Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____   _____
Anna Kaitlyn Knight   Date

Epigenetic and Transcriptomic Signatures of Maternal and Neonatal Risk for Adverse Pregnancy Outcomes

By

Anna Kaitlyn Knight

Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences

Genetics and Molecular Biology Program

_____
Alicia K. Smith, Ph.D.
Advisor

_____
Karen N. Conneely, Ph.D.
Committee Member

_____
Anne L. Dunlop, M.D.
Committee Member

_____
Jennifer G. Mulle, Ph.D., M.H.S.
Committee Member

_____
Paula M. Vertino, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Epigenetic and Transcriptomic Signatures of Maternal and Neonatal Risk for Adverse Pregnancy Outcomes

By

Anna Kaitlyn Knight
B.S, North Carolina State University

Advisor: Alicia K. Smith, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences

Genetics and Molecular Biology Program

2018

Abstract

Epigenetic and Transcriptomic Signatures of Maternal and Neonatal Risk for Adverse Pregnancy Outcomes

By Anna Kaitlyn Knight

Pregnancy and delivery complications are remarkably common and put both the mother and the neonate at increased risk for both acute and chronic adverse health outcomes. Despite these well-known risks, the underlying etiology of most pregnancy and delivery complications is not clear. The first step in understanding these etiologies is to examine changes over uncomplicated, full term pregnancies. We identify patterns of gene expression changes over pregnancy that are consistent with physiological changes, including changes in oxygen transport, immune factors, and response to microbes. In complicated pregnancies, we identify DNA methylation changes predictive of fetal intolerance of labor, a common pregnancy complication that may indicate the need for a Cesarean section. Thus, understanding the gene expression and gene regulatory changes associated with pregnancy can allow for better prediction of maternal risk.

In addition to maternal DNA methylation changes, neonatal DNA methylation can also serve as a proxy for neonatal risk. We developed a predictor of gestational age at birth based on DNA methylation data from neonatal blood spots and cord blood. The difference between predicted age and clinically estimated gestational age, termed gestational age acceleration, is associated with a variety of factors related to developmental maturity. We show that neonates requiring oxygen, steroids, and surfactant in the neonatal intensive care unit have a lower developmental maturity than those who do not. Additionally, neonates with a lower developmental maturity are more likely to develop bronchopulmonary dysplasia. Gestational age acceleration, therefore, is a useful tool for both clinical and research applications to better quantify neonatal risk.

Epigenetic and Transcriptomic Signatures of Maternal and Neonatal Risk for Adverse Pregnancy Outcomes

By

Anna Kaitlyn Knight
B.S, North Carolina State University

Advisor: Alicia K. Smith, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences

Genetics and Molecular Biology Program

2018

ACKNOWLEDGEMENTS

I am extremely grateful to my advisor, Alicia Smith, who has provided immense support and patience, invaluable opportunities, and lasting friendship over the last four and a half years. She has been a fantastic role model and has helped me become a better scientist, public speaker and problem solver, and has left all of her students better trained than we realize. I could not have imagined a better advisor for my dissertation and calming influence throughout my graduate school career.

I would like to express my deepest appreciation to the members of my committee: Karen Conneely, Anne Dunlop, Jennifer Mulle, and Paula Vertino for their guidance and mentorship and sharing their expertise, without which this dissertation would not have been possible.

I am very appreciative to the past and present members of the Smith Lab, especially Sasha Parets and Shruti Iyer for their friendship and always pushing me to do my best, Dawayland Cobb for the countless hours of extractions and his help and advice, Varun Kilaru for fixing all of my coding problems, Sarah Curtis, Seyma Katrlini, and Stefanie Pilkay for sharing an office with me, Samantha Resin for her positive demeanor and willingness to help with any task, Sabrina Gerkowicz and Misti Patel for their clinical expertise and camaraderie, and all of the other members of the Smith Lab and Department of Gynecology and Obstetrics for their expertise and insight. I would also like to thank the Genetics and Molecular Biology Program and Molecules to Mankind for their feedback, support, and encouragement.

I also had the great pleasure of working with many wonderful collaborators, including Jeffery Craig, Christiane Theda, Michael Kobor, and the entire Emory University African American Microbiome in Pregnancy working group. I truly appreciate the mentorship of Dave

TABLE OF CONTENTS

Table of Tables

Table of Figures

Chapter 1

Maternal and Neonatal Risk Associated with Pregnancy Complications

Portions of this chapter have been adapted from the below publications:

1. Parets SE, **Knight AK**, Smith AK. Insights into genetic susceptibility in the etiology of spontaneous preterm birth. *Appl Clin Genet*. 2015;8:283-90. doi: 10.2147/TACG.S58612. PubMed PMID: 26715857; PubMed Central PMCID: PMC4685889.
2. **Knight AK**, Smith AK. Epigenetic Biomarkers of Preterm Birth and Its Risk Factors. *Genes (*Basel). 2016;7(4). doi: 10.3390/genes7040015. PubMed PMID: 27089367; PubMed Central PMCID: PMC4846845.

Introduction

Pregnancy and delivery complications put both the mother and the neonate at increased risks for acute and long-term adverse health outcomes [1-5]. Such complications are remarkably common, with complications such as preterm birth occurring in approximately 10% of all pregnancies in the United States, [6] and are the leading cause of infant death and major disability [5]. Risk factors for pregnancy and delivery complications include having a low socioeconomic status, African American race, smoking, stress, pre-pregnancy diabetes or hypertension, and having a previous pregnancy complication [7-10]. Despite these well-known risk factors and substantial research efforts, the etiology of pregnancy complications is not well understood. Studies of both complicated and uncomplicated pregnancies are required to better understand the underlying etiologies of pregnancy complications.

Physiological Changes Over Pregnancy

<u>Immune Activation and Inflammation</u>

Pregnancy is associated with shifts in the maternal inflammatory state, with the first trimester typically characterized as being proinflammatory, the second trimester characterized by a more quiescent state, and the third trimester is characterized by a return to the proinflammatory state for parturition to occur [11]. As part of the inflammatory response to pregnancy, changes in the proportions of T helper (Th) cell subtypes has been observed, with the Type I Th cells being associated with the production of pro-inflammatory cytokines such as TNF-α, IFN- γ, and IL-2 and type 2 Th cells being associated with anti-inflammatory cytokines including IL-4, IL-5, IL-6, IL-10, and IL-12 [12]. However, the inflammatory state in pregnancy cannot be solely attributed to a simple shift in the Th1/Th2 ratio, as other immune cell types and regulatory systems have

also been shown to be important in maternal tolerance of the fetus [13]. The placenta also plays a major role in mediating inflammation, and can stimulate the production of various cytokines in response to an infection and stress [11].

The immune system during pregnancy is a precarious balance between tolerating the fetus and protecting the mother from infection and other threats [14]. As such, a disturbance in this balance has been associated with a variety of negative outcomes for both the mother and fetus [15-17]. Negative maternal outcomes associated with general inflammation include delivering preterm (and the subsequent accompanying lifetime risks), future preterm delivery, hypertensive disorders, high blood pressure long-term, preeclampsia, gestational diabetes, and potentially perinatal depression [18-21]. Increased inflammation associates with negative outcomes for the fetus as well, including placental dysfunction, being born preterm, and potential brain damage [22-26].

<u>Cardiopulmonary Changes</u>

Among the vast physiological changes during pregnancy are changes related to hemodynamics. Blood volume increases by approximately 50% and is accompanied by a decrease in hemoglobin and hematocrit. Additionally, during pregnancy there are higher rates of erythropoiesis and coagulation, which promotes a healthy pregnancy by helping to limit complications like post-partum hemorrhage. Abnormal changes in coagulation have been associated with complications like preeclampsia, highlighting their importance of the regulation of hematological changes during pregnancy [27-30].

In addition to hematological changes, cardiac output increases, accompanied by vasodilation, increases in maternal heart rate, and decreases in blood pressure until the end of the third trimester [31, 32]. Abnormal cardiovascular changes are associated with hypertensive

disorders of pregnancy, including preeclampsia [33]. Changes to the respiratory system are also expected during normal pregnancy as oxygen demand, minute volume, and tidal volume increase and functional residual capacity and expiratory reserve volume decrease. Poor pulmonary function increases risk for adverse outcomes, including preterm delivery and growth restriction [34].

Adverse Health Outcomes Associated with Pregnancy Complications

Maternal Outcomes

Complicated pregnancies and deliveries carry significant risks for the mother (Figure 1), the most severe of which is maternal death. For the purposes of this chapter, complications due to preterm delivery (<37 weeks gestation), preeclampsia, and gestational diabetes will be considered. Preterm delivery carries significant acute maternal risks. Women delivering between 23-27 weeks gestation have the highest risk for severe complications including hemorrhage and infection, with women delivering by Cesarean-section having increased risks compared to women delivering vaginally [35]. Additionally, women delivering preterm are more likely to have post-partum anxiety, fatigue, and less contact with their neonate [36]. Long-term risks associated with preterm delivery include increased risk for developing cardiovascular disease, type II diabetes, and breast cancer [1, 3, 37].

Preeclampsia, which is characterized by pregnancy-associated hypertension, proteinuria, and organ/uteroplacental dysfunction, is also associated with significant maternal risks, with the only cure being delivery. If left untreated, preeclampsia can progress to eclampsia where high blood pressure causes seizures, which can be fatal. The leading cause of death in women with preeclampsia is cerebral hemorrhage [38]. Long-term consequences of preeclampsia include

increased risks for cardiovascular disease and high blood pressure, white matter lesions, and visual disturbances [38, 39].

Gestational diabetes, or diabetes diagnosed during pregnancy, like preterm delivery and preeclampsia, is also associated with an increased risk of cardiovascular disease. Additionally, gestational diabetes is associated with metabolic syndrome and type II diabetes [40]. Acute consequences of gestational diabetes include increased risk of other pregnancy complications, including preeclampsia [41]. Understanding the etiology of pregnancy complications may allow for the development of novel treatments or better identification of women most at risk.

Neonatal Outcomes

Neonatal consequences of pregnancy complications are also severe (Figure 1). For the purposes of this chapter, complications due to preterm birth, preeclampsia, and gestational diabetes will be considered. Neonates born preterm are more likely to have a range of complications in the perinatal period, including neonatal death, respiratory distress and the development of bronchopulmonary dysplasia and retinopathy of prematurity [42-46]. Children born preterm are more likely to have cerebral palsy, sensory deficits, learning disabilities, and respiratory illnesses [47, 48]. Preterm birth also increases the risk of being hospitalized with infections during childhood [49]. Among school aged children, those born preterm show diminished cognitive performance and increased externalizing and internalizing behaviors, and are more likely to develop ADHD [4]. Preterm birth and reduced fetal growth have also been linked to a number of important chronic diseases of adulthood such as type 2 diabetes [50, 51].

Neonates born to mothers with preeclampsia are also more likely to be born preterm, as delivery is the most effective intervention for preeclampsia [52]. Additionally, these neonates are at a higher risk for perinatal death, growth restriction, and oligohydramnios. In childhood and

adolescence, neonates born to mothers with preeclampsia are at increased risk for higher blood pressure, followed by an increased risk of hypertension as adults [38]. Cognitive functioning throughout life has also been shown to be more likely to be impaired in children whose mothers developed preeclampsia [38].

Finally, neonates born to mothers with gestational diabetes are more likely to develop macrosomia, which is associated with birth injuries and respiratory distress [53]. These children are also more likely to be overweight, and have poor glucose tolerance and/or renal disease [40, 53]. As pregnancy complications may result in severe adverse outcomes for neonates, future research should focus on how to decrease morbidity and mortality from pregnancy complications.

Genetic, Epigenetic, and Transcriptomic Signatures of Pregnancy Complications

The physiological changes over pregnancy are well established, however, the etiology of pregnancy complications remains unclear. Clinical advancements can help mitigate the morbidity and mortality for the mother and neonate associated with pregnancy complications, but the overall pregnancy complication rate continues to rise [54, 55]. Known risk factors, such as smoking and low socioeconomic status, can explain some of the risk of pregnancy and delivery complications, but the physiological cause of these complications is largely unknown [56-60]. We will address changes in gene expression over uncomplicated, term pregnancies in Chapter 2. By evaluating changes in gene expression over uncomplicated, term pregnancies, we will be better able to understand changes associated with pregnancy complications.

One of the greatest risk factors for pregnancy complications is familial history. For example, the heritability of preterm birth is estimated at 17-35%, suggesting a genetic

component that may increase risk of pregnancy complications [61]. However, there is also a substantial environmental component involved in the development of pregnancy complications.

Genetic Studies

Genetic studies of pregnancy complications have not yet identified causal genes with large effect sizes. The majority of genetic studies of pregnancy complications have focused on candidate genes which can test a specific hypothesis, but are limited to research questions surrounding known biological processes associated with pregnancy complications, such as inflammation [61]. While these studies have identified associations between genetic variation and pregnancy complications, they fail to explain the vast majority of pregnancy complication risk [62-66]. Additionally, these studies have yet to identify genetic variants that can be used to screen women at risk for pregnancy complications.

Genome-wide association studies (GWAS) allow for a hypothesis-free approach, but require large sample sizes and careful phenotyping, and are computationally intensive. Therefore, there are limited GWAS of pregnancy complications. Several recent GWAS and whole exome studies have identified single nucleotide polymorphisms (SNPs) associated with preterm birth [67, 68] and preeclampsia [69, 70], but they have yet to be replicated and validated. The inconsistent and somewhat sporadic findings associated with genetic studies of pregnancy complications suggest that there may be other mechanisms that explain a greater proportion of risk associated with pregnancy complications.

Epigenetic Studies

As genetic studies have identified few associations with pregnancy complications, epigenetics has been proposed as a potential mechanism of pregnancy complications that may reflect a combination of genetic and environmental factors [71, 72]. Epigenetics refers to the

regulation of gene expression without changes in the underlying sequence of DNA. One commonly studied epigenetic mark is DNA methylation at the 5' position of cytosine in a cytosine-guanine dinucleotide. Previous studies have identified associations between DNA methylation and a variety of pregnancy complications and risk factors [71, 73-77].

For example, three studies have identified CpG sites related to preterm birth, identifying between 29 and 1,555 CpG sites associated with preterm birth in neonatal blood [78-80]. These findings support that there are substantial epigenetic changes that can be detected between term and preterm neonates at birth, and suggest that there are likely to be epigenetic differences present before birth that may be detectable in fetal or maternal tissues. Such differences may elucidate genes and pathways involved in preterm birth and may serve as biomarkers to identify women at risk for preterm delivery and neonates most likely to be poorly impacted by preterm birth.

For this dissertation, we will evaluate changes in DNA methylation associated with fetal intolerance of labor, which is a common delivery complication and indication for an emergency Cesarian-section, in Chapter 3.

The development and utility of biomarkers for PC

A biomarker is a biological measure that is predictive of a normal or pathogenic process or response. In clinical practice, biomarkers can be used for risk assessment, early detection or onset of a disease or chronic illness. Once a diagnosis is established, they can also be used as an indicator of symptom severity or response to treatment [81-83]. In general, a biomarker candidate must be reproducible and have sufficient sensitivity and specificity to provide clinically-relevant information [84]. Epigenetic-based biomarkers have become increasingly common as cost-effect methods for assessing the epigenome have been developed.

One example of reproducible and validated epigenetic biomarker comes from studies of smoking. Smoking is a well-known, preventable, significant contributor to neonatal and maternal morbidity and mortality, yet 12.3% of pregnant women smoked in 2010 [85]. Smoking may cause complications for the fetus due to exposure to tobacco toxins, poor umbilical blood flow, oxidative stress, and changes in gene expression [59, 86, 87]. Conditions associated with maternal smoking in pregnancy for the neonate include PTB, fetal growth restriction, sudden infant death syndrome, stillbirth, paraventricular leukomalacia, bronchopulmonary dysplasia, intraventricular hemorrhage, placenta-associated syndrome as well as reduced academic performance and elevated blood pressure in adolescence [57, 85, 86, 88-93].

In adults, DNA methylation differences in the blood of smokers versus non-smokers have been reported [75, 94]. Maternal smoking during pregnancy may change the DNA methylation profile of neonates, and this change in methylation may mediate neonatal birthweight and immune function [95-98]. Stroud and colleagues found that maternal smoking during pregnancy was associated with decreased infant salivary cortisol levels in the first post-natal month, and that placental *NR3C1* methylation was decreased in exposed fetuses [95]. They suggest that decreased placental methylation of the *NR3C1* promoter mediates the lowered cortisol levels seen in neonates [95]. Other groups have found associations between maternal smoking in pregnancy and methylation through epigenome-wide association studies [96-98]. The association between maternal smoking and methylation in the offspring can be seen through adolescence, suggesting that there is the potential for long-term effects of this behavior on the child [99, 100]. One group, Kupers and colleagues, performing an EWAS with methylation data from cord blood and maternal smoking was able to establish the mediating effect of three CpG sites associated with growth factor independent 1 transcription repressor (*GFI1*) on low birth weight, concluding

that these CpGs could explain 12-19% of the lowered birth weight [98]. These studies serve as an example for future studies of epigenetic biomarkers for adverse acute and long-term outcomes. Having similar biomarkers for pregnancy and delivery complications may allow for early identification of mothers and neonates most at risk for adverse outcomes to allow for early interventions, monitoring, and personalized care.

Age Acceleration

Recently, DNA methylation of a select group of CpG sites has been used to generate a promising new biomarker related to aging. It is widely accepted that aging influences DNA methylation across the genome, and several recent studies have taken advantage of age-related methylation changes to build a predictor of DNA methylation age (DNAm age) [101-103]. A DNAm age that is higher than a person's chronological age may indicate accelerated aging, which is a potential metric of stress and general health, shown by recent studies of DNAm age in developmental and neurocognitive outcomes, as well as all-cause mortality [104, 105]. Measuring DNAm age has potential for use in predicting maternal risk during pregnancy, though this relationship has not yet been examined.

One study has examined the relationship between age acceleration and a range of phenotypes in a large, longitudinal cohort. Overall, they found that DNAm age became more correlated with most phenotypes as a child aged, suggesting that this predictor may not be as informative for neonates and young children. Only maternal smoking and Cesarian-section associated with accelerated age at birth (p<.05), though birthweight and some maternal characteristics are associated with positive (BMI) and negative (selenium exposure, cholesterol level) accelerated aging in childhood and/or adolescence [106].

Placental and cord blood samples were included when the DNAm age predictor was initially developed, but the age of all of these samples was set to 0. Due to this, DNAm age, as initially operationalized, is not accurate enough to discriminate weeks of gestation. Gestational age has previously been associated with changes in methylation at various CpG sites [78, 107-109], which could be used to develop a predictor accurate for neonates. Further studies on DNA methylation in neonatal cord blood are required for the development of a neonatal gestational age predictor, which we present in Chapter 4. Chapter 5 extends work on this neonatal gestational age predictor to studies of outcomes and interventions in the neonatal intensive care unit.

Gene Expression Studies

Typically, an increase in DNA methylation at promoter regions correlates with a decrease in expression of that gene, though exceptions to this are commonly documented [110]. Intragenic DNA methylation is also important to regulate alternative promoters and enhancers that define a variety of alternative transcripts and promote gene expression [111]. In light of these associations between DNA methylation and gene expression, it is widely accepted that methylation changes that are correlated with gene expression are more likely to be biologically relevant. One recent study found that there were 16,327 expression-associated cytosine guanine dinucleotide sites (CpG sites) in a large cohort, which both provides a resource for interpreting epigenetic associations that fall outside of genes and demonstrates that CpGs interacting with gene expression are most often found in enhancer regions [112].

However, associations are also commonly identified with gene expression that are not associated with DNA methylation, as there are many mechanisms regulating RNA transcription and degradation [113, 114]. These studies are also essential for identifying genes and pathways involved in pregnancy and delivery complications. However, to date, there have been very few

genome-wide studies of gene expression and pregnancy complications. It is vital to understand how gene expression changes over normal pregnancy to put changes in gene expression associated with pregnancy and delivery complications into context.

Conclusions

Pregnancy complications pose a significant health risk for both the mother and the neonate. Genome-wide studies of epigenetics and gene expression have the potential to provide insight into novel genes and pathways involved in pregnancy complications, and may identify potential biomarkers that could be used in clinical practice, and could substantially improve our ability to identify and treat women at the highest risk for adverse outcomes. Targeted interventions could reduce multiple health burdens associated with pregnancy complications for the mother and the neonate. Such a biomarker does not yet exist, in part due to the complexity of the intrauterine environment. Gene regulation varies over the course of pregnancy and that regulation is likely influenced by a number of independent environmental factors. Thus, potential biomarkers would need to be capable of distinguishing health outcomes within diverse contexts. Studies that simultaneously evaluate multiple risk factors for pregnancy complications will have the most potential to identify epigenetic or expression-based predictors of pregnancy complications that can be developed into biomarkers. Recent studies have made substantial progress towards identifying and replicating epigenetic associations in diverse cohorts, and it is important that basic and epidemiological researchers develop partnerships though which the most promising results from their work can be extended into prospective and clinical studies.

In addition to individual CpG sites, summary measures that integrate information from multiple regions of the genome may also be informative. The use of the DNAm age predictor to

evaluate future health risk in adults demonstrates that this type of epigenetic predictor, which combines several hundred CpG sites, may be useful for predicting perinatal development, if optimized for use in neonates. Once appropriate biomarkers are identified, targeted and cost-effective assays could be developed to prospectively screen for adverse outcomes. Successful development and clinical implementation of such a biomarker could greatly improve clinical recommendations and allowed for more personalized treatment.

Figure 1-1: Pregnancy complications influence long and short-term maternal and neonatal risk.

References

1.   Kessous R, Shoham-Vardi I, Pariente G, Holcberg G, Sheiner E. An association between preterm delivery and long-term maternal cardiovascular morbidity. American journal of obstetrics and gynecology. 2013;209(4):368 e1-8. doi: 10.1016/j.ajog.2013.05.041. PubMed PMID: 23800639.

2.   James-Todd T, Wise L, Boggs D, Rich-Edwards J, Rosenberg L, Palmer J. Preterm birth and subsequent risk of type 2 diabetes in black women. Epidemiology. 2014;25(6):805-10. doi: 10.1097/EDE.0000000000000167. PubMed PMID: 25166879; PubMed Central PMCID: PMC4180791.

3.   Melbye M, Wohlfahrt J, Andersen AM, Westergaard T, Andersen PK. Preterm delivery and risk of breast cancer. British journal of cancer. 1999;80(3-4):609-13. doi: 10.1038/sj.bjc.6690399. PubMed PMID: 10408874; PubMed Central PMCID: PMC2362328.

4.   Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJ. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. Jama. 2002;288(6):728-37. PubMed PMID: 12169077.

5.   Callaghan WM, MacDorman MF, Rasmussen SA, Qin C, Lackritz EM. The contribution of preterm birth to infant mortality rates in the United States. Pediatrics. 2006;118(4):1566-73. doi: 10.1542/peds.2006-0860. PubMed PMID: 17015548.

6.   Hamilton BE, Martin JA, Osterman MJ, Curtin SC, Matthews TJ. Births: Final Data for 2014. Natl Vital Stat Rep. 2015;64(12):1-64. PubMed PMID: 26727629.

7.   Martin JA, Hamilton BE, Osterman MJK, Driscoll AK, Drake P. Births: Final Data for 2016. Natl Vital Stat Rep. 2018;67(1):1-55. PubMed PMID: 29775434.

8.   Adams MM, Elam-Evans LD, Wilson HG, Gilbertz DA. Rates of and factors associated with recurrence of preterm delivery. Jama. 2000;283(12):1591-6. PubMed PMID: 10735396.

9.   Mund M, Louwen F, Klingelhoefer D, Gerber A. Smoking and pregnancy--a review on the first major environmental risk factor of the unborn. Int J Environ Res Public Health. 2013;10(12):6485-99. doi: 10.3390/ijerph10126485. PubMed PMID: 24351784; PubMed Central PMCID: PMCPMC3881126.

10.   Goldenberg RL, Cliver SP, Mulvihill FX, Hickey CA, Hoffman HJ, Klerman LV, et al. Medical, psychosocial, and behavioral risk factors do not explain the increased risk for low birth weight among black women. American journal of obstetrics and gynecology. 1996;175(5):1317-24. PubMed PMID: 8942508.

11.   Mor G, Cardenas I. The immune system in pregnancy: a unique complexity. American journal of reproductive immunology. 2010;63(6):425-33. doi: 10.1111/j.1600-0897.2010.00836.x. PubMed PMID: 20367629; PubMed Central PMCID: PMC3025805.

12.   Liang PY, Diao LH, Huang CY, Lian RC, Chen X, Li GG, et al. The pro-inflammatory and anti-inflammatory cytokine profile in peripheral blood of women with recurrent implantation failure. Reprod Biomed Online. 2015;31(6):823-6. doi: 10.1016/j.rbmo.2015.08.009. PubMed PMID: 26371706.

13.   Munoz-Suano A, Hamilton AB, Betz AG. Gimme shelter: the immune system during pregnancy. Immunological reviews. 2011;241(1):20-38. doi: 10.1111/j.1600-065X.2011.01002.x. PubMed PMID: 21488887.

14.   Schminkey DL, Groer M. Imitating a stress response: a new hypothesis about the innate immune system's role in pregnancy. Medical hypotheses. 2014;82(6):721-9. doi: 10.1016/j.mehy.2014.03.013. PubMed PMID: 24698849.

15.   Christian LM. Psychoneuroimmunology in pregnancy: immune pathways linking stress with maternal health, adverse birth outcomes, and fetal development. Neuroscience and biobehavioral reviews. 2012;36(1):350-61. doi: 10.1016/j.neubiorev.2011.07.005. PubMed PMID: 21787802; PubMed Central PMCID: PMC3203997.

16.   Robinson DP, Klein SL. Pregnancy and pregnancy-associated hormones alter immune responses and disease pathogenesis. Hormones and behavior. 2012;62(3):263-71. doi: 10.1016/j.yhbeh.2012.02.023. PubMed PMID: 22406114; PubMed Central PMCID: PMC3376705.

17.   Cotechini T, Graham CH. Aberrant maternal inflammation as a cause of pregnancy complications: A potential therapeutic target? Placenta. 2015;36(8):960-6. doi: 10.1016/j.placenta.2015.05.016. PubMed PMID: 26094029.

18.   Yang H, Tang W, Zhu C, Guo C. Platelets, inflammation, and prediction of the hypertension disorders of pregnancy. The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet. 2012;25(1):99-103. doi: 10.3109/14767058.2011.560985. PubMed PMID: 21405944.

19.   Catov JM, Lewis CE, Lee M, Wellons MF, Gunderson EP. Preterm birth and future maternal blood pressure, inflammation, and intimal-medial thickness: the CARDIA study. Hypertension. 2013;61(3):641-6. doi: 10.1161/HYPERTENSIONAHA.111.00143. PubMed PMID: 23319540; PubMed Central PMCID: PMC3583341.

20.   Hackney DN, Tirumala R, Salamone LJ, Miller RK, Katzman PJ. Do placental histologic findings of chorion-decidual hemorrhage or inflammation in spontaneous preterm birth influence

outcomes in the subsequent pregnancy? Placenta. 2014;35(1):58-63. doi: 10.1016/j.placenta.2013.11.001. PubMed PMID: 24262136.

21. Osborne LM, Monk C. Perinatal depression--the fourth inflammatory morbidity of pregnancy?: Theory and literature review. Psychoneuroendocrinology. 2013;38(10):1929-52. doi: 10.1016/j.psyneuen.2013.03.019. PubMed PMID: 23608136; PubMed Central PMCID: PMC4376311.

22. Fleiss B, Tann CJ, Degos V, Sigaut S, Van Steenwinckel J, Schang AL, et al. Inflammation-induced sensitization of the brain in term infants. Developmental medicine and child neurology. 2015;57 Suppl 3:17-28. doi: 10.1111/dmcn.12723. PubMed PMID: 25800488.

23. Ho M, Faye-Petersen OM, Goldenberg RL, Carlo WA, Cliver SP, Andrews WW. Elevated midtrimester alpha-fetoprotein and delivery markers of inflammation in a preterm population. The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet. 2012;25(11):2424-7. doi: 10.3109/14767058.2012.700342. PubMed PMID: 22681575.

24. Girard S, Heazell AE, Derricott H, Allan SM, Sibley CP, Abrahams VM, et al. Circulating cytokines and alarmins associated with placental inflammation in high-risk pregnancies. American journal of reproductive immunology. 2014;72(4):422-34. doi: 10.1111/aji.12274. PubMed PMID: 24867252; PubMed Central PMCID: PMC4369138.

25. Romero R, Miranda J, Chaiworapongsa T, Korzeniewski SJ, Chaemsaithong P, Gotsch F, et al. Prevalence and clinical significance of sterile intra-amniotic inflammation in patients with preterm labor and intact membranes. American journal of reproductive immunology.

2014;72(5):458-74. doi: 10.1111/aji.12296. PubMed PMID: 25078709; PubMed Central PMCID: PMC4192099.

26.   Taylor BD, Holzman CB, Fichorova RN, Tian Y, Jones NM, Fu W, et al. Inflammation biomarkers in vaginal fluid and preterm delivery. Human reproduction. 2013;28(4):942-52. doi: 10.1093/humrep/det019. PubMed PMID: 23416276; PubMed Central PMCID: PMC3600841.

27.   Saha D, Patgaonkar M, Shroff A, Ayyar K, Bashir T, Reddy KV. Hemoglobin expression in nonerythroid cells: novel or ubiquitous? Int J Inflam. 2014;2014:803237. doi: 10.1155/2014/803237. PubMed PMID: 25431740; PubMed Central PMCID: PMC4241286.

28.   Townsley DM. Hematologic complications of pregnancy. Semin Hematol. 2013;50(3):222-31. doi: 10.1053/j.seminhematol.2013.06.004. PubMed PMID: 23953339; PubMed Central PMCID: PMC3748382.

29.   Han L, Liu X, Li H, Zou J, Yang Z, Han J, et al. Blood coagulation parameters and platelet indices: changes in normal and preeclamptic pregnancies and predictive values for preeclampsia. PloS one. 2014;9(12):e114488. doi: 10.1371/journal.pone.0114488. PubMed PMID: 25464515; PubMed Central PMCID: PMC4252147.

30.   Mallone R, Mannering SI, Brooks-Worrell BM, Durinovic-Bello I, Cilio CM, Wong FS, et al. Isolation and preservation of peripheral blood mononuclear cells for analysis of islet antigen-reactive T cell responses: position statement of the T-Cell Workshop Committee of the Immunology of Diabetes Society. Clin Exp Immunol. 2011;163(1):33-49. doi: 10.1111/j.1365-2249.2010.04272.x. PubMed PMID: 20939860; PubMed Central PMCID: PMC3010910.

31.   Soma-Pillay P, Nelson-Piercy C, Tolppanen H, Mebazaa A. Physiological changes in pregnancy. Cardiovasc J Afr. 2016;27(2):89-94. doi: 10.5830/CVJA-2016-021. PubMed PMID: 27213856; PubMed Central PMCID: PMCPMC4928162.

32.   Ouzounian JG, Elkayam U. Physiologic changes during normal pregnancy and delivery. Cardiol Clin. 2012;30(3):317-29. doi: 10.1016/j.ccl.2012.05.004. PubMed PMID: 22813360.

33.   Magee LA, Pels A, Helewa M, Rey E, von Dadelszen P, Canadian Hypertensive Disorders of Pregnancy Working G, et al. The hypertensive disorders of pregnancy (29.3). Best Pract Res Clin Obstet Gynaecol. 2015;29(5):643-57. doi: 10.1016/j.bpobgyn.2015.04.001. PubMed PMID: 26141795.

34.   Grindheim G, Toska K, Estensen ME, Rosseland LA. Changes in pulmonary function during pregnancy: a longitudinal cohort study. BJOG. 2012;119(1):94-101. doi: 10.1111/j.1471-0528.2011.03158.x. PubMed PMID: 22004272.

35.   Reddy UM, Rice MM, Grobman WA, Bailit JL, Wapner RJ, Varner MW, et al. Serious maternal complications after early preterm delivery (24-33 weeks' gestation). American journal of obstetrics and gynecology. 2015;213(4):538 e1-9. doi: 10.1016/j.ajog.2015.06.064. PubMed PMID: 26164696; PubMed Central PMCID: PMC4587283.

36.   Henderson J, Carson C, Redshaw M. Impact of preterm birth on maternal well-being and women's perceptions of their baby: a population-based survey. BMJ open. 2016;6(10):e012676. doi: 10.1136/bmjopen-2016-012676. PubMed PMID: 27855105; PubMed Central PMCID: PMCPMC5073632.

37.   Lykke JA, Paidas MJ, Damm P, Triche EW, Kuczynski E, Langhoff-Roos J. Preterm delivery and risk of subsequent cardiovascular morbidity and type-II diabetes in the mother. BJOG. 2010;117(3):274-81. doi: 10.1111/j.1471-0528.2009.02448.x. PubMed PMID: 20015308.

38.   Bokslag A, van Weissenbruch M, Mol BW, de Groot CJ. Preeclampsia; short and long-term consequences for mother and neonate. Early Hum Dev. 2016;102:47-50. doi: 10.1016/j.earlhumdev.2016.09.007. PubMed PMID: 27659865.

39.   Pauli JM, Repke JT. Preeclampsia: Short-term and Long-term Implications. Obstet Gynecol Clin North Am. 2015;42(2):299-313. doi: 10.1016/j.ogc.2015.01.007. PubMed PMID: 26002168.

40.   Malcolm J. Through the looking glass: gestational diabetes as a predictor of maternal and offspring long-term health. Diabetes Metab Res Rev. 2012;28(4):307-11. doi: 10.1002/dmrr.2275. PubMed PMID: 22228678.

41.   Wendland EM, Torloni MR, Falavigna M, Trujillo J, Dode MA, Campos MA, et al. Gestational diabetes and pregnancy outcomes--a systematic review of the World Health Organization (WHO) and the International Association of Diabetes in Pregnancy Study Groups (IADPSG) diagnostic criteria. BMC pregnancy and childbirth. 2012;12:23. doi: 10.1186/1471-2393-12-23. PubMed PMID: 22462760; PubMed Central PMCID: PMCPMC3352245.

42.   Knight AK, Smith AK, Conneely KN, Dalach P, Loke YJ, Cheong JL, et al. Relationship between Epigenetic Maturity and Respiratory Morbidity in Preterm Infants. The Journal of pediatrics. 2018;198:168-73 e2. doi: 10.1016/j.jpeds.2018.02.074. PubMed PMID: 29705119.

43.   Manuck TA, Rice MM, Bailit JL, Grobman WA, Reddy UM, Wapner RJ, et al. Preterm neonatal morbidity and mortality by gestational age: a contemporary cohort. American journal of obstetrics and gynecology. 2016;215(1):103 e1- e14. doi: 10.1016/j.ajog.2016.01.004. PubMed PMID: 26772790; PubMed Central PMCID: PMC4921282.

44.   Outcome at 2 years of children 23-27 weeks' gestation born in Victoria in 1991-92. The Victorian Infant Collaborative Study Group. J Paediatr Child Health. 1997;33(2):161-5. PubMed PMID: 9145362.

45.   Gimenez LG, Krupitzki HB, Momany AM, Gili JA, Poletta FA, Campana H, et al. Maternal and neonatal epidemiological features in clinical subtypes of preterm birth. The journal

of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet. 2015:1-9. doi: 10.3109/14767058.2015.1118035. PubMed PMID: 26701680.

46.   World Health O. Born too soon: the global action report on preterm birth. 2012.

47.   Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, Requejo JH, et al. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. Bull World Health Organ. 2010;88(1):31-8. Epub 2010/04/30. doi: 10.2471/BLT.08.062554. PubMed PMID: 20428351; PubMed Central PMCID: PMC2802437.

48.   Behrman RE, Butler AS, Institute of Medicine (U.S.). Committee on Understanding Premature Birth and Assuring Healthy Outcomes. Preterm birth : causes, consequences, and prevention. Washington, D.C.: National Academies Press; 2007. xvi, 772 p. p.

49.   Yuan W, Basso O, Sorensen HT, Olsen J. Indicators of fetal growth and infectious disease in childhood--a birth cohort with hospitalization as outcome. Eur J Epidemiol. 2001;17(9):829-34. Epub 2002/06/26. PubMed PMID: 12081101.

50.   Burdge GC, Hanson MA, Slater-Jefferies JL, Lillycrop KA. Epigenetic regulation of transcription: a mechanism for inducing variations in phenotype (fetal programming) by differences in nutrition during early life? Br J Nutr. 2007;97(6):1036-46. Epub 2007/03/27. doi: S0007114507682920 [pii]

10.1017/S0007114507682920. PubMed PMID: 17381976; PubMed Central PMCID: PMC2211525.

51.   Kajantie E, Osmond C, Barker DJ, Eriksson JG. Preterm birth--a risk factor for type 2 diabetes? The Helsinki birth cohort study. Diabetes Care. 2010;33(12):2623-5. Epub 2010/09/09. doi: dc10-0912 [pii]

10.2337/dc10-0912. PubMed PMID: 20823347; PubMed Central PMCID: PMC2992201.

52.   Davies EL, Bell JS, Bhattacharya S. Preeclampsia and preterm delivery: A population-based case-control study. Hypertens Pregnancy. 2016;35(4):510-9. doi: 10.1080/10641955.2016.1190846. PubMed PMID: 27322489.

53.   Mitanchez D, Yzydorczyk C, Siddeek B, Boubred F, Benahmed M, Simeoni U. The offspring of the diabetic mother--short- and long-term implications. Best Pract Res Clin Obstet Gynaecol. 2015;29(2):256-69. doi: 10.1016/j.bpobgyn.2014.08.004. PubMed PMID: 25267399.

54.   Fridman M, Korst LM, Chow J, Lawton E, Mitchell C, Gregory KD. Trends in maternal morbidity before and during pregnancy in California. American journal of public health. 2014;104 Suppl 1:S49-57. doi: 10.2105/AJPH.2013.301583. PubMed PMID: 24354836; PubMed Central PMCID: PMCPMC4011109.

55.   Edwards JE, Hanke JC. An update on maternal mortality and morbidity in the United States. Nurs Womens Health. 2013;17(5):376-88. doi: 10.1111/1751-486X.12061. PubMed PMID: 24138657.

56.   Kramer MS, Seguin L, Lydon J, Goulet L. Socio-economic disparities in pregnancy outcome: why do the poor fare so poorly? Paediatric and perinatal epidemiology. 2000;14(3):194-210. PubMed PMID: 10949211.

57.   Kyrklund-Blomberg NB, Granath F, Cnattingius S. Maternal smoking and causes of very preterm birth. Acta obstetricia et gynecologica Scandinavica. 2005;84(6):572-7. doi: 10.1111/j.0001-6349.2005.00848.x. PubMed PMID: 15901269.

58.   Needham BL, Adler N, Gregorich S, Rehkopf D, Lin J, Blackburn EH, et al. Socioeconomic status, health behavior, and leukocyte telomere length in the National Health and Nutrition Examination Survey, 1999-2002. Social science & medicine. 2013;85:1-8. doi: 10.1016/j.socscimed.2013.02.023. PubMed PMID: 23540359; PubMed Central PMCID: PMC3666871.

59.   Shinjo A, Ventura W, Koide K, Hori K, Yotsumoto J, Matsuoka R, et al. Maternal smoking and placental expression of a panel of genes related to angiogenesis and oxidative stress in early pregnancy. Fetal diagnosis and therapy. 2014;35(4):289-95. doi: 10.1159/000357704. PubMed PMID: 24642658.

60.   Straughen JK, Caldwell CH, Young AA, Jr., Misra DP. Partner support in a cohort of African American families and its influence on pregnancy outcomes and prenatal health behaviors. BMC pregnancy and childbirth. 2013;13:187. doi: 10.1186/1471-2393-13-187. PubMed PMID: 24134672; PubMed Central PMCID: PMC3924355.

61.   Parets SE, Knight AK, Smith AK. Insights into genetic susceptibility in the etiology of spontaneous preterm birth. Appl Clin Genet. 2015;8:283-90. doi: 10.2147/TACG.S58612. PubMed PMID: 26715857; PubMed Central PMCID: PMC4685889.

62.   Chen D, Hu Y, Wu B, Chen L, Fang Z, Yang F, et al. Tumor necrosis factor-alpha gene G308A polymorphism is associated with the risk of preterm delivery. Beijing Da Xue Xue Bao. 2003;35(4):377-81. PubMed PMID: 12947552.

63.   Engel SA, Erichsen HC, Savitz DA, Thorp J, Chanock SJ, Olshan AF. Risk of spontaneous preterm birth is associated with common proinflammatory cytokine polymorphisms. Epidemiology. 2005;16(4):469-77. PubMed PMID: 15951664.

64.   Haataja R, Karjalainen MK, Luukkonen A, Teramo K, Puttonen H, Ojaniemi M, et al. Mapping a new spontaneous preterm birth susceptibility gene, IGF1R, using linkage, haplotype sharing, and association analysis. PLoS genetics. 2011;7(2):e1001293. doi: 10.1371/journal.pgen.1001293. PubMed PMID: 21304894; PubMed Central PMCID: PMCPMC3033387.

65.   Hartel C, Finas D, Ahrens P, Kattner E, Schaible T, Muller D, et al. Polymorphisms of genes involved in innate immunity: association with preterm delivery. Molecular human reproduction. 2004;10(12):911-5. doi: 10.1093/molehr/gah120. PubMed PMID: 15516360.

66.   Romero R, Velez Edwards DR, Kusanovic JP, Hassan SS, Mazaki-Tovi S, Vaisbuch E, et al. Identification of fetal and maternal single nucleotide polymorphisms in candidate genes that predispose to spontaneous preterm labor with intact membranes. American journal of obstetrics and gynecology. 2010;202(5):431 e1-34. doi: 10.1016/j.ajog.2010.03.026. PubMed PMID: 20452482; PubMed Central PMCID: PMC3604889.

67.   Rappoport N, Toung J, Hadley D, Wong RJ, Fujioka K, Reuter J, et al. A genome-wide association study identifies only two ancestry specific variants associated with spontaneous preterm birth. Scientific reports. 2018;8(1):226. doi: 10.1038/s41598-017-18246-5. PubMed PMID: 29317701; PubMed Central PMCID: PMCPMC5760643.

68.   Huusko JM, Karjalainen MK, Graham BE, Zhang G, Farrow EG, Miller NA, et al. Whole exome sequencing reveals HSPA1L as a genetic risk factor for spontaneous preterm birth. PLoS genetics. 2018;14(7):e1007394. doi: 10.1371/journal.pgen.1007394. PubMed PMID: 30001343; PubMed Central PMCID: PMCPMC6042692 Inc.

69.   Gray KJ, Kovacheva VP, Mirzakhani H, Bjonnes AC, Almoguera B, DeWan AT, et al. Gene-Centric Analysis of Preeclampsia Identifies Maternal Association at PLEKHG1.

Hypertension. 2018;72(2):408-16. doi: 10.1161/HYPERTENSIONAHA.117.10688. PubMed PMID: 29967039; PubMed Central PMCID: PMCPMC6043396.

70. Hansen AT, Bernth Jensen JM, Hvas AM, Christiansen M. The genetic component of preeclampsia: A whole-exome sequencing study. PloS one. 2018;13(5):e0197217. doi: 10.1371/journal.pone.0197217. PubMed PMID: 29758065; PubMed Central PMCID: PMCPMC5951572.

71. Knight AK, Smith AK. Epigenetic Biomarkers of Preterm Birth and Its Risk Factors. Genes (Basel). 2016;7(4). doi: 10.3390/genes7040015. PubMed PMID: 27089367; PubMed Central PMCID: PMC4846845.

72. Parets SE, Bedient CE, Menon R, Smith AK. Preterm birth and its long-term effects: methylation to mechanisms. Biology. 2014;3(3):498-513. doi: 10.3390/biology3030498. PubMed PMID: 25256426; PubMed Central PMCID: PMC4192624.

73. Bianco-Miotto T, Lopez CR, Leemaqz S, Buckberry S, McCullough D, Zhuang Z, et al. DNA methylation biomarkers for predicting pregnancy complications. Placenta. 2015;36(9):A38-A9. doi: http://dx.doi.org/10.1016/j.placenta.2015.07.302.

74. Enquobahrie DA, Moore A, Muhie S, Tadesse MG, Lin S, Williams MA. Early Pregnancy Maternal Blood DNA Methylation in Repeat Pregnancies and Change in Gestational Diabetes Mellitus Status-A Pilot Study. Reprod Sci. 2015;22(7):904-10. doi: 10.1177/1933719115570903. PubMed PMID: 25676578; PubMed Central PMCID: PMC4565480.

75. Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, et al. Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans.

Human genetics. 2013;132(9):1027-37. doi: 10.1007/s00439-013-1311-6. PubMed PMID: 23657504; PubMed Central PMCID: PMC3744600.

76. Virani S, Rentschler KM, Nishijo M, Ruangyuttikarn W, Swaddiwudhipong W, Basu N, et al. DNA methylation is differentially associated with environmental cadmium exposure based on sex and smoking status. Chemosphere. 2016;145:284-90. doi: 10.1016/j.chemosphere.2015.10.123. PubMed PMID: 26688266; PubMed Central PMCID: PMC5047795.

77. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PloS one. 2013;8(5):e63812. doi: 10.1371/journal.pone.0063812. PubMed PMID: 23691101; PubMed Central PMCID: PMC3656907.

78. Parets SE, Conneely KN, Kilaru V, Fortunato SJ, Syed TA, Saade G, et al. Fetal DNA Methylation Associates with Early Spontaneous Preterm Birth and Gestational Age. PloS one. 2013;8(6):e67489. doi: 10.1371/journal.pone.0067489. PubMed PMID: 23826308; PubMed Central PMCID: PMC3694903.

79. Cruickshank MN, Oshlack A, Theda C, Davis PG, Martino D, Sheehan P, et al. Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. Genome medicine. 2013;5(10):96. doi: 10.1186/gm500. PubMed PMID: 24134860; PubMed Central PMCID: PMC3978871.

80. Fernando F, Keijser R, Henneman P, van der Kevie-Kersemaekers AM, Mannens MM, van der Post JA, et al. The idiopathic preterm delivery methylation profile in umbilical cord blood DNA. BMC genomics. 2015;16:736. doi: 10.1186/s12864-015-1915-4. PubMed PMID: 26419829; PubMed Central PMCID: PMC4588235.

81.   Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89-95. Epub 2001/03/10. doi: S0009-9236(01)63448-9 [pii] 10.1067/mcp.2001.113989. PubMed PMID: 11240971.

82.   Holsboer F. How can we realize the promise of personalized antidepressant medicines? Nat Rev Neurosci. 2008;9(8):638-46. doi: 10.1038/nrn2453. PubMed PMID: 18628772.

83.   Peedicayil J. Epigenetic biomarkers in psychiatric disorders. Br J Pharmacol. 2008;155(6):795-6. doi: 10.1038/bjp.2008.254. PubMed PMID: 18574461; PubMed Central PMCID: PMC2597243.

84.   Alliance NBD. Biomarker Validation 2015 [cited 2015]. Available from: http://nbdabiomarkers.org/about/what-we-do/pipeline-overview/biomarker-validation.

85.   Tong VT, Dietz PM, Morrow B, D'Angelo DV, Farr SL, Rockhill KM, et al. Trends in Smoking Before, During, and After Pregnancy - Pregnancy Risk Assessment Monitoring System, United States, 40 Sites, 2000-2010. Mmwr Surveill Summ. 2013;62(6):1-19. PubMed PMID: WOS:000327810400001.

86.   Aliyu MH, Salihu HM, Wilson RE, Alio AP, Kirby RS. The risk of intrapartum stillbirth among smokers of advanced maternal age. Arch Gynecol Obstet. 2008;278(1):39-45. doi: 10.1007/s00404-007-0529-8. PubMed PMID: 18236064.

87.   Rua Ede A, Porto ML, Ramos JP, Nogueira BV, Meyrelles SS, Vasquez EC, et al. Effects of tobacco smoking during pregnancy on oxidative stress in the umbilical cord and mononuclear blood cells of neonates. Journal of biomedical science. 2014;21:105. doi: 10.1186/s12929-014-0105-z. PubMed PMID: 25547987; PubMed Central PMCID: PMC4302517.

88.   Marufu TC, Ahankari A, Coleman T, Lewis S. Maternal smoking and the risk of still birth: systematic review and meta-analysis. BMC public health. 2015;15:239. doi:

10.1186/s12889-015-1552-5. PubMed PMID: 25885887; PubMed Central PMCID: PMC4372174.

89.   Isayama T, Shah PS, Ye XY, Dunn M, Da Silva O, Alvaro R, et al. Adverse Impact of Maternal Cigarette Smoking on Preterm Infants: A Population-Based Cohort Study. American journal of perinatology. 2015. doi: 10.1055/s-0035-1548728. PubMed PMID: 25825966.

90.   Aliyu MH, Lynch O, Wilson RE, Alio AP, Kristensen S, Marty PJ, et al. Association between tobacco use in pregnancy and placenta-associated syndromes: a population-based study. Arch Gynecol Obstet. 2011;283(4):729-34. doi: 10.1007/s00404-010-1447-8. PubMed PMID: 20354707.

91.   Smith LK, Draper ES, Evans TA, Field DJ, Johnson SJ, Manktelow BN, et al. Associations between late and moderately preterm birth and smoking, alcohol, drug use and diet: a population-based case-cohort study. Archives of disease in childhood Fetal and neonatal edition. 2015. doi: 10.1136/archdischild-2014-307265. PubMed PMID: 25972442.

92.   Anthopolos R, Edwards SE, Miranda ML. Effects of maternal prenatal smoking and birth outcomes extending into the normal range on academic performance in fourth grade in North Carolina, USA. Paediatric and perinatal epidemiology. 2013;27(6):564-74. doi: 10.1111/ppe.12081. PubMed PMID: 24134528.

93.   Hogberg L, Cnattingius S, Lundholm C, D'Onofrio BM, Langstrom N, Iliadou AN. Effects of maternal smoking during pregnancy on offspring blood pressure in late adolescence. Journal of hypertension. 2012;30(4):693-9. doi: 10.1097/HJH.0b013e32835168f4. PubMed PMID: 22388229; PubMed Central PMCID: PMC3664185.

94.   Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and

Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Human molecular genetics. 2013;22(5):843-51. doi: 10.1093/hmg/dds488. PubMed PMID: 23175441.

95.   Stroud LR, Papandonatos GD, Rodriguez D, McCallum M, Salisbury AL, Phipps MG, et al. Maternal smoking during pregnancy and infant stress response: test of a prenatal programming hypothesis. Psychoneuroendocrinology. 2014;48:29-40. doi: 10.1016/j.psyneuen.2014.05.017. PubMed PMID: 24999830; PubMed Central PMCID: PMC4136384.

96.   Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environmental health perspectives. 2012;120(10):1425-31. doi: 10.1289/ehp.1205412. PubMed PMID: 22851337; PubMed Central PMCID: PMC3491949.

97.   Ivorra C, Fraga MF, Bayon GF, Fernandez AF, Garcia-Vicent C, Chaves FJ, et al. DNA methylation patterns in newborns exposed to tobacco in utero. Journal of translational medicine. 2015;13(1):25. doi: 10.1186/s12967-015-0384-5. PubMed PMID: 25623364; PubMed Central PMCID: PMC4312439.

98.   Kupers LK, Xu X, Jankipersadsing SA, Vaez A, la Bastide-van Gemert S, Scholtens S, et al. DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. International journal of epidemiology. 2015. doi: 10.1093/ije/dyv048. PubMed PMID: 25862628.

99.   Flom J, Ferris J, Gonzalez K, Santella R, Terry MB. Prenatal tobacco smoke exposure and genomewide methylation in adulthood. Cancer Epidemiology Biomarkers & Prevention. 2011;20(4):720-.

100. Lee KW, Richmond R, Hu P, French L, Shin J, Bourdon C, et al. Prenatal exposure to maternal cigarette smoking and DNA methylation: epigenome-wide association in a discovery sample of adolescents and replication in an independent cohort at birth through 17 years of age. Environmental health perspectives. 2015;123(2):193-9. doi: 10.1289/ehp.1408614. PubMed PMID: 25325234; PubMed Central PMCID: PMC4314251.

101. Horvath S. DNA methylation age of human tissues and cell types. Genome biology. 2013;14(10):R115. doi: 10.1186/gb-2013-14-10-r115. PubMed PMID: 24138928; PubMed Central PMCID: PMC4015143.

102. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Molecular cell. 2013;49(2):359-67. doi: 10.1016/j.molcel.2012.10.016. PubMed PMID: 23177740; PubMed Central PMCID: PMC3780611.

103. Koch CM, Wagner W. Epigenetic-aging-signature to determine age in different tissues. Aging. 2011;3(10):1018-27. PubMed PMID: 22067257; PubMed Central PMCID: PMC3229965.

104. Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D, et al. Accelerated epigenetic aging in Down syndrome. Aging cell. 2015. doi: 10.1111/acel.12325. PubMed PMID: 25678027.

105. Boks MP, van Mierlo HC, Rutten BP, Radstake TR, De Witte L, Geuze E, et al. Longitudinal changes of telomere length and epigenetic age related to traumatic stress and post-traumatic stress disorder. Psychoneuroendocrinology. 2015;51:506-12. doi: 10.1016/j.psyneuen.2014.07.011. PubMed PMID: 25129579.

106. Simpkin AJ, Hemani G, Suderman M, Gaunt TR, Lyttleton O, McArdle WL, et al. Prenatal and early life influences on epigenetic age in children: A study of mother-offspring pairs from two cohort studies. Human molecular genetics. 2015. doi: 10.1093/hmg/ddv456. PubMed PMID: 26546615.

107. Schroeder JW, Conneely KN, Cubells JC, Kilaru V, Newport DJ, Knight BT, et al. Neonatal DNA methylation patterns associate with gestational age. Epigenetics : official journal of the DNA Methylation Society. 2011;6(12):1498-504. doi: 10.4161/epi.6.12.18296. PubMed PMID: 22139580; PubMed Central PMCID: PMC3256334.

108. Simpkin AJ, Suderman M, Gaunt TR, Lyttleton O, McArdle WL, Ring SM, et al. Longitudinal analysis of DNA methylation associated with birth weight and gestational age. Human molecular genetics. 2015. doi: 10.1093/hmg/ddv119. PubMed PMID: 25869828.

109. Lee H, Jaffe AE, Feinberg JI, Tryggvadottir R, Brown S, Montano C, et al. DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. International journal of epidemiology. 2012;41(1):188-99. doi: 10.1093/ije/dyr237. PubMed PMID: 22422452; PubMed Central PMCID: PMC3304532.

110. Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. Science. 2010;330(6004):612-6. doi: 10.1126/science.1191078. PubMed PMID: 21030644; PubMed Central PMCID: PMC3772643.

111. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010;466(7303):253-7. Epub 2010/07/09. doi: nature09165 [pii]

10.1038/nature09165. PubMed PMID: 20613842.

112.   Kennedy EM, Goehring G, Nichols M, Robins C, Mehta D, Klengel T, et al. An integrated -omics analysis of the epigenetic landscape of gene expression in human blood cells. BMC genomics. 2018.

113.   Ptashne M. Regulation of transcription: from lambda to eukaryotes. Trends in biochemical sciences. 2005;30(6):275-9. doi: 10.1016/j.tibs.2005.04.003. PubMed PMID: 15950866.

114.   Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. Gene regulation by the act of long non-coding RNA transcription. BMC biology. 2013;11:59. doi: 10.1186/1741-7007-11-59. PubMed PMID: 23721193; PubMed Central PMCID: PMCPMC3668284.

Chapter 2:

Characterization of Gene Expression Changes Over Healthy Term Pregnancies

Introduction

Pregnancy is characterized by extensive physiological changes including increased blood volume, elevated levels of estrogen and progesterone, changes in metabolism, and shifts in the maternal immune system in order to accommodate the demands of the growing fetus [1, 2]. Despite these carefully regulated physiological changes that occur throughout the 40 weeks of an average pregnancy, no study has yet described accompanying changes in maternal blood gene expression among healthy women with full term, uncomplicated pregnancies.

To date, studies of gene expression focus on comparisons of pregnant women who are healthy to those with autoimmune disorders. This has been prompted, in part, by the observation that women with some autoimmune disorders report alleviation of their symptoms during pregnancy [3]. For example, a recent study of peripheral blood from 20 women with rheumatoid arthritis and 5 healthy controls identified 4,710 genes that differed in their expression over pregnancy and the postpartum period. The genes identified were enriched for a variety of pathways including immune pathways, signal transduction, and disease-related pathways. Interestingly, several of the genes identified by this study, are members of the alpha defensin family and are involved in immune and defense responses [4]. Another study identified 1,286 transcripts whose expression levels changed over three timepoints in pregnancy and the postpartum period for women with rheumatoid arthritis and controls. These transcripts were enriched for a variety of pathways including hematopoietic cell lineage and toll-like receptor signaling. The authors concluded that the identified pathways may contribute to immunomodulation and thus a reduction in rheumatoid arthritis symptoms during pregnancy with these changes in gene expression being largely reversed postpartum [5]. Gilli and colleagues sampled multiple sclerosis patients and healthy controls before pregnancy, during

each trimester, and post-partum to identify 404 transcripts whose expression differed between multiple sclerosis patients and healthy controls. A refined signature of 347 transcripts was then used to evaluate gene expression in patients and controls in the ninth month of pregnancy, at which time the signature could no longer distinguish patients from controls [6]. These studies indicate that gene expression does change over pregnancy, potentially leading to alleviation of some of the symptoms of autoimmune disorders.

Although such studies have been informative for identifying pathways that change in women with autoimmune disorders during pregnancy, if the groups are not sampled at comparable stages of gestation, the findings may be more difficult to interpret or replicate in independent studies. This study sought to characterize gene expression changes over pregnancy in a cohort of healthy women with uncomplicated term deliveries. The pathways identified provide further insight into the connection between gene expression and physiological changes that occur over the course of a healthy, full term pregnancy, and will serve as a resource for ongoing studies of pregnancy complications and adverse pregnancy outcomes.

Methods

Study Subjects

Pregnant African American women were recruited, enrolled, and underwent data collection as part of the Emory University African American Vaginal, Oral, and Gut Microbiome in Pregnancy Cohort Study, as described previously [7]. To summarize, women were recruited from March 2014 through August 2016 at outpatient prenatal care clinics affiliated with two Atlanta metro area hospitals, Emory University Midtown Hospital and Grady Memorial Hospital. Women were eligible for inclusion if they self-identified as African American, were

between 18-40 years of age, had a singleton pregnancy, had less than four previous births, and were able to understand written and spoken English. Additional exclusion criteria for this analysis included the following indicators of high-risk pregnancy or pregnancy complications: gestational diabetes, hypertension, intrauterine growth restriction, preterm birth, preeclampsia, eclampsia, hemolysis, elevated liver enzymes, low platelet count (HELLP) syndrome, hyperemesis gravidarum, oligohydramnios, chorioamnionitis, macrosomia, preterm premature rupture of membranes (pPROM), or fetal intolerance of labor. Fetal death before labor and congenital abnormalities of the fetus were criteria for post-enrollment exclusion. Demographic data were collected through self-report questionnaires. Clinical obstetrical data (including estimated due date, gestational age at delivery, pregnancy complications, labor and delivery course) were ascertained via abstraction of the prenatal and labor and delivery medical chart by a qualified physician. All participants gave written informed consent. This study was approved by the Emory University Institutional Review Board.

Biological Sample Collection

63 women contributed two samples each over the course of their pregnancy. The first sample was collected at 6-15 weeks, and the second sample was collected at 22-33 weeks. During each prenatal visit, an additional 12 mL of venous blood was drawn using the same needle stick as for the routine blood draws. PBMCs were isolated from whole blood using a Ficoll density gradient and were stored in AllProtect Buffer (Qiagen) at -80 °C until a simultaneous DNA and RNA extraction using the AllPrep RNA/DNA Mini Kit (Qiagen) was performed according to manufacturer's instructions. DNA quantification and quality was assessed using the Quant-it Pico Green kit (Invitrogen). RNA quantification and quality was assessed using the Aligent RNA Nano 6000 Kit and Bioanalyzer 2100.

RNA Expression Analysis

For each subject, gene expression was assessed for ~47,000 transcripts using the HumanHT-12 v4 BeadChip (Illumina). Briefly, 750 ng of RNA was directly hybridized to the BeadChip according to manufacturer's instructions. The BeadChips were scanned using the iScan scanner, and the raw data was analyzed using the Expression Module of GenomeStudio Software (Illumina). Two samples with detection p values >.01 for more than 90% of probes were excluded. Probes detected in less than 10% of samples were also excluded; 16,311 probes passed quality control. Data was then quantile-normalized and log2 transformed prior to association testing. RNA expression data can be accessed through NCBI's Gene Expression Omnibus, GSE107437.

Cell type composition estimation

DNA methylation was interrogated for each subject using either the HumanMethylation450 or MethylationEPIC BeadChip, which measures methylated and unmethylated signal for >450,000 and >850,000 CpG sites across the genome, respectively. Initial data quality control was performed using the R package CpGassoc [8]. Any CpG site with low signal or missing data for greater than 10% of samples was removed, and any sample with missing data for greater than 5% of CpG sites was removed. Cross-reactive probes were removed [9]. Following quality control, 449,094 probes were included in subsequent analyses. Beta values (β) were calculated for each CpG site as the ratio of methylated (M) to methylated and unmethylated (U) signal: $\beta=M/(M+U)$. Beta-mixture quantile normalization was performed as previously described [10]. Cell type proportions (CD8+T, CD4+T, natural killer, B cell, monocytes, and granulocytes) were estimated as previously described from DNA methylation data [11]. Associations between cell type proportions and gestational age were examined using a

linear-mixed effects model with cellular proportions included as fixed effects and a unique identifier for each person included as a random effect. DNA methylation data can be accessed through NCBI's Gene Expression Omnibus, GSE107459.

Whole Transcriptome Analysis

We used linear mixed-effects modeling implemented in the R package "nlme" to interrogate associations between gene expression and weeks gestation at sample collection [12].The R package sva was used to estimate surrogate variables to control for potentially confounding factors, including cell type [13]. Surrogate variable analysis was used instead of a covariate adjustment as only a few cellular subtypes are well defined using current methods, and changes in the composition of unmeasured cell types may have a large influence during pregnancy. The 15 significant surrogate variables were included as covariates in the model as fixed effects (Equation 1).

Eq. 1: $Y_{ij} = \beta_0 + \beta_1 X_{1_{ij}} + \beta_2 SV_{1_{ij}} + \beta_3 SV_{2_{ij}} + \cdots + \beta_{16} SV_{15_{ij}} + u_i + \varepsilon_{ij}; \ u_i \sim N(0, \tau^2), \varepsilon_{ij} \sim N(0, \sigma^2)$

Where SV$_{1-15}$ represent each included surrogate variable, which model unmeasured factors inferred from the genome-wide methylation signatures (Fig 2-1). X represents the independent variable. β_{1-16} represent fixed effects parameters. u is the individual specific error term, and i represents the unique subject identifier, and j represents the observation number. A random effects term was included in the model to account for repeated sampling of the same person. ε refers to random error. τ^2 and σ^2 are the variances of the person specific and random error terms, respectively. A Bonferroni correction was applied to account for multiple testing. Pathway analysis was performed using DAVID for both the entire set of associated genes as well as for upregulated and downregulated genes separately [14]. Gene-ontology enrichment p-values presented used.

Results

Study Subjects

63 pregnant African American women provided venous blood samples between 6-15 weeks and again between 22-33 weeks gestation, representing the late first/early second trimesters and the late second/early third trimesters. The clinical and demographic features of these women are summarized in Table 2-1. At the time of enrollment, 7 women (12%) report using tobacco within the last 30 days, and 4 women (7%) report drinking alcohol within the last 30 days. Despite this, no woman delivered a neonate that was small for gestational age or experienced any complications over their pregnancies (see Methods for additional details).

Changes in cell proportions over pregnancy

Studies have reported changes in a range of cell types and immune activation profiles over pregnancy [15-17]. Consistent with those previous reports, we observed changes in the proportion of cell types in maternal PBMCs over the course of pregnancy. There was a significant increase in the proportion of monocytes (p=.001), along with a significant decrease in the proportion of B cells (p=.03), and natural killer cells (p=.004) based on week of gestation (Fig 2-1. Estimates of cell composition used for this study are not comprehensive, and there may be additional changes in cell composition or function that are not currently captured. To account for such changes in cell type and the potential for other known (i.e. age) or unknown confounders, we used surrogate variable analysis (SVA), which identifies and estimates sources of expression heterogeneity to increase power to detect true and replicable associations. Post-hoc evaluation of surrogate variables suggests that they reflect changes in cell composition, age, and control for technical variables such as batch (Fig 2-2).

Gene expression changes over pregnancy

Of the 16,311 transcripts that were evaluated in this transcriptome-wide analysis (Fig 2-3), 439 associated with weeks of gestation at sample collection, after adjusting for multiple testing. The majority (69.6%) of these genes increased in expression over the course of pregnancy. Genes whose expression changed were enriched for multiple biological processes (Table 2-2). Of note, many of these genes are in involved in oxygen transport or the hemoglobin complex (e.g. *AHSP*, *HBD*, *HBM*, and *HBQ1*). In addition to increased expression of genes involved in oxygen transport, other associated biological processes emphasize changes in the immune system known to occur over pregnancy. Several of these immune processes are associated with response to microbes, including the antibacterial humoral response and the innate immune response in mucosa (Table 2-2). Key genes in these pathways are members of the alpha defensin family, including *DEFA1*, *DEFA4*, and *DEFA1B* (Fig 2-4). Similar pathways were identified for genes that increased over pregnancy, though there was no enrichment for biological processes in genes whose expression decreased over pregnancy (Table 2-3).

In contextualizing our study findings with the existing literature, we attempted to compare the genes whose expression levels change in healthy pregnancies to those that have been reported to change in women with autoimmune disorders. In some cases, complete gene lists were not provided [5, 6], limiting our ability to make direct comparisons. For example, Weix and colleagues reported 19 selected candidate genes that differed in pregnancy and differed to a greater degree in pregnant women with rheumatoid arthritis. Of those candidate genes, only 1 (*STAT1*) differed in our transcriptome-wide analysis. The study by Mittal and colleagues reported the change in expression of 256 genes over pregnancy in women with rheumatoid arthritis and controls [4], 179 of which were evaluated in our study (Fig 2-5). Of these, 55.3%

change over pregnancy in our cohort of healthy women. This is a substantially higher number than would be expected by chance (99/179; $p<2.2x10^{-16}$). Of note, both this study and the one conducted by Mittal and colleagues report changes in genes involved in oxygen transport (e.g. *AHSP* and *HBD*) and immune response to microbes (*CEACAM8* and *DEFA4*).

Discussion

In this study, we identified 439 transcripts that were associated with weeks gestation at sample collection in uncomplicated term pregnancies. These transcripts were enriched for several key pathways that provide insight into the mechanisms underlying the physiological changes necessary to support a healthy pregnancy.

A widely accepted phenomenon in pregnancy is a change in the maternal immune system. In general, pregnancy is associated with a shift from a pro-inflammatory state in the first trimester to an anti-inflammatory state in the second trimester, with renewed inflammation during the third trimester and at parturition [2]. The pro-inflammatory state in the first trimester is likely a result of implantation and placentation with the anti-inflammatory state of the second trimester being a period of rapid fetal growth and development during which a more symbiotic relationship between mother, fetus, and placenta exists. Finally, in the third trimester, renewed inflammation leads to the processes which can initiate labor and delivery. Consistent with this and other studies, our results suggest that monocytes increase over the course of pregnancy, while B cell and natural killer cell proportions decrease [18, 19]. This shift has been further explored in studies of autoimmune disorders and pregnancy, with several pathways involved being identified that may contribute to reductions, or in other cases, increases in autoimmune symptoms during pregnancy [3-6]. However, such studies often do not control for differences in cellular proportions, increasing the chance that they will identify genes whose expression is not

entirely attributable to the disorder of interest or are otherwise difficult to interpret. Thus, it is vital for future studies to estimate and adjust for cellular heterogeneity, as cell type changes over pregnancy.

In this longitudinal analysis of transcriptome-wide changes across gestational week, we identified gene expression changes that supports functional differences in the immune system during pregnancy. Such immune system changes must be carefully regulated so that the mother is protected from bacterial and viral infection without negatively impacting the body's tolerance of the fetus. However, the immune systems of pregnant women are less able to appropriately respond to several types of bacterial infections and related complications, including *Listeria monocytogenes* and *Neisseria gonorrhoeae* [20-22]. Bacterial infections have been associated with preterm birth and spontaneous abortion [23-25]. Evidence of host-microbe interactions in pregnancy is demonstrated in both the gene-level results and through enrichment of genes involved in antibacterial humoral response and innate immunity in the gene ontology analysis. Specific genes related to bacterial response include several alpha defensin genes. These genes encode antimicrobial peptides (AMPs) that have been associated with microbicidal activity and host defense, and are present in the female reproductive tract during pregnancy [26, 27]. Understanding changes in the immune system and host-microbe interactions during pregnancy may provide future insight into how the immune system acts to protect both the mother and the fetus thus allowing us to more effectively treat and prevent potential maternal and fetal effects of systemic infection.

This study also identified genes involved in oxygen transport that change over pregnancy, which support the physiological changes identified in previous studies [28-31]. Some of the most pronounced adaptations in pregnancy are related to increased blood volume, higher rates of

erythropoiesis, and increased coagulation. Maternal physiology tends toward a more hypercoagulable state during pregnancy, likely in an effort to limit delivery complications such as post-partum hemorrhage. Previous studies have linked abnormal coagulation, specifically within the placental complex, with preeclampsia, demonstrating the importance of regulation of coagulation for healthy pregnancies [32, 33].

Erythropoiesis increases over the course of pregnancy, potentially to accommodate larger blood volumes and prepare for the acute loss of this volume which occurs at the time of delivery. However, this study evaluated PBMCs, which do not include mature erythrocytes [34]. Still, recent studies have reported the expression of hemoglobin in non-erythroid cells, such as macrophages, epithelial, mesangial, cervical, and endometrial cells, and have also reported functions of hemoglobin other than oxygen transport including antioxidant defense, nitrite reduction, and reactive oxygen species scavenging [35, 36]. We hypothesize that the hemoglobin-associated genes identified in this study reflect non-canonical expression in response to the physiological strain that pregnancy places on the body, potentially resulting in further demands for such alternative functions.

This is the first study to comprehensively characterize transcriptome-wide changes longitudinally over the course of uncomplicated pregnancies. This cohort has been prospectively recruited and characterized, resulting is an invaluable asset for studying pregnancy progression [7]. However, this study does have limitations. Though it is the largest longitudinal study of gene expression changes in pregnancy, it still has a modest sample size. A larger group of subjects is likely to identify more genes that change over the course of pregnancy. Despite this, we were able to identify hundreds of genes whose expression change in PBMCs after accounting for confounding factors and multiple testing. However, we cannot extrapolate these changes to

other cell types that may be relevant for pregnancy. This study also uses array-based technology, which does not allow for characterization of novel or alternatively spliced transcripts that may play vary during pregnancy. Finally, we did not have access to samples spanning the entirety of pregnancy, including before 8 weeks and after 33 weeks gestation.

Despite these limitations, we show that gene expression changes significantly over the course of pregnancy, especially in pathways related to oxygen transcript and response to microbes. Our results provide a framework for future studies on changes in gene expression over pregnancy. Knowledge of genes associated with normal changes may potentially allow for identification of abnormal patterns of gene expression associated with pregnancy and delivery complications that could be tested as potential biomarkers. Future studies should examine gene expression over pregnancy in the context of pregnancy complications and other pre-existing conditions, while acknowledging that very subtle changes in the timing of sample collection may confound the results or complicate their interpretation.

Table 2-1: Demographic table of 63 paired gene expression samples

	Mean ± SD
Maternal Age (years)	25.4 ± 4.5
Parity	1.2 ± 1.1
Gravidity	2.8 ± 1.6
Length of Gestation (weeks)	39.4 ± 1.0
Birthweight (grams)	3318 ± 411.8
	N (%)
Delivery Type	
Vaginal	55 (87)
C-section	8 (13)
Insurance Type	
Medicaid	46 (73)
Private	17 (37)
Education	
Some High School	11 (17)
High School Graduate	17 (27)
Some College	25 (40)
College Graduate	7 (11)
Graduate School	3(5)

Table 2-2: Biological Processes associated with genes whose expression changes over pregnancy.

Term	GO Identifier	Count	p-value*
oxygen transport	GO:0015671	6	0.0074
defense response to fungus	GO:0050832	7	0.0081
antibacterial humoral response	GO:0019731	8	0.012
leukocyte migration	GO:0050900	13	0.012
killing of cells of other organism	GO:0031640	5	0.041
innate immune response in mucosa	GO:0002227	6	0.044

* p-values presented after a Benjamini-Hochberg correction for multiple testing.

Table 2-3: Biological Processes associated with genes whose expression increases over pregnancy.

Term	GO Identifier	Count	p-value*
antibacterial humoral response	GO:0019731	8	0.00081
defense response to fungus	GO:0050832	7	0.00086
oxygen transport	GO:0015671	6	0.0015
leukocyte migration	GO:0050900	11	0.002
innate immune response in mucosa	GO:0002227	6	0.0047
killing of cells of other organism	GO:0031640	5	0.0062
heme biosynthetic process	GO:0006783	5	0.029
blood coagulation	GO:0007596	11	0.033

Figure 2-1: Cellular composition changes over pregnancy**.** Monocytes increase over pregnancy, whereas B cells and natural killer cells decrease over pregnancy. Other evaluated cell types do not change. The x-axis represents the weeks of gestation at sample collection and the y-axis represents cell proportions.
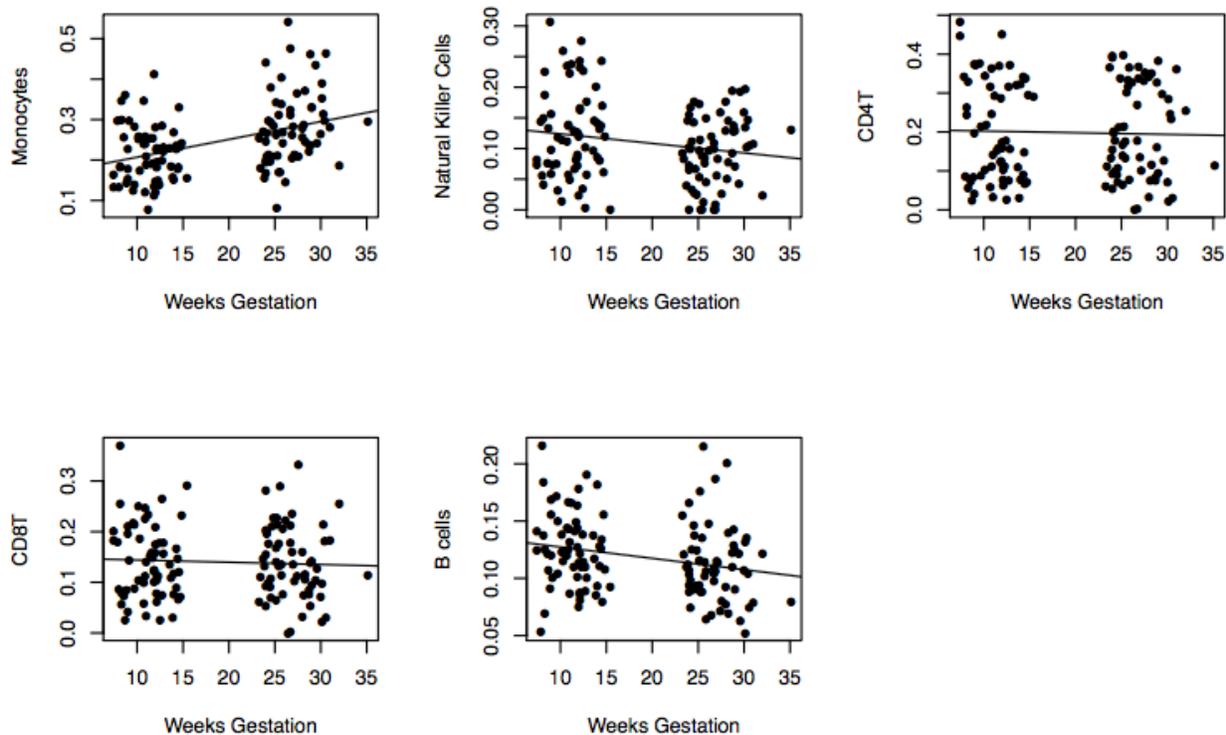
Figure 2-2: Association of surrogate variables and variables typically controlled for in EWAS. The grid consists of the correlation coefficient (r) for each pair on top with the p-value indicating significance of the correlation below. ExpressionRound indicates batch. Comp.1 indicates the first principal component for ancestry. Maternal age was measured in years. Cell composition (NK, CD4$^+$ T cells, monocytes, B cells, granulocytes and CD8$^+$ T cells) were estimated as described in the Methods. The intensity of the shading represents the correlation coefficient, with darker shading being associated with a higher correlation coefficient. Red shading represents a positive correlation coefficient and green shading represents a negative correlation coefficient.

Surrogate variable relationships
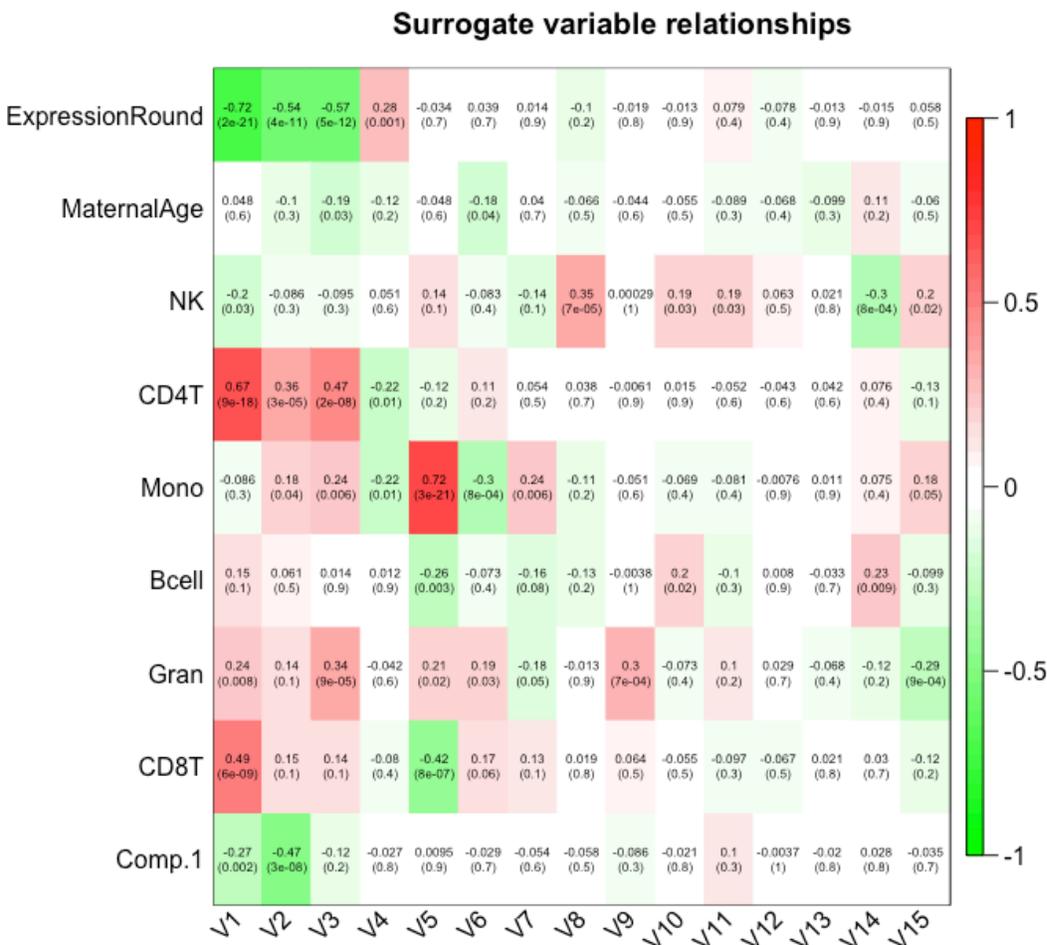
Figure 2-3: Transcriptome-wide distribution of transcripts that change over pregnancy. Volcano plot depicting associations with gene transcripts that change over pregnancy. The x-axis represents the effect size. The y-axis represents the significance level for each test. Light blue points represent experiment-wide significance.
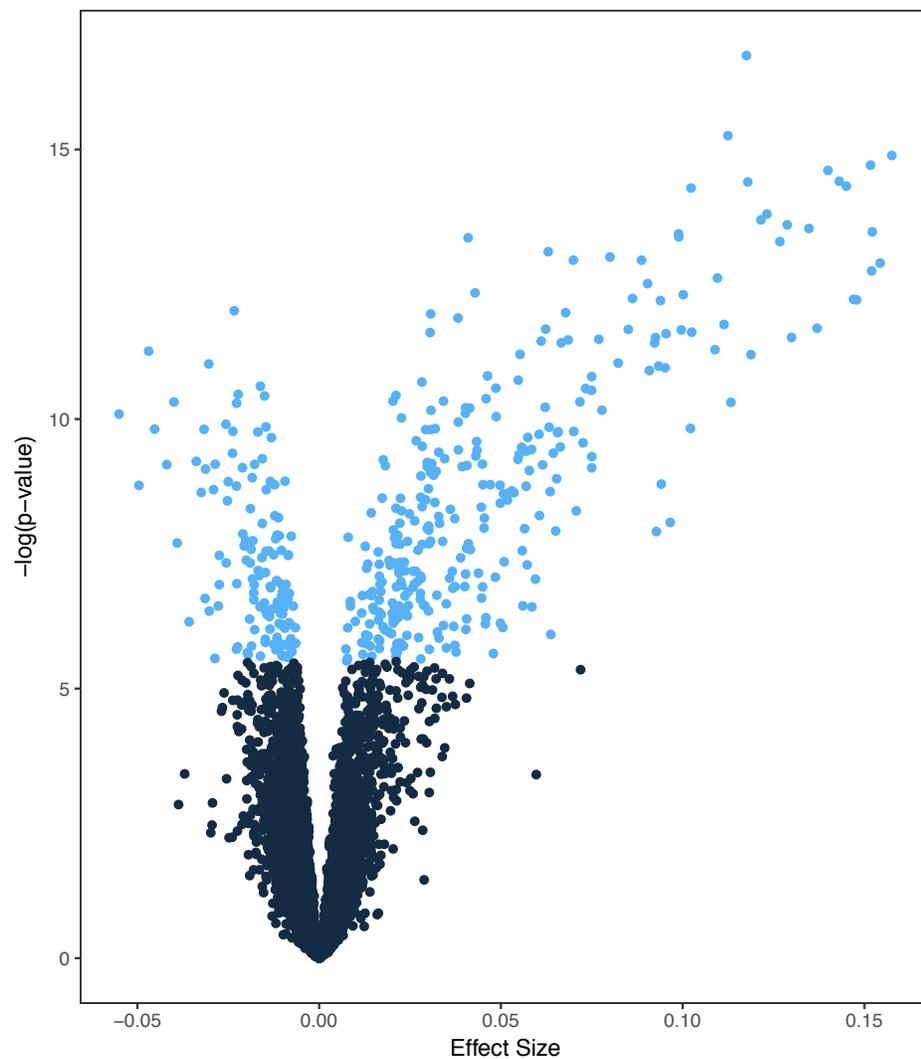
Figure 2-4: Genes in the alpha defensing family that change over pregnancy.

The x-axis represents weeks gestation at sample collection, and the y-axis represents the log2-transfomed expression levels for each transcript. a) *DEFA4* (ILMN_1753347; p=2.45x10^{-15}) b) *CEACAM8* (ILMN_1806056; p=1.57x10^{-14}) c) *DEFA1* (ILMN_1679357; p=3.37x10^{-14}) d) *DEFA1B* (ILMN_2102721; p=1.79x10^{-13}).
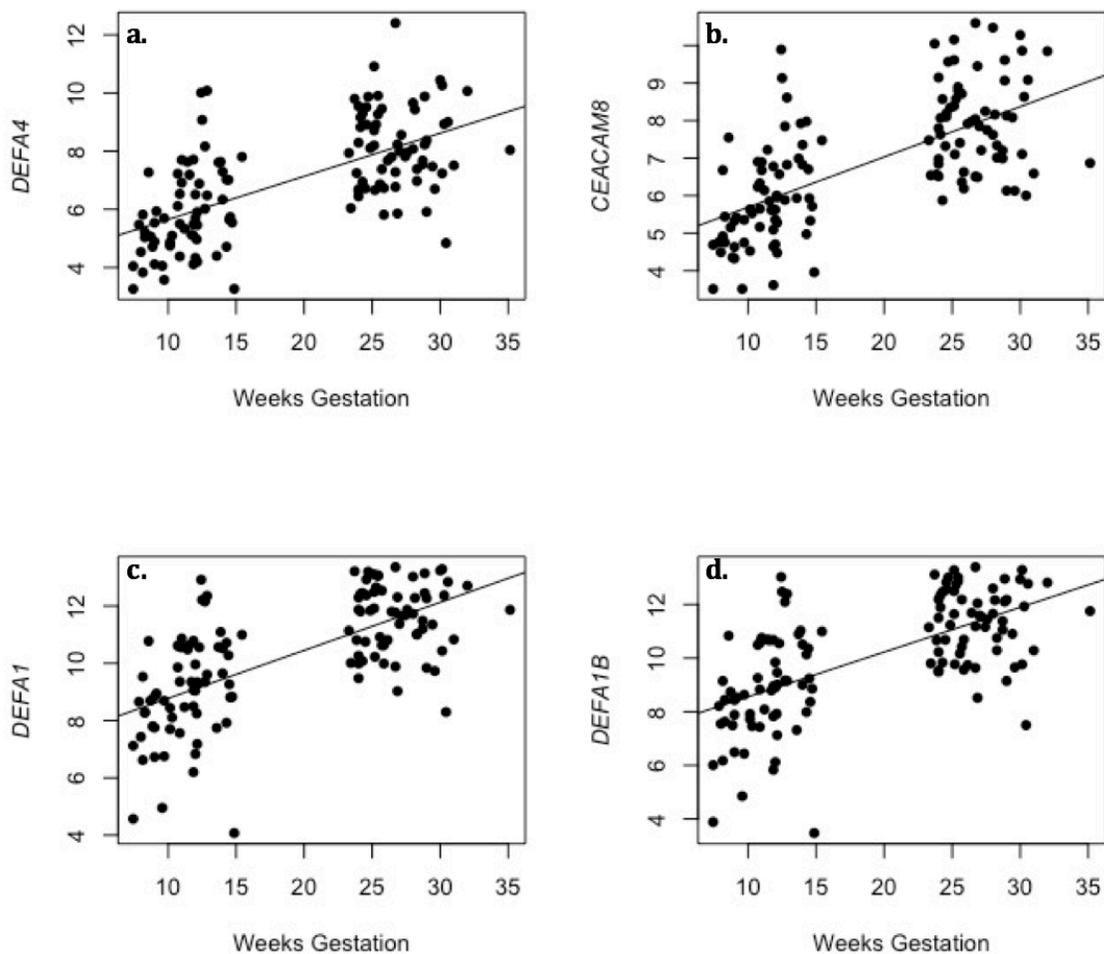
Figure 2-5: Venn diagram showing the overlapping genes between this and previous studies [4, 5].
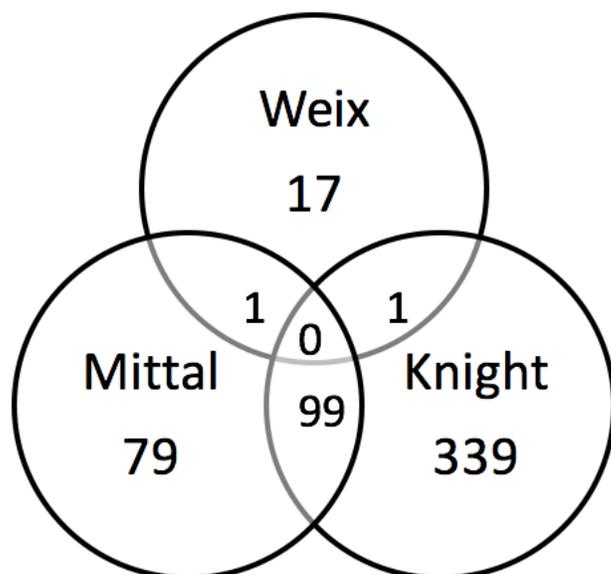
References

1.   Moya J, Phillips L, Sanford J, Wooton M, Gregg A, Schuda L. A review of physiological and behavioral changes during pregnancy and lactation: potential exposure factors and data gaps. J Expo Sci Environ Epidemiol. 2014;24(5):449-58. doi: 10.1038/jes.2013.92. PubMed PMID: 24424408.

2.   Mor G, Cardenas I. The immune system in pregnancy: a unique complexity. American journal of reproductive immunology. 2010;63(6):425-33. doi: 10.1111/j.1600-0897.2010.00836.x. PubMed PMID: 20367629; PubMed Central PMCID: PMC3025805.

3.   Piccinni MP, Lombardelli L, Logiodice F, Kullolli O, Parronchi P, Romagnani S. How pregnancy can affect autoimmune diseases progression? Clin Mol Allergy. 2016;14:11. doi: 10.1186/s12948-016-0048-x. PubMed PMID: 27651750; PubMed Central PMCID: PMC5025626.

4.   Mittal A, Pachter L, Nelson JL, Kjaergaard H, Smed MK, Gildengorin VL, et al. Pregnancy-Induced Changes in Systemic Gene Expression among Healthy Women and Women with Rheumatoid Arthritis. PloS one. 2015;10(12):e0145204. doi: 10.1371/journal.pone.0145204. PubMed PMID: 26683605; PubMed Central PMCID: PMC4684291.

5.   Weix J, Forger F, Haupl T, Surbek D, Ostensen M, Villiger PM. Influence of pregnancy on the adipocytokine and peroxisome proliferator-activated receptor pathways in peripheral blood mononuclear cells from healthy donors and rheumatoid arthritis patients. Arthritis Rheum. 2012;64(7):2095-103. doi: 10.1002/art.34375. PubMed PMID: 22231457.

6.   Gilli F, Lindberg RL, Valentino P, Marnetto F, Malucchi S, Sala A, et al. Learning from nature: pregnancy changes the expression of inflammation-related genes in patients with multiple

sclerosis. PloS one. 2010;5(1):e8962. doi: 10.1371/journal.pone.0008962. PubMed PMID: 20126412; PubMed Central PMCID: PMC2813302.

7.   Corwin EJ, Hogue CJ, Pearce B, Hill CC, Read TD, Mulle J, et al. Protocol for the Emory University African American Vaginal, Oral, and Gut Microbiome in Pregnancy Cohort Study. BMC pregnancy and childbirth. 2017;17(1):161. doi: 10.1186/s12884-017-1357-x. PubMed PMID: 28571577; PubMed Central PMCID: PMCPMC5455081.

8.   Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. Bioinformatics. 2012;28(9):1280-1. doi: 10.1093/bioinformatics/bts124. PubMed PMID: 22451269; PubMed Central PMCID: PMC3577110.

9.   Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics : official journal of the DNA Methylation Society. 2013;8(2):203-9. doi: 10.4161/epi.23470. PubMed PMID: 23314698; PubMed Central PMCID: PMC3592906.

10.   Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189-96. doi: 10.1093/bioinformatics/bts680. PubMed PMID: 23175756; PubMed Central PMCID: PMC3546795.

11.   Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC

bioinformatics. 2012;13:86. doi: 10.1186/1471-2105-13-86. PubMed PMID: 22568884; PubMed Central PMCID: PMC3532182.

12.   Pinheiro J BD, DebRoy S, Sarkar D, R Core Team nlme: Linear and Nonlinear Mixed Effects Models. 2016.

13.   Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS genetics. 2007;3(9):1724-35. doi: 10.1371/journal.pgen.0030161. PubMed PMID: 17907809; PubMed Central PMCID: PMC1994707.

14.   Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 2007;35(Web Server issue):W169-75. doi: 10.1093/nar/gkm415. PubMed PMID: 17576678; PubMed Central PMCID: PMC1933169.

15.   Burke SD, Seaward AV, Ramshaw H, Smith GN, Virani S, Croy BA, et al. Homing receptor expression is deviated on CD56+ blood lymphocytes during pregnancy in Type 1 diabetic women. PLoS One. 2015;10(3):e0119526. doi: 10.1371/journal.pone.0119526. PubMed PMID: 25793768; PubMed Central PMCID: PMC4368780.

16.   Kuo PL, Guo HR. Nucleated red blood cells in maternal blood during pregnancy. Obstet Gynecol. 1999;94(3):464-8. PubMed PMID: 10472878.

17.   Wegienka G, Havstad S, Bobbitt KR, Woodcroft KJ, Zoratti EM, Ownby DR, et al. Within-woman change in regulatory T cells from pregnancy to the postpartum period. J Reprod Immunol. 2011;88(1):58-65. doi: 10.1016/j.jri.2010.06.157. PubMed PMID: 20961621; PubMed Central PMCID: PMC3008215.

18.   Faas MM, de Vos P. Maternal monocytes in pregnancy and preeclampsia in humans and in rats. Journal of reproductive immunology. 2017;119:91-7. doi: 10.1016/j.jri.2016.06.009. PubMed PMID: 27396500.

19.   Veenstra van Nieuwenhoven AL, Bouman A, Moes H, Heineman MJ, de Leij LF, Santema J, et al. Cytokine production in natural killer cells and lymphocytes in pregnant women compared with women in the follicular phase of the ovarian cycle. Fertil Steril. 2002;77(5):1032-7. PubMed PMID: 12009363.

20.   Gellin BG, Broome CV, Bibb WF, Weaver RE, Gaventa S, Mascola L. The epidemiology of listeriosis in the United States--1986. Listeriosis Study Group. American journal of epidemiology. 1991;133(4):392-401. PubMed PMID: 1899779.

21.   Centers for Disease C, Prevention. Vital signs: Listeria illnesses, deaths, and outbreaks--United States, 2009-2011. MMWR Morbidity and mortality weekly report. 2013;62(22):448-52. PubMed PMID: 23739339.

22.   Bardin T. Gonococcal arthritis. Best Pract Res Clin Rheumatol. 2003;17(2):201-8. PubMed PMID: 12787521.

23.   Leitich H, Bodner-Adler B, Brunbauer M, Kaider A, Egarter C, Husslein P. Bacterial vaginosis as a risk factor for preterm delivery: a meta-analysis. American journal of obstetrics and gynecology. 2003;189(1):139-47. PubMed PMID: 12861153.

24.   Donders GG, Van Calsteren K, Bellen G, Reybrouck R, Van den Bosch T, Riphagen I, et al. Predictive value for preterm birth of abnormal vaginal flora, bacterial vaginosis and aerobic vaginitis during the first trimester of pregnancy. BJOG. 2009;116(10):1315-24. doi: 10.1111/j.1471-0528.2009.02237.x. PubMed PMID: 19538417.

25.   Hyman RW, Fukushima M, Jiang H, Fung E, Rand L, Johnson B, et al. Diversity of the vaginal microbiome correlates with preterm birth. Reprod Sci. 2014;21(1):32-40. doi: 10.1177/1933719113488838. PubMed PMID: 23715799; PubMed Central PMCID: PMC3857766.

26.   De Smet K, Contreras R. Human antimicrobial peptides: defensins, cathelicidins and histatins. Biotechnol Lett. 2005;27(18):1337-47. doi: 10.1007/s10529-005-0936-5. PubMed PMID: 16215847.

27.   Yarbrough VL, Winkle S, Herbst-Kralovetz MM. Antimicrobial peptides in the female reproductive tract: a critical component of the mucosal immune barrier with physiological and clinical implications. Hum Reprod Update. 2015;21(3):353-77. doi: 10.1093/humupd/dmu065. PubMed PMID: 25547201.

28.   Silversides CK, Colman JM. Physiological changes in pregnancy. Heart Disease in Pregnancy, 2nd Edition. 2007:6-17. doi: DOI 10.1002/9780470994955.ch2. PubMed PMID: WOS:000298401200003.

29.   Jwa SC, Fujiwara T, Yamanobe Y, Kozuka K, Sago H. Changes in maternal hemoglobin during pregnancy and birth outcomes. BMC pregnancy and childbirth. 2015;15. doi: ARTN 80 10.1186/s12884-015-0516-1. PubMed PMID: WOS:000352346300001.

30.   Heilmann L. Blood rheology and pregnancy. Baillieres Clin Haematol. 1987;1(3):777-99. PubMed PMID: 3327565.

31.   Gaillard R, Eilers PH, Yassine S, Hofman A, Steegers EA, Jaddoe VW. Risk factors and consequences of maternal anaemia and elevated haemoglobin levels during pregnancy: a population-based prospective cohort study. Paediatric and perinatal epidemiology. 2014;28(3):213-26. doi: 10.1111/ppe.12112. PubMed PMID: 24506330.

32. Townsley DM. Hematologic complications of pregnancy. Semin Hematol. 2013;50(3):222-31. doi: 10.1053/j.seminhematol.2013.06.004. PubMed PMID: 23953339; PubMed Central PMCID: PMC3748382.

33. Han L, Liu X, Li H, Zou J, Yang Z, Han J, et al. Blood coagulation parameters and platelet indices: changes in normal and preeclamptic pregnancies and predictive values for preeclampsia. PloS one. 2014;9(12):e114488. doi: 10.1371/journal.pone.0114488. PubMed PMID: 25464515; PubMed Central PMCID: PMC4252147.

34. Mallone R, Mannering SI, Brooks-Worrell BM, Durinovic-Bello I, Cilio CM, Wong FS, et al. Isolation and preservation of peripheral blood mononuclear cells for analysis of islet antigen-reactive T cell responses: position statement of the T-Cell Workshop Committee of the Immunology of Diabetes Society. Clin Exp Immunol. 2011;163(1):33-49. doi: 10.1111/j.1365-2249.2010.04272.x. PubMed PMID: 20939860; PubMed Central PMCID: PMC3010910.

35. Saha D, Patgaonkar M, Shroff A, Ayyar K, Bashir T, Reddy KV. Hemoglobin expression in nonerythroid cells: novel or ubiquitous? Int J Inflam. 2014;2014:803237. doi: 10.1155/2014/803237. PubMed PMID: 25431740; PubMed Central PMCID: PMC4241286.

36. Butcher JT, Johnson T, Beers J, Columbus L, Isakson BE. Hemoglobin alpha in the blood vessel wall. Free Radic Biol Med. 2014;73:136-42. doi: 10.1016/j.freeradbiomed.2014.04.019. PubMed PMID: 24832680; PubMed Central PMCID: PMC4135531.

Chapter 3

SLC9B1 Methylation Predicts Fetal Intolerance of Labor

Introduction

Fetal intolerance of labor, which is also referred to as fetal distress and non-reassuring fetal status, is the most common indication for emergency Caesarean section (C-section) [1, 2]. It is characterized by the presence of an abnormal fetal heart rate pattern, a category III tracing, detected through electronic fetal heart rate monitoring during labor [3, 4]. Category III fetal heart rate patterns indicative of fetal intolerance of labor include absent baseline fetal heart rate variability and recurrent late and/or variable decelerations and/or bradycardia or the presence of a sinusoidal pattern, typically after the onset of contractions during the second stage of labor [4, 5].

A category III tracing is considered abnormal as studies have demonstrated that these heart rate patterns are associated with an increased risk of fetal hypoxia and metabolic acidemia, though such tracings are not absolutely indicative of fetal hypoxia and/or acidemia [6]. In the case of fetal intolerance of labor, expedited delivery, often through a C-section, is indicated to avoid fetal hypoxia, acidemia, and their subsequent consequences [6-10]. Fetal hypoxia and acidemia can have drastic consequences for the fetus in the perinatal period and throughout life including severe brain damage [11], and has been previously associated with diagnosis of cerebral palsy [12]. Accurate recognition and prompt management of fetal intolerance of labor is essential for decreasing the risk of fetal hypoxia and acidemia and thereby providing the best possible pregnancy outcome.

Early identification of pregnant women at elevated risk for having a pregnancy complicated by fetal intolerance of labor would provide clinical benefits including maternal preparations to deliver at hospitals with the required resources to perform advanced monitoring and an emergency C-section, if required [1]. Another potential benefit of early identification of

those at risk for fetal intolerance of labor is a decrease in the time from decision to deliver until emergency C-section, which might promote adherence to the American College of Obstetricians and Gynecologists (ACOG) recommendation of a 30-minute timeframe, though previous studies have not demonstrated substantial differences in outcome due to decision to delivery time greater than 30 minutes [1, 2, 13]. Beyond the immediate clinical utility, identification of biological pathways underlying the development of fetal intolerance of labor may provide insight for development of novel treatments and preventive strategies. Previous studies have identified potential biomarkers associated with fetal intolerance of labor including pregnancy-associated plasma protein-A (PAPP-A) and combinations of fetal indices, such as estimated fetal weight, serum placental growth factor, and soluble fms-like tyrosine kinase-1 [14, 15]. These studies suggest an early pathogenesis for fetal intolerance that can be detected by screening prior to labor.

Previous studies have also reported associations between DNA methylation and pregnancy complications, including preeclampsia and gestational diabetes [16, 17]. DNA methylation, the addition of a methyl group to the 5' position of cytosine in a cytosine-guanine dinucleotide (CpG site), serves as a mechanism to regulate gene expression. This study utilizes an epigenome-wide association study to assess the relationship between individual CpG sites and fetal intolerance of labor.

Methods

Study Subjects

Subjects included in this study (N=69) are being enrolled into an ongoing pregnancy cohort study investigating the microbiome and epigenome and the outcome of preterm birth (R01NR014800, R01MD009064) for which pregnant African American women are being

recruited from outpatient prenatal care clinics affiliated with two Atlanta metro area hospitals: Emory University Midtown Hospital and Grady Memorial Hospital. These two hospitals represent private and public hospitals, respectively, that serve women of a wide range of socioeconomic status. Eligible women for this study were African American by self-report, between 18-35 years of age, having a singleton pregnancy, having fewer than four previous births, are able to understand written and spoken English, without chronic medical conditions, enrolled between March 2014 and August 2015 and experiencing labor (spontaneous or induced). Not experiencing labor, fetal death before labor and congenital abnormalities of the fetus were criteria for post-enrollment exclusion. Demographic and clinical obstetrical data (including estimated date of confinement, gestational age at delivery, pregnancy complications, labor and delivery course, and blood pressure) were collected through self-report questionnaires and prenatal, labor, and delivery medical chart abstraction under the supervision of a qualified physician (ALD). The diagnosis of fetal intolerance of labor was based upon the medical record documentation by the attending obstetrician or midwife determining the presence of a category III fetal heart rate tracing during labor, and after delivery each chart was reviewed by a qualified physician to confirm the diagnosis. All subjects provided written informed consent. This study was approved by the Emory Institutional Review Board.

Biological Sample Collection and DNA extraction

Venous blood samples were collected during each of two prenatal visits (between 7-15 weeks and 24-32 weeks), when an additional 12 mL of peripheral blood was drawn into a tube containing EDTA using the same needle stick as for the routine blood draws. Blood was transferred into SepMate tubes with a Ficoll density gradient to isolate peripheral blood mononuclear cells (PBMCs) from whole blood. PBMCs were stored in AllProtect Buffer

(Qiagen) at -80 °C until a simultaneous DNA and RNA extraction using the AllPrep RNA/DNA Mini Kit (Qiagen) was performed according to manufacturer's instructions. DNA quantification and quality was assessed using the Quant-it Pico Green kit (Invitrogen). RNA quantification was assessed using the Agilent 2100 Bioanalyzer. The average RIN score was 8.6, with a standard deviation of 1.4.

DNA Methylation

DNA methylation from samples of maternal PBMCs was interrogated for each subject using the HumanMethylation450 BeadChip, which evaluates >450,000 CpG sites across the genome. Briefly, 1 µg of DNA was processed and hybridized to the HumanMethylation450 BeadChip (Illumina) according to manufacturer's instructions. Initial data quality control was performed using the R package CpGassoc [18]. Any CpG site with low signal or missing data for greater than 10% of samples was removed, and any sample with missing data for greater than 5% of CpG sites was removed. Cross-reactive probes were removed [19]. Following quality control, 449,094 probes were included in subsequent analyses. One sample collected between 7-15 weeks gestation failed quality control. Beta values (β) were calculated for each CpG site as the ratio of methylated (M) to methylated and unmethylated (U) signal: β=M/M+U. Beta-mixture quantile normalization was performed as previously described [20]. Briefly, BMIQ involves fitting a three-state beta-mixture model, transforming the distribution of type 2 probes to match the type 1 distribution, followed by a dilation transformation [20]. DNA methylation data can be accessed through NCBI's Gene Expression Omnibus, GEO107459.

RNA Expression

RNA expression from maternal PBMCs was interrogated for a subset of subjects for which RNA was available. Briefly, 750 ng of RNA was directly hybridized to the HumanHT-12

v4 BeadChip (Illumina, San Diego, CA) according to manufacturer's instructions. The BeadChips were scanned using the iScan scanner, and the raw data was analyzed using the Expression Module of GenomeStudio Software (Illumina, San Diego, CA). Samples with probe detection rates <90% or with an average intensity of <50% of the experiment-wide sample mean were excluded, resulting in one sample collected between 24-32 weeks being dropped due to quality control. Probes with detection p-values >.01 in >90% of the samples were excluded. Data was then quantile-normalized and log2 transformed. Following quality control, 18,634 expression probes were included in subsequent analyses. RNA expression data can be accessed through NCBI's Gene Expression Omnibus, GEO107437.

Statistical Analysis

Demographic characteristics were compared between women with and without pregnancies complicated by fetal intolerance of labor using Student's t tests. The R package CpGassoc was used to perform an epigenome-wide association study (EWAS) to assess the associations between maternal DNA methylation at each CpG site on the array and fetal intolerance of labor for samples collected between 24-32 weeks gestation. For each CpG site, the methylation proportion was regressed on an indicator for fetal intolerance and covariates, which included chip, maternal age, and cell type proportions (CD8+T, monocytes, B cells, natural killer), estimated using the referenced dataset developed by Reinius and colleagues and implemented used the approach described by Houseman and colleagues [21-23]. CpG sites that were significantly associated with fetal intolerance for labor in samples collected between 24-32 weeks were assessed for associations earlier in pregnancy, between 7-15 weeks gestation, using a linear regression that controlled for maternal age, chip, and cellular heterogeneity as above. The false discovery rate (FDR) was controlled at .05. Gene symbols and probe locations were

assigned using the HumanMethylation450 BeadChip annotation file distributed by Illumina. To determine whether methylation of these CpG sites was influenced by cellular heterogeneity, linear regressions were performed that modeled methylation at each CpG site as a function of estimated cell type proportions. Pearson's correlation coefficient was calculated for each pair of CpG sites, and between each CpG site at the two timepoints. Predictive accuracy was assessed using the area under the curve (AUC), sensitivity, and specificity calculations, as determined by the R package pROC. Longitudinal associations between each CpG site and gene expression were assessed using a linear mixed-effects model that regressed the log2 expression signal for each gene on methylation of a single CpG site with subject as a random effect and adjustment for maternal age and cellular heterogeneity. Similarly, a linear mixed-effects model was used to interrogate associations between fetal intolerance and gene expression over pregnancy. Subject samples were included in this analysis if participants contributed blood at 24-32 weeks, a subset of which also contributed blood at 7-15 weeks (N=54). As an additional confirmation of the associations between fetal intolerance and DNA methylation, we performed differential methylated region (DMR) analysis using the R package ChAMP [24].

To assess the associations between fetal intolerance and gene expression, 65 women, recruited from the same study, were used. The log2 of the gene expression signal was regressed on fetal intolerance of labor after adjusting for cellular heterogeneity and maternal age. Subjects with available measures for blood pressure values were used to assess associations between gene expression and these cardiovascular measures. Blood pressure was dichotomized as being above or below 90 for diastolic blood pressure at least once during pregnancy and above or below 140 for systolic blood pressure at least once during pregnancy. Each outcome was regressed on the log2 gene expression signal, after adjusting for maternal age and cellular heterogeneity. Gene

expression was interrogated by two probes that differed by only 2 base pairs, thus, data from the probe with the highest number of samples that met the detection p-value threshold (ILMN_1724931) is presented.

<u>Replication Cohort and Analysis</u>

A replication cohort of 43 women was recruited as part of the Johns Hopkins Prospective PPD cohort recruited at the Women's Mood Disorders Center at Johns Hopkins. Subjects were prospectively followed during pregnancy and after delivery in order to identify genetic and clinical characteristics that precede the development of a postpartum depressive episode. The average age of the participants was 31.3 years. This cohort was ethnically diverse; 65% of participants were Caucasian, 23% of participants were African American, and the remainder participants identified as Hispanic (2%), Asian or Pacific Islander (5%), or other (5%). In this cohort, 8 women experienced pregnancies complicated by fetal intolerance to labor (21%), and 4 had missing data. Epigenetic data were generated on the HumanMethylation450 BeadChip as described previously [25]. The presence of fetal intolerance of labor was determined through medical record abstraction. Briefly, sample quality assessment and microarray analysis were conducted at The Sidney Kimmel Cancer Center Microarray Core Facility at Johns Hopkins University. Images were processed in Illumina's iScan scanner and data were extracted using Methylation Module of GenomeStudio v1.0 Software. Background and Illumina probe type correction and normalization were performed by the Dasen function in the wateRmelon package in R.[26] The association between methylation at each of the CpG sites to be replicated and fetal intolerance was assessed using a linear regression. Pearson's correlation coefficients were calculated for each pair of CpG sites.

Results

Of the 57 women included in this study, 12 had deliveries complicated by fetal intolerance of labor (21%). Maternal age, smoking, gestational hypertension, and gestational age at birth and gestational age at sample collection did not differ between those subjects who did and did not experience deliveries complicated by fetal intolerance of labor (.59<p<.98). However, the group experiencing fetal intolerance of labor was more likely to undergo delivery by C-section (p=2.70×10^{-5}; Table 3-1), which would be expected given that fetal intolerance of labor is an indication for C-section delivery.

Methylation of *SLC9B1* predicts fetal intolerance of labor

Four CpG sites associated with fetal intolerance of labor in maternal samples collected between 24 and 32 weeks gestation, after adjusting for maternal age and PBMC cell type (FDR<.05; Fig 3-1, Table 3-2). All four sites were annotated to the CpG island of solute carrier family 9, subfamily B, member 1 (*SLC9B1*), alternatively known as *NHEDC1*, a NA^+/ H^+ exchanger. Methylation of these four sites was highly correlated (r=.86-.98, Fig 3-1), and was not associated with PBMC cell type or maternal age (p>.05). In a separate analysis of these four CpG sites in samples collected between 7-15 weeks gestation (N=45), each site associated with fetal intolerance of labor to a lesser degree (.001<p< .003), suggesting that the methylation differences associated with fetal intolerance of labor may be detectable even earlier in early pregnancy. Differential methylation region analysis (DMR) identified 18 differentially methylated regions, including chr4:103940711-103941205, which contains the four CpG sites identified above. In the Johns Hopkins Prospective PPD cohort these four CpG sites were also highly correlated (r=.91-.97), and associated with fetal intolerance of labor (.036<p<.048), replicating our initial finding (Fig 3-2-Fig 3-5).

Methylation of *SLC9B1* associates with gene expression

We next sought to assess whether methylation differences of these CpG sites associate with expression of *SLC9B1*, and whether gene expression alone can predict fetal intolerance of labor. We therefore examined paired methylation and expression data for a subset of subjects with available expression data. Expression of *SLC9B1* was represented by two probes, (ILMN_1673417 and ILMN_1724931), that passed quality control. Changes in DNA methylation over time were associated with changes in gene expression for all CpG sites and *SLC9B1* expression (.003<p<.03, Fig 3-2-Fig 3-5). Gene expression data was available for 65 women who also had DNA methylation data at one or more timepoints. Gene expression alone did not predict fetal intolerance of labor between 24 and 32 weeks gestation (p=NS), or over the whole of pregnancy (p= NS), suggesting that the CpG sites identified in *SLC9B1* are better predictors of fetal intolerance of labor.

Predictive Accuracy

Each of the four CpG sites was interrogated for its predictive accuracy of fetal intolerance of labor in maternal PBMC samples collected between 24-32 weeks gestation. The area under the receiver operating characteristic curve ranged from .85 to .87 for the four sites, representing a predictive accuracy of 85-87% (Fig 3-2-Fig 3-5).

Gene expression and cardiovascular remodeling

All of the 56 women with gene expression data available underwent blood pressure measurement between 24-32 weeks. Gene expression, but not DNA methylation, was associated with both high systolic (>140; p=.003) and diastolic (>90; p=.003) blood pressure (ILMN_1724931).

Discussion

We identify four CpG sites in *SLC9B1* whose DNA methylation levels detected in the late second and early third trimesters are predictive of fetal intolerance of labor at delivery and was replicated in an independent sample. Each of these CpG sites was highly correlated with one another, indicating that any of these sites could serve as a proxy for the other three. As an additional line of evidence, *SLC9B1* was also identified as being associated with fetal intolerance of labor in a differentially methylated region analysis. This region, and the other DMRs identified should be further investigated as potential contributors to the etiology of fetal intolerance of labor. Women enrolled in this study were medically low-risk as they had singleton pregnancies with early initiation of prenatal care and did not have chronic health conditions. Thus, we believe these results are generalizable to other low risk populations.

Few studies have examined the role of *SLC9B1*, also known as *NHEDC1*. The SLC9 family of genes encodes Na+/H+ exchangers (NHE) that play a role in regulating pH and cell volume [27, 28]. A recent study suggests that DNA methylation of *SLC9B1* regulates its expression [29], which is consistent with our finding that methylation was associated with gene expression of *SLC9B1* over pregnancy. This gene was previously thought to be expressed specifically in the testes based on a limited panel of 18 tissues, which did not include immune cells [28, 30]. Data from the Genotype-Tissue Expression (GTEx) Project shows low-level expression of *SLC9B1* in a variety of tissues, including whole blood [31]. Localization to the mitochondria, as shown in the Human Protein Atlas [32], suggests a role for this gene in processes associated with the electron transport chain, which produce the energy required for the cell, as well as reactive oxygen species (ROS) as byproducts.

Early in normal pregnancy, a relative hypoxic state is essential for proper placentation and embryogenesis, but oxygen requirements increase around 11-12 weeks gestation [33].

Oxygen tension is regulated by transcription factors, including the hypoxia-inducible factor 1 (HIF-1). In normal pregnancy, HIF-1 is rapidly degraded [33]. However, increased ROS production by the mitochondria can inhibit the degradation of HIF-1, which may impair placental function and lead to pregnancy complications [34, 35]. Additionally, ROS production may result in insufficient energy generation and sodium ion influx into the cell, which has been associated with both cardiac pathology and mitochondrial function [34, 36]. We hypothesize that alterations in the methylation and expression of sodium-hydrogen transporters that contribute to proper mitochondrial function, including *SLC9B1,* may contribute to cardiac pathology and changes in pH, which are both risk factors for fetal intolerance of labor [37, 38]. This hypothesis is further supported by our findings that gene expression of *SLC9B1* is associated with high maternal systolic and diastolic blood pressure, and with a report from the literature that showed preeclampsia, which is characterized by high blood pressure, is associated with abnormal fetal heart rate [39]. Previous studies have also shown increased ROS production in both the mother and fetus is associated with fetal intolerance of labor [40]. Future studies should further examine the role of *SLC9B1* in human fertility and pregnancy outcomes, especially in the context of cardiovascular dysfunction.

This study has several limitations. First, the difference in methylation between women with pregnancies complicated by fetal intolerance of labor and those not complicated by fetal intolerance of labor is relatively small, making the development of a targeted assay for methylation of these CpG sites difficult due to the limited discriminatory power available for common techniques. Additionally, although methylation of these CpG sites was associated with fetal intolerance of labor in the PPD cohort, the range of methylation at these sites was higher in this cohort, potentially due to differences in processing or quality control, or inherent differences

in methylation between PBMCs in the original cohort and whole blood in the Johns Hopkins Prospective PPD cohort. Finally, type III fetal heart rate tracings are not absolutely indicative of fetal hypoxia or metabolic acidemia, but there are not currently standards for further evaluation of fetuses intra-labor. Other methods to diagnose fetal intolerance of labor are not routinely performed and lack sufficient clinical evidence for routine implementation. Despite these limitations, receiver operator characteristic curves of the 4 CpG sites in *SLC9B1* suggested that the positive and negative predictive values, sensitivity, and specificity, are all well within the range for potential development into a clinically useful diagnostic test (Table 3-3).

DNA methylation patterns in maternal blood at four CpG sites in *SLC9B1* are predictive of fetal intolerance of labor during the late second and early third trimester. These sites have highly positive and negative predictive values, indicating that they may be clinically relevant for the detection and management of fetal intolerance of labor. Future studies should work to develop a robust targeted assay to measure DNA methylation at one or more of these CpG sites so that the clinical utility of DNA methylation at these sites can be further evaluated for its predictive power in other studies throughout pregnancy.

Table 3-1: Cohort demographics for women with samples collected between 24-32 weeks

gestation (N=57). Those experiencing fetal intolerance of labor were more likely to have a

Cesarean section. No other characteristic differed by group. SD: Standard deviation

	Fetal Intolerance Mean±SD	Control Mean±SD	P value
Maternal Age	25.97±4.1	25.2±4.5	NS
GA at sample collection	26.8±2.2	27.3±2.3	NS
GA at birth	38.6±1.8	38.6±2.9	NS
	Fetal Intolerance (%) (N=12)	Control (%) (N=45)	P value
Cesarean Section	83.3	11.1	2.70×10^{-5}
Hypertension	16.7	13.3	NS
Maternal Smoking[a]	8.3	6.81	NS

[a]Missing data for one subject

Table 3-2: CpG sites significantly associated with fetal intolerance of labor. FDR was computed using the Benjamini-Hochberg method. Coordinates are based on GRCh37/hg19.

CpG site	Gene Symbol	Position	T statistic	P-value	FDR
cg21197425	*SLC9B1*	chr4: 103940854	7.70	2.69×10^{-10}	9.58×10^{-05}
cg05605371	*SLC9B1*	chr4: 103940876	7.29	1.25×10^{-9}	1.40×10^{-04}
cg19672271	*SLC9B1*	chr4: 103940878	7.48	6.05×10^{-10}	9.58×10^{-05}
cg06999381	*SLC9B1*	chr4: 103940936	7.47	6.40×10^{-10}	9.58×10^{-05}

Table 3-3: Threshold, sensitivity, specificity, positive predicate value (PPV), and negative

predictive value (NPV) for each CpG site.

CpG site	Threshold (% methylation)	Sensitivity	Specificity	PPV	NPV
cg06999381	.038	.67	.98	.89	.92
cg21197425	.097	.67	.98	.89	.92
cg05605371	.063	.67	.96	.80	.91
cg19672271	.065	.67	.96	.80	.91

Figure 3-1: Manhattan plot showing the association of CpG sites across the genome. Each CpG site that passed quality control was assessed for associations with fetal intolerance of labor. Sites falling above the horizontal line indicate experiment-wide significance. The x-axis represents the chromosome number and the y axis is the negative log of the p-value, which is indicative of the significance level. The plot is further zoomed in to chromosome four, and the *SLC9B1* gene. In the gene diagram, yellow boxes represent exons and the green box represents the location of the CpG island. The heatmap indicates the correlation between CpG sites.
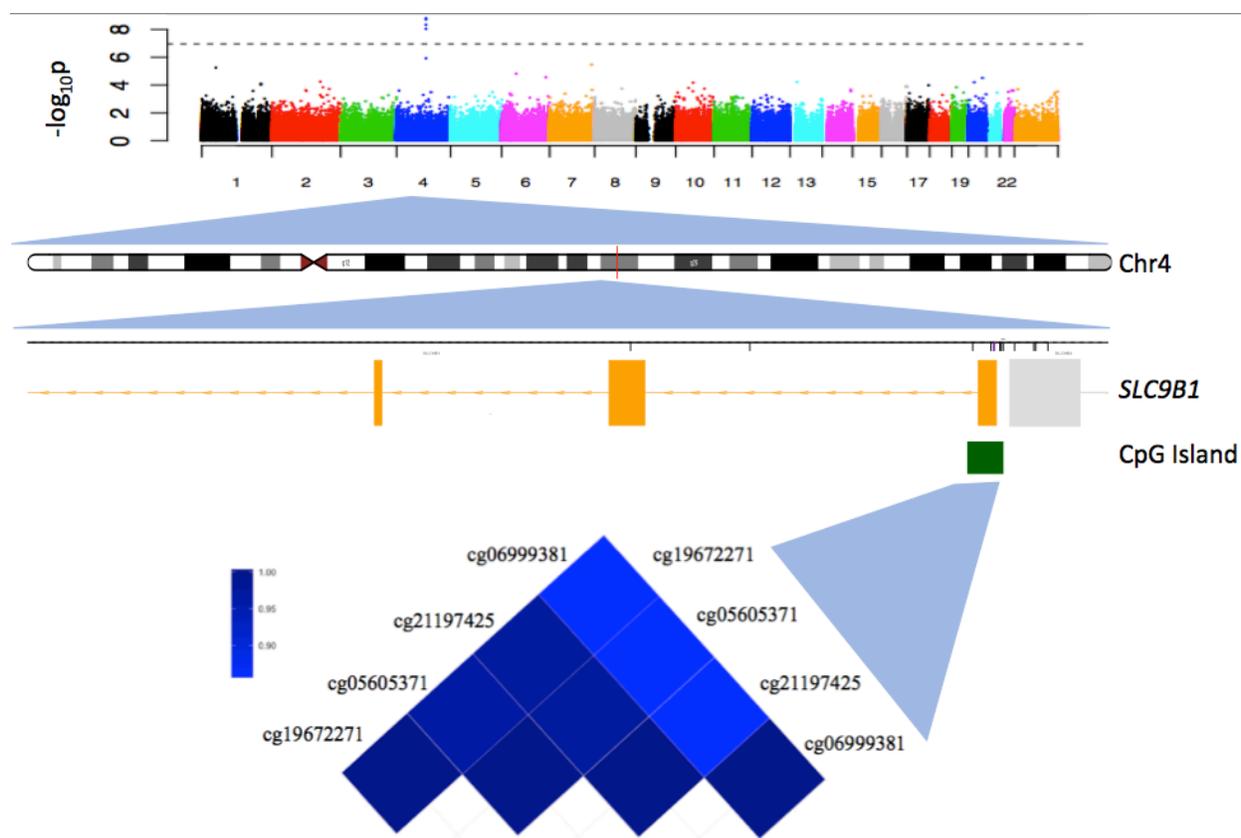
Figure 3-2: DNA methylation associates with fetal intolerance of labor. A) DNA methylation of cg06999381 associates with fetal intolerance of labor between 24-32 weeks gestation in the original cohort, B) Association with fetal intolerance of labor in the replication cohort, C) Receiver operator characteristic curve, D) DNA methylation associates with gene expression over pregnancy (ILMN_1724931). Red indicates fetal intolerance to labor. Open circles are samples from visit 1, closed circles are samples from visit 2. All associations are statistically significant (p<.05).
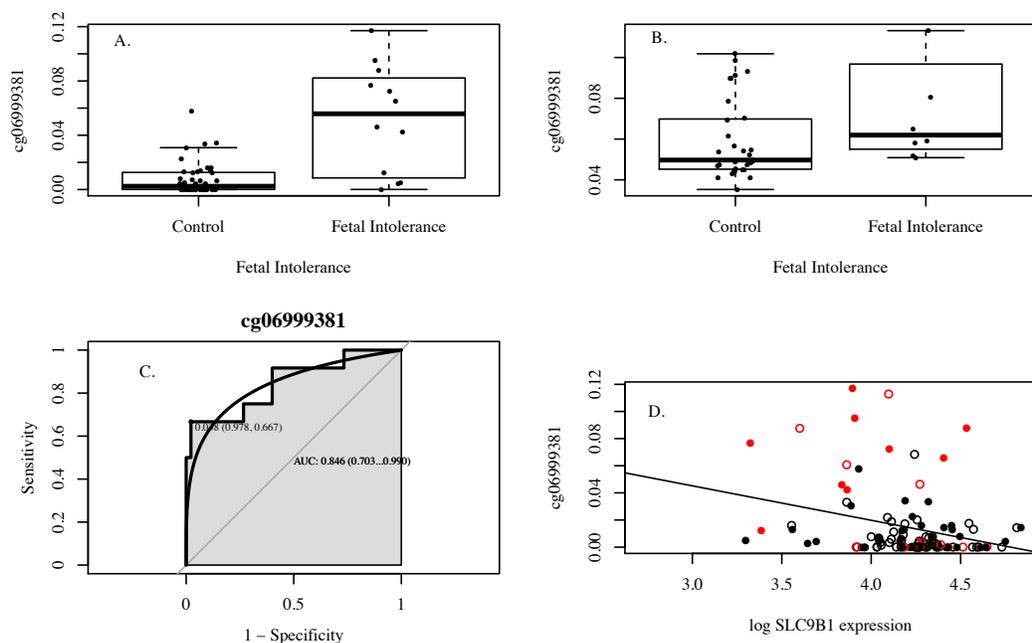
Figure 3-3: Cg21197425 associates with fetal intolerance of labor. A) DNA methylation of cg21197425 associates with fetal intolerance of labor between 24-32 weeks gestation in the original cohort, B) Association with fetal intolerance of labor in the replication cohort, C) Receiver operator characteristic curve, D) DNA methylation associates with gene expression over pregnancy (ILMN_1724931). Red indicates fetal intolerance to labor. Open circles are samples from visit 1, closed circles are samples from visit 2.
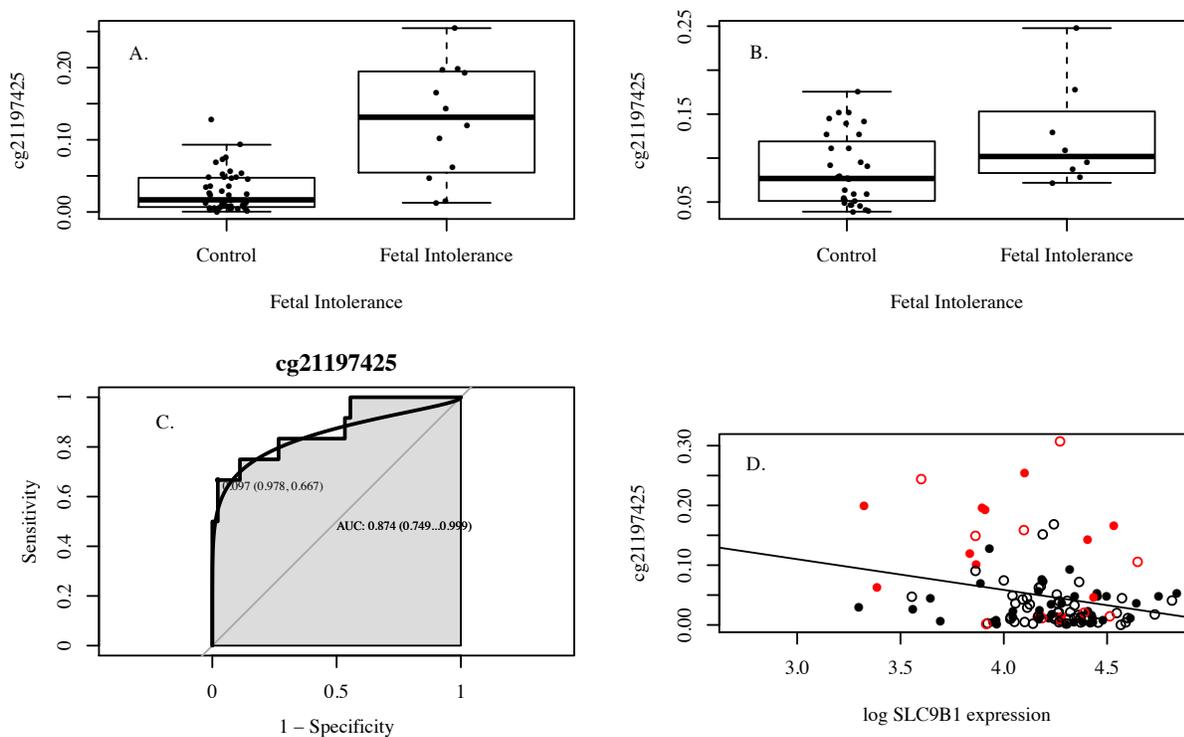
Figure 3-4: Cg05605371 associates with fetal intolerance of labor. A) DNA methylation of cg05605371 associates with fetal intolerance of labor between 24-32 weeks gestation in the original cohort, B) Association with fetal intolerance of labor in the replication cohort, C) Receiver operator characteristic curve, D) DNA methylation associates with gene expression over pregnancy (ILMN_1724931). Red indicates fetal intolerance to labor. Open circles are samples from visit 1, closed circles are samples from visit 2.
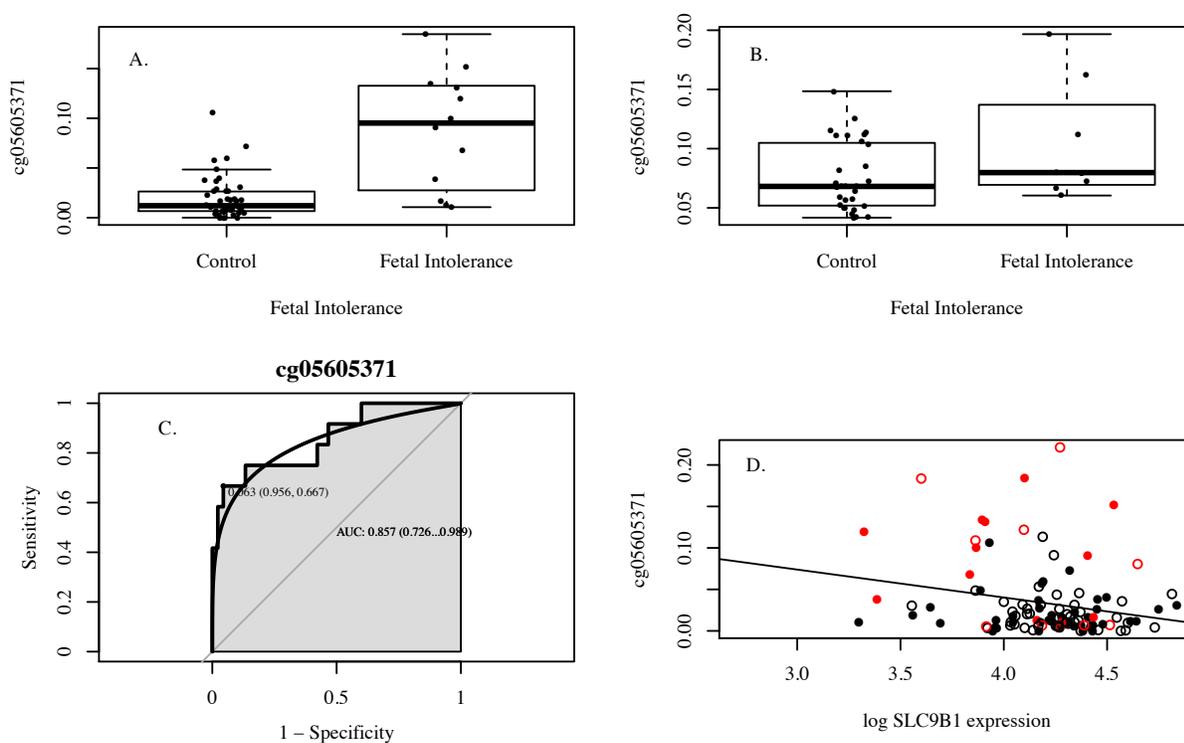
Figure 3-5: Cg19672271 associates with fetal intolerance of labor. A) DNA methylation of cg19672271 associates with fetal intolerance of labor between 24-32 weeks gestation in the original cohort, B) Association with fetal intolerance of labor in the replication cohort, C) Receiver operator characteristic curve, D) DNA methylation associates with gene expression over pregnancy (ILMN_1724931). Red indicates fetal intolerance to labor. Open circles are samples from visit 1, closed circles are samples from visit 2.
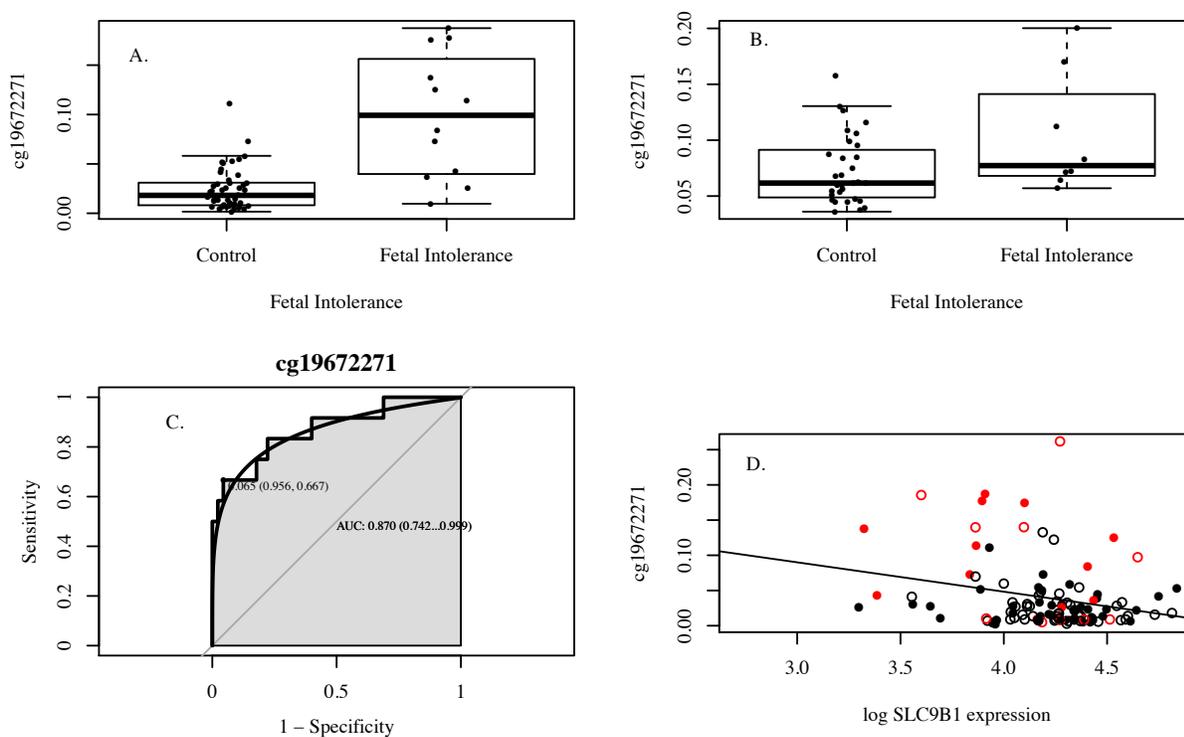
References

1.   Chauhan SP, Magann EF, Scott JR, Scardo JA, Hendrix NW, Martin JN, Jr. Cesarean delivery for fetal distress: rate and risk factors. Obstetrical & gynecological survey. 2003;58(5):337-50. doi: 10.1097/01.OGX.0000066802.19138.AE. PubMed PMID: 12719676.

2.   Schauberger CW, Chauhan SP. Emergency cesarean section and the 30-minute rule: definitions. American journal of perinatology. 2009;26(3):221-6. doi: 10.1055/s-0028-1103033. PubMed PMID: 19031352.

3.   Bergmans MG. How to diagnose intrapartum fetal distress? European journal of obstetrics, gynecology, and reproductive biology. 1998;79(2):119-21. PubMed PMID: 9720826.

4.   Penning S, Garite TJ. Management of fetal distress. Obstet Gynecol Clin North Am. 1999;26(2):259-74. PubMed PMID: 10399760.

5.   Liston R, Sawchuck D, Young D, Society of O, Gynaecologists of C, British Columbia Perinatal Health P. Fetal health surveillance: antepartum and intrapartum consensus guideline. J Obstet Gynaecol Can. 2007;29(9 Suppl 4):S3-56. PubMed PMID: 17845745.

6.   Maso G, Piccoli M, De Seta F, Parolin S, Banco R, Camacho Mattos L, et al. Intrapartum fetal heart rate monitoring interpretation in labour: a critical appraisal. Minerva Ginecol. 2015;67(1):65-79. PubMed PMID: 25411863.

7.   Wood S, Ross S, Sauve R. Cesarean Section and Subsequent Stillbirth, Is Confounding by Indication Responsible for the Apparent Association? An Updated Cohort Analysis of a Large Perinatal Database. PloS one. 2015;10(9):e0136272. doi: 10.1371/journal.pone.0136272. PubMed PMID: 26331274; PubMed Central PMCID: PMC4557984.

8.   Sperling LS, Henriksen TB, Ulrichsen H, Hedegard M, Moller H, Hansen V, et al. Indications for cesarean section in singleton pregnancies in two Danish counties with different cesarean section rates. Acta obstetricia et gynecologica Scandinavica. 1994;73(2):129-35. PubMed PMID: 8116351.

9.   Benzouina S, Boubkraoui Mel M, Mrabet M, Chahid N, Kharbach A, El-Hassani A, et al. Fetal outcome in emergency versus elective cesarean sections at Souissi Maternity Hospital, Rabat, Morocco. Pan Afr Med J. 2016;23:197. doi: 10.11604/pamj.2016.23.197.7401. PubMed PMID: 27347286; PubMed Central PMCID: PMC4907743.

10.   Silberstein T, Sheiner E, Salem SY, Hamou B, Aricha B, Baumfeld Y, et al. Fetal heart rate monitoring category 3 during the 2nd stage of labor is an independent predictor of fetal acidosis. The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet. 2016:1-4. doi: 10.3109/14767058.2016.1172064. PubMed PMID: 27023800.

11.   Jensen A, Garnier Y, Berger R. Dynamics of fetal circulatory responses to hypoxia and asphyxia. European journal of obstetrics, gynecology, and reproductive biology. 1999;84(2):155-72. PubMed PMID: 10428339.

12.   MacLennan AH, Thompson SC, Gecz J. Cerebral palsy: causes, pathways, and the role of genetic variants. American journal of obstetrics and gynecology. 2015;213(6):779-88. doi: 10.1016/j.ajog.2015.05.034. PubMed PMID: 26003063.

13.   Roy KK, Baruah J, Kumar S, Deorari AK, Sharma JB, Karmakar D. Cesarean section for suspected fetal distress, continuous fetal heart monitoring and decision to delivery time. Indian J Pediatr. 2008;75(12):1249-52. doi: 10.1007/s12098-008-0245-9. PubMed PMID: 19190880.

14.   Uccella S, Colombo GF, Bulgheroni CM, Serati M, Bogani G, Salvatore S, et al. First-trimester maternal serum screening and the risk for fetal distress during labor. American journal of obstetrics and gynecology. 2009;201(2):166 e1-6. doi: 10.1016/j.ajog.2009.05.009. PubMed PMID: 19646567.

15.   Valino N, Giunta G, Gallo DM, Akolekar R, Nicolaides KH. Biophysical and biochemical markers at 35-37 weeks' gestation in the prediction of adverse perinatal outcome. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2016;47(2):203-9. doi: 10.1002/uog.15663. PubMed PMID: 26224608.

16.   Anderson CM, Ralph JL, Wright ML, Linggi B, Ohm JE. DNA methylation as a biomarker for preeclampsia. Biological research for nursing. 2014;16(4):409-20. doi: 10.1177/1099800413508645. PubMed PMID: 24165327.

17.   Wu P, Farrell WE, Haworth KE, Emes RD, Kitchen MO, Glossop JR, et al. Maternal genome-wide DNA methylation profiling in gestational diabetes shows distinctive disease-associated changes relative to matched healthy pregnancies. Epigenetics : official journal of the DNA Methylation Society. 2016:0. doi: 10.1080/15592294.2016.1166321. PubMed PMID: 27019060.

18.   Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. Bioinformatics. 2012;28(9):1280-1. doi: 10.1093/bioinformatics/bts124. PubMed PMID: 22451269; PubMed Central PMCID: PMC3577110.

19.   Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium

HumanMethylation450 microarray. Epigenetics : official journal of the DNA Methylation Society. 2013;8(2):203-9. doi: 10.4161/epi.23470. PubMed PMID: 23314698; PubMed Central PMCID: PMC3592906.

20.   Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189-96. doi: 10.1093/bioinformatics/bts680. PubMed PMID: 23175756; PubMed Central PMCID: PMC3546795.

21.   Horvath S. DNA methylation age of human tissues and cell types. Genome biology. 2013;14(10):R115. doi: 10.1186/gb-2013-14-10-r115. PubMed PMID: 24138928; PubMed Central PMCID: PMC4015143.

22.   Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13:86. doi: 10.1186/1471-2105-13-86. PubMed PMID: 22568884; PubMed Central PMCID: PMC3532182.

23.   Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PloS one. 2012;7(7):e41361. doi: 10.1371/journal.pone.0041361. PubMed PMID: 22848472; PubMed Central PMCID: PMC3405143.

24.   Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics. 2014;30(3):428-30. doi: 10.1093/bioinformatics/btt684. PubMed PMID: 24336642; PubMed Central PMCID: PMC3904520.

25.   Guintivano J, Arad M, Gould TD, Payne JL, Kaminsky ZA. Antenatal prediction of postpartum depression with blood DNA methylation biomarkers. Molecular psychiatry. 2014;19(5):560-7. doi: 10.1038/mp.2013.62. PubMed PMID: 23689534.

26.   Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC genomics. 2013;14:293. doi: 10.1186/1471-2164-14-293. PubMed PMID: 23631413; PubMed Central PMCID: PMC3769145.

27.   Fuster DG, Alexander RT. Traditional and emerging roles for the SLC9 Na+/H+ exchangers. Pflugers Arch. 2014;466(1):61-76. doi: 10.1007/s00424-013-1408-8. PubMed PMID: 24337822.

28.   Donowitz M, Ming Tse C, Fuster D. SLC9/NHE gene family, a plasma membrane and organellar family of Na(+)/H(+) exchangers. Mol Aspects Med. 2013;34(2-3):236-51. doi: 10.1016/j.mam.2012.05.001. PubMed PMID: 23506868; PubMed Central PMCID: PMC3724465.

29.   Kumar PL, James PF. Identification and characterization of methylation-dependent/independent DNA regulatory elements in the human SLC9B1 gene. Gene. 2015;561(2):235-48. doi: 10.1016/j.gene.2015.02.050. PubMed PMID: 25701605; PubMed Central PMCID: PMC4361323.

30.   Ye G, Chen C, Han D, Xiong X, Kong Y, Wan B, et al. Cloning of a novel human NHEDC1 (Na+/H+ exchanger like domain containing 1) gene expressed specifically in testis. Mol Biol Rep. 2006;33(3):175-80. doi: 10.1007/s11033-006-0010-y. PubMed PMID: 16850186.

31.   Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013;45(6):580-5. doi: 10.1038/ng.2653. PubMed PMID: 23715323; PubMed Central PMCID: PMC4010069.

32.   Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419. doi: 10.1126/science.1260419. PubMed PMID: 25613900.

33.   Patel J, Landers K, Mortimer RH, Richard K. Regulation of hypoxia inducible factors (HIF) in hypoxia and normoxia during placental development. Placenta. 2010;31(11):951-7. doi: 10.1016/j.placenta.2010.08.008. PubMed PMID: 20869770.

34.   Maltepe E, Saugstad OD. Oxygen in health and disease: regulation of oxygen homeostasis--clinical implications. Pediatric research. 2009;65(3):261-8. doi: 10.1203/PDR.0b013e31818fc83f. PubMed PMID: 18852690.

35.   Herrera EA, Krause B, Ebensperger G, Reyes RV, Casanello P, Parra-Cordero M, et al. The placental pursuit for an adequate oxidant balance between the mother and the fetus. Front Pharmacol. 2014;5:149. doi: 10.3389/fphar.2014.00149. PubMed PMID: 25009498; PubMed Central PMCID: PMC4068002.

36.   Murphy E, Eisner DA. Regulation of intracellular and mitochondrial sodium in health and disease. Circ Res. 2009;104(3):292-303. doi: 10.1161/CIRCRESAHA.108.189050. PubMed PMID: 19213964; PubMed Central PMCID: PMC2662399.

37.   Omo-Aghoja L. Maternal and fetal Acid-base chemistry: a major determinant of perinatal outcome. Ann Med Health Sci Res. 2014;4(1):8-17. doi: 10.4103/2141-9248.126602. PubMed PMID: 24669324; PubMed Central PMCID: PMC3952302.

38.   Gravett C, Eckert LO, Gravett MG, Dudley DJ, Stringer EM, Mujobu TB, et al. Non-reassuring fetal status: Case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine. 2016;34(49):6084-92. doi: 10.1016/j.vaccine.2016.03.043. PubMed PMID: 27461459; PubMed Central PMCID: PMC5139811.

39.   Sharma KJ, Esakoff TF, Guillet A, Burwick RM, Caughey AB. Pregnancies complicated by both preeclampsia and growth restriction between 34 and 37 weeks' gestation are associated with adverse perinatal outcomes. The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet. 2016:1-4. doi: 10.1080/14767058.2016.1248394. PubMed PMID: 27756175.

40.   Raicevic S, Cubrilo D, Arsenijevic S, Vukcevic G, Zivkovic V, Vuletic M, et al. Oxidative stress in fetal distress: potential prospects for diagnosis. Oxidative medicine and cellular longevity. 2010;3(3):214-8. doi: 10.4161/oxim.3.3.12070. PubMed PMID: 20716946; PubMed Central PMCID: PMC2952080.

Chapter 4

Predicting Neonatal Gestational Age Using DNA Methylation

Introduction

Differences in gestational age as small as one week have been shown to have significant impacts on neonatal morbidity and mortality, as well as long-term outcomes [1-6]. In light of this, the American College of Obstetricians and Gynecologists (ACOG) recently recommended revising the categorization of births from term (>37 weeks gestation) and preterm (≤ 37 weeks gestation) into several subcategories (early preterm, preterm, early term, full term, late term, and post term) that better reflect the developmental differences associated with gestational age at each of these time points [7, 8]. Accurate classification systems that reflect both developmental time and maturity may improve our ability to predict neonatal risk.

Traditionally, GA is estimated using one or more of the following methods: early obstetric ultrasound, last menstrual period (LMP), or neonatal estimation [9]. Ultrasound-based methods are considered to be the gold standard and have proven to be a better predictor of delivery date [10] as LMP estimates may be influenced by uncertainty regarding LMP dates, normal variations in ovulation timing, atypical bleeding, and contraceptive use [9]. Neonatal estimation, which is based on a combination of physical appearance, muscular tone, flexibility, and reflexes, is the only available method for determining gestational age after birth but is less precise than LMP and ultrasound [9, 11, 12]. In circumstances where LMP date is uncertain and ultrasounds are not available, a more accurate method for estimating gestational age may be beneficial.

Recently, DNA methylation has been used to accurately predict chronological age in children and adults [13-16]. Later work revealed that a methylation-based prediction of age may also associate with physiological consequences in adults when a study reported that an increased methylation age relative to chronological age was associated with an increase in mortality risk

[17-22]. However, the predictors optimized in these studies were not designed to estimate GA, and did not attempt to differentiate between different GA, as samples taken at birth were either assigned an age of zero or were excluded from the model [13, 14]. Because the accuracy and precision of a prediction model is, in general, the weakest at the extremes of the distribution, a predictor developed from primarily adult samples would, by nature, be less accurate in neonates than one that is optimized for that purpose.

DNA methylation differences in specific CpG sites have been associated with GA at birth in multiple studies [23-26]. We hypothesize that a predictor designed specifically for use with umbilical cord blood or blood spots already routinely collected for newborn screening could allow for accurate neonatal estimation of GA that may also be informative of developmental stage. The objective of this study was to develop such a predictor to estimate GA from DNA methylation data using umbilical cord blood or blood spot samples and to assess its ability to predict other indicators of developmental maturity.

Methods

Description of Cohorts

Training datasets were selected to include a wide range of gestational ages and ancestries. Publically available datasets were downloaded from the Gene Expression Omnibus (GEO): GSE36642, GSE62924 [27], GSE51180 [28] and GSE30870 [29]. Methylation data for all of these datasets were generated on either the Illumina Infinium HumanMethylation27 BeadChip or Infinium HumanMethylation450 BeadChip (Table 4-1). These methods have been shown to be highly reproducible and consistent with the results of other epigenetic methods [30, 31]. For umbilical cord blood samples, gestational age was defined as the gestational age at birth. For

blood spot samples, gestational age was defined as the gestational age plus the number of days that occurred between birth and sampling.

Emory Women's Mental Health Program (WMHP)

Women with a history of neuropsychiatric illnesses who participated in prospective studies to examine the perinatal course of illness, perinatal pharmacokinetic alterations, and impact of maternal stress on offspring were screened for inclusion in the current study. Details of this study has been described elsewhere [24, 32, 33]. Each woman's obstetrician estimated GA based on the date of her last menstrual period and ultrasound dating. Birth weight in kilograms was assessed at delivery and extracted from the medical records. Umbilical cord blood samples were collected at birth, stored on ice, and processed within 2 hours of delivery. Forty samples were run on the HumanMethylation450 array (abbreviated as WMHP1), and 251 were run on the HumanMethylation27 array (abbreviated as WMHP2). All women provided written informed consent prior to study enrollment following procedures approved by the Institutional Review Board of Emory University.

Conditions Affecting Neurocognitive Development & Learning in Early Childhood (CANDLE)

Neonates were selected from the Urban Child Institute's CANDLE study, a longitudinal cohort study of human development from pregnancy to age three conducted in Shelby County, Tennessee. This cohort has been described in detail elsewhere [24, 34-36]. A combination of obstetrician report (60%) or LMP (40%) was used to estimate GA. Whole umbilical cord blood samples were stored at 4^0C and processed within 24 hours of delivery. Samples from this cohort were interrogated using the HumanMethylation27 BeadChip (N= 198).

Nashville Birth Cohort (NBC)

All subjects were recruited at Centennial Women's Hospital and the Perinatal Research Center in Nashville, TN beginning in 2003. Pregnant women were enrolled during their first clinical visit after obtaining informed consent as described previously [23]. Demographic and clinical data specific to the fetus was collected from clinical records. Gestational age of the neonate was determined by maternal reporting of the last menstrual period and corroboration by ultrasound dating. Umbilical cord blood samples were collected in EDTA tubes soon after placental delivery. Blood samples were centrifuged at 3,000 RPM to separate plasma, and buffy coats were aliquoted and stored at -80°C. Samples were processed on the HumanMethylation450 BeadChip (N=36).

Programming Research in Obesity, GRowth Environment and Social Stress (PROGRESS)

All participants were recruited at 12–24 weeks' gestation through the Mexican social security system after obtaining informed consent between 2007 and 2011. Women had to be greater than 18 years old that have an access to a telephone and a plan to reside within Mexico City for the following 3 years to be enrolled. The study was approved by the Institutional Review Boards of the participating institutions (Brigham and Women's Hospital and The National Institute of Public Health in Mexico). Gestational age was based on the difference between the birth date and the mother's report on enrollment of her last menstrual period. Umbilical cord blood samples were aliquoted and frozen until manual DNA extraction including a red blood cell lysis step followed by isopropanol and ethanol extraction of DNA from total white blood cells. Resulting DNA samples were randomized for plating and bisulfite converted and analyzed on the HumanMethylation450 BeadChip by Illumina FastTrack Services (Illumina Inc., San Diego CA), prior to preprocessing and quality control with the methylumi package (N=148).

Victorian Infant Collaborative Study (VICS)

All 298 survivors born either <1000 g or <28 weeks' gestation in the state of Victoria were enrolled in a longitudinal follow-up study [37], which was approved by the Human Research Ethics Committees at the Royal Women's Hospital, the Mercy Hospital for Women, Monash Medical Centre, and the Royal Children's Hospital, Melbourne, Australia. Gestational age was determined by ultrasound estimation before 20 weeks of gestation, or by menstrual history in the minority if no ultrasound dating was available. DNA samples were derived from dried blood spots taken for newborn screening when infants were several days of age, after obtaining permission from the participants when they were aged 18 years, or from their parents if they were younger than 18 years. 183 samples were processed on the HumanMethylation450 BeadChip and passed sample quality control.

Danish neonatal screening biobank trios (DNSBtrios)

This cohort was recruited as a subset of samples in the Lundbeck Foundation funded initiative for integrative psychiatric research (iPSYCH). Trios (Mother/Father/child) were identified for a psychiatric study where all children and one or both parents had been diagnosed with a psychiatric illness (phenotype data not disclosed for the current study). All samples were isolated within the Danish Neonatal Screening Biobank (DNSB), which stores excess blood from the Danish Neonatal Screening Program. DNSB stores samples from almost every Dane born since 1982.

DNSBtrios study inclusion criteria was known GA as determined via last menstrual period until the late 1990s after which crown rump was used. The samples were also selected based on being collected relatively shortly after birth (<39 days). The samples were collected

via heel prick onto filter paper and then stored at -20°C. DNA was extracted from two punches of 3.2mm before being processed with the HumanMethylation450 BeadChip (N=264).

Prediction and Prevention of Preeclampsia (PREDO)

The mothers participating in the PREDO study come from one of ten hospital maternity clinics participating in the PREDO project in Finland [38]. They were recruited in conjunction of the first screening ultrasound at 12+0 to 13+6 weeks of gestation, based on which gestational age was also determined. Umbilical cord samples were collected in EDTA-tubes and stored immediately at -80°C. 91 samples were processed on the HumanMethylation450 BeadChip and passed sample quality control.

EpiPrem

Longitudinal samples were collected from neonates in the NICU at the Royal Women's Hospital in Melbourne, Australia. Blood collection occurred by heel stick and was collected on Whatman paper shortly after birth at 25 weeks gestation, one day post birth, and at the equivalent of 28, 32, 36, and 40 weeks' gestation (N=2).

Folic Acid supplementation in Pregnancy (FAP)

Healthy young (18-40 years old) pregnant women were recruited from Athens Regional Midwifery Clinic (Athens, GA) at their initial prenatal visit (<12 weeks gestation). In this study, the participants were selected based on the following exclusion criteria: 1) pre-existing chronic condition including anemia, diabetes or hypertension, 2) smokers, 3) those using prescription drugs, 4) those who were carrying more than one fetus. Participants were not allowed to take any vitamins/mineral supplements other than those provided by the research team for the study. Two doses of folic acid (400ug per day, 800ug per day) were provided to participants during

gestation. The study regimen includes all the vitamins/minerals/DHA recommended for pregnant women. Cord blood samples were collected at delivery. 24 samples were processed on the HumanMethylation450 BeadChip and passed sample quality control.

Quality Control and Normalization

All analyses were performed using R version 3.1.2. Datasets used in this study underwent several quality control measures. The DNAm age predictor developed by Horvath was initially run on all samples to establish predicted age and gender [13]. Samples with gender discordance or estimated age >1.5 years were excluded from further analysis. After this initial quality control step, datasets were subjected to standard quality control through the use of the R package CpGassoc [39]. A data frame consisting of β values (methylated signal/(methylated signal + unmethylated signal)) was supplied as input to CpGassoc. Any data point with a detection p-value above .001 was set to missing. CpG sites with >5% missing data were excluded; subsequently samples with > 5% missing data were excluded. These quality control measures were performed to ensure that the predictor is built based on high quality probes and samples. Any probe missing entirely from one of the datasets was excluded from the remaining datasets, so only probes passing quality control in all training datasets, and probes present on both the HumanMethylation450 and HumanMethylation27 arrays were included, for a total of 16,838 probes. Finally, datasets were normalized according to Horvath's modified beta-mixture quantile (BMIQ) normalization [13, 40]. While the original BMIQ is a within-sample normalization method to address probe type bias by modifying the type II distribution to match that of type I probes, Horvath modified this BMIQ procedure for a different purpose: the distribution of each given array is related to that of a "gold standard" array (defined here as the mean across all of the training datasets). Thus, Horvath's modification of the BMIQ method

could be interpreted as a form of between-sample normalization. All training datasets were normalized together, as a single group. After normalization, missing values for each sample were imputed by the *k*-nearest neighbors method where *k*=10, using the R package impute so that no missing values remain in the dataset after pipeline completion [41]. Test datasets were normalized separately, following the same procedures as above. One test cohort, PROGRESS, which was processed with an out of band background correction, dye bias correction and then the original BMIQ procedure, was excluded from the quality control pipeline as raw files were not available. Principal components analysis (PCA) was used to assess the potential impact of BeadChip on the CpG sites selected for inclusion in the predictor. We did not observe clustering by chip, suggesting that chip was not a confounding factor.

Estimation of Cellular Composition

Proportions of white blood cells and nucleated red blood cells were estimated from genome-wide DNA methylation patterns using the method proposed by Houseman et al. [42], with reference samples from homogenous cell populations for white blood cells (CD4$^+$ T cells, CD8$^+$ T cells, natural killer cells, B cells, monocytes, and granulocytes), nucleated red blood cells [43], and whole blood (GSE80310).

Epigenome-wide association study

The R package CpGassoc [39] was used to perform epigenome-wide association studies (EWAS) to assess associations between gestational age and DNA methylation. Two separate EWAS were performed, with and without the inclusion of cellular composition covariates. Each EWAS was performed as a meta-analysis across all cohorts by including indicators for each study as covariates. Test statistics from the two EWAS were plotted to assess the robustness of results to potential cell type heterogeneity.

Elastic Net Regression and Age Prediction

The six training datasets (GSE36642, WMHP1, GSE62924, NBC, GSE51180 and GSE30870) were combined to perform an elastic net regression of gestational age on the 16,838 CpG probes remaining after quality control and filtering. The regression was performed using the R package glmnet to select a parsimonious set of CpG sites predictive of gestational age. Following Horvath [13], the elastic net mixing parameter, alpha, was set to .5 allowing for equal contribution of the ridge and lasso methods [44]. The lambda parameter was chosen through a 10-fold cross validation, which involves randomly partitioning the training dataset into 10 equally-sized subsamples. The cross-validation procedure is then performed 10 times, retaining a different subsample as a validation dataset each time   In the procedure, data from the other nine subsamples is used to build a predictor based on a particular value of lambda, and the fit of the predictor is then tested in the omitted validation set. The mean squared error is calculated for the validation set in each iteration, and then averaged over the 10 subsamples. This procedure is performed for a sequence of lambda values to determine the lambda that yields the minimum mean squared error. No additional covariates were included in the analysis, consistent with the development of the DNAm age predictor by Horvath [13]. The training coefficient values and CpG probes selected from this regression were used to fit a linear model to generate predicted values of GA, based on a modified version of the R code in the DNAm age tutorial published by Horvath [13].   The accuracy of predicted values of gestational age was determined from correlation coefficients obtained through linear regression of DNAm GA and clinical GA.

Analysis of GA acceleration

GA acceleration was calculated as the residual from a linear regression of DNAm GA on clinical estimates of GA for the combined testing dataset. Analysis of DNAm GA with birth

weight and birth weight percentile was then conducted using linear regression of birthweight and birthweight percentile on GA acceleration and covariates for race, estimated cell type proportions, and cohort. Clinically estimated GA was included as a covariate in the analysis for birthweight, but not birthweight percentile as birthweight percentile is already adjusted for clinically estimated gestational age. Maternal insurance status (as a proxy for income) was analyzed in the CANDLE cohort through logistic regression of maternal insurance status on GA acceleration, adjusting for estimated clinical GA, race (African American v Caucasian), and estimated cell type proportions.

Enrichment tests

To assess whether the CpG sites selected for the DNAm GA predictor were more likely than others to be located in functionally relevant regions, two approaches were used. First, CpG positions were intersected the with the hg19 CpG island annotation track from UCSC Genome Browser (http://genome.ucsc.edu), to define whether each site was located in a CpG island, CpG shore (+/- 1.5 kb from island) or CpG shelf (+/- 1.5 kb from shore). Second, the CpG positions were intersected with ENCODE's ChromHMM annotation for lymphoblastoid cell line GM12878, which uses a hidden Markov model to assign genomic features based on the combinatorial pattern of various chromatin marks [45]. The ChromHMM annotation allowed identification of CpGs located in promoters and enhancers. Fisher's exact test was used to assess whether there was significant enrichment of each feature in CpG sites selected for the predictor compared to the full set of 16,838 sites included in the elastic net model. A similar analysis was preformed to assess whether these CpG sites were enriched for sites containing a genetic variant in the 50-bp probe (using annotation derived from the Thousand Genomes Project in [46]), or

sites previously reported to associate with race [36]. DAVID was used to evaluate whether CpG sites used to estimate DNAm GA were located genes enriched for any biological pathways [47].

Results

DNA methylation data from 1,434 neonates, representing 15 independent cohorts, were used for this study. For each sample, HumanMethylation27 or HumanMethylation450 BeadChips (Table 4-1,4-2) were used to generate data from DNA extracted from umbilical cord blood or blood spots. Of the 16,676 CpG sites that passed quality control in the testing and training datasets referenced in Table 4-1, 3,155 (19%) were at least nominally associated with GA in an epigenome-wide association study ($p<.05$, Fig 4-1), and adjustment for proportions of white blood cell subtypes and nucleated red blood cells had little effect on the results (Fig 4-2). Associated CpGs were enriched for a range of biological processes, including cell proliferation and chordate embryonic development (Table 4-3).

Predicting DNAm GA in neonates

To train the DNAm GA predictor, six independent cohorts were selected to sample a wide range of gestational ages and ancestries. Consistent with the approach described by Horvath [13], elastic net regression was used to select a set of 148 CpG sites predictive of GA from a set of 16,838 CpG sites that were available in all training datasets. Although some of the individual studies report associations between the perinatal environment and DNA methylation, no CpG site reported to associate with environmental exposures in these cohorts were among the sites selected for this predictor [27, 32, 33, 48]. Overall, 90 out of 148 CpG sites selected for the predictor (61%) showed some evidence for association with gestational age in the cell-type-adjusted epigenome-wide association study ($p<.05$). In the training datasets, correlation between

the resulting predictor (DNAm GA) and clinically estimated GA was .99 (Figure 4-3A), indicating a strong fit of the model. The 148 CpG sites selected by the elastic net were uniformly distributed across the genome, and were not located in genes more likely to be represented among specific biological pathways (data not shown). They were more likely to reside in CpG island shores than the remaining 16,690 CpG sites that were eligible for inclusion in the predictor (OR=1.73; p=.00096) and less likely to reside in CpG islands (OR=0.53; p=.00019) or active promoters (OR=0.59; p=.0028). The 148 sites showed no significant enrichment or depletion for CpG island shelves or enhancers (Table 4-4). They were also not enriched or depleted for sites with genetic variants located in the probe sequence, or sites previously reported to associate with African American or Caucasian race (Table 4-4) [36, 46, 47].

Predictive accuracy of the model was tested in 1,135 samples from 6 independent datasets. The testing and training datasets had comparable GA distributions (Fig 4-4). In the testing datasets, overall correlation between DNAm GA and GA was .91 (p $< 2.20 \times 10^{-16}$; Figure 4-3B). Within individual test datasets, correlation between GA and DNAm GA remained high (.52 $<$ r $<$ .65; Fig 4-5) though appeared lower than in the combined dataset due to lower sample sizes and GA range. We were not able to obtain similar predictive power using the DNAm age predictor proposed by Horvath, which has a highly significant but much weaker correlation with GA (r=.14, p= 4.89×10^{-6}; Fig 4-6). This correlation coefficient is similar to that observed for prenatal brain samples (r=0.15) [49].

We did not evaluate the Hannum predictor [14] since it is less accurate than the Horvath predictor in children [14, 50]. Of note, only 6 CpG sites included in the DNAm GA predictor overlap with CpG sites in the predictor designed by Horvath, and no sites overlap with the

predictor designed by Hannum. However, one would not necessarily expect overlap. Elastic   net regression selects a parsimonious set of the full list of CpG sites, and among highly correlated CpG sites only one may be chosen, introducing an element of chance into CpG selection. Moreover, the late gestational period is associated with unique developmental changes that cannot be discriminated by the adult predictor, which did not include measures of gestational age in its training dataset.   Thus, this lack of overlap may indicate that the CpG sites predictive of gestational age in neonates are distinct from CpG sites predicting age in adults because of their association with changes specific to gestational development.

The average absolute difference between DNAm GA and gestational age in test samples was 1.49 weeks, with a standard deviation of 1.16 weeks. The median absolute difference ('median error') between DNAm GA and GA was 1.24 weeks. This falls well within the range of error for clinical estimates of GA based on either LMP or ultrasound, as each of these clinical measures has an inherent variability due to recall bias and natural phenotypic variation associated with development [9, 10, 51]. However, it was interesting to note that DNAm GA correlated more strongly with clinical GA estimates based on ultrasound that on those based exclusively on LMP (Fig 4-7). Error rates for ultrasound range from 5-7 days if performed during the first trimester to 3.0-4.3 weeks when performed in the third trimester. This predictor is closer to clinical estimates of GA than post-birth measures using neonatal estimation, which can overestimate the GA of preterm neonates by up to 2.57 weeks [52-56].

The accuracy of this predictor is consistent with that of established clinical methods for estimating GA, though its accuracy can only be interpreted in context of the available clinical measurements. Predictive accuracy was not influenced by neonatal sex as there was no difference between the median errors in males versus females (p=.76).   The median error

between DNAm GA and clinically-estimated GA was 1.07 for the cord blood datasets and 1.57 for blood spot datasets. This discrepancy may be due to differences in precision of GA, as sample collection for blood spots was performed up to 39 days after birth (Fig 4-4 and Fig 4-8). It is also possible that there may be differences in sample quality, as some blood spot samples were stored for more than 30 years, although there was no difference in the number of samples that failed quality control between the cord blood and blood spot datasets. The correlation between DNAm GA and clinically estimated GA was .94 (m.e.=1.4) for samples processed on the HumanMethylation450 array and .55 (m.e.=1.02) for samples processed on the HumanMethylation27 array (Fig 4-9). This discrepancy is likely due to the differences in gestational age range between samples run on the two arrays (19.3 and 11.1 weeks, respectively). Finally, the partial correlation from regressions of DNAm GA on clinically estimated GA did not substantially change when cell composition covariates were included, suggesting that the accuracy of the predictor is not confounded by cellular heterogeneity ($r_{original}$=.91, $r_{cell\ type\ adjusted}$=.81).

To limit concerns regarding the potential for overfitting of the models, we next validated the predictor in a second testing dataset, comprised of 92 samples from three cohorts (FAP, GSE66459, and GSE69633) that were not included in the initial testing or training sets. Cohort demographics are provided in Table 4-2. The correlation in these datasets is similar to that of the first testing dataset (r=.89, m.e.=.89; Fig 4-10), further indicating that this model fits well when applied to novel datasets and should be generalizable to other studies.

Accuracy of DNAm GA in the same subjects

Serial blood sampling was conducted from two neonates admitted to the Neonatal Intensive Care Unit (NICU), independent of the testing and training samples. Seven timepoints

were collected between birth at 25 weeks and discharge at 40 weeks. DNAm GA increased as expected over time from birth until term equivalency (Figure 4-11). These pilot data demonstrate that the predictor has the sensitivity required to detect changes in DNAm GA in the magnitude of days or weeks, and that methylation patterns change from birth to term equivalency in a predictable manner.

<u>DNAm GA as a measure of developmental age</u>

In adults, the difference between DNA methylation-based age estimates and chronological age associates with all cause mortality, HIV status, and Down syndrome [17, 21, 57]. This difference is usually described as age acceleration [13]. We calculated a similar measure, which we will subsequently refer to as GA acceleration, in our cohorts by using the residual of a linear model regressing DNAm GA on clinically estimated GA. Because accelerated GA may indicate increased developmental maturity, we sought to evaluate whether GA acceleration associated with perinatal measures of health and development in the cohorts with available data.

Birthweight is widely used as a proxy of developmental maturity in studies assessing the association between the prenatal environment and short-term or long-term neonatal risk, with those born at the lowest birthweight generally having the highest risk for mortality over the first year of life and for cardio-metabolic conditions as adults [58, 59]. Birthweight is positively correlated with GA so birthweight percentile, which is calculated based on birthweight averages for a given GA corrected for fetal sex, is commonly used as an indicator of perinatal health [60, 61]. Previous studies have shown that infants in the lowest birthweight percentiles have an increased risk of perinatal death and other adverse outcomes, and are often defined as growth restricted [62, 63]. In cohorts with available data, GA acceleration significantly predicted

birthweight percentile ($p=4.5\times10^{-4}$; Fig 4-12A) and birthweight ($p=.033$; Fig 4-12B) after correcting for clinically estimated GA, race, estimated cell type proportions, and cohort. Consistent with the idea that DNAm GA may reflect maturity, the fitted regression model predicts approximately the fiftieth percentile to have GA acceleration of 0. Thus, neonates falling in the lowest birthweight percentiles are lower, while neonates falling in the highest percentiles show higher, or accelerated GA. There was no association between GA acceleration calculated using the DNAm age predictor of Horvath [13] and either birthweight or birthweight percentile (Fig 4-13).

One study by Appleton and colleagues suggests that socioeconomic adversity promotes adverse health outcomes through epigenetic programming of neonatal DNA methylation [64]. We hypothesize that factors related to early life adversity might influence the developmental age of the neonate. One such factor is socioeconomic status, which is essential to examine as children born into socioeconomically disadvantaged families, often operationalized by insurance status (Medicaid versus private health insurance), have poorer health in childhood and early adulthood [65, 66]. In the most socioeconomically diverse cohort (CANDLE), developmental aging associated with maternal Medicaid status ($p=.023$) after adjusting for race, clinically estimated GA, and estimated cell type proportions (Fig 4-14). Specifically, methylation-based estimates of GA were lower than clinical estimates for the neonates of women on Medicaid compared with women with private health insurance. This association supports the hypothesis that prenatal adversity associates with changes in neonatal methylation consistent with a delayed developmental age, which may have consequences later in life.

Discussion

GA can be accurately predicted between 24 and 44 weeks gestation using DNA methylation values obtained from both umbilical cord blood and blood spot samples. DNAm GA is more concordant with GA based estimates performed with the gold standard of ultrasound with estimates based on LMP. However, the question remains as to whether GA acceleration is truly a measure of maturity versus a reflection of the relative accuracy of DNAm GA compared clinical estimates. We consider three possibilities for interpreting the difference between DNAm GA and clinically-estimated GA. First, an accelerated GA may reflect differences in physiological development of the neonate such that neonates with a higher DNAm GA are more developmentally mature than their chronological age suggests. A second possibility is that the differences between DNAm GA and chronological GA reflect epigenetic programming by early life environmental exposures, such as maternal prenatal stress or pregnancy disorders, which may affect neonatal outcomes and development [67]. Finally, any difference may simply be reflective of the variable nature of clinical GA estimations; evaluation of DNAm GA in neonates conceived through in vitro fertilization (IVF) would be helpful for delineating these different possibilities. These models may be interrelated, such that the true interpretation is likely a combination of these possibilities. A future study examining other prediction methods, including the use of non-linear models or transformations, may facilitate this interpretation by further delineating developmental differences between early and late gestational ages. Overall, our results suggest that DNAm GA and GA acceleration are promising tools for evaluating neonatal developmental maturity.

A targeted assay of the CpG sites necessary to compute DNAm GA could provide a rapid and robust estimator of GA at birth, and the framework described in this paper could be used to develop and validate a predictor based on other tissues that may be sampled prior to delivery,

such as chorionic villi or amniotic fluid. Our results suggest that DNAm GA is highly reproducible and can predict measures of developmental maturity, such as birthweight, better than clinical estimates of GA alone. As such, it has the potential to serve as a biomarker for GA and the rate of neonatal development. Recent studies of GA and DNA methylation [23-26] support that shifts in methylation underlie the aging process, further supporting the development of methylation-based biomarkers. DNA methylation is a convenient molecular marker for GA in that umbilical cord blood and blood sampling are routinely performed to monitor neonatal health in humans, and it can be readily sampled repeatedly in the same person, as demonstrated by the time course data in the subjects from preterm birth through term equivalency.

As a biomarker, DNAm GA and GA acceleration would have numerous clinical, research and forensic applications. It would serve as a molecular marker of GA that complements clinical estimates, when available, and provides additional information when clinical estimates are unavailable or unreliable. For example, it could be used to estimate GA in women who seek prenatal care late in pregnancy, are unsure of the date of their last menstrual period, or did not have ultrasounds performed early in pregnancy. DNAm GA is more precise than the estimation methods typically performed at birth, which rely on biometric measurements. Precise knowledge of GA would be most informative for neonates born extremely preterm, when parents and clinicians are confronted with decisions regarding active intensive care interventions versus providing comfort care. GA based on an epigenetic developmental profile may also complement clinical estimates of GA, providing a screening tool to identify children who may benefit from additional monitoring and care. Studies to explore the extent to which DNAm GA reflects developmental maturity, and thus may be a more reliable predictor of outcomes after preterm birth compared to time or growth-based methods, are needed.

DNAm GA may also serve as a surrogate marker for developmental maturity in research studies of neonatal development, interventions, and disease. Our results already demonstrate that it will be fruitful to study antenatal and perinatal factors that associate with DNAm GA and GA acceleration, and to determine whether these metrics are better prognosticators of neonatal well-being than conventional measures. Future studies should evaluate the effects of maternal stress, nutrition, and interventions such as vitamin supplementation that are highly relevant to fetal development and pregnancy outcomes. Future research could also explore whether GA acceleration relates to risk of developing pediatric disorders such as autism, and whether it can predict health outcomes later in life. Finally, establishing precise gestational age is important for forensic, anthropologic or other medico-legal investigations. Indeed, DNA methylation-based predictors of adult age are already under investigation for forensic applications [68]. In summary, we have identified a potential biomarker for GA with an abundance of applications that warrant further investigation and development.

Table 4-1: Description of testing and training cohorts. Training datasets and test datasets were chosen to represent a similar range of gestational ages.

Training Datasets								
Dataset	**N**	**GA Range**	**GA mean ± SD**	**%Male**	**Race**	**Nationality**	**Source**	**Array**
GSE36642	51	32-38	36.3±1.7	56.9	White	Australian	Cord	27k
WMHP1	40	31-41	37.9±2.3	47.5	98% White	American	Cord	450k
GSE62924	38	34-41	39.1±1.4	42.1	White	Mexican	Cord	450k
NBC	36	24-41	36.0±5.4	47.2	Black	American	Cord	450k
GSE51180	23	25-42	32.7±6.6	69.6	White	Australian	Spot	450k
GSE30870	19	34-41	38.9±2.1	NA*	White	Spanish	Cord	450k
Test Datasets								
Dataset	**N**	**GA Range**	**GA mean ± SD**	**%Male**	**Race**	**Nationality**	**Source**	
DNSBtrios	264	28-44	40.3±1.9	64.9	White	Danish	Spot	450k
WMHP2	251	33-43	38.7±1.4	51.0	80% White	American	Cord	27k
CANDLE	198	32-41	39±1.3	52.0	51% Black, 40.4% White	American	Cord	27k
VICS	183	24-35	28.0±2.1	42.1	89% White	Australian	Spot	450k
PROGRESS	148	30-43	38.6±1.7	52.0	White	Mexican	Cord	450k
PREDO	91	31-42	39.6±1.5	54.9	White	Finnish	Cord	450k

* Indicates data not available.

Table 4-2: Description of additional testing cohorts. Additional cohorts were identified to further validate the accuracy of the predictor.

Testing Datasets							
Dataset	N	GA Range	GA mean ± SD	%Male	Race	Nationality	Source
FAP	24	38-43	40.1±1.2	45.8	83% White	American	Cord
GSE66459	22	26-42	35.2±5.2	50.0	White	Dutch	Cord
GSE69633	46	36-41	38.9±1.3	50.0	White	Mexican	Cord

Table 4-3: Enrichment for the top 20 biological processes in genes containing GA-associated

CpG sites. The p-value is adjusted using a Benjamini- Hochberg correction.

Biological Process	Count	%	p-value
Biological adhesion	172	6.3	7.60E-07
Cell adhesion	172	6.3	1.40E-06
Regulation of cell proliferation	184	6.7	6.60E-06
Regulation of phosphorylation	116	4.2	1.00E-04
Cell motion	118	4.3	1.00E-04
Regulation of phosphorus metabolic process	120	4.4	1.20E-04
Regulation of phosphate metabolic process	120	4.4	1.20E-04
Cell migration	78	2.8	1.20E-04
Cell motility	83	3	1.80E-04
Localization of cell	83	3	1.80E-04
Vasculature development	71	2.6	1.90E-04
Intracellular signaling cascade	260	9.5	1.90E-04
Chemical homeostasis	123	4.5	2.00E-04
Regulation of cell motion	58	2.1	2.50E-04
Blood vessel development	69	2.5	2.50E-04
Regulation of cell adhesion	45	1.6	3.30E-04
Positive regulation of cell proliferation	102	3.7	4.70E-04
Regulation of cell death	177	6.4	5.20E-04
Chordate embryonic development	85	3.1	6.10E-04
Cell activation	76	2.8	6.30E-04

Table 4-4: Enrichment tests for 148 CpG sites selected by elastic net regression**.**

Significant enrichment (OR > 1) or depletion (OR < 1) after adjustment for 12 tests (p < .05/12)

is indicated by **bold text**.

Enrichment/depletion for:	OR	95% C.I.	p-value
CpG islands	0.53	(0.37, 0.75)	**.00019**
CpG island shores	1.73	(1.24, 2.43)	**.00096**
CpG island shelves	0.76	(0.34, 1.49)	.54
Promoter regions	0.70	(0.50, 0.99)	.038
- Active promoters	0.59	(0.41, 0.84)	**.0028**
- Weak promoters	1.26	(0.74, 2.03)	.36
- Poised promoters	1.26	(0.65, 2.24)	.42
Enhancer regions	1.38	(0.76, 2.33)	.23
- Strong enhancers	1.79	(0.75, 3.65)	.14
- Weak enhancers	1.09	(0.46, 2.21)	.71
Race-associated CpG sites	0.97	(0.57, 1.57)	.99
CpG sites with genetic variants in probe	0.78	(0.45, 1.27)	.37

OR: odds ratio.
C. I.: confidence interval

Figure 4-1: Correlation between DNAm GA and GA. (A) DNAm GA and estimated clinical GA (EGA) are highly correlated in the training dataset: r=.99, median error (m.e.) =.35. (B) DNAm GA and estimated clinical GA were also highly correlated in the testing dataset: r =.91, median error = 1.24. Solid line = regression line; dotted line indicates equivalence.
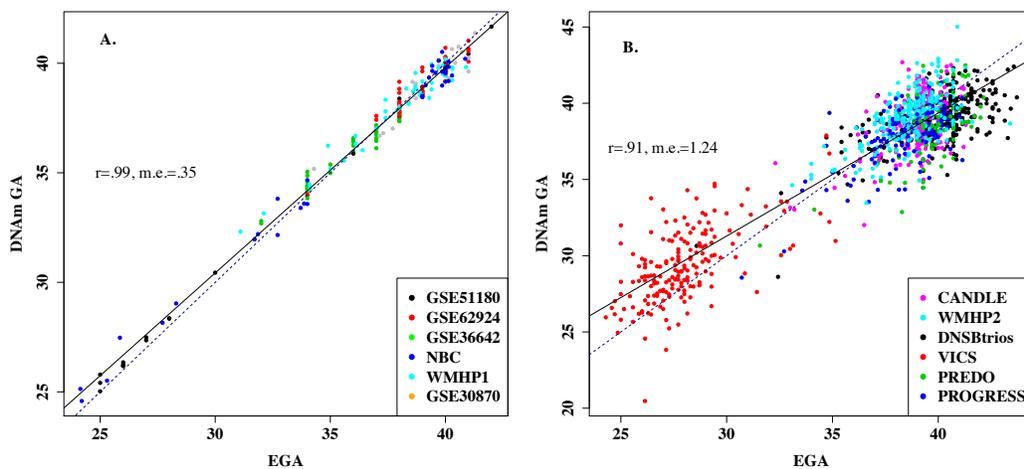
Figure 4-2: Manhattan plot showing the distribution of GA-associated CpG sites across the genome**.** Points falling above the dashed horizontal line indicate experiment-wide significance FDR<.05; points above the solid horizontal line are significant according to a more conservative step-down Bonferroni adjustment for 16,676 CpG sites.
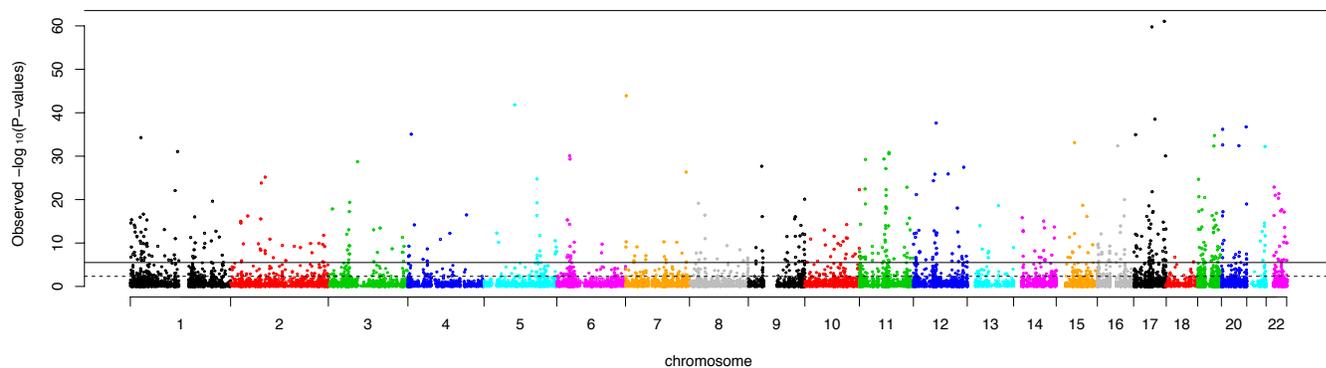
Figure 4-3: Comparison of t-statistics from two epigenome-wide associations studies (EWAS) of gestational age to assess robustness of results to cell type heterogeneity. The x-axis shows t-statistics from an EWAS for gestational age adjusting for estimated cellular composition (proportions of six white blood cell subtypes and nucleated red blood cell counts) while the y-axis shows t-statistics unadjusted for cell composition.
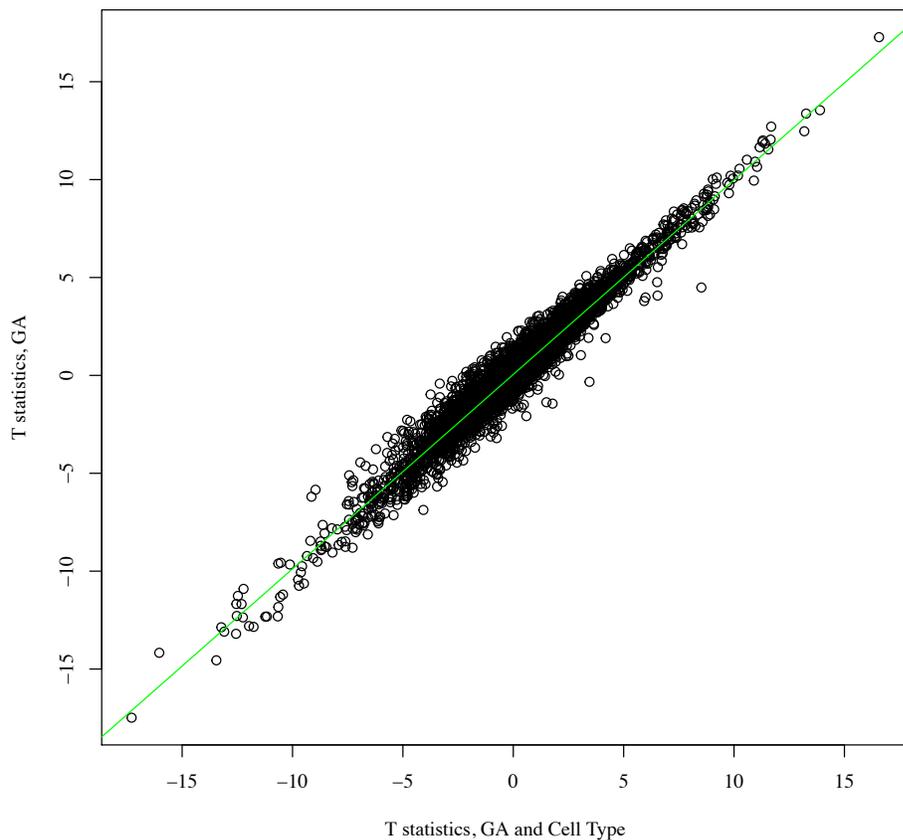
Figure 4-4: Distribution of clinically estimated gestational age (EGA) ranges in the training (A) and testing (B) datasets**.**
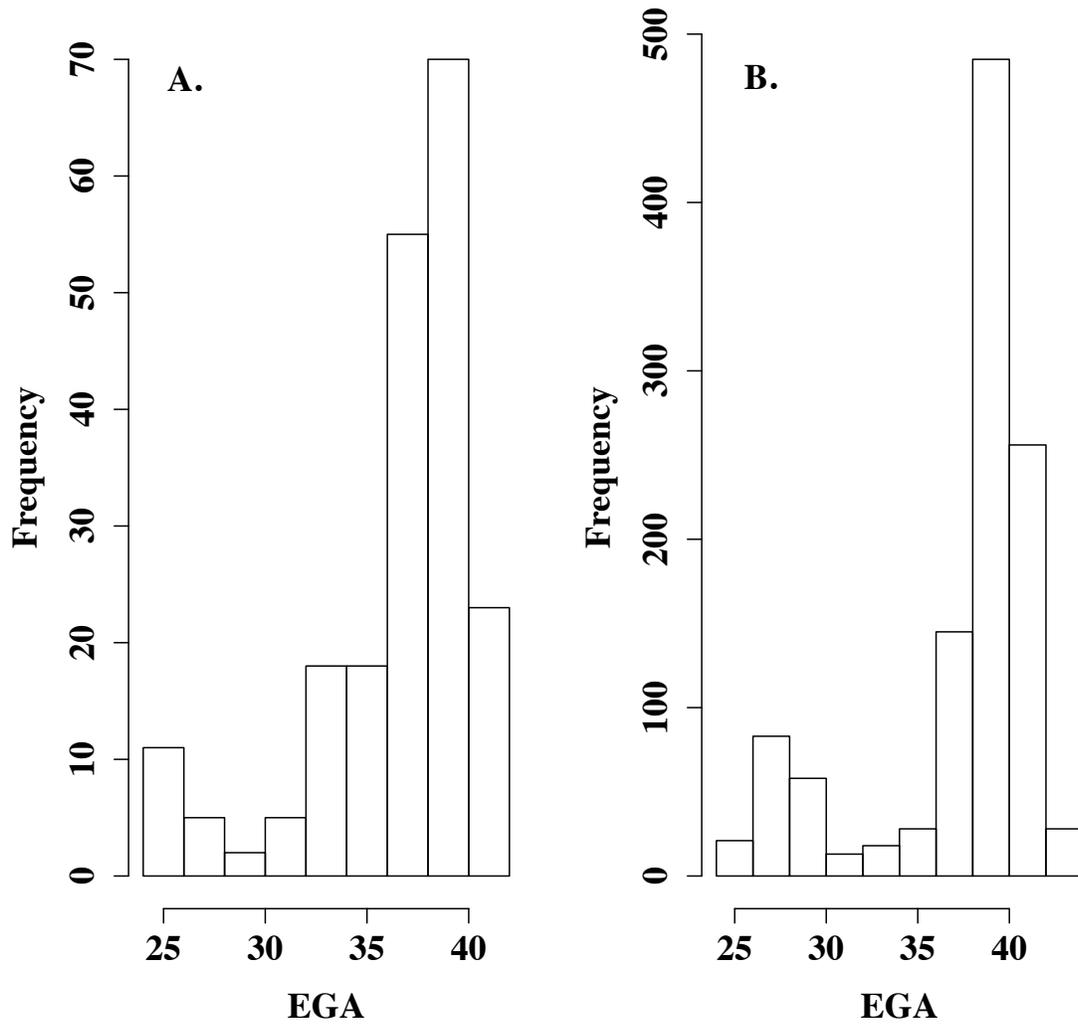
Figure 4-5: Correlation between clinically estimated gestational age (EGA) and DNAm GA for each testing dataset. Solid line = regression line; dotted line indicates equivalence. Median absolute difference ('median error') = m.e.
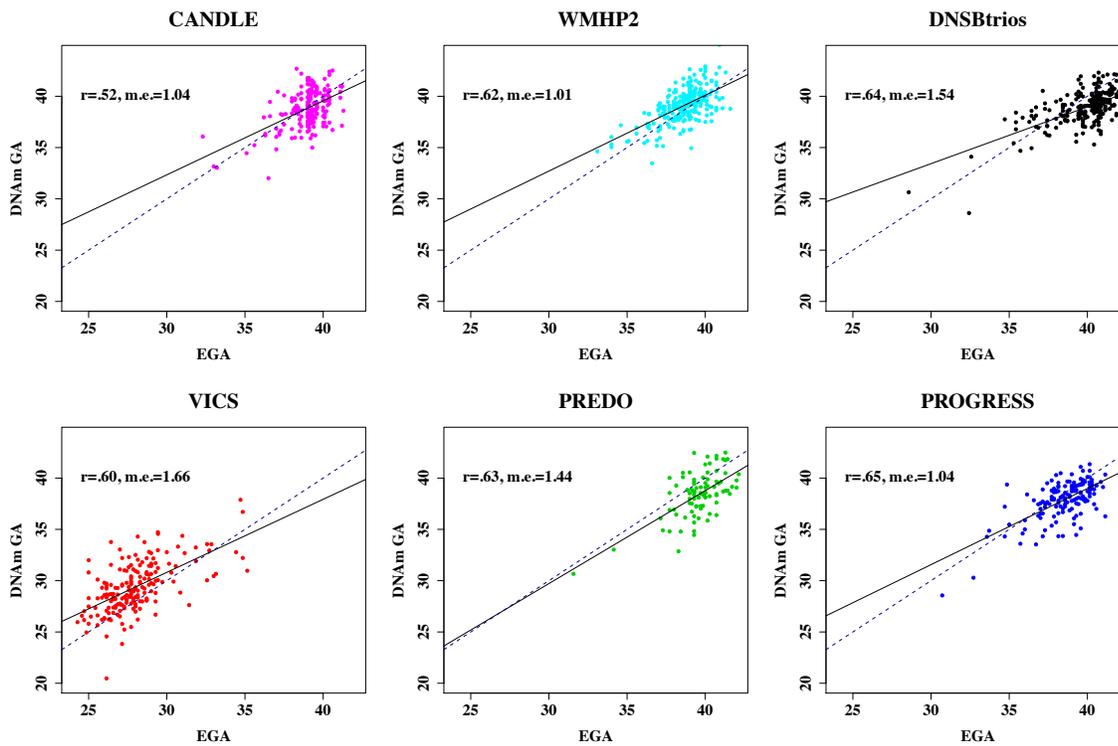
Figure 4-6: Correlation between clinically estimated GA (EGA) and DNAm age estimated using the Horvath predictor [2] (r=.14, p= 4.89×10^{-6}, median error=9.19). DNAm age is represented in equivalent weeks gestation. Red= VICS, fuchsia= CANDLE, cyan= WMHP2, black= DNSBtrios, green= PREDO, blue= PROGRESS. Solid line = regression line; dotted line indicates equivalence.
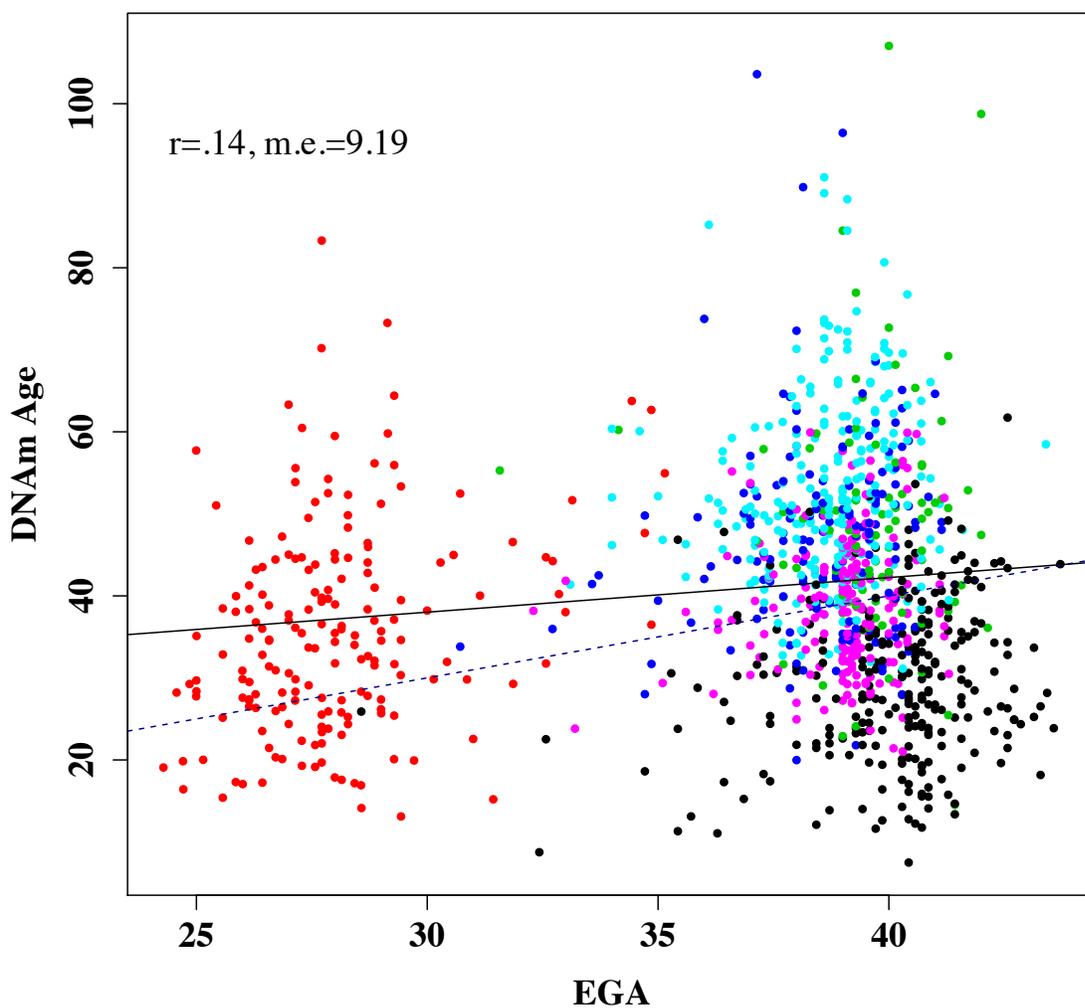
Figure 4-7: Comparison of DNAm GA with clinical GA estimates based on LMP (last menstrual period) or ultrasound in WMHP2. (r_{LMP}=.41, p_{LMP}= 5.5×10^{-8}, median error$_{LMP}$=.93; $r_{Ultrasound}$=.54, $p_{Ultrasound}$= 4.79×10^{-14}, median error$_{Ultrasound}$=1.01).
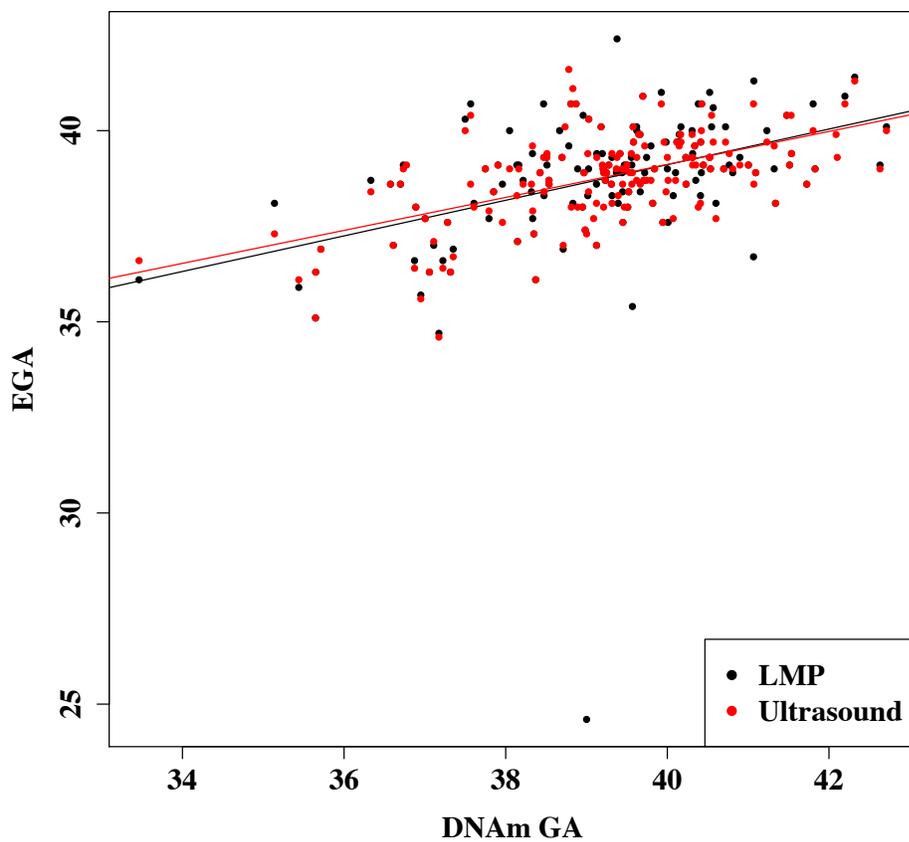
Figure 4-8: Predictive accuracy of DNAm GA in cord blood and blood spot samples. Clinically estimated GA (EGA) is depicted on the x-axis. $r_{CordBlood} = .57$, $p_{CordBlood} < 2.2 \times 10^{-16}$, median error$_{CordBlood} = 1.57$; $r_{BloodSpot} = .95$, $p_{BloodSpot}$ $p < 2.2 \times 10^{-16}$, median error$_{Ultrasound} = 1.07$
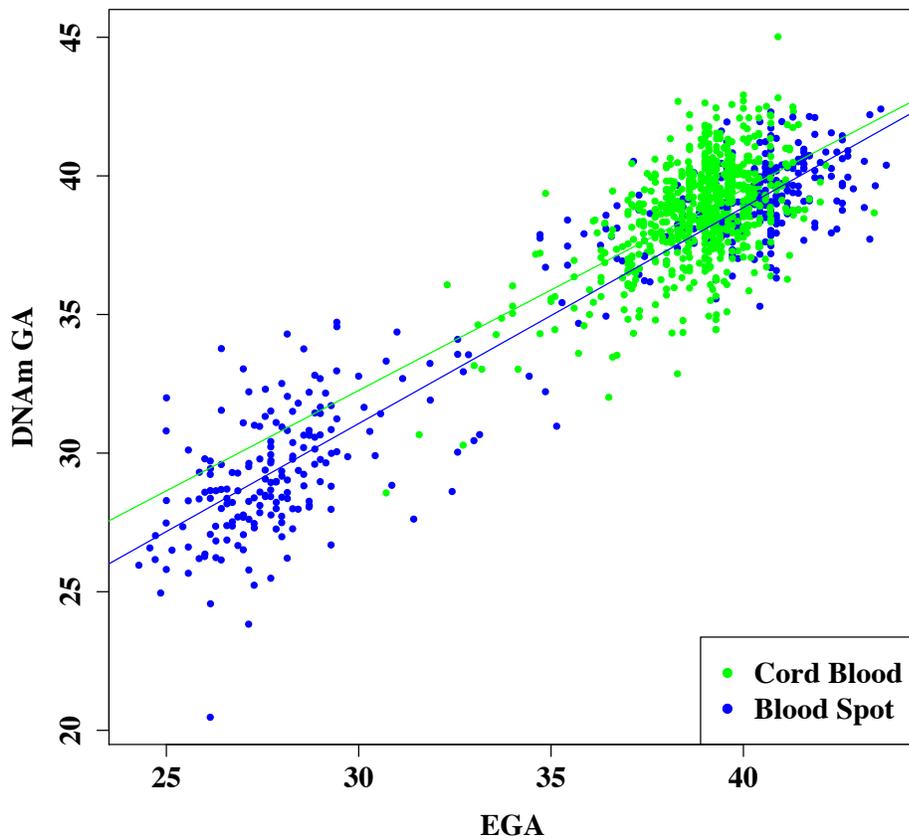
Figure 4-9: Predictive accuracy of DNAm GA on the HumanMethylation27 array (A) and HumanMethylation450 array (B). The x-axis represents clinically estimated gestational age (EGA) and the y-axis represents DNA methylation gestational age (DNAm GA).
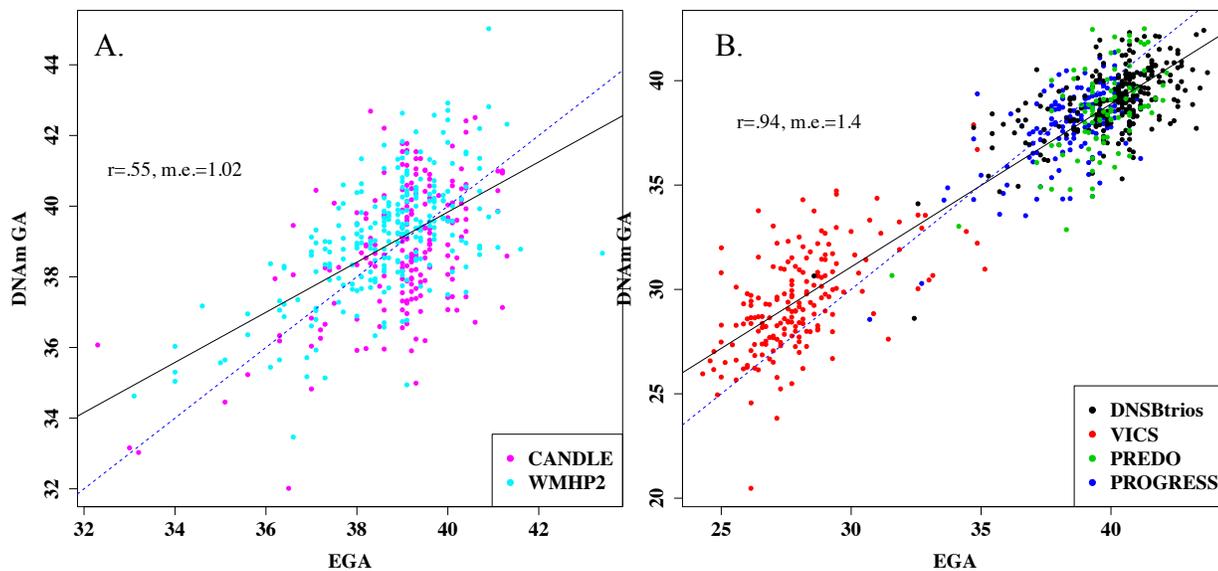
Figure 4-10: Predictive accuracy of DNAm GA on three additional testing datasets (FAP, GSE66459, GSE69633). The x-axis represents clinically estimated gestational age (EGA) and the y-axis represents DNA methylation gestational age (DNAm GA).
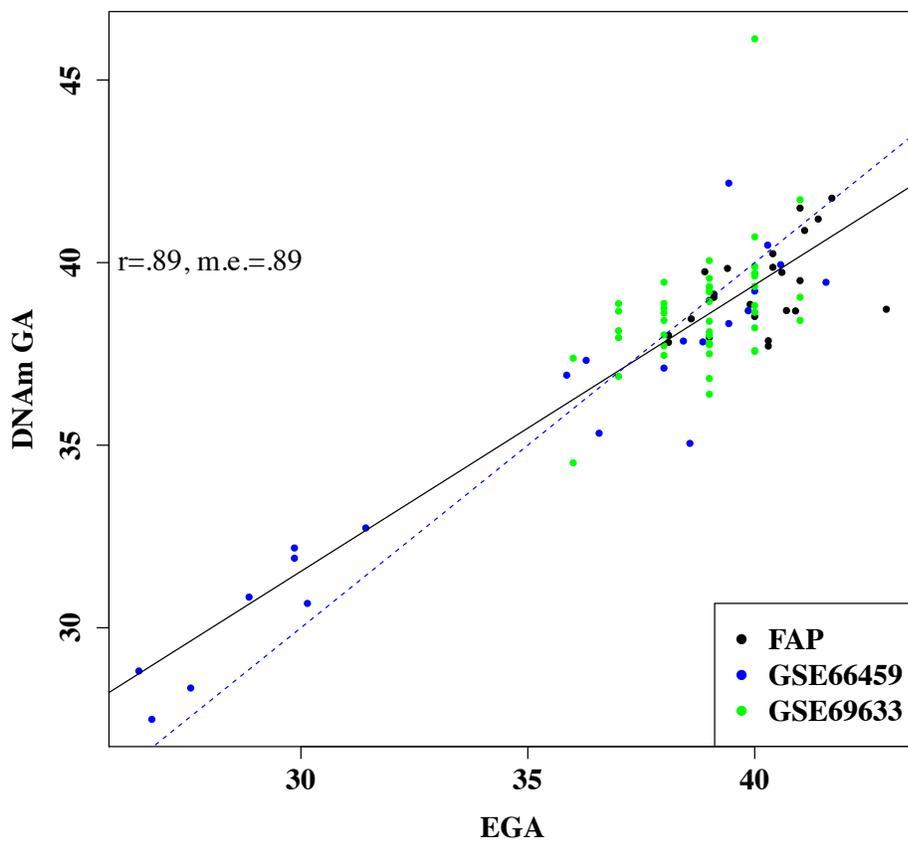
Figure 4-11: Reproducibility of DNAm GA. DNAm GA from birth until term equivalency for 2 subjects recruited through the EpiPrem study, gestational age at birth 25 weeks. DNAm GA increases appropriately with gestational age in weeks. Change in DNAm GA over equivalent weeks gestation.
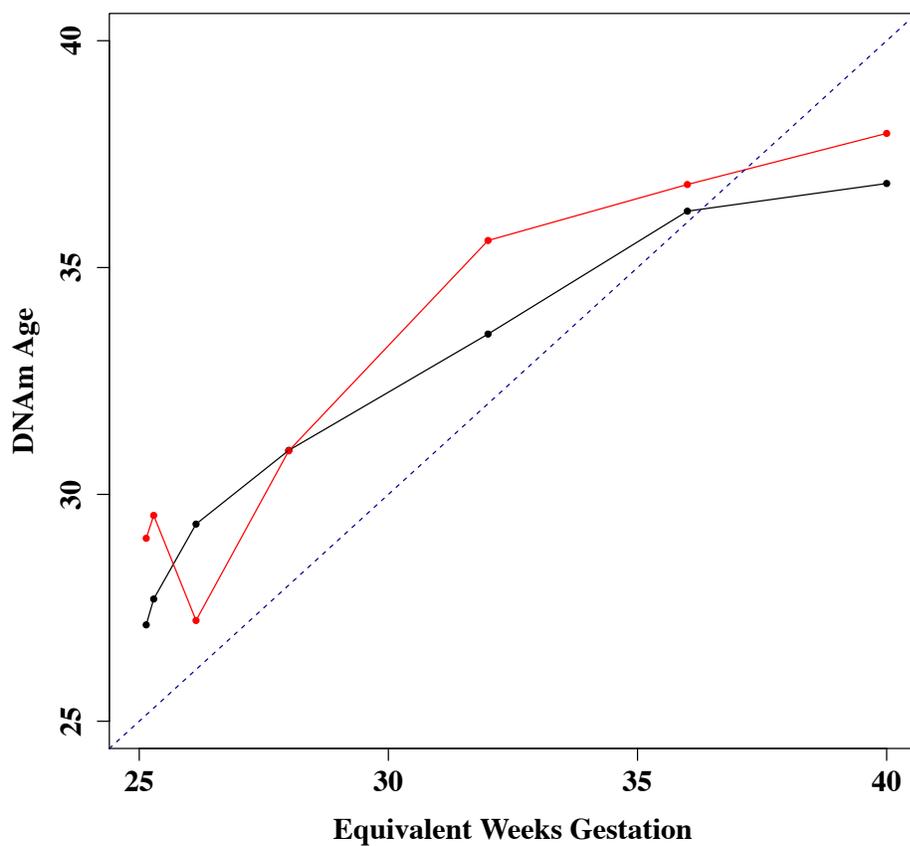
Figure 4-12: GA acceleration associates with birthweight. The association between GA acceleration and (A) birthweight percentile (p=4.5×10^{-4}) or (B) birthweight (p=.033) adjusted for race, cellular composition, cohort and gestational age in CANDLE, WMHP, and PROGRESS. Solid line = regression line.
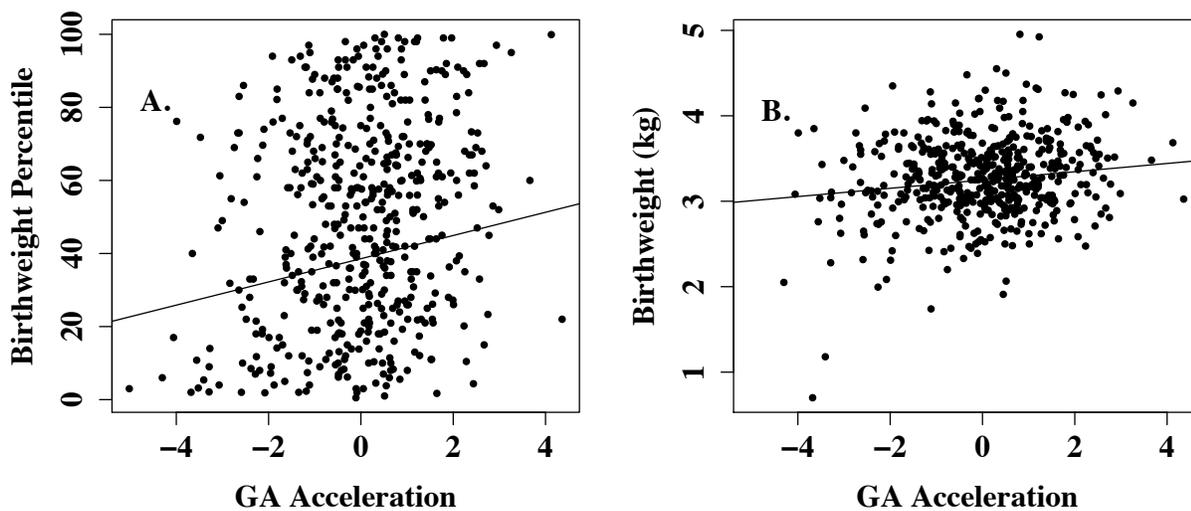
Figure 4-13: Association between birthweight percentile (A) and birthweight (B) and age

acceleration calculated using the DNAm age predictor **[2].** $r_{birthweight}$= .0014, $p_{birthweight}$ = .79;

$r_{birthweightpercentile}$= .05, $p_{birthweightpercentile}$= .28 Solid line = regression line.
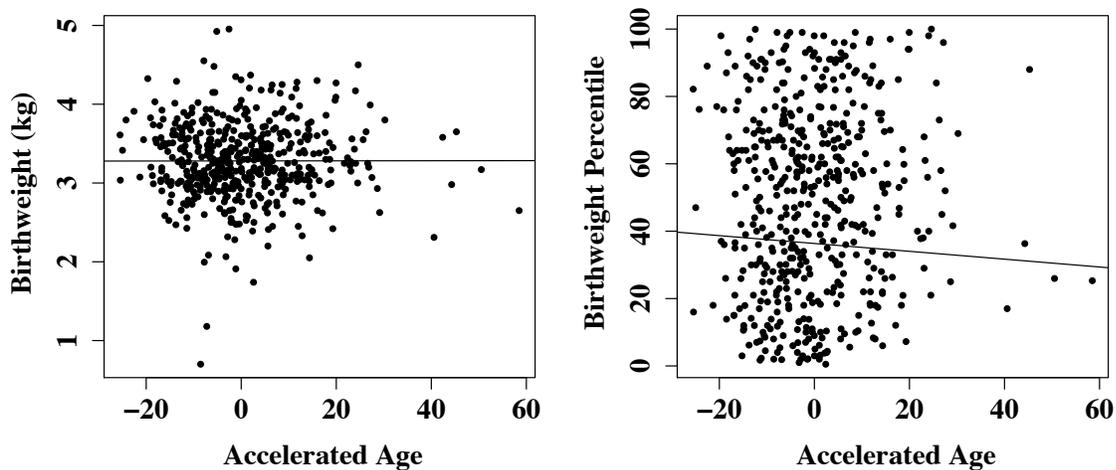
Figure 4-14: Maternal insurance status and GA acceleration. Neonates born to mothers with private insurance have higher GA acceleration than neonates born to mothers on Medicaid (p=.023) after adjusting for race, gestational age, and cellular composition.
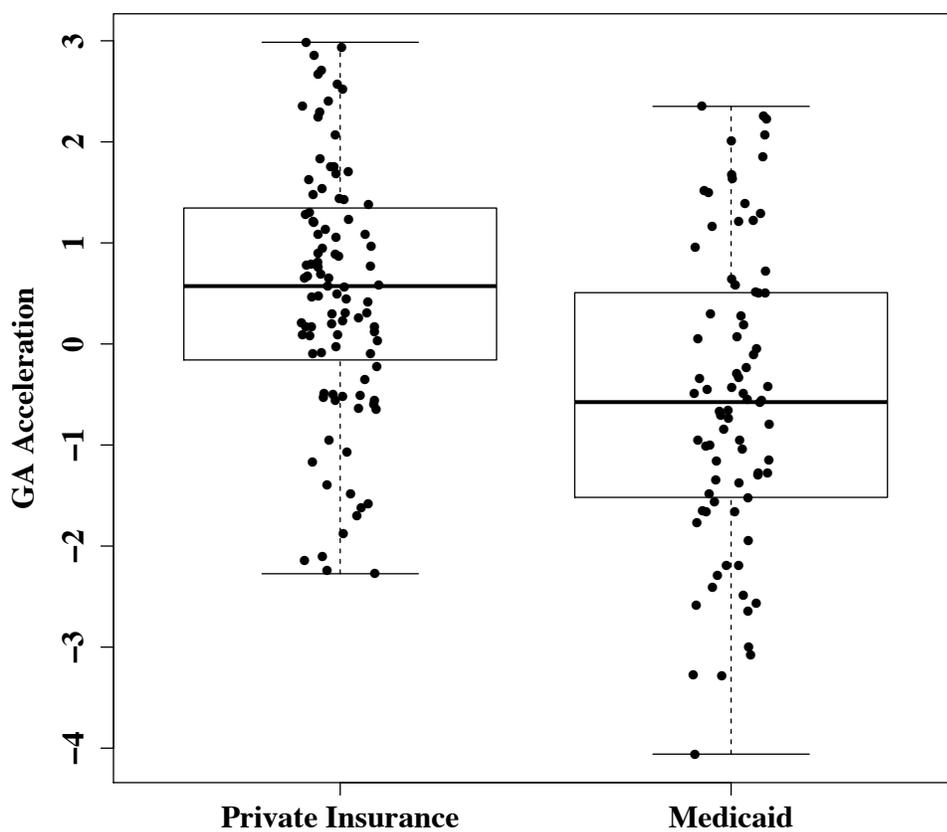
References

1.   Young PC, Glasgow TS, Li X, Guest-Warnick G, Stoddard G. Mortality of late-preterm (near-term) newborns in Utah. Pediatrics. 2007;119(3):e659-65. doi: 10.1542/peds.2006-2486. PubMed PMID: 17332185.

2.   Engle WA. Morbidity and mortality in late preterm and early term newborns: a continuum. Clinics in perinatology. 2011;38(3):493-516. doi: 10.1016/j.clp.2011.06.009. PubMed PMID: 21890021.

3.   Yang S, Platt RW, Kramer MS. Variation in child cognitive ability by week of gestation among healthy term births. American journal of epidemiology. 2010;171(4):399-406. doi: 10.1093/aje/kwp413. PubMed PMID: 20080810; PubMed Central PMCID: PMC3435092.

4.   Davis EP, Buss C, Muftuler LT, Head K, Hasso A, Wing DA, et al. Children's Brain Development Benefits from Longer Gestation. Frontiers in psychology. 2011;2:1. doi: 10.3389/fpsyg.2011.00001. PubMed PMID: 21713130; PubMed Central PMCID: PMC3111445.

5.   Hansen AK, Wisborg K, Uldbjerg N, Henriksen TB. Risk of respiratory morbidity in term infants delivered by elective caesarean section: cohort study. Bmj. 2008;336(7635):85-7. doi: 10.1136/bmj.39405.539282.BE. PubMed PMID: 18077440; PubMed Central PMCID: PMC2190264.

6.   Parikh LI, Reddy UM, Mannisto T, Mendola P, Sjaarda L, Hinkle S, et al. Neonatal outcomes in early term birth. American journal of obstetrics and gynecology. 2014;211(3):265 e1- e11. doi: 10.1016/j.ajog.2014.03.021. PubMed PMID: 24631438; PubMed Central PMCID: PMC4149822.

7.	Taylor BD, Holzman CB, Fichorova RN, Tian Y, Jones NM, Fu W, et al. Inflammation biomarkers in vaginal fluid and preterm delivery. Human reproduction. 2013;28(4):942-52. doi: 10.1093/humrep/det019. PubMed PMID: 23416276; PubMed Central PMCID: PMC3600841.

8.	Gynecologists TACoOa. Early Preterm Birth 2011 [cited 2015]. Available from: http://www.acog.org/Patients/FAQs/Early-Preterm-Birth.

9.	Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. Paediatric and perinatal epidemiology. 2007;21 Suppl 2:86-96. doi: 10.1111/j.1365-3016.2007.00865.x. PubMed PMID: 17803622.

10.	Mongelli M, Wilcox M, Gardosi J. Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates. American journal of obstetrics and gynecology. 1996;174(1 Pt 1):278-81. PubMed PMID: 8572021.

11.	Dubowitz LM, Dubowitz V, Goldberg C. Clinical assessment of gestational age in the newborn infant. The Journal of pediatrics. 1970;77(1):1-10. PubMed PMID: 5430794.

12.	Ballard JL, Khoury JC, Wedig K, Wang L, Eilers-Walsman BL, Lipp R. New Ballard Score, expanded to include extremely premature infants. The Journal of pediatrics. 1991;119(3):417-23. PubMed PMID: 1880657.

13.	Horvath S. DNA methylation age of human tissues and cell types. Genome biology. 2013;14(10):R115. doi: 10.1186/gb-2013-14-10-r115. PubMed PMID: 24138928; PubMed Central PMCID: PMC4015143.

14.	Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Molecular cell. 2013;49(2):359-67. doi: 10.1016/j.molcel.2012.10.016. PubMed PMID: 23177740; PubMed Central PMCID: PMC3780611.

15.   Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, et al. Epigenetic predictor of age. PloS one. 2011;6(6):e14821. doi: 10.1371/journal.pone.0014821. PubMed PMID: 21731603; PubMed Central PMCID: PMC3120753.

16.   Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome biology. 2014;15(2):R24. doi: 10.1186/gb-2014-15-2-r24. PubMed PMID: 24490752; PubMed Central PMCID: PMC4053864.

17.   Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, et al. DNA methylation age of blood predicts all-cause mortality in later life. Genome biology. 2015;16:25. doi: 10.1186/s13059-015-0584-6. PubMed PMID: 25633388; PubMed Central PMCID: PMC4350614.

18.   Christiansen L, Lenart A, Tan Q, Vaupel JW, Aviv A, McGue M, et al. DNA methylation age is associated with mortality in a longitudinal Danish twin study. Aging cell. 2016;15(1):149-54. doi: 10.1111/acel.12421. PubMed PMID: 26594032; PubMed Central PMCID: PMC4717264.

19.   Levine ME, Lu AT, Bennett DA, Horvath S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. Aging. 2015;7(12):1198-211. PubMed PMID: 26684672; PubMed Central PMCID: PMC4712342.

20.   Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. International journal of epidemiology. 2015;44(4):1388-96. doi: 10.1093/ije/dyu277. PubMed PMID: 25617346; PubMed Central PMCID: PMC4588858.

21.   Horvath S, Levine AJ. HIV-1 Infection Accelerates Age According to the Epigenetic Clock. The Journal of infectious diseases. 2015;212(10):1563-73. doi: 10.1093/infdis/jiv277. PubMed PMID: 25969563; PubMed Central PMCID: PMC4621253.

22.   Breitling LP, Saum KU, Perna L, Schottker B, Holleczek B, Brenner H. Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. Clinical epigenetics. 2016;8:21. doi: 10.1186/s13148-016-0186-5. PubMed PMID: 26925173; PubMed Central PMCID: PMC4768341.

23.   Parets SE, Conneely KN, Kilaru V, Fortunato SJ, Syed TA, Saade G, et al. Fetal DNA Methylation Associates with Early Spontaneous Preterm Birth and Gestational Age. PloS one. 2013;8(6):e67489. doi: 10.1371/journal.pone.0067489. PubMed PMID: 23826308; PubMed Central PMCID: PMC3694903.

24.   Schroeder JW, Conneely KN, Cubells JC, Kilaru V, Newport DJ, Knight BT, et al. Neonatal DNA methylation patterns associate with gestational age. Epigenetics : official journal of the DNA Methylation Society. 2011;6(12):1498-504. doi: 10.4161/epi.6.12.18296. PubMed PMID: 22139580; PubMed Central PMCID: PMC3256334.

25.   Simpkin AJ, Suderman M, Gaunt TR, Lyttleton O, McArdle WL, Ring SM, et al. Longitudinal analysis of DNA methylation associated with birth weight and gestational age. Human molecular genetics. 2015. doi: 10.1093/hmg/ddv119. PubMed PMID: 25869828.

26.   Lee H, Jaffe AE, Feinberg JI, Tryggvadottir R, Brown S, Montano C, et al. DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. International journal of epidemiology. 2012;41(1):188-99. doi: 10.1093/ije/dyr237. PubMed PMID: 22422452; PubMed Central PMCID: PMC3304532.

27.   Rojas D, Rager JE, Smeester L, Bailey KA, Drobna Z, Rubio-Andrade M, et al. Prenatal arsenic exposure and the epigenome: identifying sites of 5-methylcytosine alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes. Toxicological sciences : an official journal of the Society of Toxicology. 2015;143(1):97-106. doi: 10.1093/toxsci/kfu210. PubMed PMID: 25304211; PubMed Central PMCID: PMC4274382.

28.   Cruickshank MN, Oshlack A, Theda C, Davis PG, Martino D, Sheehan P, et al. Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. Genome medicine. 2013;5(10):96. doi: 10.1186/gm500. PubMed PMID: 24134860; PubMed Central PMCID: PMC3978871.

29.   Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. Distinct DNA methylomes of newborns and centenarians. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(26):10522-7. doi: 10.1073/pnas.1120658109. PubMed PMID: 22689993; PubMed Central PMCID: PMC3387108.

30.   Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, et al. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. BMC Res Notes. 2012;5:210. doi: 10.1186/1756-0500-5-210. PubMed PMID: 22546179; PubMed Central PMCID: PMCPMC3420245.

31.   Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium((R)) assay. Epigenomics. 2009;1(1):177-200. Epub 2009/10/01. doi: 10.2217/epi.09.14. PubMed PMID: 22122642.

32.   Smith AK, Conneely KN, Newport DJ, Kilaru V, Schroeder JW, Pennell PB, et al. Prenatal antiepileptic exposure associates with neonatal DNA methylation differences.

Epigenetics : official journal of the DNA Methylation Society. 2012;7(5):458-63. doi: 10.4161/epi.19617. PubMed PMID: 22419127; PubMed Central PMCID: PMC3368809.

33. Schroeder JW, Smith AK, Brennan PA, Conneely KN, Kilaru V, Knight BT, et al. DNA methylation in neonates born to women receiving psychiatric care. Epigenetics : official journal of the DNA Methylation Society. 2012;7(4):409-14. doi: 10.4161/epi.19551. PubMed PMID: 22419064; PubMed Central PMCID: PMC3368823.

34. Adkins RM, Tylavsky FA, Krushkal J. Newborn umbilical cord blood DNA methylation and gene expression levels exhibit limited association with birth weight. Chem Biodivers. 2012;9(5):888-99. doi: 10.1002/cbdv.201100395. PubMed PMID: 22589090.

35. Mozhui K, Smith AK, Tylavsky FA. Ancestry dependent DNA methylation and influence of maternal nutrition. PloS one. 2015;10(3):e0118466. doi: 10.1371/journal.pone.0118466. PubMed PMID: 25742137; PubMed Central PMCID: PMC4350920.

36. Adkins RM, Krushkal J, Tylavsky FA, Thomas F. Racial differences in gene-specific DNA methylation levels are present at birth. Birth defects research Part A, Clinical and molecular teratology. 2011;91(8):728-36. doi: 10.1002/bdra.20770. PubMed PMID: 21308978; PubMed Central PMCID: PMC3429933.

37. Roberts G, Cheong J, Opie G, Carse E, Davis N, Duff J, et al. Growth of extremely preterm survivors from birth to 18 years of age compared with term controls. Pediatrics. 2013;131(2):e439-45. doi: 10.1542/peds.2012-1135. PubMed PMID: 23296433.

38. Raikkonen K, Pesonen AK, O'Reilly JR, Tuovinen S, Lahti M, Kajantie E, et al. Maternal depressive symptoms during pregnancy, placental expression of genes regulating glucocorticoid and serotonin function and infant regulatory behaviors. Psychol Med. 2015;45(15):3217-26. doi: 10.1017/S003329171500121X. PubMed PMID: 26096463.

39.   Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. Bioinformatics. 2012;28(9):1280-1. doi: 10.1093/bioinformatics/bts124. PubMed PMID: 22451269; PubMed Central PMCID: PMC3577110.

40.   Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189-96. doi: 10.1093/bioinformatics/bts680. PubMed PMID: 23175756; PubMed Central PMCID: PMC3546795.

41.   Hastie TT, Robert ; Narasimhan, Balasubramanian; Chu, Gilbert. impute: impute: Imputation for microarray data. R package version 1.42.0.

42.   Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13:86. doi: 10.1186/1471-2105-13-86. PubMed PMID: 22568884; PubMed Central PMCID: PMC3532182.

43.   de Goede OM, Razzaghian HR, Price EM, Jones MJ, Kobor MS, Robinson WP, et al. Nucleated red blood cells impact DNA methylation and expression analyses of cord blood hematopoietic cells. Clinical epigenetics. 2015;7(1):95. doi: 10.1186/s13148-015-0129-6. PubMed PMID: 26366232; PubMed Central PMCID: PMC4567832.

44.   Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software. 2010;33(1):1-22. PubMed PMID: 20808728; PubMed Central PMCID: PMC2929880.

45.   Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473(7345):43-9. doi: 10.1038/nature09906. PubMed PMID: 21441907; PubMed Central PMCID: PMC3088773.

46.   Barfield RT, Almli LM, Kilaru V, Smith AK, Mercer KB, Duncan R, et al. Accounting for population stratification in DNA methylation studies. Genetic epidemiology. 2014;38(3):231-41. doi: 10.1002/gepi.21789. PubMed PMID: 24478250; PubMed Central PMCID: PMC4090102.

47.   Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57. doi: 10.1038/nprot.2008.211. PubMed PMID: 19131956.

48.   Nemoda Z, Massart R, Suderman M, Hallett M, Li T, Coote M, et al. Maternal depression is associated with DNA methylation changes in cord blood T lymphocytes and adult hippocampi. Translational psychiatry. 2015;5:e545. doi: 10.1038/tp.2015.32. PubMed PMID: 25849984.

49.   Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CC, O'Donovan MC, et al. Methylomic trajectories across human fetal brain development. Genome research. 2015;25(3):338-52. doi: 10.1101/gr.180273.114. PubMed PMID: 25650246; PubMed Central PMCID: PMC4352878.

50.   Simpkin AJ, Hemani G, Suderman M, Gaunt TR, Lyttleton O, McArdle WL, et al. Prenatal and early life influences on epigenetic age in children: A study of mother-offspring pairs from two cohort studies. Human molecular genetics. 2015. doi: 10.1093/hmg/ddv456. PubMed PMID: 26546615.

51.   Nguyen TH, Larsen T, Engholm G, Moller H. Evaluation of ultrasound-estimated date of delivery in 17,450 spontaneous singleton births: do we need to modify Naegele's rule? Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 1999;14(1):23-8. doi: 10.1046/j.1469-0705.1999.14010023.x. PubMed PMID: 10461334.

52.   Moore KA, Simpson JA, Thomas KH, Rijken MJ, White LJ, Dwell SL, et al. Estimating Gestational Age in Late Presenters to Antenatal Care in a Resource-Limited Setting on the Thai-Myanmar Border. PloS one. 2015;10(6):e0131025. doi: 10.1371/journal.pone.0131025. PubMed PMID: 26114295; PubMed Central PMCID: PMC4482646.

53.   Donovan EF, Tyson JE, Ehrenkranz RA, Verter J, Wright LL, Korones SB, et al. Inaccuracy of Ballard scores before 28 weeks' gestation. National Institute of Child Health and Human Development Neonatal Research Network. The Journal of pediatrics. 1999;135(2 Pt 1):147-52. PubMed PMID: 10431107.

54.   Sanders M, Allen M, Alexander GR, Yankowitz J, Graeber J, Johnson TR, et al. Gestational age assessment in preterm neonates weighing less than 1500 grams. Pediatrics. 1991;88(3):542-6. PubMed PMID: 1881734.

55.   Karunasekera KA, Sirisena J, Jayasinghe JA, Perera GU. How accurate is the postnatal estimation of gestational age? Journal of tropical pediatrics. 2002;48(5):270-2. PubMed PMID: 12405168.

56.   Dubowitz L, Ricciw D, Mercuri E. The Dubowitz neurological examination of the full-term newborn. Mental retardation and developmental disabilities research reviews. 2005;11(1):52-60. doi: 10.1002/mrdd.20048. PubMed PMID: 15856443.

57.   Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D, et al. Accelerated epigenetic aging in Down syndrome. Aging cell. 2015. doi: 10.1111/acel.12325. PubMed PMID: 25678027.

58.   Wilcox AJ, Russell IT. Birthweight and perinatal mortality: II. On weight-specific mortality. Int J Epidemiol. 1983;12(3):319-25. PubMed PMID: 6685112.

59.   Godfrey KM, Barker DJ. Fetal nutrition and adult disease. Am J Clin Nutr. 2000;71(5 Suppl):1344S-52S. PubMed PMID: 10799412.

60.   Coory M. Does gestational age in combination with birthweight provide better statistical adjustment of neonatal mortality rates than birthweight alone? Paediatr Perinat Epidemiol. 1997;11(4):385-91. PubMed PMID: 9373861.

61.   Hertz-Picciotto I, Din-Dzietham R. Comparisons of infant mortality using a percentile-based method of standardization for birthweight or gestational age. Epidemiology. 1998;9(1):61-7. PubMed PMID: 9430270.

62.   Boulet SL, Alexander GR, Salihu HM, Kirby RS, Carlo WA. Fetal growth risk curves: defining levels of fetal growth restriction by neonatal death risk. American journal of obstetrics and gynecology. 2006;195(6):1571-7. doi: 10.1016/j.ajog.2006.03.069. PubMed PMID: 16769013.

63.   Patterson RM, Prihoda TJ, Gibbs CE, Wood RC. Analysis of birth weight percentile as a predictor of perinatal outcome. Obstetrics and gynecology. 1986;68(4):459-63. PubMed PMID: 3748491.

64.   Appleton AA, Armstrong DA, Lesseur C, Lee J, Padbury JF, Lester BM, et al. Patterning in placental 11-B hydroxysteroid dehydrogenase methylation according to prenatal

socioeconomic adversity. PloS one. 2013;8(9):e74691. doi: 10.1371/journal.pone.0074691. PubMed PMID: 24040322; PubMed Central PMCID: PMC3764127.

65.   Silva LM. Fetal origins of socioeconomic inequalities in early childhood health. The Generation R Study, Erasmus University Rotterdam, The Netherlands. 2009.

66.   Kramer MS, Seguin L, Lydon J, Goulet L. Socio-economic disparities in pregnancy outcome: why do the poor fare so poorly? Paediatric and perinatal epidemiology. 2000;14(3):194-210. PubMed PMID: 10949211.

67.   Ching T, Ha J, Song MA, Tiirikainen M, Molnar J, Berry MJ, et al. Genome-scale hypomethylation in the cord blood DNAs associated with early onset preeclampsia. Clinical epigenetics. 2015;7(1):21. doi: 10.1186/s13148-015-0052-x. PubMed PMID: 25806090; PubMed Central PMCID: PMC4371797.

68.   Yi SH, Xu LC, Mei K, Yang RZ, Huang DX. Isolation and identification of age-related DNA methylation markers for forensic age-prediction. Forensic Sci Int Genet. 2014;11:117-25. doi: 10.1016/j.fsigen.2014.03.006. PubMed PMID: 24727429.

Chapter 5

Predicting Perinatal Risk Using DNA Methylation

Introduction

Preterm infants, especially those born very (<32 weeks' gestation) or extremely (<28 weeks' gestation) preterm are at an increased risk of developing acute and long-term health complications, including acute respiratory distress and failure [1-3]. Most require respiratory interventions such as mechanical ventilation, continuous positive airway pressure (CPAP), high flow oxygen, or treatment with surfactant or postnatal corticosteroids [3, 4]. Advances in these supportive therapies have improved neonatal survival, but many infants develop bronchopulmonary dysplasia (BPD) as a consequence of the developmental immaturity of their lungs and the respiratory interventions required [3, 5, 6]. BPD is a chronic lung disorder most often defined as a requirement for supplemental oxygen at 36 weeks' postmenstrual age [3, 7, 8]. Additionally, preterm infants may also face long-term neurodevelopmental and pulmonary morbidities [9, 10].

Gestational age (GA) is a major determinant of neonatal morbidity and mortality [11], including the risk of developing BPD. GA is clinically determined either by the date of the last menstrual period or, more accurately, by early obstetrical ultrasound assessment [12, 13]. However, individual variations in outcome are observed for infants with the same GA, for reasons not clearly understood; biological variability in development and maturity between babies of the same GA may be a contributing factor.

Epigenetic modifications such as variations in DNA methylation are important biological mechanisms regulating developmental processes [14]. Methylation levels at cytosine-guanosine dinucleotides (CpG) vary at hundreds of thousands of locations genome wide and can be determined by array technologies [15]. We have recently developed a method to predict GA at birth using DNA methylation (DNAm GA) at 148 CpG sites across the genome [16]. DNAm

GA is highly correlated with clinically estimated gestational age. The residual between DNAm age and clinical GA, known as GA acceleration, has previously been associated with birth weight, which may reflect developmental maturity [16]. The aim of this study was to explore the relationships between GA acceleration and respiratory interventions and outcomes after birth in infants born extremely preterm. As higher GA acceleration may indicate increased developmental maturity, we hypothesized that it may be associated with less respiratory intervention in the neonatal intensive care unit and lower likelihood of developing BPD.

Methods

All surviving preterm babies born at less than 28 weeks' gestation by best obstetrical estimate in the state of Victoria, Australia, during 1991 and 1992 were enrolled in a longitudinal follow-up study. The study was approved by the Human Research Ethics Committees of the Royal Women's Hospital, the Mercy Hospital for Women, Monash Medical Centre, and the Royal Children's Hospital, Melbourne, Australia (HREC No. 23034C). Details of this cohort have been previously described [17].

Blood samples and clinical data

The 143 subjects included in this analysis represent probands born in the state of Victoria at less than 28 weeks' gestational age in 1991-1992 and who survived to 18 years of age, and for whom consent was obtained for the current study. Initial enrollment into the above-mentioned study was performed with parental consent. The study probands had reached adulthood and gave informed consent for the analysis of their neonatal blood spots, which were collected at an average of 9.9 days after birth. Blood spots were stored at ambient temperatures after collection. Data relevant to this study were recorded during the perinatal period, including neonatal sex,

surfactant administration, use of postnatal corticosteroids, duration of assisted ventilation (including intermittent positive pressure ventilation via an endotracheal tube or nasal continuous positive airway pressure), and incidence of BPD (defined as receiving oxygen at 36 weeks' postmenstrual age). Decisions to give surfactant (Exosurf was the only surfactant available) or postnatal corticosteroids (predominantly dexamethasone) were made by the treating clinicians.

DNA methylation analysis

DNA was extracted from dried blood spot cards using the ZR DNA Card Extraction Kit manufactured by Zymo Research (Irvine, California USA). Sample quality and DNA concentration were assessed through spectrophotometry (Nanodrop). Samples with sufficient quality and concentration were processed on the Illumina Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA USA) per manufacturer's recommendation as in our previous work [18]. Data were subjected to standard quality control through the use of the R package CpGassoc [19]. Data frames of the raw signal values were supplied as input to CpGassoc. Any data point with a detection p-value above 0.001 was set to missing. CpG sites with >5% missing data and subsequently, samples with >5% missing data were excluded. After quality control, β values were determined by the ratio of methylation to total signal. Cellular heterogeneity (proportions of CD4, CD8, CD14, CD19, CD56, neutrophils, and eosinophils) was estimated as previously described [20].

Determination of DNAm GA and GA acceleration

DNA methylation gestational age (DNAm GA) was estimated using our previously published method [16], implemented in the publicly available software (https://github.com/akknight/PredictGestationalAge). Briefly, we estimate gestational age based on DNA methylation at 148 CpG sites across the genome. Prior to gestational age estimation, a

data frame of all CpG sites passing quality control is supplied as input for the predictive algorithm, and undergoes a modified beta-mixture quantile normalization as described previously [16, 21, 22]. DNAm GA was compared with clinically estimated gestational age (GA). GA acceleration was determined as the residual of a linear model regressing all DNAm GA values on clinically determined GA values, after adjusting for time to sample collection (9.9 ± standard deviation (SD) of 3.9 days) [16]. Table 5-1 gives an overview of the definitions and descriptions of all measures used to describe GA and GA acceleration.

Statistical analyses

To assess associations between GA acceleration and common neonatal complications, logistic regressions were performed to determine the associations between each binary outcome (surfactant administration, use of postnatal corticosteroids, and incidence of BPD) and GA acceleration, adjusting for cellular heterogeneity and neonatal sex. Linear regression was performed to evaluate the association between duration of assisted ventilation and GA acceleration, adjusting for cellular heterogeneity and neonatal sex. To assess the additional contribution of GA acceleration to models including GA for neonatal outcome predictions, linear regressions, adjusting for neonatal sex and cell type, with and without the inclusion of the GA acceleration term were compared. As male infants have been shown to fare worse, interaction terms were created for GA acceleration by multiplying GA acceleration and an indicator variable for sex. Models with significant interaction terms were further investigated in a subgroup analysis, where linear regressions were performed on male subjects and female subjects separately. Confidence intervals reflect the change in slope. To ensure individual CpG sites were not driving the associations between the outcomes and DNAm age, linear regressions were performed for each of the 148 CpG sites included in the predictor and each outcome with the

false discovery rate controlled at 5%. All analyses were performed using R version 3.3.0 and the R package ppcor [23].

Results

Cohort characteristics

A total of 225 infants born at less than 28 weeks' gestational age in the state of Victoria during 1991 and 1992 survived to 18 years of age. Cohort demographics for the 143 extremely preterm survivors who participated in this study are presented in Table 5-2. Almost one-half received exogenous surfactant, just over one third were treated with postnatal corticosteroids, and 41% developed BPD.

DNAm age and GA acceleration calculations

DNAm GA and subsequently GA acceleration were determined for each subject. GA was correlated with DNAm GA (Fig 5-1, r=0.43, p=8.7×10^{-8}), but not GA acceleration (Fig 5-2, r=2.0×10^{-17}, p=0.99).

Associations of GA acceleration with respiratory outcomes and interventions

Extremely preterm infants who received surfactant had lower GA acceleration values compared with those who did not (mean difference -0.057, 95% CI [-0.099,-0.015] weeks, p=0.009; Fig 5-3A); GA acceleration was lower in infants who were treated with postnatal corticosteroids compared with those who did not receive postnatal corticosteroids (mean difference -0.056 weeks, 95% CI [-0.098,-0.015], p=0.008; Fig 5-3B), and was negatively associated with days of assisted ventilation (-1.79 days per week of GA acceleration, 95% CI [-3.28,-0.30], p=0.02; Fig 5-3C). Additionally, infants who developed BPD had lower GA acceleration (mean difference -0.055 weeks, 95% CI [-0.098, -0.012], p=0.01; Fig 5-3D).

For each model assessing the association of GA and each respiratory outcome and intervention, the inclusion of the GA acceleration added significantly to the amount of variance explained by the model when compared with GA alone, indicating that the models including GA acceleration are better predictors of outcome than GA alone (Table 5-3).

GA acceleration variation between the sexes

Previous studies have shown sex-specific differences in outcomes after preterm birth, prompting us to consider whether GA acceleration was associated with infant sex. Males had a mean GA acceleration of -0.40 and females had a mean GA acceleration of 0.35, indicating that females were potentially developmentally more mature than males (mean difference=0.75 weeks, 95% CI [0.14, 1.36], p=0.02; Fig 5-4); this is further illustrated in Fig 5-5, which shows the distribution of GA acceleration in the overall cohort. The interaction term for sex was significant only for BPD (p=0.007). In subsequent subgroup analysis stratified by sex, development of BPD was significantly related to lower GA acceleration only in males (mean difference -0.154 weeks, 95% CI [-0.230, -0.079], p<0.008) but not in females (mean difference 0.27 weeks, 95% CI [-.05,.05], p=0.9).

Analysis of individual CpG sites

To ensure results were not driven by CpG sites comprising the predictor that were also independently associated with the outcome or intervention of interest, we used linear models to assess associations with each of the 148 CpG sites. Two CpG sites were experiment-wide significant. One site (cg15856055, *ZNF511*) was associated with BPD (p=0.002), and one site was associated with days of assisted ventilation (p=0.0001, cg27258399, *HTRA4*). No individual CpG site was significantly associated with surfactant or postnatal corticosteroid administration,

supporting our hypothesis that the relationship between the outcome and GA acceleration was not due to associations with individual CpG sites.

Discussion

In our study of infants born at <28 weeks' gestational age, increased DNAm age was strongly related to reduced respiratory morbidity and fewer respiratory interventions after birth. Importantly, GA acceleration explained these outcomes over and above knowledge of gestational age at birth alone. To our knowledge, this is the first study to use a metric based on DNA methylation (which we postulate represents developmental maturity in utero and up to ten days after birth) to assess associations between the postulated measure of developmental maturity and respiratory outcomes that is independent of gestational age.

The clinical interventions selected for inclusion in this study mark several critical benchmarks in treating respiratory distress in infants. The demographic characteristics and frequency of respiratory interventions and BPD in our cohort are similar to other cohorts from this era [24, 25]. Infants admitted to the Neonatal Intensive Care Unit (NICU) often require respiratory support [26] and the risk for BPD and long term adverse effects on pulmonary function is generally higher the earlier the infant is born [11].

The decision to treat an infant with surfactant was made by a neonatologist based on clinical criteria. Surfactant was administered in this cohort only as rescue treatment and infants had to be intubated, receiving intermittent positive airway pressure respiratory support, and require >50% oxygen. Our result, which reported infants who received surfactant had a lower GA acceleration than infants not administered surfactant, is consistent with developmental

physiology which implies that more immature preterm infants are expected to be more surfactant deficient [27].

Infants experiencing more severe respiratory failure may require postnatal corticosteroids during their NICU course [28]. Postnatal corticosteroids were widely used in the early 1990s after they had been shown to improve respiratory function and facilitate extubation of very immature preterm babies [28, 29]. This is consistent with our observation that infants who received postnatal corticosteroids had a lower GA acceleration than those who were never given corticosteroids. Practice changed in the late 1990s when it became evident that postnatal corticosteroid use carried significant long-term risks [28, 30, 31]. They are now used with caution and are generally only administered to a select group of babies after prolonged periods of assisted ventilation [28, 30]; the effect of this practice change is not captured in our cohort due to recruitment in the early 1990s.

Finally, this study examined two markers of long-term pulmonary outcomes associated with extremely preterm birth: duration of assisted ventilation and development of BPD. Extremely preterm infants require disproportionately more assisted ventilation with diminishing maturity; for each week of decrease in gestational age, survivors born <28 weeks' gestation required 13 days more assisted ventilation after birth in one study [32]. In the current study, we found that longer duration of assisted ventilation was associated with lower GA acceleration. Similarly, we also found that GA acceleration was lower in infants who developed BPD.

All four indicators of preterm birth-associated respiratory mortality were substantially correlated with GA acceleration. This supports our hypothesis that GA acceleration is a marker of developmental maturity, including that of the lung. Although there is some overlap between subjects with multiple interventions, these four indicators are not collinear. Fig 5-6, shows

overlap between three respiratory interventions. This correlation of respiratory outcomes (-0.44 < r < 0.65) in this cohort is likely to reduce our power to detect statistically significant associations between GA acceleration and each outcome individually. However, even assuming a stringent Bonferroni correction for the number of outcomes tested (0.05 / 4 = 0.0125), inclusion of GA acceleration in models predicting each of the 4 outcomes remains significant (Table 5-2).

Further research is needed to determine if these associations remain if samples are taken closer to birth, such as in cord blood, or even before birth, if in utero fetal blood sampling is performed. One recent study reported associations between GA acceleration estimated from cord blood methylation and several pregnancy-associated outcomes, including maternal preeclampsia and prenatal betamethasone treatment [33]. Antenatal betamethasone was associated with increased GA acceleration, which is consistent with antenatal corticosteroids being used to accelerate maturity before birth [33].

Interestingly, males had lower GA acceleration than females. Males are more likely to experience complications due to respiratory disorders, have poorer psychomotor development when born preterm, and have higher mortality rates than females [34-38]. Thus, this observation is consistent with our hypothesis that DNAm age is reflective of neonatal morbidity. Our previous study[16] did not show a significant sex difference in accuracy based on the median error, but did not examine the association between sex and GA acceleration. However, a recent study also showed increased GA acceleration in females [33]. Lack of associations with neonatal interventions in the analyses stratified by sex could be due to a reduction in power from decreased sample size in our study.

We assessed if any of the 148 CpG sites were associated with the outcome variables. Only two variables showed associations, one with BPD and one with days of assisted ventilation.

No individual CpG site was significantly associated with surfactant or postnatal corticosteroid administration. We conclude that associations between CpG sites and variables examined did thus not adversely affect our study outcome.

Our study has several limitations. As this cohort was recruited between 1991 and 1992, clinical care given to these infants may vary from current practice. Surfactant therapy has "dramatically reduced mortality and morbidity [28]"; as surfactant was available as therapy to our 1991-1992 cohort, we feel that this major change in clinical practice was captured in our study. In our cohort surfactant use was limited to infants requiring supplemental oxygen at over 50%; thus, surfactant was given to babies with more severe respiratory distress syndrome. Today, surfactant is given at a lower oxygen requirement; the potential effect of this change in practice could only be determined by repeating our study with a more contemporary cohort.

With a rate of 73% for our 1991-1992 cohort, the antenatal corticosteroid rate was very similar to many current day cohorts [6]. Since the 1990s, ventilation strategies have changed and new ventilator modes have been introduced, but we agree with Owen et al. [28] that "most interventions have had little effect on the risk of bronchopulmonary dysplasia". We therefore feel that results from our 1991-1992 cohort remain relevant until superseded by more contemporary data.

A further limitation was that DNA methylation was measured from blood spots, which were collected at an average of 9.9 days after birth. Although we have adjusted for postnatal age in our analyses, methylation may have been affected by the days and care received between birth and sample collection, and thus may not have been fully representative of developmental maturity at birth alone. Further information on how rapidly the epigenome changes after birth would also be most relevant; we have initiated a prospective study (The EpiPrem Study) that

addresses longitudinal changes in genome wide DNA methylation and GA acceleration over several weeks of NICU care with multiple analyses between birth and term equivalence. We hope to identify unique epigenetic signatures which could assist in identifying patients who are at risk for specific acute and long-term morbidities associated with preterm birth.

We anticipate that assessment of GA acceleration will be informative in studies of long-term outcomes. As GA acceleration is independent of gestational age at birth, this metric provides another clinical and research tool to evaluate developmental maturity, regardless of the gestation at which the infant was born. Future studies should investigate the development and use of GA acceleration during pregnancy based on cells obtained from amniotic fluid or other fetal cells, which could assist to inform clinical decisions at the limits of viability.

If further studies determine that use of DNAm GA has clinical utility in predicting more neonatal morbidity, targeted assays to rapidly return results of the assessment of DNA methylation-based developmental maturity to clinicians would be needed; technically, it should be possible to perform the DNA methylation analysis and provide results to clinicians within 48 hours.

In summary, a measure of GA acceleration based on DNA methylation in preterm babies' blood correlates with respiratory interventions and morbidities and may reflect a preterm infant's developmental maturity. The use of surfactant, postnatal corticosteroids, and assisted ventilation, and the rate of BPD were all lower in epigenetically more mature extremely preterm infants. The potential clinical utility of such a DNA methylation based assessment of developmental maturity and its potential contribution to personalized neonatal care interventions and outcome predictions should be further investigated.

Table 5-1: Definitions and description of measures used to quantify gestational age and gestational age acceleration.

1. **Clinically estimated gestational age (GA):** Gestational age at birth as determined in routine clinical care through measurements on ultrasound before 20 weeks, or menstrual history if no ultrasound dating was available

2. **DNA methylation gestational age (DNAm GA):** Gestational age at sample collection predicted from DNA methylation. This measure is highly correlated with GA, but differences between GA and DNAm GA are observed and may reveal an infant's developmental maturity.

3. **Gestational Age acceleration (GA acceleration):** Calculated as residual of an individual proband's DNAm GA onto the regression curve of GA and DNAm GA for the complete cohort. This measure is used to assess an infant's developmental maturity. A positive GA acceleration indicates a more mature infant than their GA suggests and a negative GA acceleration indicates a less mature infant than their GA suggests. It is important to note that GA acceleration is independent of GA.

Table 5-2 – VICS cohort demographics **(n=143)**

General Characteristics	
Male, N (%)	66 (46.2)
Gestational Age (weeks), mean ± SD	25.8 ± 1.1
Male	25.64 ± 1.1
Female	25.94 ± 1.1
Postmenstrual Age at sampling*, mean ± SD	27.2 ± 1.2
Male	27.1 ± 1.1
Female	27.3 ± 1.3
DNAm Age (weeks), mean ± SD	28.8 ± 2.1
Male	28.3 ± 1.7
Female	29.2 ± 2.4
Birth weight (g), mean ± SD	883 ± 178
Respiratory Outcomes/Interventions	
Surfactant Administration, N (%)	67 (46.8)
Male, N (%)	35 (63.6)
Corticosteroid Administration, N (%)	55 (38.4)
Male, N (%)	34 (50.7)
Days of Assisted Ventilation, mean ± SD	23.2 ± 17.3
Male	28.6 ± 18.6
Female	18.56 ± 14.8
Bronchopulmonary Dysplasia, N (%)	58 (40.6)
Male, N (%)	30 (44.8)

*Fractional week when blood sample collected plus GA

Table 5-3: Coefficients of determination with and without inclusion of GA acceleration in the model. P-value for comparison of the models**.**

	Adjusted r*	Adjusted r, GA acceleration	p-value for addition of GA acceleration
Surfactant Administration	0.35	0.40	0.004
Corticosteroid Administration	0.36	0.42	0.002
Days Assisted Ventilation	0.61	0.64	0.009
BPD	0.34	0.40	0.004

* r is computed as the square root of R^2 for linear regression models, and the square root of

McFadden's pseudo-R^2 for logistic regression models.

Figure 5-1: Correlation between clinically estimated gestational age (GA), adjusted for days to sample collection, and DNA methylation based gestational age (DNAm GA). Both measured in weeks. Solid line represents the regression line; dotted line indicates equivalence.
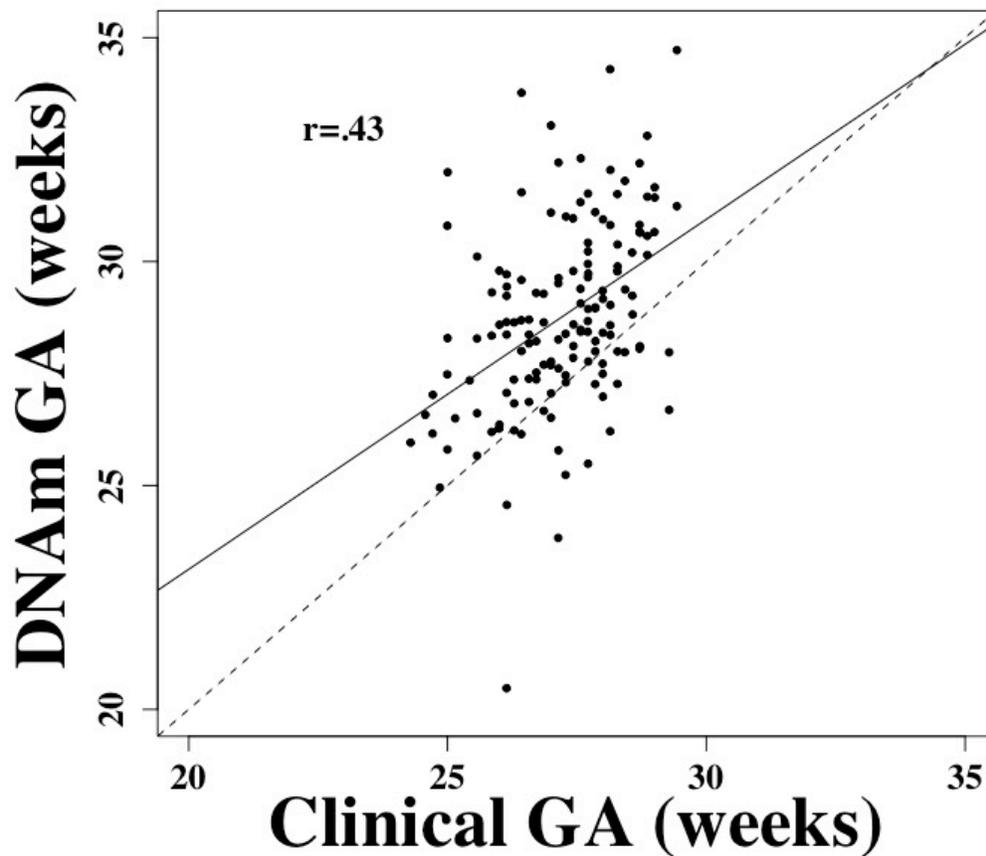
Figure 5-2: GA acceleration is not associated with GA (p>.05).
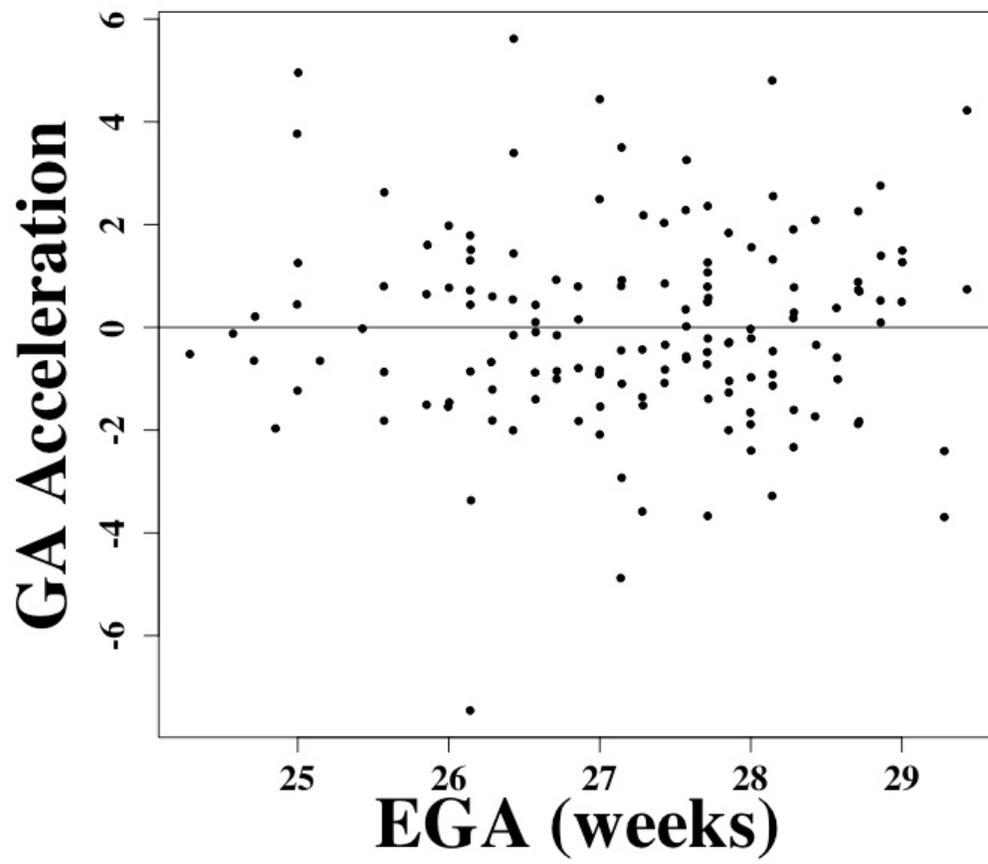
Figure 5-3: Respiratory interventions and outcomes of 143 preterm infants born at less than 28 weeks' gestation. (A) Infants administered surfactant have a lower GA acceleration (p=0.009); (B) Infants administered postnatal corticosteroids have a lower GA acceleration (p=0.008). (C) GA acceleration is negatively associated with days of associated ventilation (p=0.02. (D) Infants who develop Bronchopulmonary dysplasia (BPD) have a lower GA acceleration (p=0.01).
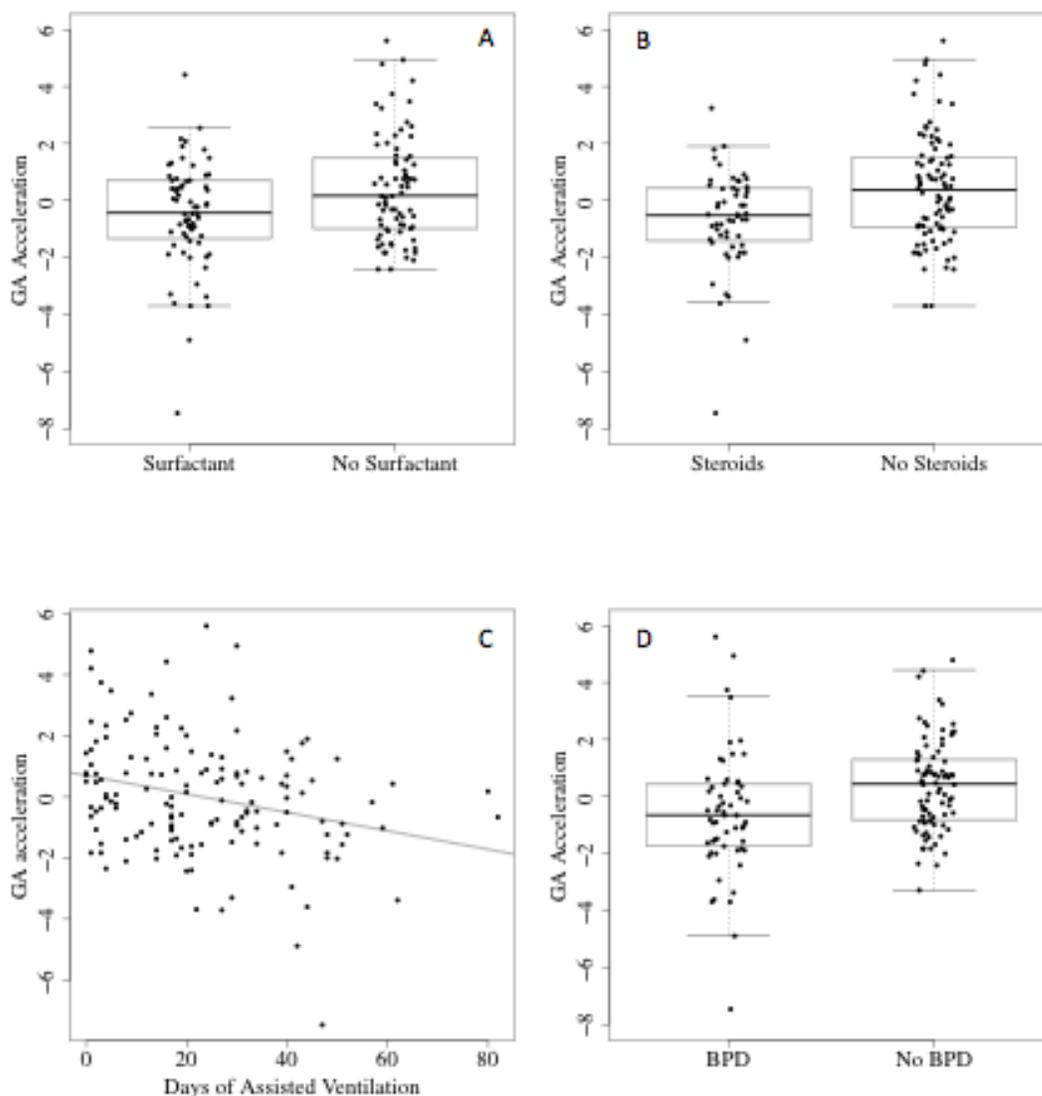
Figure 5-4: GA acceleration is associated with neonatal sex (p=0.02), with males having lower gestational age acceleration.
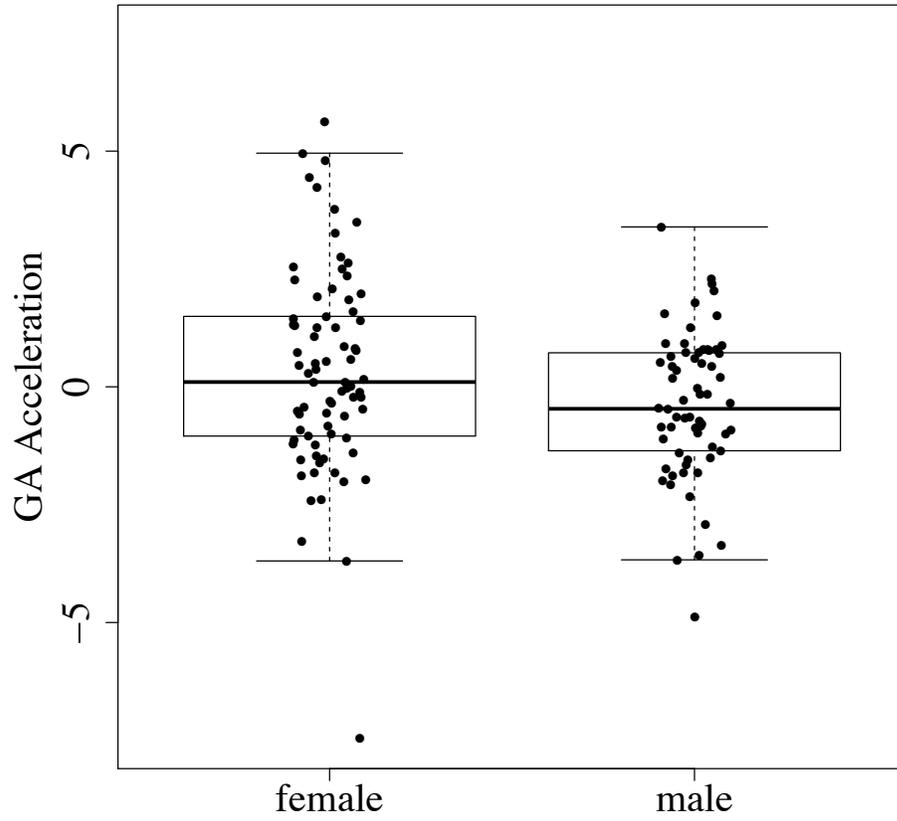
Figure 5-5: Histogram of the distribution of GA acceleration (GAacc) for the entire cohort and separately for males and females.
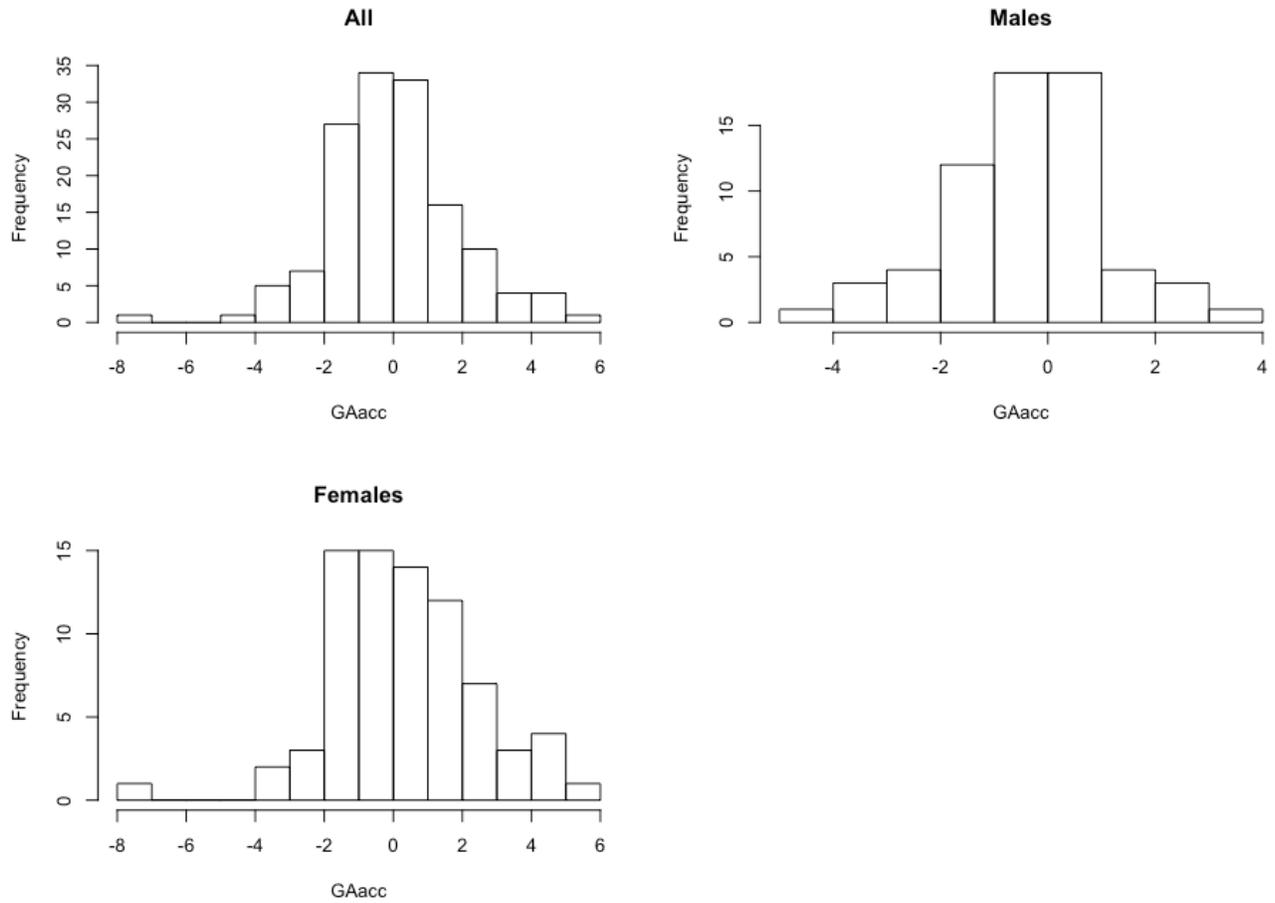
Figure 5-6: Venn diagram demonstrating the overlap between three respiratory interventions.
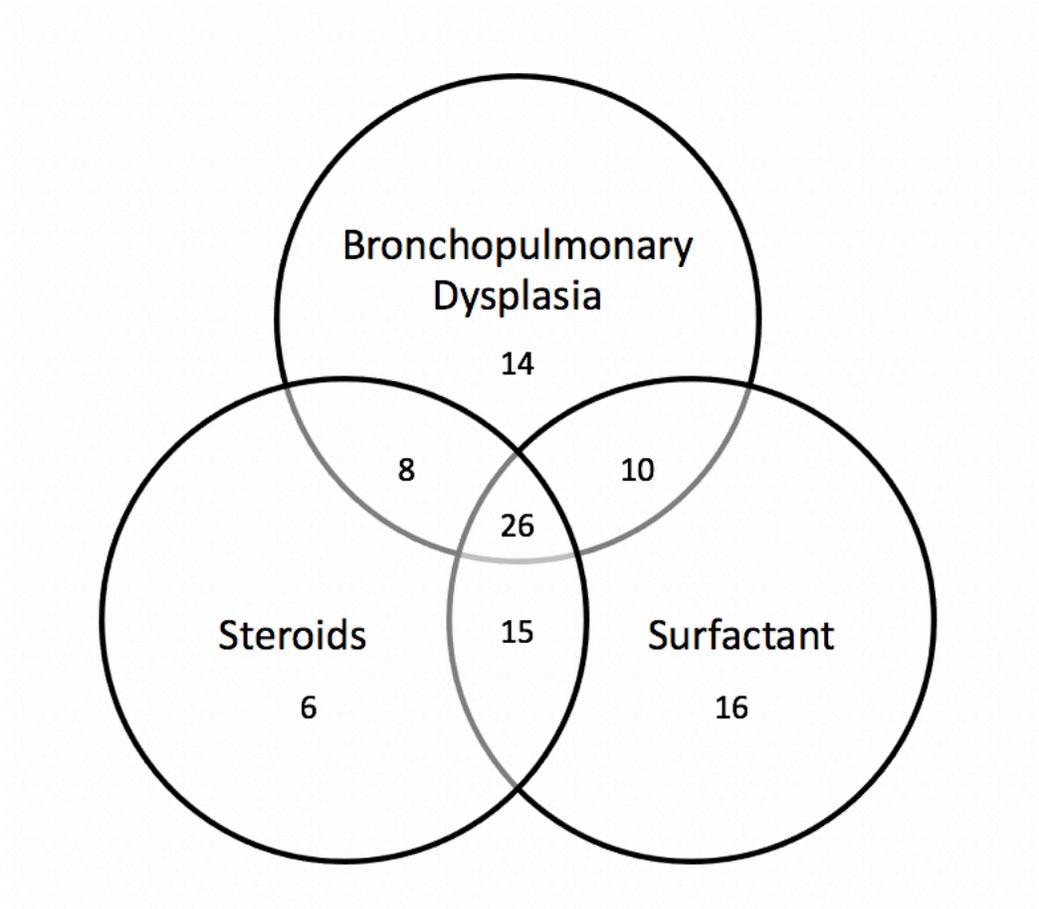
References

[1] Natile M, Ventura ML, Colombo M, Bernasconi D, Locatelli A, Plevani C, et al. Short-term respiratory outcomes in late preterm infants. Ital J Pediatr. 2014;40:52.

[2] Sung JH, Kim SH, Kim YM, Kim JH, Kim MN, Lee HR, et al. Neonatal outcomes of twin pregnancies delivered at late-preterm versus term gestation based on chorionicity and indication for delivery. J Perinat Med. 2016;44:903-11.

[3] Gleason CAJ, S.E. Avery's diseases of the newborn. . Philadelphia PA USA: Saunders; 2012.

[4] Roberts CT, Owen LS, Manley BJ, Froisland DH, Donath SM, Dalziel KM, et al. Nasal High-Flow Therapy for Primary Respiratory Support in Preterm Infants. N Engl J Med. 2016;375:1142-51.

[5] Doyle LW, Davis P, Dharmalingam A, Bowman E. Assisted ventilation and survival of extremely low birthweight infants. J Paediatr Child Health. 1996;32:138-42.

[6] Doyle LW, Adams AM, Robertson C, Ranganathan S, Davis NM, Lee KJ, et al. Increasing airway obstruction from 8 to 18 years in extremely preterm/low-birthweight survivors born in the surfactant era. Thorax. 2017;72:712-9.

[7] Shahzad T, Radajewski S, Chao CM, Bellusci S, Ehrhardt H. Pathogenesis of bronchopulmonary dysplasia: when inflammation meets organ development. Mol Cell Pediatr. 2016;3:23.

[8] Jobe AH, Bancalari E. Bronchopulmonary dysplasia. Am J Respir Crit Care Med. 2001;163:1723-9.

[9] Anderson PJ, Doyle LW. Neurodevelopmental outcome of bronchopulmonary dysplasia. Semin Perinatol. 2006;30:227-32.

[10] Bhandari A, Panitch HB. Pulmonary outcomes in bronchopulmonary dysplasia. Semin Perinatol. 2006;30:219-26.

[11] Manuck TA, Rice MM, Bailit JL, Grobman WA, Reddy UM, Wapner RJ, et al. Preterm neonatal morbidity and mortality by gestational age: a contemporary cohort. Am J Obstet Gynecol. 2016;215:103 e1- e14.

[12] Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. Paediatric and perinatal epidemiology. 2007;21 Suppl 2:86-96.

[13] Mongelli M, Wilcox M, Gardosi J. Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates. American journal of obstetrics and gynecology. 1996;174:278-81.

[14] Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14:204-20.

[15] Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6:692-702.

[16] Knight AK, Craig JM, Theda C, Baekvad-Hansen M, Bybjerg-Grauholm J, Hansen CS, et al. An epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol. 2016;17:206.

[17] Roberts G, Cheong J, Opie G, Carse E, Davis N, Duff J, et al. Growth of extremely preterm survivors from birth to 18 years of age compared with term controls. Pediatrics. 2013;131:e439-45.

[18] Cruickshank MN, Oshlack A, Theda C, Davis PG, Martino D, Sheehan P, et al. Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. Genome Med. 2013;5:96.

[19] Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. Bioinformatics. 2012;28:1280-1.

[20] Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13:86.

[21] Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14:R115.

[22] Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29:189-96.

[23] Kim S. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. Commun Stat Appl Methods. 2015;22:665-74.

[24] Berger TM, Fontana M, Stocker M. The journey towards lung protective respiratory support in preterm neonates. Neonatology. 2013;104:265-74.

[25] Zysman-Colman Z, Tremblay GM, Bandeali S, Landry JS. Bronchopulmonary dysplasia - trends over three decades. Paediatr Child Health. 2013;18:86-90.

[26] Chow SLM, R.; Hossain, S.; Haslam, R.; Lui, K. Report of the Australian and New Zealand Neonatal Network 2013. . ANZNN. 2015.

[27] Whitsett JA, Weaver TE. Alveolar development and disease. Am J Respir Cell Mol Biol. 2015;53:1-7.

[28] Owen LS, Manley BJ, Davis PG, Doyle LW. The evolution of modern respiratory care for preterm infants. Lancet. 2017;389:1649-59.

[29] Mammel MC, Green TP, Johnson DE, Thompson TR. Controlled trial of dexamethasone therapy in infants with bronchopulmonary dysplasia. Lancet. 1983;1:1356-8.

[30] Doyle LW, Ehrenkranz RA, Halliday HL. Early (< 8 days) postnatal corticosteroids for preventing chronic lung disease in preterm infants. Cochrane Database Syst Rev. 2014:CD001146.

[31] Yeh TF, Lin YJ, Huang CC, Chen YJ, Lin CH, Lin HC, et al. Early dexamethasone therapy in preterm infants: a follow-up study. Pediatrics. 1998;101:E7.

[32] Doyle LW, Murton LJ, Kitchen WH. Escalating consumption of nursery resources by extremely immature infants. Aust N Z J Obstet Gynaecol. 1987;27:201-4.

[33] Girchenko P, Lahti J, Czamara D, Knight AK, Jones MJ, Suarez A, et al. Associations between maternal risk factors of adverse pregnancy and birth outcomes and the offspring epigenetic clock of gestational age at birth. Clinical epigenetics. 2017;9:49.

[34] Skiold B, Alexandrou G, Padilla N, Blennow M, Vollmer B, Aden U. Sex differences in outcome and associations with neonatal brain morphology in extremely preterm children. J Pediatr. 2014;164:1012-8.

[35] Iturri P, Bairam A, Soliz J. Efficient breathing at neonatal ages: A sex and Epo-dependent issue. Respir Physiol Neurobiol. 2016.

[36] Mage DT, Donner M. Female resistance to hypoxia: does it explain the sex difference in mortality rates? J Womens Health (Larchmt). 2006;15:786-94.

[37] Romeo DM, Brogna C, Sini F, Romeo MG, Cota F, Ricci D. Early psychomotor development of low-risk preterm infants: Influence of gestational age and gender. Eur J Paediatr Neurol. 2016;20:518-23.

[38] Neubauer V, Griesmaier E, Ralser E, Kiechl-Kohlendorfer U. The effect of sex on outcome of preterm infants - a population-based survey. Acta Paediatr. 2012;101:906-11.

Chapter 6

Conclusions and Recommendations for Future Studies

Pregnancy and delivery complications are associated with a wide range of adverse health outcomes for both the mother and her neonate. For example, women delivering preterm are more likely to develop cardiovascular disease and breast cancer [1-4]. Neonates born preterm are more likely to experience respiratory distress and admission to the neonatal intensive care unit in the perinatal period and are more likely to develop chronic disorders in adulthood [5, 6]. Thus, the morbidity and mortality associated with pregnancy complications make this a vital area of research to improve health for women and their children.

African Americans are at an increased risk of developing pregnancy and delivery complications, even when controlling for known risk factors such as low socioeconomic status [7-9], and thus would benefit most from a reduction in adverse pregnancy outcomes. African Americans are also traditionally understudied in biomedical research. Therefore, we chose to focus our studies of maternal pregnancy and delivery complications on a cohort of pregnant African American women, the Emory University African American Microbiome in Pregnancy (AAMP) cohort. While this may limit the generalizability of these results to another population, this high-risk cohort may also better powered to detect associations between DNA methylation or gene expression and pregnancy complications while controlling for potential confounding by race. Replication in cohorts of other racial/ethnic backgrounds will be essential to assess generalizability of these results.

The overarching objective of these studies is to improve maternal and neonatal risk prediction that may subsequently be used to develop strategies and treatments to promote healthy, full term pregnancies. There are two key ways these studies advance the goal of improving maternal risk. First, identifying changes in gene expression in uncomplicated pregnancies lays the groundwork for identifying gene expression changes in complicated

pregnancies. Changes in gene expression associated with pregnancy complications will provide novel targets for interventions and support the development of clinically-useful tests. Expression-based biomarkers of pregnancy and delivery complications are uncommon in the field, and there is no previous study we are aware of examining expression changes over pregnancy in healthy women. Our results are somewhat consistent with those of a small study in women with rheumatoid arthritis [10], but the genes and pathways we identified that change over pregnancy are largely novel. As the gene expression array used in these studies has been discontinued by the company, other techniques such as RNA-seq will need to be employed for future studies. Studies using RNA-seq would allow investigation of other research questions in addition to potentially validating these findings, including the role of alternative transcripts and transcripts not present on the gene expression array.

Second, DNA methylation-based signatures of delivery complications can be used to predict risk prior to the onset of labor, allowing for adequate time for a detailed and personalized treatment plan to be implemented. The use of DNA methylation-based biomarkers in predicting pregnancy complications, such as preterm birth, has largely focused on candidate gene studies to date. While there are several small epigenome-wide studies of preterm birth, large cohort-based studies are only beginning to be performed. Our epigenome-wide study of associations with fetal intolerance resulting in the identification of *SLC9B1* as a potential biomarker sets the stage for other epigenome-wide association studies in the AAMP cohort, which will provide greater power than previous studies and has been carefully phenotyped. Unfortunately, at the time of this dissertation, we did not have sufficient power to assess other pregnancy and delivery complications, but recruitment is ongoing and future studies in AAMP will examine complications such as preterm birth and preeclampsia. The issue of power is related to the

number of subjects enrolled as well as the heterogeneity of the phenotype. We were adequately powered to study fetal intolerance of labor as a substantial proportion of women experienced this delivery complication, which was well defined by set criteria. However, other pregnancy complications are less common and have more heterogeneous presentations. For example, in AAMP, preterm birth can be spontaneous, medically-indicated, or the result of premature rupture of membranes. These complexities will be further explored in future studies.

Of note, the DNA methylation arrays used in this study, and in our studies of neonatal development, the HumanMethylation450 and HumanMethylation27 BeadChip have also been discontinued and replaced with the MethylationEPIC BeadChip. These changes in technology are to be expected with a rapidly-advancing field. The EPIC array is much more similar to the 450K array than the gene expression array is to RNA-seq, but there are still challenges associated with its adoption. While there is a high correlation overall between probes on the 450K and EPIC arrays, individual CpG sites are much less correlated [11]. This will require careful replication and scrutiny of results generated on both platforms. However, the EPIC array interrogates approximately double the number of CpG sites that were present on the 450K array, including gene regulatory regions not previously examined, so this change in technology is likely to benefit the field.

These studies also advance the goal of promoting neonatal health by providing the research community with a tool to study the impact of risk factors, such as smoking and low socioeconomic status, and novel interventions on neonatal development. After birth, knowledge of neonatal developmental maturity could be used by clinicians and parents to make individualized treatment decisions and better understand acute and long-term risks of a neonate who is born preterm.

One significant advancement in the field of epigenetics has been the development and use of "epigenetic clocks" to assess age acceleration. The first epigenetic clock was developed by Steve Horvath [12] and was promptly followed by the development of several other epigenetic clocks [13-15]. Follow-up studies using this metric show associations between age acceleration and numerous adverse health outcomes, including psychiatric disorders, cancer, and all-cause mortality. However, the current predictors do not have the necessary specificity for use in neonates as the error rates range from three to greater than seven years [12, 13]. Recognizing the need for a neonatal predictor of age, we developed a DNA methylation gestational age (DNAm age) predictor [16]. We optimized this predictor based on previous work in the adult literature, utilizing 15 cohorts with the widest possible gestational ages for testing and training datasets. Unlike our previous studies focused on African Americans, these cohorts were collected around the world to represent as many races/ethnicities as possible.

Since the development of this predictor, we have begun to study the impact of developmental maturity on neonatal health after birth. In the Victorian Infants Collaborative Study, developmental maturity is associated with neonatal outcomes and interventions in the neonatal intensive care unit (NICU), including the need for respiratory support and the development of bronchopulmonary dysplasia [17]. However, this cohort was recruited from 1991-1992, and current NICU interventions have since advanced. This provides an opportunity for future studies to examine the impact of modern interventions and biobehavioral risk factors on developmental maturity.

This work has already begun using the Prediction and Prevention of Preeclampsia and Intrauterine Growth Restriction (PREDO) cohort, where studies have identified associations with

maternal biobehavioral risk factors such as depression [18] and childhood psychiatric outcomes with developmental maturity [18]. As the field continues to expand, there are applications of this work for surveillance of preterm birth and determining cause of death in low and middle income countries as well as for other public health applications.

The approach of using combinations of CpG sites to predict age can be applied to numerous other phenotypes, which may also be relevant for studies of pregnancy and delivery complications. If few predictive CpG sites are epigenome-wide significant, combining multiple CpG sites to predict a phenotype may serve as a more robust predictor. The framework presented in this study can be applied to other facets of maternal and neonatal health research.

Overall, this dissertation outlines the potential utility of gene expression and DNA methylation-based biomarkers to improve maternal and neonatal health. The results of these studies add to the body of knowledge surrounding pregnancy and delivery complications, and their potential consequences in the perinatal period. These studies suggest that an interplay between epigenetics, gene expression, and the environment likely contributes to pregnancy and delivery complications and neonatal health. Future studies should examine relationships between these factors to better understand the complex etiologies of pregnancy and delivery complications.

References

1.   Melbye M, Wohlfahrt J, Andersen AM, Westergaard T, Andersen PK. Preterm delivery and risk of breast cancer. British journal of cancer. 1999;80(3-4):609-13. doi: 10.1038/sj.bjc.6690399. PubMed PMID: 10408874; PubMed Central PMCID: PMC2362328.

2.   Kessous R, Shoham-Vardi I, Pariente G, Holcberg G, Sheiner E. An association between preterm delivery and long-term maternal cardiovascular morbidity. American journal of obstetrics and gynecology. 2013;209(4):368 e1-8. doi: 10.1016/j.ajog.2013.05.041. PubMed PMID: 23800639.

3.   Lykke JA, Paidas MJ, Damm P, Triche EW, Kuczynski E, Langhoff-Roos J. Preterm delivery and risk of subsequent cardiovascular morbidity and type-II diabetes in the mother. BJOG. 2010;117(3):274-81. doi: 10.1111/j.1471-0528.2009.02448.x. PubMed PMID: 20015308.

4.   Ngo AD, Chen JS, Figtree G, Morris JM, Roberts CL. Preterm birth and future risk of maternal cardiovascular disease - is the association independent of smoking during pregnancy? BMC pregnancy and childbirth. 2015;15:144. doi: 10.1186/s12884-015-0571-7. PubMed PMID: 26141292; PubMed Central PMCID: PMC4491219.

5.   Behrman RE, Butler AS, Institute of Medicine (U.S.). Committee on Understanding Premature Birth and Assuring Healthy Outcomes. Preterm birth : causes, consequences, and prevention. Washington, D.C.: National Academies Press; 2007. xvi, 772 p. p.

6.   Knight AK, Smith AK. Epigenetic Biomarkers of Preterm Birth and Its Risk Factors. Genes (Basel). 2016;7(4). doi: 10.3390/genes7040015. PubMed PMID: 27089367; PubMed Central PMCID: PMC4846845.

7.   Burris HH, Collins JW, Jr. Race and preterm birth--the case for epigenetic inquiry. Ethnicity & disease. 2010;20(3):296-9. PubMed PMID: 20828105.

8.   Kramer MR, Hogue CR. What causes racial disparities in very preterm birth? A biosocial perspective. Epidemiologic reviews. 2009;31:84-98. doi: 10.1093/ajerev/mxp003. PubMed PMID: 19477907; PubMed Central PMCID: PMC4361938.

9.   Engelhardt KA, Hisle-Gorman E, Gorman GH, Dobson NR. Lower Preterm Birth Rates but Persistent Racial Disparities in an Open-Access Health Care System. Mil Med. 2018. doi: 10.1093/milmed/usy012. PubMed PMID: 29547967.

10.   Mittal A, Pachter L, Nelson JL, Kjaergaard H, Smed MK, Gildengorin VL, et al. Pregnancy-Induced Changes in Systemic Gene Expression among Healthy Women and Women with Rheumatoid Arthritis. PloS one. 2015;10(12):e0145204. doi: 10.1371/journal.pone.0145204. PubMed PMID: 26683605; PubMed Central PMCID: PMC4684291.

11.   Logue MW, Smith AK, Wolf EJ, Maniates H, Stone A, Schichman SA, et al. The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. Epigenomics. 2017;9(11):1363-71. doi: 10.2217/epi-2017-0078. PubMed PMID: 28809127.

12.   Horvath S. DNA methylation age of human tissues and cell types. Genome biology. 2013;14(10):R115. doi: 10.1186/gb-2013-14-10-r115. PubMed PMID: 24138928; PubMed Central PMCID: PMC4015143.

13.   Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Molecular cell. 2013;49(2):359-67. doi: 10.1016/j.molcel.2012.10.016. PubMed PMID: 23177740; PubMed Central PMCID: PMC3780611.

14.   Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging. 2018;10(4):573-91. doi: 10.18632/aging.101414. PubMed PMID: 29676998; PubMed Central PMCID: PMCPMC5940111.

15.   Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. Aging. 2018;10(7):1758-75. doi: 10.18632/aging.101508. PubMed PMID: 30048243; PubMed Central PMCID: PMCPMC6075434.

16.   Knight AK, Craig JM, Theda C, Baekvad-Hansen M, Bybjerg-Grauholm J, Hansen CS, et al. An epigenetic clock for gestational age at birth based on blood methylation data. Genome biology. 2016;17(1):206.

17.   Knight AK, Smith AK, Conneely KN, Dalach P, Loke YJ, Cheong JL, et al. Relationship between Epigenetic Maturity and Respiratory Morbidity in Preterm Infants. The Journal of pediatrics. 2018;198:168-73 e2. doi: 10.1016/j.jpeds.2018.02.074. PubMed PMID: 29705119.

18.   Girchenko P, Lahti J, Czamara D, Knight AK, Jones MJ, Suarez A, et al. Associations between maternal risk factors of adverse pregnancy and birth outcomes and the offspring epigenetic clock of gestational age at birth. Clinical epigenetics. 2017;9:49. doi: 10.1186/s13148-017-0349-z. PubMed PMID: 28503212; PubMed Central PMCID: PMCPMC5422977.