit bids for the pricier formulas (Marcus, 2010). Formulas submitted for consideration contain additives that, according to manufacturers, optimize formula to more closely resemble human milk. As a result, an additional cost of $91 million dollars is spent yearly on additive-fueled formula that is provided at no cost to some of our nation's most vulnerable infants (Marcus, 2010).

Peer Counseling

An inexpensive and widely accepted approach to alleviate poor breastfeeding rates is peer counseling. Peer counseling is a community-driven public health practice that has consistently improved breastfeeding rates for WIC participants, including black (Caulfield et al., 1998) and adolescent populations (Volpe & Bear, 2000; Wambach et al., 2010). Peer counseling provides one- on-one support by mothers who have breastfed for at least 6 months, though group counseling programs are also common. In 2004, Best Start Social Marketing, Tampa, FL released results of a peer counseling program model that was contracted with the Department of Agriculture's Food and Nutrition Service to develop a breastfeeding peer counseling program model. This extensive project included a literature review, an assessment of current practices, and semistructured telephone interviews of WIC staff and WIC peer counselors. This project was specifically designed to meet the needs of WIC participants and staff to implement and expand breastfeeding peer counseling programs. Many barriers were identified as a result of the Best Start Program, including but not limited to: insufficient resources for program initiation, funding stream discontinuity, inadequate or lack of compensation for counselors, and lack of program structure consistency. Overall, the lack of funding was a common thread underlying many of the obstacles identified that prevented the implementation and sustainability of a peer counselor program (Best Start Social Marketing, 2004). Walker and Avis (1999) support these conclusions in their review of peer education and state that peer counseling programs often fail because of numerous factors including an absence of defined program goals, insufficient training, and inadequate funding. When used well, peer counselors are seen as "... filling a unique and vital role in the WIC program" (Best Start Social Marketing, 2004).

WIC participants who receive peer counseling support experience higher breastfeeding rates. Gross (2009) conducted a cross-sectional study that examined Maryland WIC participant breastfeeding initiation rates in three groups: peer counselor, lactation consultant, and standard care. Breastfeeding initiation was significantly higher among those who received peer counseling, but not in the lactation consultant or standard care group. Peer counselors were trained using an International Board Certified Lactation Consultant (IBCLC)-designed curriculum, and Maryland protocol requires that peer counselors contact all pregnant clients upon referral, at 1 month and 2 weeks before their due date, and near their delivery date. Postpartum contact is structured, and breastfeeding clients are called regularly.

In a large (N = 328) randomized controlled trial, Pugh et al. (2010) showed that breastfeeding rates could be improved in the WIC population using a nurse-peer counselor model. WIC mothers were randomized to an intervention or usual care group and followed for 24 weeks. The intervention was performed by a community health nurse and peer counselor (breast-feeding support team) who provided hospital and home visits, telephone support, and 24-hour pager access. Results showed statistically significant higher breastfeeding rates at 6 weeks postpartum (intervention group = 66.7% breastfeeding vs. usual care = 56.9%; odds ratio = 1.71 [95% confidence interval]). In addition, breastfeeding at 12 weeks was higher in the intervention group (49.4%) vs. the usual care group (40.6%), although this was not statistically significant.

Yun et al.'s (2009) retrospective study also uncovered a positive response to peer counselors among WIC participants in Missouri. Breastfeeding initiation rates among WIC agencies that provided peer counseling were higher than agencies without peer counseling programs. In peer counseling agencies, participation length was positively associated with the likelihood of initiation. Gross et al. (2009) and Yun et al. (2009) both revealed a positive impact of peer counseling programs on breastfeeding initiation among WIC participants. Despite this benefit, a survey conducted by Evans, Labbok, and Abrahams (2011) distributed to WIC directors in North Carolina uncovered a racial/ethnic disparity in breastfeeding rates and support services available. WIC offices located in areas with a higher black population were significantly less likely to provide breastfeed support services, including peer counseling. Although the study is older, Kistin, Abramson, and Dublin (1994) showed that black, urban, low-income women experienced a breast-feeding duration longer than 6 weeks among women with a peer counselor compared with 28% among those without peer counselor support.

Review of 4 WIC peer counselor programs revealed that 24% of counselors received no monetary compensation (Bronner, Barber, Vogelhut, & Resnik, 2001). Additionally, fundamental components of peer counselor programs were absent; there were inconsistent policies, a failure to match counselor demographics with new mothers, and an inability to provide adequate counselor training programs. Both Best Start and Bronner et al. (2001) revealed that most of the barriers identified by Walker and Avis (1999) still exist, preventing the successful implementation of breast-feeding peer counselor programs. Although WIC acknowledges the value of peer counselors and frequently cites their potential use in a position paper (National WIC Association Position Paper, 2012), there is no mention of counselors in their recently published and detailed Breastfeeding Strategic Plan (National

WIC Association National Breastfeeding Strategic Plan, 2012). Perhaps most concerning is that, despite the advantages of peer counselor programs on breast-feeding among disadvantaged women, only 16.7% of WIC service delivery sites offer this support (Walker & Avis, 1999).

Discussion

WIC provides a vital public health nutrition service and should be a safe, supportive, and proactive venue for breastfeeding. Unfortunately, breastfeeding initiatives for the most vulnerable dyads are grossly outspent by profitable formula corporations. Despite research that has revealed consistent success with peer counseling programs, WIC allocates little money to sustain these successful programs. Rebates included, WIC spends 25 times more on formula than on breastfeeding initiatives for mothers who experience some of the lowest breastfeeding rates and subsequent infant health consequences.

Structural barriers to breastfeeding exist that do not implicate WIC. The potential for breastfeeding success is optimized when new mothers have the support of their partner and workplace. These support systems are often absent for low-income women, and poor breastfeeding rates persist. The Temporary Aid to Needy Family Program requires that those who receive welfare benefits be employed, often at hourly or entry-level jobs. Employers are often not amenable to providing facilities for breastfeeding mothers. As a result, many new mothers decide against breastfeeding because they view breastfeeding while employed as stressful. Factors regarded as ideal for-breastfeeding success include a private space with a locking door, time to pump, and adequate storage facilities (Stewart-Glenn, 2008). Entry-level or hourly jobs often do not provide adequate breastfeeding support, and new mothers are frequently without pumping facilities or a place to store milk. The recent passage of Section 4207 of the Patient Protection and Affordable Care Act is a good start to support new mothers because it requires employers to provide "... reasonable unpaid break time and a private, non-bathroom place for non-exempt employees who are nursing mothers to express milk during the work day" (United States Breastfeeding Committee, 2008).

Summary and Recommendations

WIC provides valuable services to a large portion of our country's population who are at nutritional risk. Despite other positive health outcomes, breastfeeding rates have not increased among WIC participants. Breastfeeding is a public health issue that should be a targeted lifestyle practice because it improves health outcomes and minimizes health cost spending. Despite numerous WIC policy statements that support breastfeeding, funding is overwhelmingly spent on formula with only a small fractional portion allocated toward peer counseling programs. The evidence is clear that peer counseling programs are an economically feasible option for providing breastfeeding support, and the implementation of such programs is associated with improved breastfeeding.

The AAN urges our government partners to re-evaluate how taxpayer dollars are spent. Ferguson (2001) addresses the unique perspective nurses contribute to healthcare policy development. Nursing experts are instrumental in policy change (Ferguson, 2001), and the

AAN Expert Panel on Breastfeeding is comprised of lactation experts who firmly believe that WIC is medically and ethically obliged to improve breastfeeding efforts. We strongly recommend that the funding source for WIC state agencies, the Food and Nutrition Service, re-evaluate money allocation and consider mandates to ensure that all WIC programs have a robust and structured peer counseling program. The AAN Expert Panel on Breastfeeding challenges the traditional formula contractual obligations in favor of investment in peer counseling programs that would result in increased breastfeeding rates. A change in funding allocation and subsequent WIC practice is needed to meet the breastfeeding needs of the vulnerable families they serve.

Acknowledgments

Supported by the National Institute of Nursing Research (T32NR009759) and the Corrine M. Barnes Award.

REFERENCES

Best Start Social Marketing. Using Loving Support© to Implement best practices in peer counseling: Final research brief. 2004. FNS contract 58-3198-1-050

Bronner Y, Barber T, Vogelhut J, Resnik A. Breastfeeding peer counseling: Results from the national WIC survey. Journal of Human Lactation. 2001; 17(2):119–125. [PubMed: 11847825]

Calfield L, Gross S, Bentley M, Bronner Y, Kessler L, Jensen J, Paige D. WIC-based interventions to promote breastfeeding among African-American Women in Baltimore: Effects of breastfeeding initiation and continuation. Journal of Human Lactation. 1998; 14(1):15–22. [PubMed: 9543954]

Evans K, Labbok M, Abrahams SW. WIC and breastfeeding support services: Does the mix of services offered vary with race and ethnicity? Breastfeeding Medicine. 2011; 6(6):401–406. [PubMed: 21453123]

Ferguson SL. An activist looks at nursing's role in health policy development. Journal of Obstetric, Gynecologic, and Neonatal Nursing. 2001; 30(5):546–551.

Gross S, Resnik A, Cross-Barnet C, Nanda J, Augustyn M, Paige D. The differential impact of WIC peer counseling programs on breastfeeding initiation across the state of Maryland. Journal of Human Lactation. 2009; 25(4):435–443. [PubMed: 19652195]

Kistin N, Abramson R, Dublin P. Effect of peer counselors on breastfeeding initiation, exclusivity, and duration among low-income urban women. Journal of Human Lactation. 1994; 10(1):11–15. [PubMed: 7619241]

Lawrence R. Lower breastfeeding rates among supplemental nutrition program for women, infants, and children participants: A call for action. Pediatrics. 2006; 117:1432–1433. [PubMed: 16585343]

Li R, Darling N, Maurice E, Barker L, Grummer-Strawn L. Breastfeeding rates in the United States by characteristics of the child, mother, or family: The 2002 National Immunization Survey. Pediatrics. 2005; 115:31–37.

Marcus, R. Lobbying fight over infant formula highlights budget gridlock. The Washington Post. 2010. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2010/07/13/AR2010071304634.html

National WIC Association Position Paper. Retrieved from http://www.paramountcommunication.com/nwica/BreastfeedingPromotionandSupportintheWICProgram.pdf

National WIC Association National Breastfeeding Strategic Plan. Retrieved rom http://www.paramountcommunication.com/nwica/NationalBreastfeedingStrategicPlan.pdf

Neuberger, Z. WIC food package should be based on science: Foods with new functional ingredients should be provided only if they deliver health or nutritional benefits. Center on Budget and Policy Priorities; Washington, DC: 2010.

Pugh LC, Serwint JR, Frick KD, Nanda JP, Sharps PW, Spatz DL, Milligan RA. A randomized controlled community-based trial to improve breastfeeding rates among urban low-income mothers. Academic Pediatrics. 2010; 10(1):14–20. [PubMed: 19854119]

Ryan A, Rush D, Krieger F, Lewandowski G. Recent declines in breastfeeding in the United States. Pediatrics. 1991; 88:719–727. [PubMed: 1896274]

Ryan A, Zhou W. Lower breastfeeding rates persist among the special supplemental nutrition program for women, infants and children participants, 1978-2003. Pediatrics. 2006; 117:1136–1146. [PubMed: 16585308]

Stewart-Glenn J. Knowledge, perceptions, and attitudes of managers, coworkers, and employed breastfeeding mothers. American Association of Occupational Health Nurses Journal. 2008; 56(10):423–429.

United States Breastfeeding Committee. Protect, Promote, and Support Breastfeeding Through Legislation. 2008. Retrieved from http://www.usbreastfeeding.org/Portals/0/Publications/State-Legislation-White-Paper-2008-USBC.pdf

USDA. Food and Nutrition Service. 2012a. Retrieved from http://www.fns.usda.gov/ora/WICFoodCosts/FY2005/FY2005.pdf

USDA. Nutrition Program Facts Food and Nutrition Services. 2012b. Retrieved from http://www.fns.usda.gov/wic/wic-fact-sheet.pdf

USDA. Food and Nutrition Service. 2012c. Retrieved from http://www.fns.usda.gov/wic/breastfeeding/bfrequirements.HTM

Volpe E, Bear M. Enhancing breastfeeding initiation in adolescent mothers through the Breastfeeding Educated and Supported Teen (BEST) club. Journal of Human Lactation. 2000; 16(3):196–200. [PubMed: 11153152]

Walker SA, Avis M. Common reasons why peer education fails. Journal of Adolescence. 1999; 22:573–577. [PubMed: 10469520]

Wambach K, Aaronson L, Breedlove G, Domian EW, Rojjanasrirat W, Yeh HW. A randomized controlled trial of breastfeeding support and education for adolescent mothers. Western Journal of Nursing Research. 2010; 33(4):486–505. [PubMed: 20876551]

Weimer, J. The economic benefits of breastfeeding: A review and analysis (Food AssistanceNutr Res Rep No. [FANRR13]:20). USDA; 2001.

Yun S, Qian L, Mertzlufft K, Kruse C, White M, Fuller P, Zhu B. Evaluation of the Missouri WIC (Special Supplemental Nutrition Program for Women, Infants and Children) breast-feeding peer counseling programme. Public Health Nutrition. 2009; 13(2):229–237. [PubMed: 19607746]

APPENDIX J

LICENSE AGREEMENT FOR MANUSCRIPT #4

ELSEVIER LICENSE
TERMS AND CONDITIONS

Apr 16, 2015

This is a License Agreement between Kelley L Baumgartel ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Kelley L Baumgartel
Customer address	1619 Three Degree Rd MARS, PA 16046
License number	3610860092416
License date	Apr 16, 2015
Licensed content publisher	Elsevier
Licensed content publication	Nursing Outlook
Licensed content title	WIC (The Special Supplemental Nutrition Program for Women, Infants, and Children): Policy versus practice regarding breastfeeding
Licensed content author	None
Licensed content date	November–December 2013
Licensed content volume number	61
Licensed content issue number	6
Number of pages	5
Start Page	466
End Page	470
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No

Title of your thesis/dissertation	The Impact of Promoter Polymorphisms on Cytokine Concentration in Preterm Breast Milk and Subsequent Infant Outcomes
Expected completion date	May 2015
Estimated size (number of pages)	100
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described

in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or

enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
    - via their non-commercial person homepage or blog
    - by updating a preprint in arXiv or RePEc with the accepted manuscript
    - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
    - directly by providing copies to their students or to research collaborators for their personal use
    - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
    - via non-commercial hosting platforms such as their institutional repository
    - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final

record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third

party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access

BIBLIOGRAPHY

American Academy of Pediatrics, A. (2012). Breastfeeding and the Use of Human Milk. *Pediatrics, 115*(2), 496.

Araujo, E. D., Goncalves, A. K., Cornetta Mda, C., Cunha, H., Cardoso, M. L., Morais, S. S., & Giraldo, P. C. (2005). Evaluation of the secretory immunoglobulin A levels in the colostrum and milk of mothers of term and pre-term newborns. *Braz J Infect Dis, 9*(5), 357-362. doi: /S1413-86702005000500002

Arrandale, V., Koehoorn, M., NacNaby, Y., Kennedy, S. (2006). How to use SAS® Proc Traj and SAS® Proc Glimmix in Respiratory Epidemiology. In U. o. B. Columbia (Ed.).

Arthur, P. G., Kent, J. C., & Hartmann, P. E. (1994). Metabolites of lactose synthesis in milk from diabetic and nondiabetic women during lactogenesis II. *J Pediatr Gastroenterol Nutr, 19*(1), 100-108.

Aydemir, O., Aydemir, C., Sarikabadayi, Y. U., Emre Canpolat, F., Erdeve, O., Biyikli, Z., & Dilmen, U. (2012). Fecal calprotectin levels are increased in infants with necrotizing enterocolitis. *J Matern Fetal Neonatal Med, 25*(11), 2237-2241. doi: 10.3109/14767058.2012.684172

Bachour, P., Yafawi, R., Jaber, F., Choueiri, E., & Abdel-Razzak, Z. (2012). Effects of smoking, mother's age, body mass index, and parity number on lipid, protein, and secretory immunoglobulin A concentrations of human milk. *Breastfeed Med, 7*(3), 179-188. doi: 10.1089/bfm.2011.0038

Baier, R. J. (2006). Genetics of perinatal brain injury in the preterm infant. *Front Biosci, 11*, 1371-1387.

Ballabio, C., Bertino, E., Coscia, A., Fabris, C., Fuggetta, D., Molfino, S., . . . Restani, P. (2007). Immunoglobulin-A profile in breast milk from mothers delivering full term and preterm infants. *Int J Immunopathol Pharmacol, 20*(1), 119-128.

Banks, W. A., Kastin, A. J., & Broadwell, R. D. (1995). Passage of cytokines across the blood-brain barrier. *Neuroimmunomodulation, 2*(4), 241-248.

Barbosa, L., Butte, N. F., Villalpando, S., Wong, W. W., & Smith, E. O. (1997). Maternal energy balance and lactation performance of Mesoamerindians as a function of body mass index. *Am J Clin Nutr, 66*(3), 575-583.

Barsam, F. J., Borges, G. S., Severino, A. B., de Mello, L. M., da Silva, A. S., & Nunes, A. A. (2013). Factors associated with community-acquired pneumonia in hospitalised children and adolescents aged 6 months to 13 years old. *Eur J Pediatr, 172*(4), 493-499. doi: 10.1007/s00431-012-1909-z

Berenhauser, A. C., Pinheiro do Prado, A. C., da Silva, R. C., Gioielli, L. A., & Block, J. M. (2012). Fatty acid composition in preterm and term breast milk. *Int J Food Sci Nutr, 63*(3), 318-325. doi: 10.3109/09637486.2011.627843

Bhandari, V., Buhimschi, C. S., Han, C. S., Lee, S. Y., Pettker, C. M., Campbell, K. H., . . . Buhimschi, I. A. (2011). Cord blood erythropoietin and interleukin-6 for prediction of intraventricular hemorrhage in the preterm neonate. *J Matern Fetal Neonatal Med, 24*(5), 673-679. doi: 10.3109/14767058.2010.520048

Bisquera, J. A., Cooper, T. R., & Berseth, C. L. (2002). Impact of necrotizing enterocolitis on length of stay and hospital charges in very low birth weight infants. *Pediatrics, 109*(3), 423-428.

Bosviel, R., Garcia, S., Lavediaux, G., Michard, E., Dravers, M., Kwiatkowski, F., . . . Bernard-Gallon, D. J. (2012). BRCA1 promoter methylation in peripheral blood DNA was identified in sporadic breast cancer and controls. *Cancer Epidemiol, 36*(3), e177-182. doi: 10.1016/j.canep.2012.02.001

Bottcher, M. F., Jenmalm, M. C., Garofalo, R. P., & Bjorksten, B. (2000). Cytokines in breast milk from allergic and nonallergic mothers. *Pediatr Res, 47*(1), 157-162.

Browne, E. P., Punska, E. C., Lenington, S., Otis, C. N., Anderton, D. L., & Arcaro, K. F. (2011). Increased promoter methylation in exfoliated breast epithelial cells in women with a previous breast biopsy. *Epigenetics, 6*(12), 1425-1435. doi: 10.4161/epi.6.12.18280

Cabantous, S., Poudiougou, B., Oumar, A. A., Traore, A., Barry, A., Vitte, J., . . . Dessein, A. J. (2009). Genetic evidence for the aggravation of Plasmodium falciparum malaria by interleukin 4. *J Infect Dis, 200*(10), 1530-1539. doi: 10.1086/644600

Calhoun, D. A., Lunoe, M., Du, Y., Staba, S. L., & Christensen, R. D. (1999). Concentrations of granulocyte colony-stimulating factor in human milk after in vitro simulations of digestion. *Pediatr Res, 46*(6), 767-771.

Campeotto, F., Kalach, N., Lapillonne, A., Butel, M. J., Dupont, C., & Kapel, N. (2007). Time course of faecal calprotectin in preterm newborns during the first month of life. *Acta Paediatr, 96*(10), 1531-1533. doi: 10.1111/j.1651-2227.2007.00457.x

Capasso, M., Avvisati, R. A., Piscopo, C., Laforgia, N., Raimondi, F., de Angelis, F., & Iolascon, A. (2007). Cytokine gene polymorphisms in Italian preterm infants: association between interleukin-10 -1082 G/A polymorphism and respiratory distress syndrome. *Pediatr Res, 61*(3), 313-317. doi: 10.1203/pdr.0b013e318030d108

Carlson, S. E., Werkman, S. H., Peeples, J. M., Cooke, R. J., & Tolley, E. A. (1993). Arachidonic acid status correlates with first year growth in preterm infants. *Proc Natl Acad Sci U S A, 90*(3), 1073-1077.

Carroll, J. F., Franks, S. F., Smith, A. B., & Phelps, D. R. (2009). Visceral adipose tissue loss and insulin resistance 6 months after laparoscopic gastric banding surgery: a preliminary study. *Obes Surg, 19*(1), 47-55. doi: 10.1007/s11695-008-9642-4

Catalano, P. M., Farrell, K., Thomas, A., Huston-Presley, L., Mencin, P., de Mouzon, S. H., & Amini, S. B. (2009). Perinatal risk factors for childhood obesity and metabolic dysregulation. *Am J Clin Nutr, 90*(5), 1303-1313. doi: 10.3945/ajcn.2008.27416

Christensen, R. D., Jensen, J., Maheshwari, A., & Henry, E. (2010). Reference ranges for blood concentrations of eosinophils and monocytes during the neonatal period defined from over 63 000 records in a multihospital health-care system. *J Perinatol, 30*(8), 540-545. doi: 10.1038/jp.2009.196

Cochery-Nouvellon, E., Nguyen, P., Attaoua, R., Cornillet-Lefebvre, P., Mercier, E., Vitry, F., & Gris, J. C. (2009). Interleukin 10 gene promoter polymorphisms in women with pregnancy loss: preferential association with embryonic wastage. *Biol Reprod, 80*(6), 1115-1120. doi: 10.1095/biolreprod.108.072215

Collado, M. C., Laitinen, K., Salminen, S., & Isolauri, E. (2012). Maternal weight and excessive weight gain during pregnancy modify the immunomodulatory potential of breast milk. *Pediatr Res, 72*(1), 77-85. doi: 10.1038/pr.2012.42

Connor Gorber, S., Tremblay, M., Moher, D., & Gorber, B. (2007). A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev, 8*(4), 307-326. doi: 10.1111/j.1467-789X.2007.00347.x

Corpeleijn, W. E., Kouwenhoven, S. M., Paap, M. C., van Vliet, I., Scheerder, I., Muizer, Y., . . . Vermeulen, M. J. (2012). Intake of own mother's milk during the first days of life is associated with decreased morbidity and mortality in very low birth weight infants during the first 60 days of life. *Neonatology, 102*(4), 276-281. doi: 10.1159/000341335

Cristofalo, E. A., Schanler, R. J., Blanco, C. L., Sullivan, S., Trawoeger, R., Kiechl-Kohlendorfer, U., . . . Abrams, S. (2013). Randomized trial of exclusive human milk versus preterm formula diets in extremely premature infants. *J Pediatr, 163*(6), 1592-1595 e1591. doi: 10.1016/j.jpeds.2013.07.011

Currie, A. J., Curtis, S., Strunk, T., Riley, K., Liyanage, K., Prescott, S., . . . Burgner, D. (2011). Preterm infants have deficient monocyte and lymphocyte cytokine responses to group B streptococcus. *Infect Immun, 79*(4), 1588-1596. doi: 10.1128/IAI.00535-10

Cury, D. B., Mizsputen, S. J., Versolato, C., Miiji, L. O., Pereira, E., Delboni, M. A., . . . Moss, A. C. (2013). Serum calprotectin levels correlate with biochemical and histological markers of disease activity in TNBS colitis. *Cell Immunol, 282*(1), 66-70. doi: 10.1016/j.cellimm.2013.04.004

Dabritz, J., Jenke, A., Wirth, S., & Foell, D. (2012). Fecal phagocyte-specific S100A12 for diagnosing necrotizing enterocolitis. *J Pediatr, 161*(6), 1059-1064. doi: 10.1016/j.jpeds.2012.06.003

Dempsey, E., & Miletin, J. (2010). Banked preterm versus banked term human milk to promote growth and development in very low birth weight infants. *Cochrane Database Syst Rev*(6), CD007644. doi: 10.1002/14651858.CD007644.pub2

Engel, S. A., Olshan, A. F., Savitz, D. A., Thorp, J., Erichsen, H. C., & Chanock, S. J. (2005). Risk of small-for-gestational age is associated with common anti-inflammatory cytokine polymorphisms. *Epidemiology, 16*(4), 478-486.

Erbagci, A. B., Cekmen, M. B., Balat, O., Balat, A., Aksoy, F., & Tarakcioglu, M. (2005). Persistency of high proinflammatory cytokine levels from colostrum to mature milk in preeclampsia. *Clin Biochem, 38*(8), 712-716. doi: 10.1016/j.clinbiochem.2005.05.004

Etem Piskin, I., Nur Karavar, H., Arasli, M., & Ermis, B. (2012). Effect of maternal smoking on colostrum and breast milk cytokines. *Eur Cytokine Netw, 23*(4), 187-190. doi: 10.1684/ecn.2013.0324

Ewaschuk, J. B., Unger, S., O'Connor, D. L., Stone, D., Harvey, S., Clandinin, M. T., & Field, C. J. (2011). Effect of pasteurization on selected immune components of donated human breast milk. *J Perinatol, 31*(9), 593-598. doi: 10.1038/jp.2010.209

Fetherston, C. M., Wells, J. I., & Hartmann, P. E. (2006). Severity of mastitis symptoms as a predictor of C-reactive protein in milk and blood during lactation. *Breastfeed Med, 1*(3), 127-135. doi: 10.1089/bfm.2006.1.127

Fields, D. A., & Demerath, E. W. (2012). Relationship of insulin, glucose, leptin, IL-6 and TNF-alpha in human breast milk with infant growth and body composition. *Pediatr Obes, 7*(4), 304-312. doi: 10.1111/j.2047-6310.2012.00059.x

Fituch, C. C., Palkowetz, K. H., Goldman, A. S., & Schanler, R. J. (2004). Concentrations of IL-10 in preterm human milk and in milk from mothers of infants with necrotizing enterocolitis. *Acta Paediatr, 93*(11), 1496-1500. doi: 10.1080/08035250410022314

Friguls, B., Joya, X., Garcia-Algar, O., Pallas, C. R., Vall, O., & Pichini, S. (2010). A comprehensive review of assay methods to determine drugs in breast milk and the safety of breastfeeding when taking drugs. *Anal Bioanal Chem, 397*(3), 1157-1179. doi: 10.1007/s00216-010-3681-0

Fu, L. H., Ma, C. L., Cong, B., Li, S. J., Chen, H. Y., & Zhang, J. G. (2011). Hypomethylation of proximal CpG motif of interleukin-10 promoter regulates its expression in human rheumatoid arthritis. *Acta Pharmacol Sin, 32*(11), 1373-1380. doi: 10.1038/aps.2011.98

Garcia-Rodriguez, S., Callejas-Rubio, J. L., Ortego-Centeno, N., Zumaquero, E., Rios-Fernandez, R., Arias-Santiago, S., . . . Zubiaur, M. (2012). Altered AKT1 and MAPK1 gene expression on peripheral blood mononuclear cells and correlation with T-helper-transcription factors in systemic lupus erythematosus patients. *Mediators Inflamm, 2012*, 495934. doi: 10.1155/2012/495934

Gibertoni, D., Corvaglia, L., Vandini, S., Rucci, P., Savini, S., Alessandroni, R., . . . Faldella, G. (2015). Positive effect of human milk feeding during NICU hospitalization on 24 month neurodevelopment of very low birth weight infants: an Italian cohort study. *PLoS One, 10*(1), e0116552. doi: 10.1371/journal.pone.0116552

Gidrewicz, D. A., & Fenton, T. R. (2014). A systematic review and meta-analysis of the nutrient content of preterm and term breast milk. *BMC Pediatr, 14*, 216. doi: 10.1186/1471-2431-14-216

Goepfert, A. R., Andrews, W. W., Carlo, W., Ramsey, P. S., Cliver, S. P., Goldenberg, R. L., & Hauth, J. C. (2004). Umbilical cord plasma interleukin-6 concentrations in preterm infants and risk of neonatal morbidity. *Am J Obstet Gynecol, 191*(4), 1375-1381. doi: 10.1016/j.ajog.2004.06.086

Groer, M., Kane, B., Williams, N. . (2013). Multiplexing of Human Preterm and Term Milk Cytokine. *The FASEB Journal, 629*(7).

Groer, M. W., & Shelton, M. M. (2009). Exercise is associated with elevated proinflammatory cytokines in human milk. *J Obstet Gynecol Neonatal Nurs, 38*(1), 35-41. doi: 10.1111/j.1552-6909.2008.00303.x

Hadsell, D. L., Olea, W., Lawrence, N., George, J., Torres, D., Kadowaki, T., & Lee, A. V. (2007). Decreased lactation capacity and altered milk composition in insulin receptor substrate null mice is associated with decreased maternal body mass and reduced insulin-dependent phosphorylation of mammary Akt. *J Endocrinol, 194*(2), 327-336. doi: 10.1677/JOE-07-0160

Hegazy, S. K., & El-Bedewy, M. M. (2010). Effect of probiotics on pro-inflammatory cytokines and NF-kappaB activation in ulcerative colitis. *World J Gastroenterol, 16*(33), 4145-4151.

Hoppu, U., Isolauri, E., Laakso, P., Matomaki, J., & Laitinen, K. (2012). Probiotics and dietary counselling targeting maternal dietary fat intake modifies breast milk fatty acids and cytokines. *Eur J Nutr, 51*(2), 211-219. doi: 10.1007/s00394-011-0209-0

Jones, B., Nagin, D., Roeder, K. (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods Rsch, 29*(3), 374-393.

Josefsson, S., Bunn, S. K., & Domellof, M. (2007). Fecal calprotectin in very low birth weight infants. *J Pediatr Gastroenterol Nutr, 44*(4), 407-413. doi: 10.1097/MPG.0b013e3180320643

Kapel, N., Campeotto, F., Kalach, N., Baldassare, M., Butel, M. J., & Dupont, C. (2010). Faecal calprotectin in term and preterm neonates. *J Pediatr Gastroenterol Nutr, 51*(5), 542-547. doi: 10.1097/MPG.0b013e3181e2ad72

Kapel, N., Roman, C., Caldari, D., Sieprath, F., Canioni, D., Khalfoun, Y., . . . Ruemmele, F. M. (2005). Fecal tumor necrosis factor-alpha and calprotectin as differential diagnostic markers for severe diarrhea of small infants. *J Pediatr Gastroenterol Nutr, 41*(4), 396-400.

Kassal, R., Anwar, M., Kashlan, F., Smulian, J., Hiatt, M., & Hegyi, T. (2005). Umbilical vein interleukin-6 levels in very low birth weight infants developing intraventricular hemorrhage. *Brain Dev, 27*(7), 483-487. doi: 10.1016/j.braindev.2004.12.003

Kelishadi, R., Hadi, B., Iranpour, R., Khosravi-Darani, K., Mirmoghtadaee, P., Farajian, S., & Poursafa, P. (2012). A study on lipid content and fatty acid of breast milk and its association with mother's diet composition. *J Res Med Sci, 17*(9), 824-827.

Kwon, N. H., Kim, J. S., Lee, J. Y., Oh, M. J., & Choi, D. C. (2008). DNA methylation and the expression of IL-4 and IFN-gamma promoter genes in patients with bronchial asthma. *J Clin Immunol, 28*(2), 139-146. doi: 10.1007/s10875-007-9148-1

Laine, O. K., Koskela, S. M., Outinen, T. K., Joutsi-Korhonen, L., Huhtala, H., Vaheri, A., . . . Mustonen, J. T. (2014). Plasma pentraxin-3 and coagulation and fibrinolysis variables during acute Puumala hantavirus infection and associated thrombocytopenia. *Blood Coagul Fibrinolysis, 25*(6), 612-617. doi: 10.1097/MBC.0000000000000117

Lawrence, R. A., & Lawrence, R. M. (2005). *Breastfeeding: A Guide for the Medical Profession*: Elsevier Mosby.

Llanos, A. R., Moss, M. E., Pinzon, M. C., Dye, T., Sinkin, R. A., & Kendig, J. W. (2002). Epidemiology of neonatal necrotising enterocolitis: a population-based study. *Paediatr Perinat Epidemiol, 16*(4), 342-349.

Lowe, P. R., Galley, H. F., Abdel-Fattah, A., & Webster, N. R. (2003). Influence of interleukin-10 polymorphisms on interleukin-10 expression and survival in critically ill patients. *Crit Care Med, 31*(1), 34-38. doi: 10.1097/01.CCM.0000038211.80952.F3

Lu, P., Struijs, M. C., Mei, J., Witte-Bouma, J., Korteland-van Male, A. M., de Bruijn, A. C., . . . Renes, I. B. (2013). Endoplasmic reticulum stress, unfolded protein response and altered T cell differentiation in necrotizing enterocolitis. *PLoS One, 8*(10), e78491. doi: 10.1371/journal.pone.0078491

Maayan-Metzger, A., Avivi, S., Schushan-Eisen, I., & Kuint, J. (2012). Human milk versus formula feeding among preterm infants: short-term outcomes. *Am J Perinatol, 29*(2), 121-126. doi: 10.1055/s-0031-1295652

Maheshwari, A., Schelonka, R. L., Dimmitt, R. A., Carlo, W. A., Munoz-Hernandez, B., Das, A., . . . Human Development Neonatal Research, N. (2014). Cytokines associated with necrotizing enterocolitis in extremely-low-birth-weight infants. *Pediatr Res, 76*(1), 100-108. doi: 10.1038/pr.2014.48

Malarstig, A., Lindahl, B., Wallentin, L., & Siegbahn, A. (2006). Soluble CD40L levels are regulated by the -3459 A>G polymorphism and predict myocardial infarction and the efficacy of antithrombotic treatment in non-ST elevation acute coronary syndrome. *Arterioscler Thromb Vasc Biol, 26*(7), 1667-1673. doi: 10.1161/01.ATV.0000222908.78873.36

Maningat, P. D., Sen, P., Rijnkels, M., Sunehag, A. L., Hadsell, D. L., Bray, M., & Haymond, M. W. (2009). Gene expression in the human mammary epithelium during lactation: the milk

fat globule transcriptome. *Physiol Genomics, 37*(1), 12-22. doi: 10.1152/physiolgenomics.90341.2008

Manzoni, P., Stolfi, I., Pedicino, R., Vagnarelli, F., Mosca, F., Pugni, L., . . . Prevention of Neonatal Fungal Infections, I. S. o. N. (2013). Human milk feeding prevents retinopathy of prematurity (ROP) in preterm VLBW neonates. *Early Hum Dev, 89 Suppl 1*, S64-68. doi: 10.1016/S0378-3782(13)70019-7

Marek, A., Zagierski, M., Liberek, A., Aleksandrowicz, E., Korzon, M., Krzykowski, G., . . . Szlagatys-Sidorkiewicz, A. (2009). TGF-beta(1), IL-10 and IL-4 in colostrum of allergic and nonallergic mothers. *Acta Biochim Pol, 56*(3), 411-414.

Mehta, R., & Petrova, A. (2011). Biologically active breast milk proteins in association with very preterm delivery and stage of lactation. *J Perinatol, 31*(1), 58-62. doi: 10.1038/jp.2010.68

Miralles, O., Sanchez, J., Palou, A., & Pico, C. (2006). A physiological role of breast milk leptin in body weight control in developing infants. *Obesity (Silver Spring), 14*(8), 1371-1377. doi: 10.1038/oby.2006.155

Mizuno, K., Hatsuno, M., Aikawa, K., Takeichi, H., Himi, T., Kaneko, A., . . . Itabashi, K. (2012). Mastitis is associated with IL-6 levels and milk fat globule size in breast milk. *J Hum Lact, 28*(4), 529-534. doi: 10.1177/0890334412455946

Molinari, C. E., Casadio, Y. S., Hartmann, B. T., Arthur, P. G., & Hartmann, P. E. (2013). Longitudinal analysis of protein glycosylation and beta-casein phosphorylation in term and preterm human milk during the first 2 months of lactation. *Br J Nutr, 110*(1), 105-115. doi: 10.1017/S0007114512004588

Morecroft, J. A., Spitz, L., Hamilton, P. A., & Holmes, S. J. (1994). Plasma interleukin-6 and tumour necrosis factor levels as predictors of disease severity and outcome in necrotizing enterocolitis. *J Pediatr Surg, 29*(6), 798-800.

Motil, K. J., Sheng, H. P., & Montandon, C. M. (1994). Case report: failure to thrive in a breast-fed infant is associated with maternal dietary protein and energy restriction. *J Am Coll Nutr, 13*(2), 203-208.

Nagin, D. S. (2005). What Has Been Learned from Group-Based Trajectory Modeling? Examples from Physical Aggression and Other Problem Behaviors. *The Annals of the American Academy of Political and Social Science, 602*(1), 82-117.

Nguyen, D. P., Genc, M., Vardhana, S., Babula, O., Onderdonk, A., & Witkin, S. S. (2004). Ethnic differences of polymorphisms in cytokine and innate immune system genes in pregnant women. *Obstet Gynecol, 104*(2), 293-300. doi: 10.1097/01.AOG.0000133486.85400.5e

Nikniaz, L., Jr., Mahdavi, R., Arefhoesseini, S. R., & Sowti Khiabani, M. (2009). Association between fat content of breast milk and maternal nutritional status and infants' weight in tabriz, iran. *Malays J Nutr, 15*(1), 37-44.

Nile, C. J., Read, R. C., Akil, M., Duff, G. W., & Wilson, A. G. (2008). Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. *Arthritis Rheum, 58*(9), 2686-2693. doi: 10.1002/art.23758

Ochiai, S., Shimojo, N., Morita, Y., Tomiita, M., Arima, T., Inoue, Y., . . . Kohno, Y. (2013). Cytokine biomarker candidates in breast milk associated with the development of atopic dermatitis in 6-month-old infants. *Int Arch Allergy Immunol, 160*(4), 401-408. doi: 10.1159/000342995

Ordovas, J. M., & Smith, C. E. (2010). Epigenetics and cardiovascular disease. *Nat Rev Cardiol, 7*(9), 510-519. doi: 10.1038/nrcardio.2010.104

Peng, Y., Zhou, T., Wang, Q., Liu, P., Zhang, T., Zetterstrom, R., & Strandvik, B. (2009). Fatty acid composition of diet, cord blood and breast milk in Chinese mothers with different dietary habits. *Prostaglandins Leukot Essent Fatty Acids, 81*(5-6), 325-330. doi: 10.1016/j.plefa.2009.07.004

Piyathilake, C. J., Badiga, S., Alvarez, R. D., Partridge, E. E., & Johanning, G. L. (2013). A lower degree of PBMC L1 methylation is associated with excess body weight and higher HOMA-IR in the presence of lower concentrations of plasma folate. *PLoS One, 8*(1), e54544. doi: 10.1371/journal.pone.0054544

Platenburg, G. J., Vollebregt, E. J., Karatzas, C. N., Kootwijk, E. P., De Boer, H. A., & Strijker, R. (1996). Mammary gland-specific hypomethylation of Hpa II sites flanking the bovine alpha S1-casein gene. *Transgenic Res, 5*(6), 421-431.

Poralla, C., Hertfelder, H. J., Oldenburg, J., Muller, A., Bartmann, P., & Heep, A. (2012). Elevated interleukin-6 concentration and alterations of the coagulation system are associated with the development of intraventricular hemorrhage in extremely preterm infants. *Neonatology, 102*(4), 270-275. doi: 10.1159/000341266

Poralla, C., Traut, C., Hertfelder, H. J., Oldenburg, J., Bartmann, P., & Heep, A. (2012). The coagulation system of extremely preterm infants: influence of perinatal risk factors on coagulation. *J Perinatol, 32*(11), 869-873. doi: 10.1038/jp.2011.182

Prokesova, L., Lodinova-Zadnikova, R., Zizka, J., Kocourkova, I., Novotna, O., Petraskova, P., & Sterzl, I. (2006). Cytokine levels in healthy and allergic mothers and their children during the first year of life. *Pediatr Allergy Immunol, 17*(3), 175-183. doi: 10.1111/j.1399-3038.2006.00395.x

Qaddourah, R. H., Magdoud, K., Saldanha, F. L., Mahmood, N., Mustafa, F. E., Mahjoub, T., & Almawi, W. Y. (2014). IL-10 gene promoter and intron polymorphisms and changes in IL-10 secretion in women with idiopathic recurrent miscarriage. *Hum Reprod, 29*(5), 1025-1034. doi: 10.1093/humrep/deu043

Qin, W., Zhang, K., Kliethermes, B., Ruhlen, R. L., Browne, E. P., Arcaro, K. F., & Sauter, E. R. (2012). Differential expression of cancer associated proteins in breast milk based on age at first full term pregnancy. *BMC Cancer, 12*, 100. doi: 10.1186/1471-2407-12-100

Rawson, J. B., Sun, Z., Dicks, E., Daftary, D., Parfrey, P. S., Green, R. C., . . . Bapat, B. (2012). Vitamin D intake is negatively associated with promoter methylation of the Wnt antagonist gene DKK1 in a large group of colorectal cancer patients. *Nutr Cancer, 64*(7), 919-928. doi: 10.1080/01635581.2012.711418

Read, J., Bentall, R. P., & Fosse, R. (2009). Time to abandon the bio-bio-bio model of psychosis: Exploring the epigenetic and psychological mechanisms by which adverse life events lead to psychotic symptoms. *Epidemiol Psichiatr Soc, 18*(4), 299-310.

Reeves, A. A., Johnson, M. C., Vasquez, M. M., Maheshwari, A., & Blanco, C. L. (2013). TGF-beta2, a protective intestinal cytokine, is abundant in maternal human milk and human-derived fortifiers but not in donor human milk. *Breastfeed Med, 8*(6), 496-502. doi: 10.1089/bfm.2013.0017

Ribeiro, K. D., Araujo, K. F., Pereira, M. C., & Dimenstein, R. (2007). Evaluation of retinol levels in human colostrum in two samples collected at an interval of 24 hours. *J Pediatr (Rio J), 83*(4), 377-380. doi: doi:10.2223/JPED.1646

Riiskjaer, M., Nielsen, K., Steffensen, R., Erikstrup, C., Forman, A., & Kruse, C. (2011). Association of interleukin--10 promoter polymorphism and endometriosis. *Am J Reprod Immunol, 65*(1), 13-19. doi: 10.1111/j.1600-0897.2010.00888.x

Rocquelin, G., Tapsoba, S., Dop, M.C., Mbemba, F., Traissac, P., Martin-Prével, Y. (1998). Lipid content and essential fatty acid (EFA) composition of mature Congolese breast milk are influenced by mothers' nutritional status: impact on infants' EFA supply. *European J Clin Nutr, 52*(3), 164-171.

Rouge, C., Butel, M. J., Piloquet, H., Ferraris, L., Legrand, A., Vodovar, M., . . . Roze, J. C. (2010). Fecal calprotectin excretion in preterm infants during the neonatal period. *PLoS One, 5*(6), e11083. doi: 10.1371/journal.pone.0011083

Saraiva, M., & O'Garra, A. (2010). The regulation of IL-10 production by immune cells. *Nat Rev Immunol, 10*(3), 170-181. doi: 10.1038/nri2711

Sawada, K., Sato, N., & Koike, T. (1995). Inhibition of GM-CSF production by recombinant human interleukin-4: negative regulator of hematopoiesis. *Leuk Lymphoma, 19*(1-2), 33-42. doi: 10.3109/10428199509059661

Schanler, R. J. (2005). Human milk supplementation for preterm infants. *Acta Paediatr Suppl, 94*(449), 64-67. doi: 10.1080/08035320510043574

Schuster, S., Hechler, C., Gebauer, C., Kiess, W., & Kratzsch, J. (2011). Leptin in maternal serum and breast milk: association with infants' body weight gain in a longitudinal study over 6 months of lactation. *Pediatr Res, 70*(6), 633-637. doi: 10.1203/PDR.0b013e31823214ea

Selmi, C., Cantarini, L., Kivity, S., Dagaan, A., Shovman, O., Zandman-Goddard, G., . . . Shoenfeld, Y. (2015). The 2014 ACR annual meeting: a bird's eye view of autoimmunity in 2015. *Autoimmun Rev*. doi: 10.1016/j.autrev.2015.03.003

Singh, S., Vennila, J. J., Snijesh, V. P., George, G., & Sunny, C. (2015). Implying analytic measures for unraveling rheumatoid arthritis significant proteins through drug target interaction. *Interdiscip Sci*. doi: 10.1007/s12539-014-0245-6

Snijders, B. E., Stelma, F. F., Reijmerink, N. E., Thijs, C., van der Steege, G., Damoiseaux, J. G., . . . Koppelman, G. H. (2010). CD14 polymorphisms in mother and infant, soluble CD14 in breast milk and atopy development in the infant (KOALA Study). *Pediatr Allergy Immunol, 21*(3), 541-549. doi: 10.1111/j.1399-3038.2009.00939.x

Snyder, C. L., Gittes, G. K., Murphy, J. P., Sharp, R. J., Ashcraft, K. W., & Amoury, R. A. (1997). Survival after necrotizing enterocolitis in infants weighing less than 1,000 g: 25 years' experience at a single institution. *J Pediatr Surg, 32*(3), 434-437.

Sorokin, Y., Romero, R., Mele, L., Iams, J. D., Peaceman, A. M., Leveno, K. J., . . . Sibai, B. (2014). Umbilical cord serum interleukin-6, C-reactive protein, and myeloperoxidase concentrations at birth and association with neonatal morbidities and long-term neurodevelopmental outcomes. *Am J Perinatol, 31*(8), 717-726. doi: 10.1055/s-0033-1359723

Szlagatys-Sidorkiewicz, A., Martysiak-Zurowska, D., Krzykowski, G., Zagierski, M., & Kaminska, B. (2013). Maternal smoking modulates fatty acid profile of breast milk. *Acta Paediatr, 102*(8), e353-359. doi: 10.1111/apa.12276

Szlagatys-Sidorkiewicz, A., Wos, E., Aleksandrowicz, E., Luczak, G., Zagierski, M., Martysiak-Zurowska, D., . . . Kaminska, B. (2013). Cytokine profile of mature milk from smoking and nonsmoking mothers. *J Pediatr Gastroenterol Nutr, 56*(4), 382-384. doi: 10.1097/MPG.0b013e318275f95c

Takahashi, T., Yoshida, Y., Hatano, S., Sugita-Konishi, Y., Igimi, S., Yajima, M., . . . Kuwata, T. (2002). Reactivity of secretory IgA antibodies in breast milk from 107 Japanese mothers to 20 environmental antigens. *Biol Neonate, 82*(4), 238-242. doi: 65893

Thawani, N., Tam, M., & Stevenson, M. M. (2009). STAT6-mediated suppression of erythropoiesis in an experimental model of malarial anemia. *Haematologica, 94*(2), 195-204. doi: 10.3324/haematol.13422

Tischendorf, J. J., Yagmur, E., Scholten, D., Vidacek, D., Koch, A., Winograd, R., . . . Lammert, F. (2007). The interleukin-6 (IL6)-174 G/C promoter genotype is associated with the

presence of septic shock and the ex vivo secretion of IL6. *Int J Immunogenet, 34*(6), 413-418. doi: 10.1111/j.1744-313X.2007.00712.x

Trujillo, M. E., Sullivan, S., Harten, I., Schneider, S. H., Greenberg, A. S., & Fried, S. K. (2004). Interleukin-6 regulates human adipose tissue lipid metabolism and leptin production in vitro. *J Clin Endocrinol Metab, 89*(11), 5577-5582. doi: 10.1210/jc.2004-0603

Untalan, P. B., Keeney, S. E., Palkowetz, K. H., Rivera, A., & Goldman, A. S. (2009). Heat susceptibility of interleukin-10 and other cytokines in donor human milk. *Breastfeed Med, 4*(3), 137-144. doi: 10.1089/bfm.2008.0145

Ustundag, B., Yilmaz, E., Dogan, Y., Akarsu, S., Canatan, H., Halifeoglu, I., . . . Aygun, A. D. (2005). Levels of cytokines (IL-1beta, IL-2, IL-6, IL-8, TNF-alpha) and trace elements (Zn, Cu) in breast milk from mothers of preterm and term infants. *Mediators Inflamm, 2005*(6), 331-336. doi: 10.1155/MI.2005.331

Vega-Sanchez, R., Barajas-Vega, H. A., Rozada, G., Espejel-Nunez, A., Beltran-Montoya, J., & Vadillo-Ortega, F. (2010). Association between adiposity and inflammatory markers in maternal and fetal blood in a group of Mexican pregnant women. *Br J Nutr, 104*(12), 1735-1739. doi: 10.1017/S0007114510002825

Velez, D. R., Fortunato, S. J., Williams, S. M., & Menon, R. (2008). Interleukin-6 (IL-6) and receptor (IL6-R) gene haplotypes associate with amniotic fluid protein concentrations in preterm birth. *Hum Mol Genet, 17*(11), 1619-1630. doi: 10.1093/hmg/ddn049

Venkatesh, M., & Abrams, S. (2009). Can lactoferrin prevent neonatal sepsis and necrotizing enterocolitis? *Expert Rev Anti Infect Ther, 7*(5), 515-525. doi: 10.1586/eri.09.25

Villalpando, S., Del Prado, M., Lance, A., Alfonso, E., Rodriguez, M., Demmelmair, H., & Koletzko, B. (2001). [13C]linoleic acid oxidation and transfer into milk in stunted lactating women with contrasting body mass indexes. *Am J Clin Nutr, 74*(6), 827-832.

Vineis, P., Chuang, S. C., Vaissiere, T., Cuenin, C., Ricceri, F., Genair, E. C., . . . Herceg, Z. (2011). DNA methylation changes associated with cancer risk factors and blood levels of vitamin metabolites in a prospective study. *Epigenetics, 6*(2), 195-201.

Vohr, B. R., Poindexter, B. B., Dusick, A. M., McKinley, L. T., Higgins, R. D., Langer, J. C., . . . Human Development National Research, N. (2007). Persistent beneficial effects of breast milk ingested in the neonatal intensive care unit on outcomes of extremely low birth weight infants at 30 months of age. *Pediatrics, 120*(4), e953-959. doi: 10.1542/peds.2006-3227

Vorbach, C., Capecchi, M. R., & Penninger, J. M. (2006). Evolution of the mammary gland from the innate immune system? *Bioessays, 28*(6), 606-616. doi: 10.1002/bies.20423

Winter, C. D., Pringle, A. K., Clough, G. F., & Church, M. K. (2004). Raised parenchymal interleukin-6 levels correlate with improved outcome after traumatic brain injury. *Brain, 127*(Pt 2), 315-320. doi: 10.1093/brain/awh039

Wong, C. M., Anderton, D. L., Smith-Schneider, S., Wing, M. A., Greven, M. C., & Arcaro, K. F. (2010). Quantitative analysis of promoter methylation in exfoliated epithelial cells isolated from breast milk of healthy women. *Epigenetics, 5*(7), 645-655. doi: 10.4161/epi.5.7.12961

World Health Organization, W. (2014, November 2014). Preterm Birth. Retrieved May 5, 2015, from http://www.who.int/mediacentre/factsheets/fs363/en/

Wu, W., Clark, E. A., Stoddard, G. J., Watkins, W. S., Esplin, M. S., Manuck, T. A., . . . Jorde, L. B. (2013). Effect of interleukin-6 polymorphism on risk of preterm birth within population strata: a meta-analysis. *BMC Genet, 14*, 30. doi: 10.1186/1471-2156-14-30

Yaeger, R., Avila-Bront, A., Abdul, K., Nolan, P. C., Grann, V. R., Birchette, M. G., . . . Joe, A. K. (2008). Comparing genetic ancestry and self-described race in african americans born in the United States and in Africa. *Cancer Epidemiol Biomarkers Prev, 17*(6), 1329-1338. doi: 10.1158/1055-9965.EPI-07-2505

Yilmaz, V., Yentur, S. P., & Saruhan-Direskeneli, G. (2005). IL-12 and IL-10 polymorphisms and their effects on cytokine production. *Cytokine, 30*(4), 188-194. doi: 10.1016/j.cyto.2005.01.006

Yoon, J. M., Park, J. Y., Ko, K. O., Lim, J. W., Cheon, E. J., & Kim, H. J. (2014). Fecal calprotectin concentration in neonatal necrotizing enterocolitis. *Korean J Pediatr, 57*(8), 351-356. doi: 10.3345/kjp.2014.57.8.351

Yudkin, J. S., Stehouwer, C. D., Emeis, J. J., & Coppack, S. W. (1999). C-reactive protein in healthy subjects: associations with obesity, insulin resistance, and endothelial dysfunction: a potential role for cytokines originating from adipose tissue? *Arterioscler Thromb Vasc Biol, 19*(4), 972-978.

Zanardo, V., Golin, R., Amato, M., Trevisanuto, D., Favaro, F., Faggian, D., & Plebani, M. (2007). Cytokines in human colostrum and neonatal jaundice. *Pediatr Res, 62*(2), 191-194. doi: 10.1203/PDR.0b013e31809871c9

Zhao, J., Goldberg, J., & Vaccarino, V. (2013). Promoter methylation of serotonin transporter gene is associated with obesity measures: a monozygotic twin study. *Int J Obes (Lond), 37*(1), 140-145. doi: 10.1038/ijo.2012.8