The epigenetic landscape of prematurity, as it relates to health and development

By

Kenyaita Monet Hodge, MPH

Doctor of Philosophy

Genetics and Molecular Biology

Todd M. Everson, PhD

Advisor

Carmen J. Marsit, PhD

Committee Member

Karen Conneely, PhD

Committee Member

Anke Huels, PhD

Committee Member

T. Michael O'Shea, MD

Committee Member

Accepted:

_____

Kimberly Jacob Arriola, PhD, MPH
Dean of the James T. Laney School of Graduate Studies

_____

Date

EMORY UNIVERSITY

The epigenetic landscape of prematurity, as it relates to health and development

By

Kenyaita Monet Hodge

B.S., Virginia State University, 2011

MPH, University of Virginia, 2013

Advisor: Todd M. Everson, PhD

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 in Genetics and Molecular Biology
2024

ABSTRACT

The epigenetic landscape of prematurity, as it relates to health and development
By Kenyaita Monet Hodge

This dissertation investigates the intricate relationships between prenatal exposures, neonatal morbidities, and their epigenetic consequences on neurodevelopmental outcomes in infants born very preterm (VPT). The first aim evaluated how gestational age (GA) and post-menstrual age (PMA) at NICU discharge are related to DNA methylation (DNAm) in buccal cells from the Neonatal Neurobehavior Outcomes in Very Preterm Infants (NOVI) study, identifying numerous CpGs and differentially methylated regions linked to neurodevelopmental pathways. The second aim examined the association between bronchopulmonary dysplasia (BPD) severity, antenatal steroid exposure, and DNAm of the hypothalamic-pituitary-adrenal (HPA) axis genes in an Environmental influences on Child Health Outcomes (ECHO) wide analysis. Using DNAm data from buccal tissue from the NOVI Study and blood spot tissue from Extremely Low Gestational Age Newborns (ELGAN), the study found that BPD did not have a significant association with polyepigenetic glucocorticoid risk score or neonatal DNAm of HPA genes, however antenatal steroid treatment was associated with altered neonatal DNAm of HPA genes in buccal tissue. The third aim explored whether DNAm at 125 CpG sites mediates the association between neonatal morbidities and neurodevelopmental outcomes at 24 months. High-dimensional mediation analysis of infants revealed significant mediation effects, particularly for cognitive and language outcomes, with specific CpGs within genes (e.g., *SMYD3*, *TMEM245*, *FGFR1OP*) identified as potential mediators of multiple outcomes. The findings from this dissertation highlight the profound relationships between epigenetic modifications and the health and development of infants born VPT. By investigating the age-associated epigenetic changes, the effects of antenatal steroids and BPD on the epigenome, and the role of DNAm as a mediator of neurodevelopmental outcomes, we have illuminated critical pathways that could influence long-term health outcomes. Together, these aims not only enhance our understanding of the biological mechanisms underlying preterm birth outcomes but also pave the way for developing targeted interventions to improve the health and neurodevelopment of this vulnerable population.

EMORY UNIVERSITY

The epigenetic landscape of prematurity, as it relates to health and development

By

Kenyaita Monet Hodge

B.S., Virginia State University, 2011

MPH, University of Virginia, 2013

Advisor: Todd M. Everson, PhD

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 in Genetics and Molecular Biology
2024

ACKNOWLEDGMENTS

First and foremost, I would like to thank God for guiding me through this chapter of my life. This journey had its challenges, but my faith in God kept me focused on my goal of earning a doctoral degree. Next, I would like to thank my advisor Dr. Todd Everson for providing an environment conducive for learning, collaborating, and thriving as a doctoral student. Todd, you have been an exceptional mentor and my accomplishments as a doctoral student reflect your outstanding leadership! I would also like to thank my committee members, Dr. Carmen Marsit, Dr. Karen Conneely, Dr. Anke Huels, and Dr. T. Michael O'Shea, co-authors (especially the NOVI team) and lab members for their guidance and support as I worked on analysis plans, presentations for conferences, and manuscripts for publication. I will always remember your kind words of encouragement and excitement as I reached my research goals. To my cohort (GMO), you will always have a special place in my heart! The nine of you made this journey fun, exciting, and honestly bearable when times became difficult. We all became friends instantaneously and now we are family! I look forward to cheering for you all on as you reach your goals just as you have cheered for me. To my former teachers and professors, there are so many of you that have made me feel supported and encouraged to do my very best. From my third and fourth grade teacher, Mrs. Drema Turner to my high school math teacher Mr. Frank Mannino, to my undergraduate and graduate professors Dr. Milton Faison, Dr. Regina Knight-Mason, Dr. Zhifu Xie, Dr. Ralph Gatrone, Dr. Colleen Taylor and Dr. Jeanita Richardson (just to name a few), thank you for making me love learning, for helping me find my voice and for providing me with the tools that I needed to succeed! To Dr. Struan Grant, thank you for encouraging me to apply to doctoral programs! It was the nudge that I did not know I needed. I am blessed to have had all of you in my life! To my best friends, Aria Roadcloud and Dr. Madelyn Crowell, thank you for your sisterhood! Your friendship means so much to me and I am grateful for you both always having a listening ear, a shoulder to lean on and for always making me laugh. Thank you to all of my friends who have been with me at various stages of my life. To my siblings, Morgan and Stephan Taylor, I am so proud of how much you have grown over the years! You are amazing siblings and even better parents! I am grateful for your constant love and support. You both have been along for the ride from the very beginning and have always been excited for me as I have accomplished each of my goals. I know it was not easy having to drive for hours to drop me off at college or miss out on activities to attend my school functions. I greatly appreciate your many sacrifices that have gotten me here today. To my mom, Monica Taylor, thank you for always believing in me and encouraging me. You have been my cheerleader every step of the way. Thank you for your sacrifice and for being a positive role model. I am blessed and honored to have you as my mother. I love you and you will always be "My Girl". To my family, thank you for always checking on me, cheering for me and supporting me. I am blessed to have such an amazing support system. I am who I am because of the example you all have set and because of the solid foundation that has been laid throughout my life. To my pets, Kitty, Buddy Love, Granny Smith Apple, and Girlie, I am forever grateful for the companionship and love you have given me. The pandemic was very stressful but as each of you joined our family, the stress slowly began to melt away. I have truly enjoyed watching you all grow. Lastly, I would like to thank my partner, Justin Eskridge, who has been here since I first embarked on this journey. Justin, you have been nothing short of amazing as this journey took me hundreds of miles away. Thank you for staying awake with me during my all-nighters, for listening to my presentations on repeat, for always being a positive voice in my times of uncertainty. This journey was just as much yours as it has been mine. Thank you for being my person! This accomplishment belongs to us all!

With Respect and Love,

Kenyaita

TABLE OF CONTENTS

CHAPTER 1: Introduction

Preterm birth, defined as birth before 37 weeks of gestation, poses significant challenges to neonatal health and development. Advances in neonatal care have increased the survival rates of preterm infants, yet these infants remain at high risk for a range of morbidities and developmental delays with the highest risk observed among infants born very preterm (VPT, <30 weeks of gestation). Epigenetics, particularly DNA methylation, has emerged as a critical area of study to understand the mechanisms underlying these outcomes. Infants born preterm exhibit distinct epigenetic profiles compared to term infants, some of which persist into adulthood and may provide a biological link between preterm birth and adult disease[1; 2]. These epigenetic differences are likely due to disruption of the typical course of development along with premature exposure to the ex-utero environment, as well as unique environmental circumstances during the early postnatal period which is spent in a neonatal intensive care unit (NICU); all of these likely affect gene expression and developmental processes[3]. Studies have shown that very preterm infants have altered DNA methylation patterns in genes associated with neurodevelopment, immune function, and metabolic pathways[4-6]. These epigenetic modifications may contribute to the increased vulnerability of preterm infants to various health issues and developmental delays. This dissertation aims to explore the epigenetic changes among infants and young children that were born VPT, focusing on age-related methylation changes, the epigenetics of bronchopulmonary dysplasia (BPD) and the hypothalamic-pituitary-adrenal (HPA) axis, and the role of DNA methylation in mediating neurodevelopmental outcomes.

Epigenome-wide association studies (EWAS) have provided valuable insights into how age-related changes in DNA methylation may influence development. For preterm infants, the gestational age (GA) at birth and postmenstrual age (PMA), which is the estimated time from conception to the end of the NICU stay in our study, are crucial factors that are related to developmental state and associated with prospective outcomes. Studies have demonstrated that DNA methylation patterns are highly dynamic during childhood[7] and can be influenced by the perinatal environment, which in turn impacts neurodevelopment and other physiological processes. By conducting an EWAS focused on age-related

methylation changes, Aim 1 seeks to identify differentially methylated CpGs and regions (DMRs) associated with GA and PMA in very preterm infants, shedding light on the epigenetic modifications that may be related to early development in this population.

Transitioning from the broader age-related methylation changes, Aim 2 focuses on the most common morbidity among infants born preterm, BPD. BPD is a chronic lung disease in infants, particularly among those who have received prolonged mechanical ventilation and oxygen therapy[8]. We hypothesize that BPD perturbs the HPA axis and its genes, which regulates stress responses, due to the physiological stress that BPD induces on the body of developing infants and because these infants are sometimes exposed to glucocorticoid-based steroids[9]. Epigenetic modifications in HPA axis-related genes can alter their expression and potentially influence the susceptibility to and severity of prospective negative health outcomes. This aim investigates the DNA methylation patterns in HPA axis genes among preterm infants with BPD, aiming to uncover epigenetic markers that could serve as early indicators of infants at risk for negative health outcomes.

While BPD is the most common morbidity among very preterm infants, many of these infants experience several other morbidities during their NICU stay. It is essential to consider the broader effects of neonatal morbidities on neurodevelopmental outcomes for infants born preterm. Preterm infants that experience multiple neonatal morbidities are at heightened risk of delays and impairments in neurodevelopment[10], which can manifest as cognitive, motor, and language deficits. Our group has also shown that increasing numbers of morbidities are also associated with altered DNA methylation patterns[11]. This aim examines whether DNA methylation acts as a mediator between neonatal morbidities—such as BPD, serious infection, retinopathy, and brain injury—and neurodevelopmental outcomes. By applying high-dimensional mediation analysis, this study aims to identify specific methylation sites that could explain the relationship between early health challenges and later neurodevelopmental performance, offering potential pathways for therapeutic intervention.

The overall goal of this dissertation is to elucidate age-specific epigenetic changes, understand the impact of BPD and antenatal steroids on the stress response system, and unravel the mediating role of DNAm in

linking early life stressors to neurodevelopmental outcomes in infants born very preterm. Through these interconnected aims, we strive to contribute to the knowledge base that informs strategies for improving the health and development of this vulnerable population.

CHAPTER 2: Epigenetic Associations with Neonatal Age In Infants Born Very Preterm, Particularly Among Genes Involved In Neurodevelopment

Kenyaita M. Hodge[1], Amber A. Burt[1], Marie Camerota[2,3], Brian S. Carter[4], Jennifer Check[5], Karen N. Conneely[6], Jennifer Helderman[5], Julie A. Hofheimer[7], Anke Hüls[1,8,9], Elisabeth C. McGowan[10], Charles R. Neal[11], Steven L. Pastyrnak[12], Lynne M. Smith[13], Sheri A. DellaGrotta[3], Lynne M. Dansereau[3], T. Michael O'Shea[7], Carmen J. Marsit[1,8], Barry M. Lester[2,3,10] and Todd M. Everson[1,8*]

[1]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA

[2]Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, RI

[3]Brown Center for the Study of Children at Risk, Women and Infants Hospital, Providence, RI

[4]Department of Pediatrics-Neonatology, Children's Mercy Hospital, Kansas City, MO

[5]Department of Pediatrics, Wake Forest School of Medicine, Winston-Salem, NC

[6]Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA

[7]Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC

[8]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA

[9]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA

[10]Department of Pediatrics, Warren Alpert Medical School of Brown University and Women and Infants Hospital, Providence, RI

[11]Department of Pediatrics, University of Hawaii John A. Burns School of Medicine, Honolulu, HI

[12]Department of Pediatrics, Spectrum Health-Helen Devos Hospital, Grand Rapids, MI

[13]Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA

2.1 BACKGROUND

The neonatal and infancy periods are the most dynamic in human growth and development. During these periods, the nervous, respiratory, vascular, and immune systems, among others undergo vital changes. For instance, the brain begins to develop in the third week of gestation with robust changes in cortical grey and white matter beginning at 20 weeks of gestation and the volume of grey matter increasing up until the second year of life[12; 13]. Similarly, development of the respiratory system involves a series of intricate stages, from the formation of the airway tree in gestational weeks 6 and 7, to the branching and growth of bronchioles, then production of pulmonary surfactant and the formation of alveoli between gestational weeks 32 and 36[14]. Preterm birth disrupts the normal progression of these processes, and of other organ systems, which can give rise to a range of neonatal medical complications and developmental delays. Clearly, age since conception is a critical indicator of infant health and development and the timing of birth can perturb healthy development. While phenotypic presentations of aging and development are clear, age also plays an important role in molecular regulation as studies have observed widespread and robust changes in the epigenome during early postnatal development, most notably with DNA methylation[15]. However, there are few studies that focus on the relationship between age and epigenetic patterns during the neonatal period[16].

Early life stress, such as that caused by preterm birth, has been linked to changes in DNA methylation (DNAm)[17-19]. DNAm is a biological process in which methyl groups are added to the DNA at cytosine (C) bases that are followed by guanine (G) bases, referred to as CpG sites. This process influences the expression potential of genes but does not change the sequence of DNA. DNAm is an epigenetic mechanism that has been studied extensively due to its ability to capture information about exposures throughout the course of life[20]. Several studies on the development of epigenetic clocks have successfully utilized buccal tissue, which primarily consists of epithelial cells, to accurately predict age in pediatric populations. For example, the development of the Pediatric-Buccal-Epigenetic (PedBE) clock leveraged DNAm profiles from buccal epithelial cells from healthy individuals ranging 0 to 20 years of age[21]. When compared to the pan-tissue Horvath DNAm clock[22], the PedBE clock demonstrated higher

accuracy in predicting chronological age in longitudinal sampling during the first year of life and in adolescence. McEwen et al found that higher gestational age (GA) was associated with a higher PedBE age in typically developing individuals[21]. However, children with a neurodevelopmental disorder, as compared to typically developing children, had higher predicted PedBE age. This finding suggests that deviations in predicted PedBE age are associated with altered development. Our group developed the NEOage clock, an epigenetic clock that predicts chronological age in neonates born preterm[23]. Like the PedBE clock, the NEOage clock leverages DNAm profiles from buccal epithelial cells and has high accuracy in predicting neonatal age in both the training data and an independent saliva data set. These studies provide strong evidence that buccal tissue can be used to investigate the epigenetics of aging in neonatal and pediatric populations[21]. While these prior projects focused on the development of algorithms to estimate age from DNA methylation, our study aims to characterize the genes and genomic regions that are associated with age-specific epigenetic changes in infants born preterm.

In the United States approximately 1 in 10 babies are born prematurely each year[24]. Globally, prematurity is the leading cause of death in children under five years old[25], resulting in a significant public health burden. Infants born very preterm (born <30 weeks gestation) are at heightened risk of long-term health and development impairments, compared to their term-born counterparts. For example, bronchopulmonary dysplasia (BPD), the most common neonatal morbidity among preterm infants, has been associated with chronic diseases and neurodevelopmental impairments[26; 27]. Little is known about how these early life stressors impact the epigenome of infants born preterm. However, studies have shown that preterm infants have different methylation patterns than those born at term, and these differences can be detected in multiple tissues, with special emphasis on minimally invasive tissues such as placenta, cord blood and saliva[28-30]. Variations in the epigenetic landscape during early life may shed light on the interrelationships between aging, maturation, the methylome, and the health of infants born preterm.

We aimed to address this gap in knowledge by conducting an epigenome-wide association study (EWAS) to determine if there are neonatal age associated differences in DNA methylation (DNAm) in

buccal tissue of infants born very preterm, and to characterize implicated biological pathways. We also examined whether the buccal DNAm levels were correlated with brain DNAm levels of the identified differentially methylated CpGs, using an independent publicly available data source (GEO Accession: GSE111165)[31].

2.2 METHODS

2.2.1 Characterization of Study Population

The NOVI Study was conducted at 9 university affiliated NICUs in Providence, RI, Grand Rapids, MI, Kansas City, MO, Honolulu, HI, Winston Salem, NC, and Torrance and Long Beach, CA from April 2014 through June 2016. These NICUs were also Vermont Oxford Network (VON) participants. Eligibility was determined based on the following inclusion criteria: 1) birth at <30 weeks post menstrual age; 2) parental ability to read and speak English or Spanish and 3) residence within 3 hours of the NICU and follow-up clinic. Exclusion criteria included maternal age <18 years, maternal cognitive impairment, maternal death, infants with major congenital anomalies, including central nervous system, cardiovascular, gastrointestinal, genitourinary, chromosomal, and nonspecific anomalies, and NICU death. Parents of eligible infants were invited to participate in the study when survival to discharge was determined to be likely by the attending neonatologist. Written informed consent was obtained from all participants and all study procedures and protocols were carried out with approval from each of the following center's Institutional Review Board *[Children's Mercy Hospital IRB in Kansas City, MO (IRB00004750), Western Institutional Review Board in Puyallup, WA (WIRB20131387), John F. Wolf Human Subjects Committee in Los Angeles, CA (IRB00000389), Spectrum Health Systems, Inc. in Grand Rapids, MI (IRB00009435), Women & Infants Hospital in Providence, RI (IRB00000746), and Wake Forest University Health Sciences in Winston-Salem, NC (IRB00000212)].* All study activities were performed in accordance with relevant guidelines and regulations. Overall, 704 eligible infants were enrolled and written informed consent was obtained for 93% of infants for epigenomic screening via

buccal cells swabs that were collected on 624 infants. Ultimately, 538 samples were included in the study from infants for whom DNAm data, PMA, and GA, and covariate data were available.

Maternal interviews were performed to collect demographic information such as age, race and ethnicity, and educational attainment, while the Hollingshead Index was used to assess socioeconomic status (SES; collected by maternal report on education and occupation) with a Hollingshead level V indicating low SES. Infant medical records were reviewed to collect birthweight, gestational age, length of NICU stay, whether the newborn was outborn, and diagnoses of neonatal morbidities described in detail below. Outborn refers to infants that were born in a hospital without subspecialty providers of neonatal intensive care and were transferred, almost always on the day of birth, to a tertiary center for subspecialty care. Gestational age was estimated using the highest quality of available information: first using the dates of embryo retrieval or intrauterine insemination, then using fetal ultrasound, then using date of last menstrual period, and finally assigned by attending neonatologist in the absence of the above information.

2.2.2 Age Variables

In this study we investigated PMA, an estimate of the period between conception to buccal tissue collection at NICU discharge, and GA, an estimate of the period between conception and birth. PMA in NOVI was calculated by adding postnatal age at buccal collection to the estimated GA at birth which is an established process[32; 33]. GA was estimated by using the date of embryo retrieval, intrauterine insemination, fetal ultrasound, or the date of last menstrual period as described elsewhere by Everson et al[11]. Buccal cells were collected for epigenomic analyses during the week of discharge from the NICU (±3 days); thus, PMA at buccal swab collection represents the combination of GA at birth plus the length of NICU stay (±3 days).

2.2.3 Description of Confounding variables

As described in Everson et al[11], we adjusted for potential confounders, batch (EPIC array) effects and cellular heterogeneity[34] and we also adjusted for other potential confounders, such as outborn status[35] given that this variable increases the risk of neonatal morbidities and neurobehavioral deficits and is

associated with longer stays in the NICU in males[11; 36; 37]. We adjusted for sex[38] given its association with DNAm. Cell heterogeneity, the proportion of epithelial, fibroblast, and immune cells, were estimated in our buccal samples using reference methylomes[34]. As in our previous studies, we observe that epithelial cells make up 95.7% of cells in 95% of our buccal samples, with the remaining majority are immune cells[11]. Therefore, we adjusted for cell heterogeneity by using the proportions of epithelial cells as a covariate in our statistical models. Lastly, we adjusted for potential unmeasured confounding such as batch effect and unwanted sources of variation with surrogate variables[39]. We included both PMA and GA in our models (as independent variables), so we could characterize the relationships of both age metrics with DNAm, independent of each other. In all models, we adjusted for sex, whether the newborn was outborn, the proportion of estimated epithelial cells, recruitment site (6-level factor), batch (7-level), cumulative morbidity risk score (4-level factor) ranging from 0 to 3+ serious morbidities (BPD, ROP, SBI, and infection) experienced during NICU stay, and five SVs.

2.2.4 DNA methylation (DNAm) measurement, quality control, and preprocessing

To assess the methylation profiles of the NOVI study cohort we analyzed the DNAm data from buccal tissue at NICU discharge measured using the Illumina MethylationEPIC. These data have undergone established quality control and processing pipelines, including the exclusion of samples with >5% of probes yielding detection p-values > 1.0E-5, with mismatch between reported and predicted sex, or incomplete phenotype data[11]. Functional normalization and beta-mixture quantile (BMIQ) normalization were performed[40], and then we excluded probes located on the X and Y chromosomes, those with single nucleotide polymorphisms (SNP) within the binding region, those able to cross-hybridize to other regions of the genome[41], or those with low variability[42]. After these exclusions, and dropping four samples that had missing covariate data, 706,278 probes were available from 538 samples. Logit transformation of beta values were calculated to get M-values which were used in all downstream analysis[43; 44].

2.2.5 Statistical Analysis

All statistical analyses were performed in R version 4.1.0. First, we conducted SVA to identify unknown confounding within the NOVI study using the *sva*[45] package. The model matrix included GA, PMA and potential confounders, while the null model excluded GA and PMA. To identify differentially methylated CpGs associated with GA and PMA, we conducted linear regression models using the *gee*[46] package, regressing DNAm levels on GA and PMA (mutually adjusted for), while also adjusting for potential confounders and SVs. Generalized Estimating Equations (GEE) are a type of regression model that accounts for potential non-independence of observations (i.e., multiple births nested within families) as these siblings may have similar methylation patterns and other characteristics. We thus included family ID as a nesting variable in all GEE models. We accounted for multiple comparisons by using Bonferroni corrections ($\alpha = 0.05/706{,}323$). Among CpGs significantly associated with GA or PMA, we then tested whether these age-associated CpGs were in close genomic proximity and formed differentially methylated regions (DMRs) using the dmr*ff*[47] package. We defined DMRs as sets of CpG sites within 1500bp apart (*maxgap*) with each site having a Bonferroni adjusted p-value < 0.05 (*p.cutoff*) and the same direction of effect[47]. Finally, we ran pathway enrichment analysis on significant differentially methylated genes using the R package *clusterProfiler* which identified pathways defined by GO, KEGG, and Reactome databases[48-52]. The R package *methylGSA* was then used for pathway enrichment analysis to account for probe bias[53]. Lastly, these tests were adjusted for multiple testing correction using the Benjamini-Hochberg method as is common practice in gene-set enrichment analyses.

2.2.6 Brain-buccal Correlation

Much of our work in this cohort has focused on newborn neurobehavior, medical morbidities, and subsequent neurodevelopment associated with epigenetic features. These relationships have been studied using non-invasive collection of neonatal peripheral tissue via cheek swabs to determine whether any of the age-associated changes in buccal cells may be reflective of age-associated changes in the central nervous system (CNS). Since we are studying a peripheral tissue, but much of our work in this cohort focuses on neurodevelopment and neurobehavior, we aimed to understand if any of the age-associated

changes in buccal cells may be reflective of age-associated changes in the CNS. Thus, we conducted Pearson correlation tests among the CpGs significantly associated with GA and PMA using a publicly available dataset consisting of paired brain and buccal tissue samples (GEO Accession: GSE111165)[31]. Pathway enrichment analysis, as mentioned above, was conducted on the CpGs with significant brain-buccal correlations for each age metric. Statistical significance was based on CpGs with positive correlations that met the Bonferroni cutoff of ($\alpha = 0.05/706{,}323$). Significant results were then uploaded and analyzed by eFORGE (experimentally-derived Functional element Overlap analysis of ReGions from EWAS), an online tool that identifies cell type specific signaling in epigenomic data[54-56], using reference data from the Consolidated Roadmap Epigenetics- DNase I hypersensitive sites (DHSs) with a strict significant threshold set at $p < 0.01$ and marginal significant threshold set at $p < 0.05$.

2.3 RESULTS

2.3.1 Characteristics of the study population

There were 538 infants in the Neonatal Neurobehavior and Outcomes in Very Preterm Infants (NOVI) study for whom DNAm data, post-menstrual age (PMA) at NICU discharge, and GA at birth, and covariate data were available. In NOVI, GA ranged from 22.0 to 29.9 weeks (mean = 27.0 weeks) and PMA ranged from 32.1 to 51.4 weeks (mean = 39.2 weeks; Table 1). There were 115 infants (21.4%) born of multiple gestation, and 423 infants were singleton births. Of the 538 infants, 113 were born at a hospital which did not have capacity for neonatal intensive care (hereafter referred to as "outborn") (21.0%) which was associated with increasing PMA at discharge ($p = 0.0001$). The majority of infants were diagnosed with at least one neonatal morbidity including BPD (51.5%), serious brain injury (SBI; 12.8%), infection (19.0%), or severe retinopathy of prematurity (ROP; 6.3%). GA at birth was inversely associated with number of medical morbidities ($r = -0.48$, $p = 7.83 \times 10^{-33}$), which in turn was positively associated with PMA at discharge ($r = 0.59$, $p = 1.23 \times 10^{-52}$), when buccal swabs were collected.

2.3.2 Summary of GA and PMA EWAS

In our EWAS models, we adjusted for the potential confounding effects of neonatal morbidities, as well as the inverse correlation between GA and PMA by including both age variables and a cumulative morbidity risk score (described below) in our models. We also adjusted for sex, study site, batch, outborn status (i.e., requiring transfer into study site hospital), epithelial cell proportion, and the five most influential surrogate variables (SVs; described in the methods section) – SVs represent major sources of variation in the DNA methylation data and are included to account for unmeasured confounding. Overall, from our EWAS we observed 2,366 CpGs associated with GA and 14,979 CpGs associated (Bonferroni correction: $P_{Bonf.} = 0.05/706,323$) with PMA while mutually adjusting for each other, as well as other confounders and technical factors. There were 1,151 CpGs significantly associated with both GA and PMA; these coefficients had a strong inversely correlated relationship ($r = -0.95$, $p < 2.2x10^{-16}$). Approximately 68% (n = 1608) of the CpGs significantly associated with GA were hypermethylated with older GA at birth. Of the CpGs significantly associated with PMA, 70% (n = 10,568) were hypermethylated with increasing PMA. We provide the complete set of coefficients, standard errors, and p-values of the GA and PMA associations as supplementary data files (Supplementary Tables S1 and S2).

2.3.3 Associations with GA

The CpG most significantly associated with GA was cg13211971 ($p = 9.79x10^{-46}$) which was annotated to *PDE9A*. We highlight the 10 CpG associations with the smallest p-values, their relationships with GA, and their genomic annotations in Table 2. We then performed differentially methylated regions (DMR) analysis to identify spatially clustered CpGs that exhibited the same direction of coefficient associated with GA. 417 CpGs identified from the EWAS were within 160 DMRs (Supplementary Tables S3 and S4). The DMRs included *IRX4* ($P_{Bonf} = 7.24x10^{-20}$), *SLC7A5* ($P_{Bonf} = 7.20x10^{-19}$), *LINC-PINT; FLJ43663* ($P_{Bonf} = 9.74x10^{-19}$), *STAT5A* ($P_{Bonf} = 2.21x10^{-16}$), *PHLDB1* ($P_{Bonf} = 1.14x10^{-15}$), *NXN* ($P_{Bonf} = 7.60x10^{-15}$), *TPM4* ($P_{Bonf} = 8.90x10^{-14}$), *RBFOX2* ($P_{Bonf} = 1.00x10^{-13}$), *PML* ($P_{Bonf} = 1.34x10^{-13}$), and *ANP32A* ($P_{Bonf.} = 1.71x10^{-13}$), as the annotated genes with the smallest p-values associated with GA (Table 3). We highlight the GA-DMRs that were annotated to these genes, along with our overall EWAS results

with volcano and Manhattan plots (Figure 1a & 1b). Interestingly, some genes not only had multiple CpGs that showed differential methylation, but also several DMRs. *SLC7A5* had the most DMRs (6) associated with GA which contained 22 CpGs in total, followed by *IRX4* with four DMRs containing 20 CpGs. Lastly, the most CpGs within a GA-DMR was eight (cg17650747, cg18246442, cg01875784, cg25378916, cg06654549, cg20426932, cg09155145, and cg14024788) which are located 1kb away from the transcription start site of *IRX4*. The majority of the CpGs identified in highlighted DMRs are located within the gene body of their respective genes and were hypermethylated with increasing GA.

2.3.4 Associations with PMA

The CpG most significantly associated with PMA was cg06413398 which was annotated to *DDO* ($p=3.30x10^{-75}$), and we highlight the 10 CpGs with the smallest p-values, their relationships with PMA, and their genomic annotations in Table 4. From the DMR analysis, we identified 4,498 CpGs within 1,562 Bonferroni significant DMRs (Supplementary Tables S5 and 6). The top DMRs with gene annotations were *DDO* ($P_{Bonf} = 1.54x10^{-73}$), *RBPJ* ($P_{Bonf} = 1.09x10^{-52}$), *SGCE* ($P_{Bonf} = 3.29x10^{-49}$), *MN1* ($P_{Bonf} = 3.64x10^{-48}$), *DGKZ* ($P_{Bonf} = 6.38x10^{-48}$), *LINC01491* ($P_{Bonf} = 1.83x10^{-47}$), *LRRC17* ($P_{Bonf} = 1.84x10^{-43}$), *IKZF4* ($P_{Bonf} = 1.28x10^{-42}$), *CUX1* ($P_{Bonf} = 2.35x10^{-42}$), and *GGACT* ($P_{Bonf} = 5.93x10^{-42}$) (Table 5). We highlight the DMRs annotated to these genes, along with our overall EWAS results with volcano and Manhattan plots (Figure 2a & 2b). Of these top DMRs, *MN1* had the most CpGs (16) within 5 DMRs. Most of the DMRs had only two CpGs within a region however, the annotated DMR with the most CpGs was *C1QTNF1* ($P_{Bonf} = 2.03x10^{-8}$) with 24 CpGs within 1500bp. All of the significant CpGs within the DMRs of *DDO* and *GGACT* were hypomethylated with increasing PMA, whereas the CpGs within *RBPJ*, *SGCE*, *MN1*, *DGKZ*, *LRRC17*, and *IKZF4* were all hypermethylated with increasing PMA.

2.3.5 Overrepresentation of genes in biological pathways

Enrichment analysis of the genes annotated to GA- and PMA-associated CpGs were used to identify gene ontology (GO) of biological pathways (BP), cellular composition (CC), and molecular functions (MF), as well as KEGG[50-52] and Reactome pathways that were associated with the epigenetic signatures of these age metrics (Supplementary Figures S1-S4). These figures report the pathways that

had the most genes within the pathway that were significantly enriched. Interestingly, GO analysis identified two MF pathways ("the regulation of GTPase activity" and "cadherin binding"; Supplementary Figures S1b and S2b), two BP pathways ("axonogenesis" and "regulation of neuron projection development"; Supplementary Figures S1c and S2c), and one KEGG pathway ("PI3K-Akt signaling"; Supplementary Figures S3a and S4a) as being significantly over-represented amongst genes differentially methylated with both GA and PMA. The top CC and MF pathways over-represented amongst differentially methylated genes associated with PMA were cell-cell junction ($p = 3.70x10^{-18}$) and actin binding ($p = 4.74x10^{-17}$) (Supplementary Figure S2a & S2b). More specifically, pathway enrichment analysis identified many neurodevelopmental related pathways over-represented amongst genes demonstrating differential methylation associated with both age metrics. For example, neuronal cell body and neuron to neuron synapse were top CC pathways over-represented in genes demonstrating differential methylation with GA (Supplementary Figure S1a) and PMA (Supplementary Figure S2a). Axon guidance ($p = 7.34x10^{-11}$) was also among the most enriched KEGG pathways over-represented in genes demonstrating differential methylation with PMA (Supplementary Figure S4a) and the BP "neuron projection extension" ($p = 4.58x10^{-8}$) was over-represented in genes demonstrating differential methylation with GA (Supplementary Figure S1c). Pathways associated with growth and development were also significantly over-represented amongst genes with differential methylation associated with neonatal age metrics. Growth factor binding ($p = 2.24x10^{-7}$) was among the top MF pathways over-represented by genes with differential methylation associated with PMA (Supplementary Figure S2a), while cell growth ($p = 4.24x10^{-9}$), the regulation of cell growth ($p = 2.07x10^{-7}$), developmental cell growth ($p = 2.06x10^{-8}$), and the regulation of developmental growth ($p = 8.84x10^{-8}$) were among the top BP pathways enriched with GA associated differential methylation (Supplementary Figure S1c). In addition, several pathways related to both the development and function of muscle tissue and the heart were found to be associated with GA (Supplementary Table S7). Consistent with the GO analysis, the most significant Reactome pathway over-represented amongst genes with GA-associated differential methylation was neuronal systems ($p = 9.63x10^{-4}$) (Supplementary Figure S3b). The most enriched

Reactome pathways amongst genes differentially methylated with PMA were RHO GTPase cycle ($p = 7.29x10^{-13}$) and diseases of signal transduction by growth factor receptors and second messengers ($p = 1.63x10^{-7}$) (Supplementary Figure S4b). All results can be found in Supplementary Tables S7-S16.

Lastly, we utilized the gene enrichment R package *methylGSA*[53] that accounts for probe bias (Supplementary Figures S17-S22). Positive regulation of GTPase activity ($p = 1.56x10^{-5}$) remained one of the most significant PMA-associated pathways identified by GO analysis. Immune related pathways, such as natural killer cell differentiation ($p = 2.70x10^{-3}$), immunological synapse ($p = 2.74x10^{-3}$) and T cell activation ($p = 2.87x10^{-3}$) were also enriched pathways. Neurodevelopmental pathways including negative regulation of long-term synaptic potentiation ($p = 6.95x10^{-4}$), regulation of modifications to synaptic structure ($1.48x10^{-3}$) and axon guidance receptor activity ($p = 2.96 x10^{-3}$) were also identified as enriched pathways associated with PMA.

2.3.6 Brain-Buccal Correlations

We tested whether methylation levels at our GA- and PMA-associated CpGs exhibited positive correlations between brain and buccal tissues in an independent external dataset (GEO Accession: GSE111165). Briefly, this external dataset consisted of paired blood, saliva, buccal, and brain tissue samples from living donors. Samples were drawn during surgery, from patients with medically intractable epilepsy between the ages of 5 and 61. DNAm was assessed with the Infinium HumanMethylationEPIC BeadChip arrays ($n = 21$)[31]. We observed 85 of the 2,366 CpGs with differential methylation associated with GA also had positive and significant brain-buccal correlations (Supplementary Table S23), and of them, 12 CpGs were within GA-DMRs. The CpG with the strongest positive brain-buccal correlation for GA was cg22757362 which was annotated to *RASGEF1A* ($p = 3.70x10^{-7}$, cor=0.870) (Table 6). *IRX4* (cg17774559, $p=5.49x10^{-4}$, cor=0.702) was also among the top genes with positive brain-buccal correlated CpGs. The top CpGs with the most significant brain-buccal correlation also within a GA-associated DMR were cg23605961 ($p = 2.32x10^{-5}$, cor = 0.793) and cg19087971 (p = 0.001, cor = 0.676) annotated to *PRKAR1B*. We also observed that 802 of the 14,979 CpGs associated with PMA also had positive and significant ($p < 0.05$) brain-buccal correlations (Supplementary Table S24). Of the significant

and positively correlated CpGs we observed that 267 of them were also within our identified PMA-DMRs. Overall, the most significant brain-buccal correlation that was within a CpG with the strongest differential methylation with PMA was cg26330304, which was annotated to *ZNF710* ($p = 0$, cor = 0.838) (Table 6). The most significantly correlated CpG within a DMR associated with PMA was cg11193064 which was annotated to *SMAD6* ($p = 8.94 \times 10^{-7}$, cor = 0.830). Of the highlighted genes with differential methylation associated with PMA in Figure 2 and Table 5, *MN1* and *LRRC17* were annotated to CpGs with significant positive brain-buccal correlations. Last, we performed the GO-term and pathway enrichment analyses, restricting the analysis to age-associated CpGs with significant, positive brain-buccal correlations. Several pathways involving the response to and signaling of growth factor pathways were found to be associated with brain-buccal associated GA CpGs ($q < 0.05$). Additionally, we found axonogenesis, regulation of neuron projection development, and cellular adhesion pathways among the top significant GO pathways ($q < 0.05$) associated with the brain-buccal positive PMA CpGs. Finally, the significant GA and PMA brain-buccal CpGs were then run through the online tool eFORGE[54-56] to determine if any sets of these CpGs are overlapping chromatin accessibility in other tissue types. Positive brain-buccal correlated and GA-associated CpGs were found to be accessible in skin ($p = 6.36 \times 10^{-5}$, $q = 0.01$) and breast tissue ($p = 0.002$, $q = 0.202$) (Supplementary Figure S5). Fetal tissue had the most significant associations with positive brain-buccal correlated and PMA-associated CpGs: fetal lung ($p = 2.13 \times 10^{-8}$, $q = 3 \times 10^{-6}$), fetal kidney ($p = 4 \times 10^{-4}$, $q = 0.039$), fetal stomach ($p = 7.41 \times 10^{-4}$, $q = 0.041$), and fetal muscle leg ($p = 0.002$, $q = 0.085$) (Supplementary Figure S24)

2.4 DISCUSSION

Our EWAS study identified age-associated differences in the neonatal buccal tissue epigenome of infants born very preterm. Both age metrics were observed to have widespread associations with differential methylation throughout the genome. GA-associated DNAm may represent how timing of birth leaves a lasting and detectable signature in the epigenome when measured at the end of the NICU stay, while PMA-associated changes in methylation reveals how the methylome varies with aging and development between conception and discharge from the NICU. Increasing GA and PMA were both

predominately associated with increasing methylation levels; however, a few of the most significant CpGs for GA were observed to have decreased methylation levels with increasing age at birth.

A number of our findings align with those of a recent study of GA in both term-born and preterm-born infants[57]. Multiple CpGs within *IRX4* (cg18172877, cg07167946, and cg04441405) and one CpG in *ZBP1* (cg11460314), exhibited hypomethylation with increasing GA in both of our independent analyses. Consistent with our findings, Wheater et al[57] also found *IRX4* to be the most significant DMR associated with GA. Our DMR analysis identified *IRX4* and *SLC7A5* as having the strongest association with GA as well as having the most CpGs within their respective DMRs. Thus, DNA methylation patterns at these genes are undergoing dynamic changes in early development. *IRX4* is associated with several pathways related to cardiac development, and prematurity is a known risk factor for long-term issues with cardiac function[58]. *SLC7A5* has previously been observed to have multiple differentially methylated regions associated with neural function in preterm infants[59]. Interestingly, both cardiac- and neurodevelopmental processes were also over-represented in our pathway enrichment analyses.

Importantly, we identified novel neonatal age-associated changes in DNAm. For example, our overall EWAS found that cg13211971, annotated to *PDE9A*, had the strongest association with GA. This CpG and gene was not identified as age-associated by Wikenius et al[16], but this gene is involved in cGMP regulation, is highly expressed in neurons, and is being investigated as a promising target for treating neurodevelopmental[60] and neurodegenerative disorder[61]. Additionally, of the 26 CpGs within *SLC7A5* highlighted by Sparrow et al, we observed 12 with differential methylation as well as additional novel *SLC7A5* CpG sites associated with GA.

PMA-associated CpGs were annotated to genes associated with neurodevelopment. We found the CpG most strongly associated with PMA to be cg06413398 which is annotated to *DDO*. *DDO*, a known age-associated gene[62; 63], becomes increasingly hypomethylated as humans age and appears to be a marker of aging. Interestingly, we also observed an overlap between the GA-associated CpGs from Wheater et al[57] and our PMA-associated CpGs where cg04441405, cg17774559 within *IRX4*, and cg11460314 within *ZPB1* had significant opposite associations between our studies. Lastly, we observed two of the GA-

associated genes identified by Wheater et al also being associated with PMA in our study: *HEATR2* and *SMIM2*; however differential methylation was observed at different CpGs between these two studies. Overall, increasing DNA methylation was associated with increasing PMA at the majority of CpGs and this observation aligns with the findings of Wikenius et al[16] who described changes in DNAm associated with PMA between 6 and 52 postnatal weeks. Of the 42 genes identified by Wikenius et al[16] as undergoing changes in DNA methylation in early development, we observed that 29 were associated with PMA in our study. In both studies, *IKZF4* and *MN1* were among the genes that were most hyper-methylated with increasing age, however, we observed a larger number of differentially methylated CpGs within these genes in our DMR analysis. The novel findings that we observe could be due to differences in study population as we focus only on very preterm infants and thus, our range of gestational age at birth differs substantially from that of Wheater et al[57], Sparrow et al[59], and Wikenius et al[16]. Additionally, our study used the EPIC array which provides more epigenomic coverage and our larger sample size increases the power of our study. Despite these differences we captured genes and CpGs that were first identified in these prior studies, and discovered novel neonatal age-associated genes involved in cardiac development and neurodevelopment which is consistent with the developmental processes that are occurring during this critical period. Biological processes identified in our pathway enrichment analysis are fundamental to the development of neonates, specifically those born prematurely.

We also explored whether our neonatal age associated CpGs were captured by the PedBE Clock, which can be used to predict age in epithelial cells sampled from persons age 0 - 20 years[21]. We found that CpGs within *BMP4*, *UBA7*, and *AMBRA1* were captured in our data as well as the PedBE Clock. *BMP4* encodes a ligand of TGF-beta and has been found to also regulate heart development and adipogenesis. *AMBRA1* has been associated with neural tube defects and is a biomarker of multiple system atrophy[64]. The age-associated methylation changes in these genes may represent developmental processes that are dynamic during the neonatal period but remain dynamic through the pediatric years. We also screened whether our strongest neonatal age-associated CpGs are associated with aging throughout the life course. *DDO* was observed to be differentially hypomethylated as GA increases. Two of our top

CpGs within *DDO*, cg02872426 and cg06413398, have been associated with aging in adult populations[62]. Additionally, Cameron et al (2023), examined whether there were differences in methylation when comparing infants with very low birth weight (VLBW), a condition that overlaps highly with preterm birth, to healthy term birth infants, as well as their methylation status as adults[65]. Interestingly, 12 of the 17 top genes associated with VLBW birth were also found to be significantly associated with GA or PMA in our study. In adults born with VLBW, cg24263062 annotated to *EBF4* was also differentially methylated in our study and associated with PMA (coef = 0.046, $p = 3.81x10^{-9}$). These findings suggest that age at birth has a lasting impact on the methylome throughout the life course.

We acknowledge that there were some limitations to this study. We utilized non-invasive cheek swabs to collect newborns' buccal tissue to study age associated perturbations in DNAm. As such, the epigenetic changes in this tissue may not be reflective of what is occurring in different developing organ systems. Despite this limitation, we identified pathways and CpGs involved in neurodevelopment, heart development, as well as cellular processes that were also found in studies that utilize blood samples. Another potential limitation is that we are studying DNAm at NICU discharge in infants born preterm which may be capturing the changes that are occurring as these children become healthier, as well as epigenetic changes influenced by biological and environmental experiences in the NICU. We attempted to control for this by adjusting for number of neonatal morbidities (BPD, ROP, SBI, and infection) which are associated with longer NICU stays. We recognize that other unmeasured exposures and experiences in the NICU may have influenced our observed associations. Importantly however, many of our findings align with prior evidence which has shown that preterm birth and gestational age are associated with DNA methylation difference and has a lasting effect on health throughout the life course[16; 65; 66]. Additionally, our study provided a unique opportunity to study changes in DNAm solely among very preterm infants, rather than comparing preterm to term born infants. Another strength of our study is our large sample size compared to prior studies that investigated similar research questions, and thus we were well powered to address the association between neonatal age and DNAm in infants born very prematurely. Additionally, the NOVI study is an ongoing cohort study with longitudinal follow-up up to seven years. Thus, our

future research will allow us to the opportunity to study how methylation of these genes and CpGs change throughout childhood.

In conclusion, we identified CpGs and associated genes that undergo dynamic changes in methylation during early development including previously identified developmental and age associated genes like *DDO*, *IRX4* and *SLC7A5*. We found that the majority of the age-associated differentially methylated CpGs were hyper-methylated with increasing GA and PMA which is consistent with other studies that show increased methylation with aging. Among infants born very preterm, the genes with differential methylation associated with PMA and GA were involved in neurodevelopment, growth processes, cardiac development, and cellular function. Additionally, a small subset of these CpGs also exhibit significant correlations in methylation levels between brain and buccal samples, and these were enriched for pathways associated with axonogenesis, response to growth factors, and cell development.

2.5 CONCLUSION

In conclusion, we identified CpGs and associated genes that undergo dynamic changes in methylation during early development including previously identified developmental and age associated genes like *DDO*, *IRX4* and *SLC7A5*. We found that the majority of the age-associated differentially methylated CpGs were hyper-methylated with increasing GA and PMA which is consistent with other studies that show increased methylation with aging. Among infants born very preterm, the genes with differential methylation associated with PMA and GA were involved in neurodevelopment, growth processes, cardiac development, and cellular function. Additionally, a small subset of these CpGs also exhibit significant correlations in methylation levels between brain and buccal samples, and these were enriched for pathways associated with axonogenesis, response to growth factors, and cell development.

2.6 AIM 1 TABLES

Table 1: Distribution of demographic characteristics, neonatal morbidities and maternal/fetal characteristics of the NOVI study population.

Sample Characteristics (N= 538)	
	Overall
GA (weeks)	26.99 (1.92)
PMA (weeks)	39.16 (3.38)
Infant Race (%)	
White	278 (51.7)
Black	122 (22.7)
Other	51 (9.5)
Asian	41 (7.6)
Hawaiian/Pacific Islander	38 (7.0)
Native American	n < 5
Not Reported	n < 5
Infant Ethnicity (% Hispanic)	115 (21.4)
Outborn (%)	113 (21.0)
Male (%)	299 (55.6)
Maternal age at birth (years)	29.1 (6.36)
Maternal Education < HS/GED (%)	72 (13.4)
Low SES (Hollingshead Level V; %)	42 (7.8)
Number of Neonatal Medical Morbidities (%)	
0	214 (39.8)
1	196 (36.4)
2	99 (18.4)
3+	29 (5.4)
Proportion of Epithelial cells (%)	98.98 (0.03)

Table 2: The top 10 most significant (α=0.05/706323) CpGs associated with gestational age (GA) in NOVI.

CpG	Coefficient	Standard Error	p-value	$P_{Bonf.}$	Chromosome	Position	Gene	Region
cg13211971	0.2637	0.0186	9.79E-46	6.91E-40	21	44194350	*PDE9A*	Body
cg11424970	0.2264	0.0192	3.76E-32	2.66E-26	2	71969999	*2q13.2*	-
cg10202436	0.2272	0.0196	3.79E-31	2.68E-25	5	138733726	*SPATA24*	Body
cg03558436	0.2009	0.0176	2.69E-30	1.90E-24	5	173060723	*5q35.2*	-
cg18955208	0.1661	0.0146	4.07E-30	2.87E-24	6	21873871	*CASC15*	Body
cg09672187	0.2528	0.0227	9.00E-29	6.35E-23	5	1885367	*IRX4*	Body
cg23720947	-0.1766	0.0162	1.17E-27	8.28E-22	19	5967163	*RANBP3*	Body
cg03173167	0.1818	0.0169	4.28E-27	3.02E-21	4	154598228	*LOC100419170*	-
cg27258182	0.1931	0.0181	1.22E-26	8.61E-21	2	171500500	*MYO3B*	Body
cg14771313	0.1604	0.015	1.32E-26	9.35E-21	3	71161285	*FOXP1*	Body

Table 3: The most significant differentially methylated regions (DMR) containing 2 or more CpGs within 1500 base pairs of another significant CpG associated with gestational age (GA).

Gene	# of CpGs	Coefficient	Standard Error	p-value	$P_{Bonf.}$	Chromosome	Position (start)	Position (end)	Region
IRX4	3	1.5214	0.1451	1.02E-25	7.24E-20	5	1885778	1886309	Body
SLC7A5	3	1.2703	0.1237	1.01E-24	7.20E-19	16	87878292	87879698	Body
LINC-PINT; FLJ43663	2	1.2968	0.1267	1.37E-24	9.74E-19	7	130626559	130630462	Body
STAT5A	3	-0.9553	0.0985	3.11E-22	2.21E-16	17	40463425	40463806	3'UTR
6q21.2	3	1.5263	0.1587	6.57E-22	4.67E-16	12	76084308	76084549	-
PHLDB1	7	-1.1414	0.1198	1.61E-21	1.14E-15	11	118502137	118506352	Body;ExonBnd
SLC7A5	3	1.4335	0.1525	5.44E-21	3.86E-15	16	87870597	87873389	Body
NXN	6	-0.6998	0.0750	1.07E-20	7.60E-15	17	867763	872770	Body
4q31.3	2	-1.1678	0.1262	2.13E-20	1.51E-14	4	152720133	152720170	-
7q36.1	2	1.4169	0.1563	1.22E-19	8.69E-14	7	151620691	151620888	-
TPM4	3	1.1060	0.1220	1.25E-19	8.90E-14	19	16178426	16178570	5'UTR;1stExon
RBFOX2	2	-0.9662	0.1067	1.41E-19	1.00E-13	22	36278499	36278607	Body
PML	2	1.6763	0.1859	1.89E-19	1.34E-13	15	74290287	74290466	Body
ANP32A	2	-0.9526	0.1059	2.41E-19	1.71E-13	15	69082449	69084888	Body

Table 4: The top 10 most significant (α=0.05/706323) CpGs associated with post-menstrual age (PMA) in NOVI.

CpG	Coefficient	Standard Error	p-value	$P_{Bonf.}$	Chromosome	Position	Gene	Region
cg06413398	-0.1806	0.0098	3.30E-75	2.33E-69	6	110736865	*DDO*	TSS200
cg10167094	0.1008	0.0056	2.83E-73	2.00E-67	12	53567373	*CSAD*	5'UTR;Body
cg07164639	-0.1772	0.01	2.37E-70	1.67E-64	6	110736958	*DDO*	TSS1500
cg02603733	0.1712	0.0097	5.66E-70	4.00E-64	1	151447362	*1q21.3*	
cg20919287	0.089	0.0051	1.34E-68	9.44E-63	12	53567340	*CSAD*	5'UTR;Body
cg21424090	-0.1978	0.0117	1.00E-63	7.08E-58	11	118550443	*TREH*	TSS200
cg01772842	0.1599	0.0095	1.35E-63	9.54E-58	2	220265394	*2q35*	
cg04453552	-0.2411	0.0145	3.33E-62	2.35E-56	16	57394387	*CCL22*	Body
cg13861278	-0.1426	0.0087	5.15E-61	3.64E-55	1	167868783	*ADCY10*	Body
cg22786472	-0.19	0.0116	2.51E-60	1.77E-54	4	26198975	*RBPJ*	Body

Table 5: The most significant differentially methylated regions (DMR) containing 2 or more CpGs within 1500 base pairs of another significant CpG associated with post-menstrual age (PMA).

Gene	# of CpGs	Coefficient	Standard Error	p-value	$P_{Bonf.}$	Chromosome	Position (start)	Position (end)	Region
DDO	2	-1.4664	0.0777	2.08E-79	1.54E-73	6	110736772	110736865	TSS200
2q35	2	1.2173	0.0736	1.96E-61	1.45E-55	2	220265305	220265394	-
11p15.5	3	1.1819	0.0720	1.72E-60	1.27E-54	11	334561	336031	-
RBPJ	2	-1.1250	0.0697	1.47E-58	1.09E-52	4	26198975	26199269	Body
SGCE;PEG10	4	1.0177	0.0651	4.44E-55	3.29E-49	7	94279660	94284274	Body;TSS1500
MN1	3	1.0009	0.0647	4.92E-54	3.64E-48	22	28191222	28191914	Body
DGKZ	2	1.1833	0.0766	8.62E-54	6.38E-48	11	46383051	46383141	TSS200;Body
LINC01491	3	0.6394	0.0416	2.47E-53	1.83E-47	15	48137313	48137817	Body
LRRC17; FBXL13	4	1.0983	0.0744	2.49E-49	1.84E-43	7	102553437	102557016	5'UTR;1stExon
16p13.3	2	1.1127	0.0754	3.05E-49	2.26E-43	16	3045060	3048362	-
IKZF4	3	2.1913	0.1497	1.72E-48	1.28E-42	12	56414508	56414698	TSS200;5'UTR;1stExon
CUX1	3	0.8773	0.0601	3.17E-48	2.35E-42	7	101740892	101741221	Body
GGACT	2	-1.2237	0.0842	8.00E-48	5.93E-42	13	101197327	101197428	5'UTR

Table 6: The most significant CpGs associated with gestational age (GA) and post-menstrual age (PMA) were also found to have a positive and significant Brain-Buccal correlation.

GA						
CpG	Correlation	p-value	Chromosome	Position	Gene	Regulatory Region
cg22757362	0.8701	3.70E-07	10	43721445	*RASGEF1A*	5'UTR
cg23605961	0.7935	2.32E-05	7	751331	*PRKAR1B*	5'UTR
cg22158475	0.7416	1.82E-04	1	157096908	*CYCSP52; ETV3*	TSS1500; Body
cg17774559	0.7026	5.49E-04	5	1879698	*IRX4*	Body
cg19087971	0.6766	1.03E-03	7	751233	*PRKAR1B*	5'UTR
cg12227401	0.674	1.09E-03	4	185189393		
cg20544356	0.6468	1.97E-03	17	29890935		
cg18552453	0.6234	3.12E-03	5	151903736		
cg16248376	0.6169	3.53E-03	2	175472517	*WIPF1*	5'UTR
cg03688854	0.6169	3.53E-03	20	59767609		
PMA						
CpG	Correlation	p-value	Chromosome	Position	Gene	Regulatory Region
cg26330304	0.8377	0.00E+00	15	90569017	*ZNF710*	5'UTR
cg11193064	0.8299	8.94E-07	15	67008444	*SMAD6*	Body
cg02003272	0.7974	1.88E-05	13	50702719		
cg23605961	0.7935	2.32E-05	7	751331	*PRKAR1B*	5'UTR
cg21371176	0.7896	2.83E-05	1	25407112		
cg17250863	0.7649	8.13E-05	20	33451272	*GGT7*	Body
cg26576041	0.7636	8.53E-05	1	1247929	*CPSF3L*	Body
cg22817719	0.7597	9.84E-05	11	130068300	*ST14*	Body
cg13990107	0.7519	1.29E-04	15	63663946	*CA12*	Body
cg09299774	0.7429	1.74E-04	1	1251017	*CPSF3L*	Body

2.7 AIM 1 FIGURES

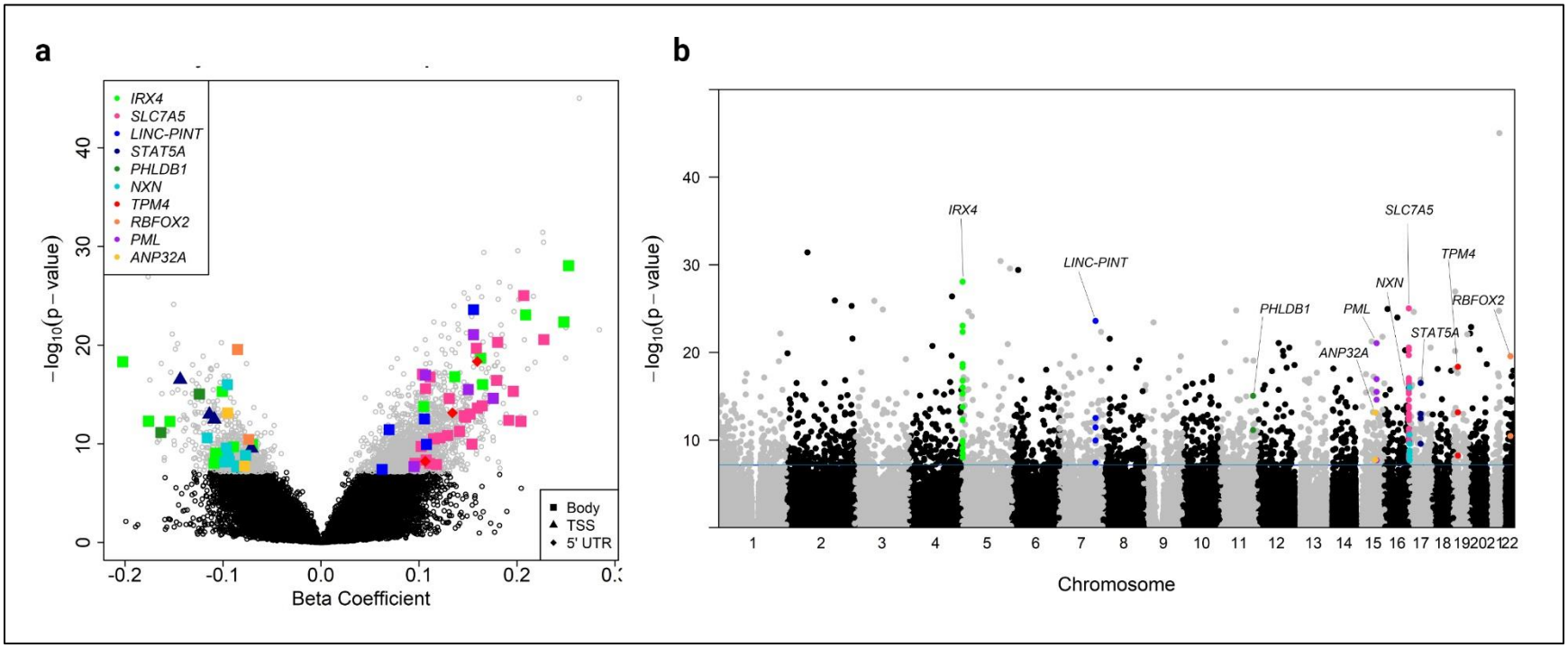


Figure 1: Volcano plot **(a)** of the beta coefficients and − log10(p values) from the epigenome-wide association study (EWAS) for GA (gray = Bonferroni, color = CpGs within the top 10 differentially methylated regions (DMR)), and Manhattan plot **(b)** of the genomic distribution of these results with gene names annotated to those CpGs within top 10 DMRs. Bonferroni significance threshold of $\alpha = 0.05/706323$ (blue horizontal line).

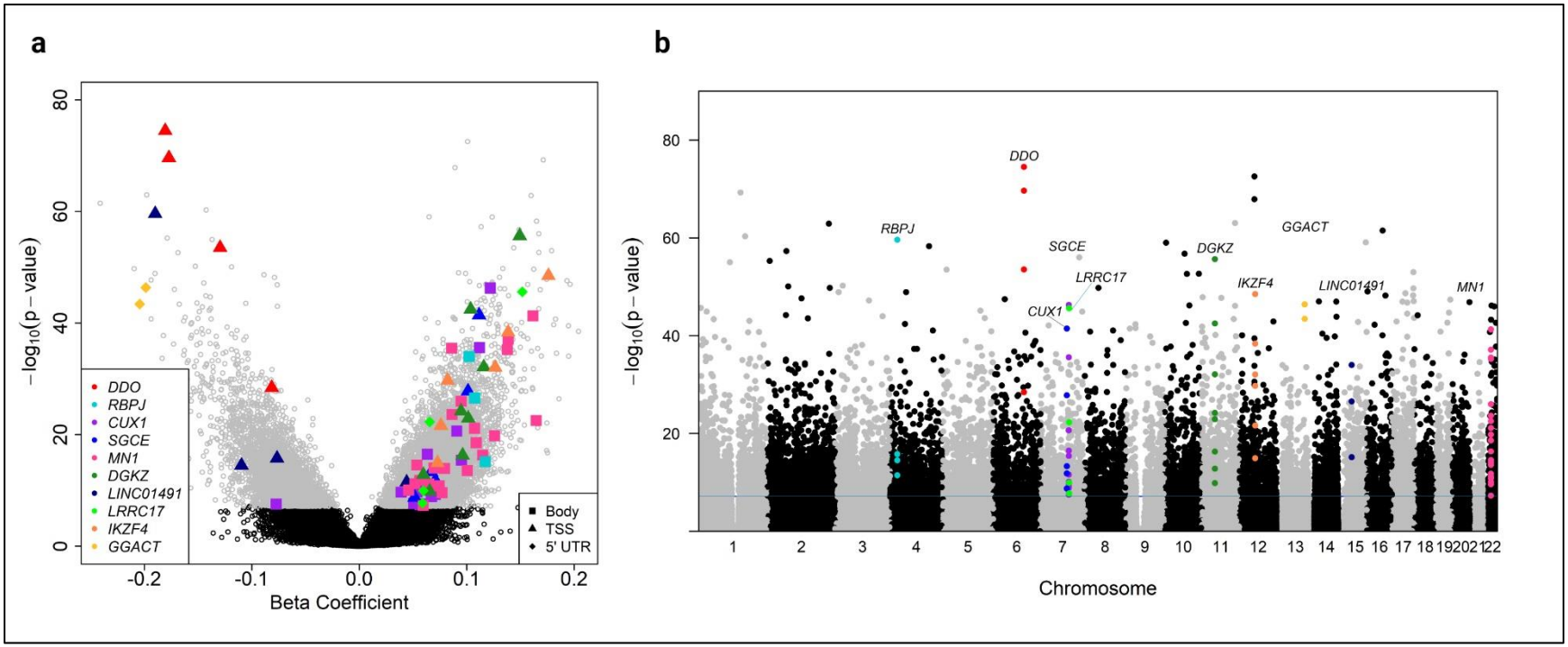
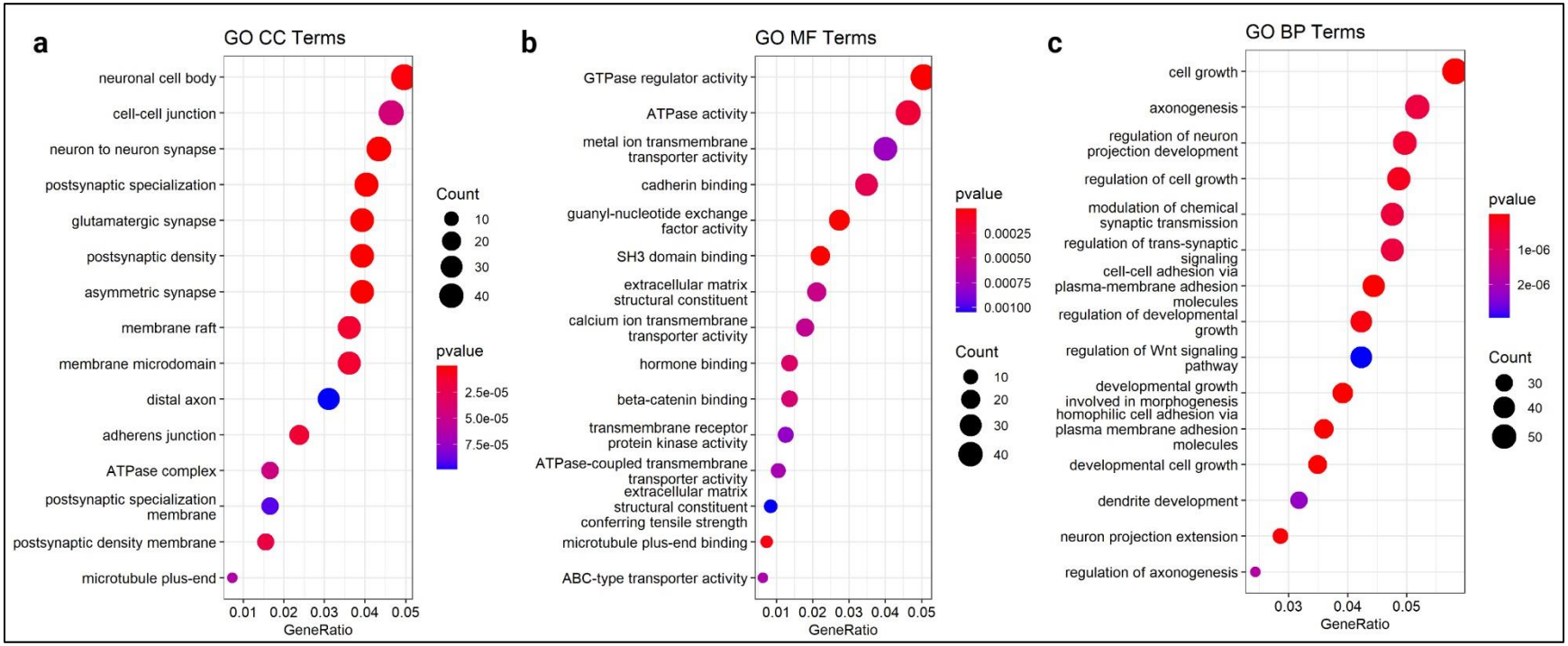


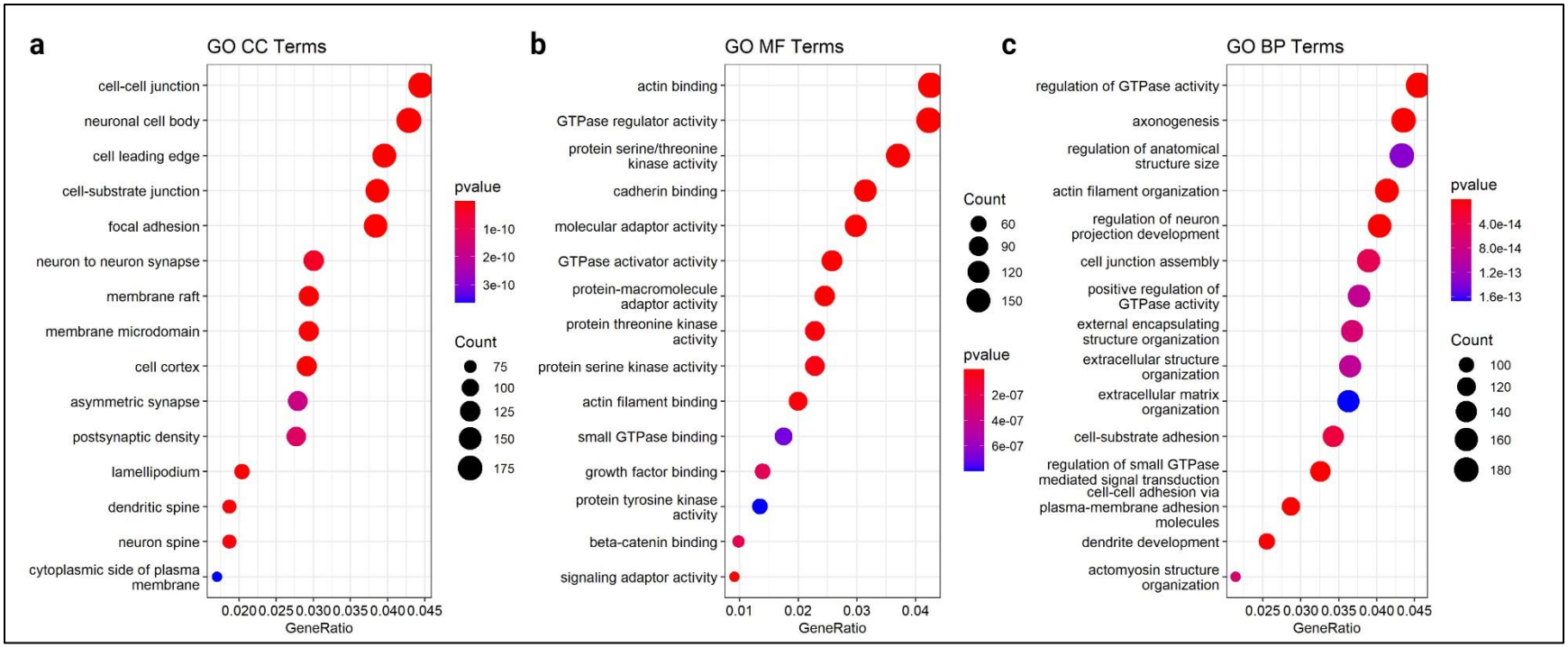
Figure 2: Volcano plot **(a)** of the beta coefficients and − log10(p values) from the epigenome-wide association study (EWAS) for PMA (gray = Bonferroni, color = CpGs within the top 10 differentially methylated regions (DMR)), and Manhattan plot **(b)** of the genomic distribution of these results with gene names annotated to those CpGs within top 10 DMRs. Bonferroni significance threshold of $\alpha = 0.05/706323$ (blue horizontal line).

2.8 AIM 1 SUPPLEMENTARY TABLES AND FIGURES

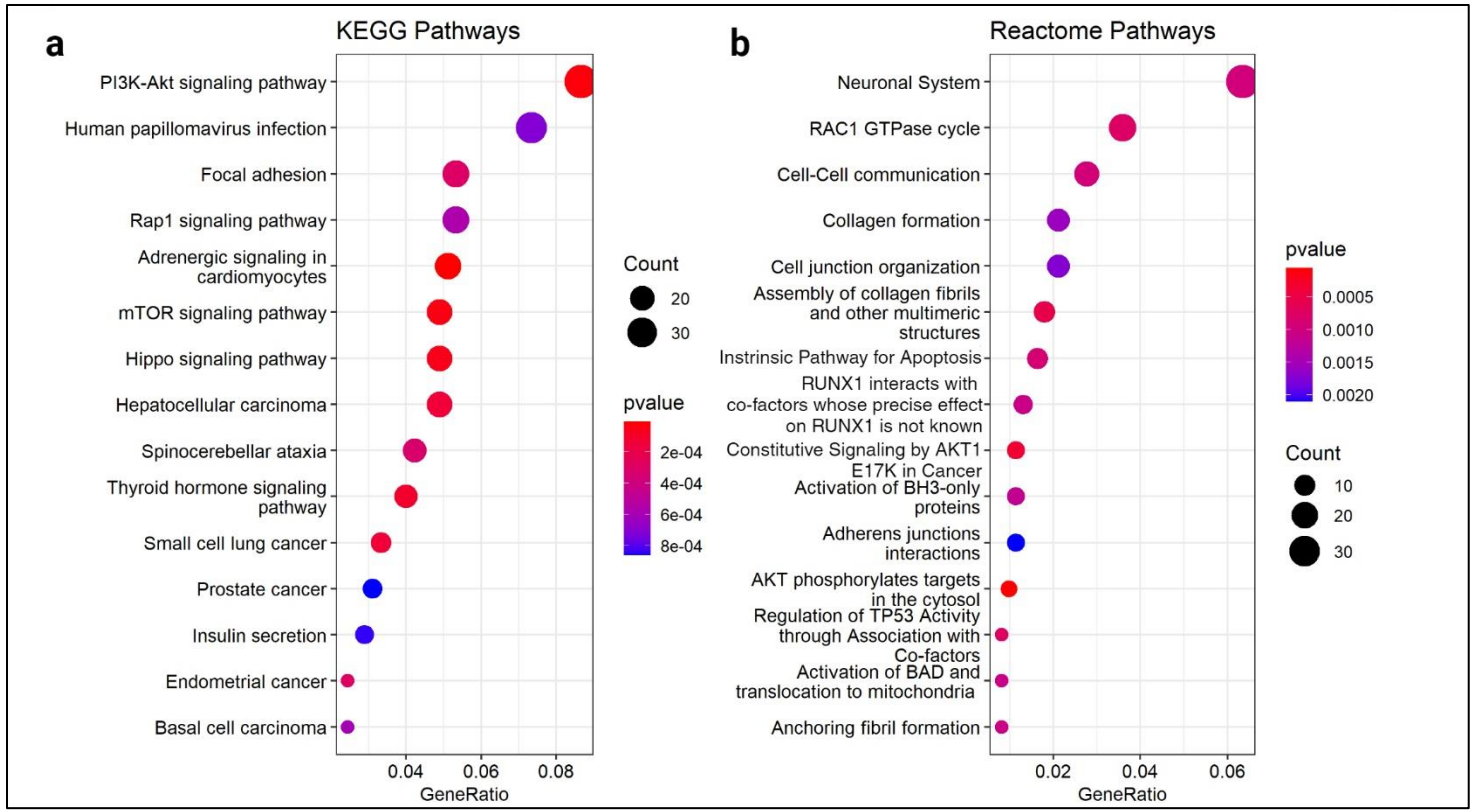
All supplementary tables can be found in Supplementary_Tables_Chapter2.xlsx file.



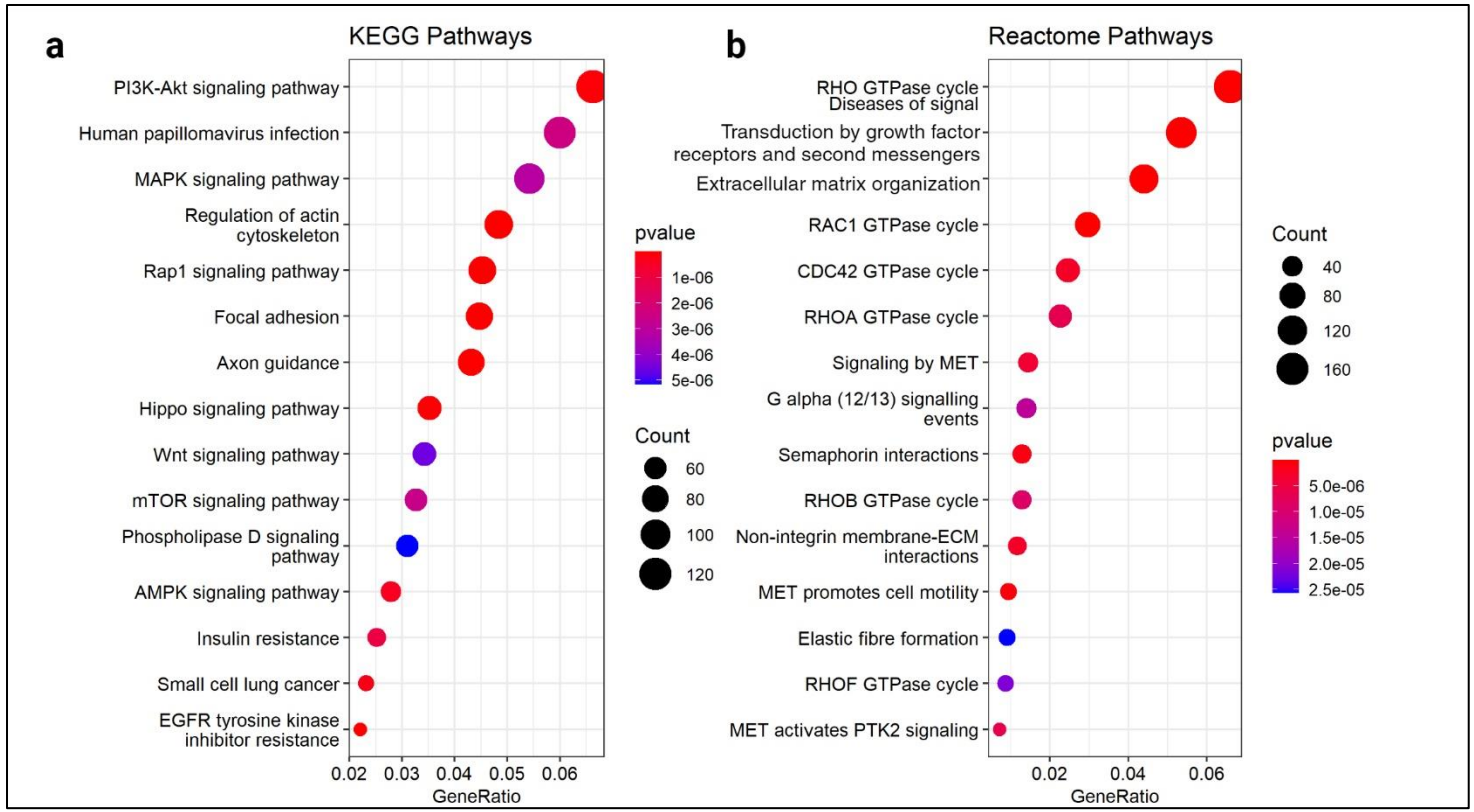
Supplementary Figure S1: Top 15 Gene Ontology (GO) enrichment results for the CpGs significantly associated with gestational age (GA) for the **(a)** cellular components (CC), **(b)** molecular function (MF), and **(c)** biological pathways (BP) associated (circle size and color indicate the number of GA-associated CpGs within a pathway and the p-value for the enrichment test, respectively).



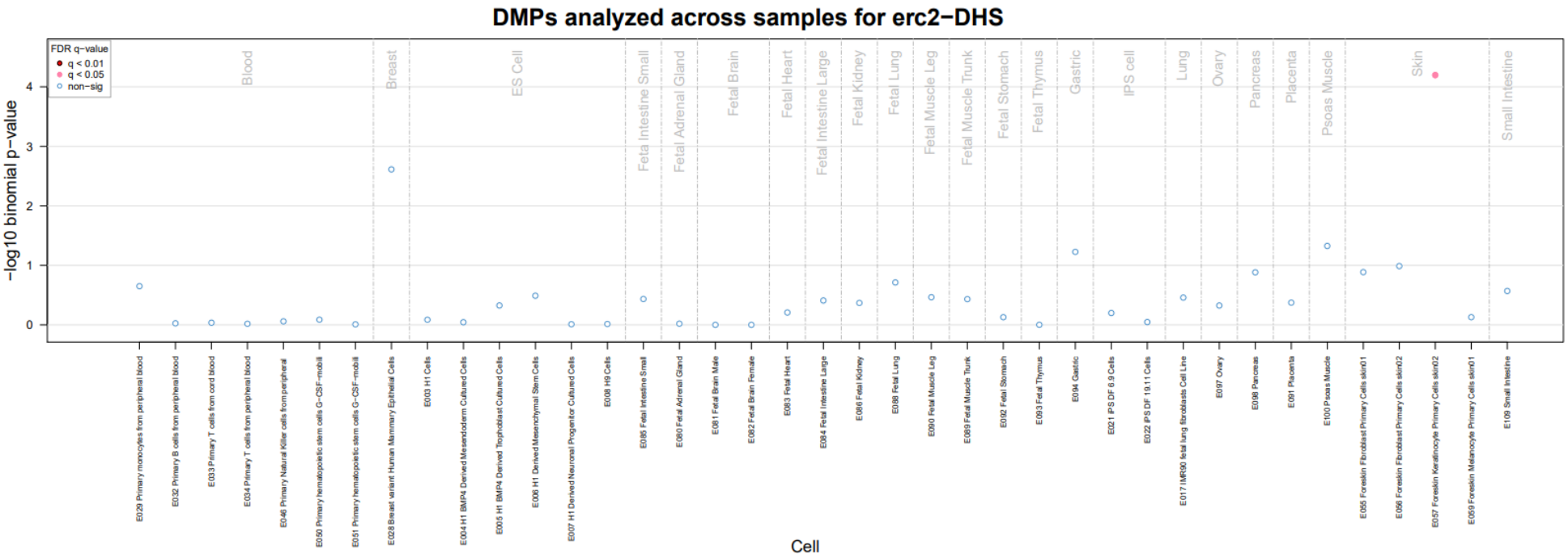
Supplementary Figure S2: Top 15 Gene Ontology (GO) enrichment results for the CpGs significantly associated with post-menstrual age (PMA) for the **(a)** cellular components (CC), **(b)** molecular function (MF), and **(c)** biological pathways (BP) associated (circle size and color indicate the number of PMA-associated CpGs within a pathway and the p-value for the enrichment test, respectively).



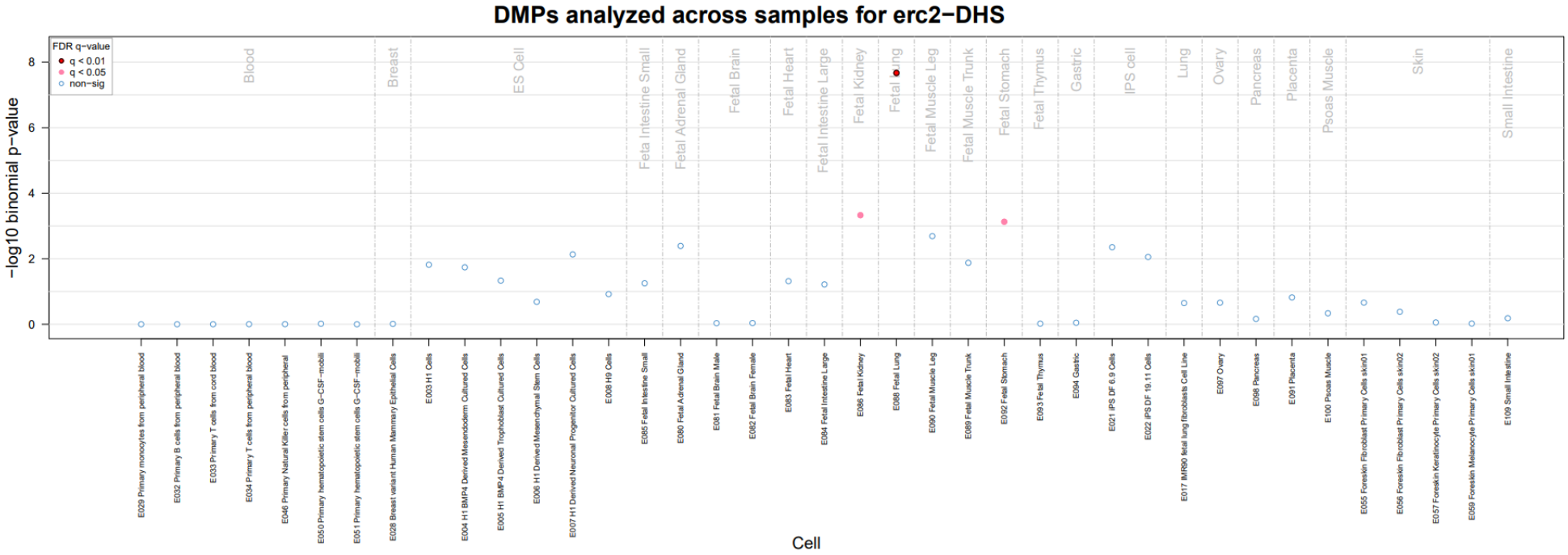
Supplementary Figure S3: Top 15 pathway enrichment results of the genes most significantly associated with gestational age (GA) for both KEGG pathways **(a)** and Reactome pathways **(b)** denoted by the number of genes within the pathway and $p < 0.05$ significance threshold.



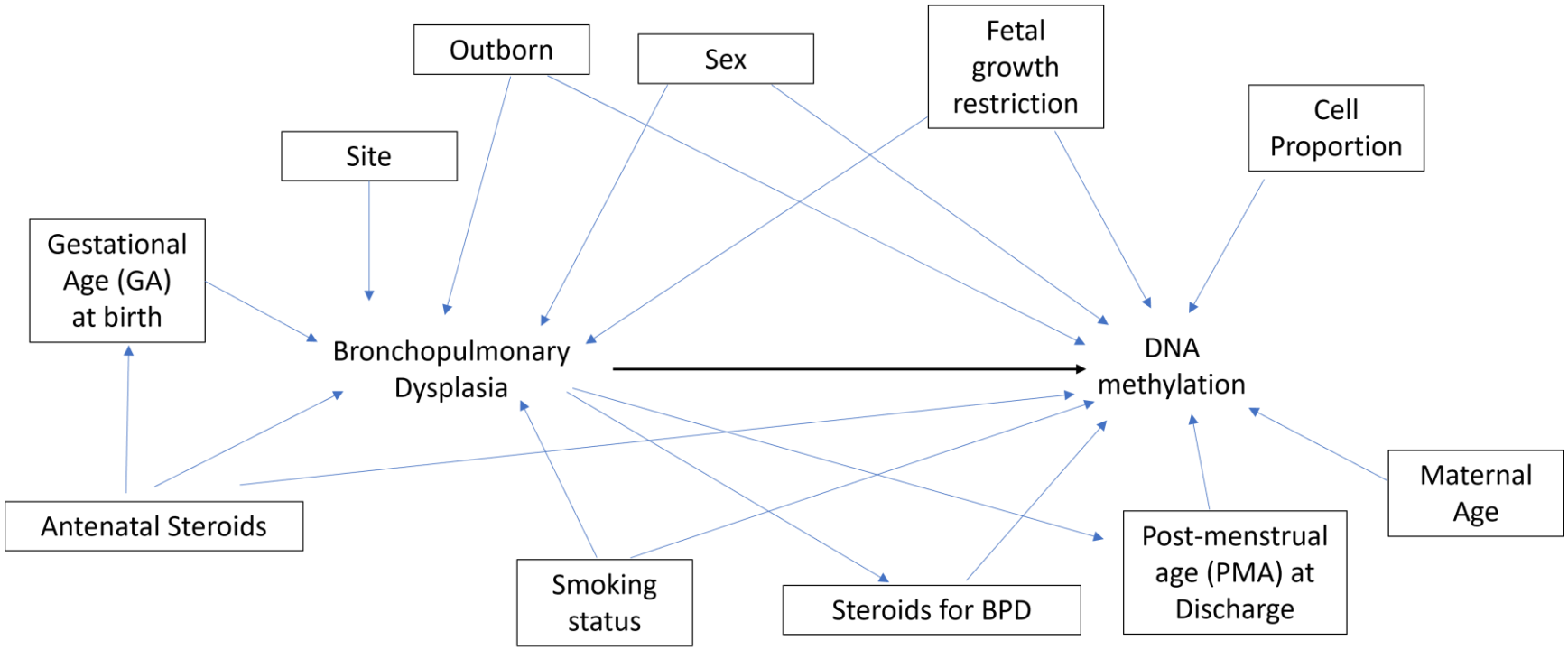
Supplementary Figure S4: Top 15 pathway enrichment results of the genes most significantly associated with post-menstrual age (PMA) KEGG pathways **(a)** and Reactome pathways (**b**) denoted by the number of genes within the pathway and $p < 0.05$ significance threshold.



Supplementary Figure S5: eFORGE results of cell type specific overlap of differentially methylated probes associated with positive brain-buccal and gestational age (GA).



Supplementary Figure S6: eFORGE results of cell type specific overlap of differentially methylated probes associated with positive brain-buccal and post-menstrual age (PMA).



Supplementary Figure S7: Directed acyclic graph to identify potential confounders.

CHAPTER 3: Epigenetic Associations in HPA Axis Genes Related To Bronchopulmonary Dysplasia and Antenatal Steroids

Kenyaita M. Hodge[1], Vasyl Zhabotynsky[2], Amber A. Burt[1], Brian S. Carter[3], Rebecca C. Fry[4,5], Jennifer Helderman[6], Julie A. Hofheimer[5], Elisabeth C. McGowan[7], Charles R. Neal[8], Steven L. Pastyrnak[9], Lynne M. Smith[10], Sheri A. DellaGrotta[11], Lynne M. Dansereau[11], Barry M. Lester[7,11,12,], Carmen J. Marsit[1,13], T. Michael O'Shea[5] and Todd M. Everson[1,13*] on behalf of program collaborators for Environmental influences on Child Health Outcomes*

[1]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA,
[2]Institute for Environmental Health Solutions, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC
[3]Department of Pediatrics-Neonatology, Children's Mercy Hospital, Kansas City, MO
[4]Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC
[5]Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC
[6]Department of Pediatrics, Wake Forest School of Medicine, Winston-Salem, NC
[7]Department of Pediatrics, Warren Alpert Medical School of Brown University and Women and Infants Hospital, Providence, RI, USA
[8]Department of Pediatrics, University of Hawaii John A. Burns School of Medicine, Honolulu, HI
[9]Department of Pediatrics, Spectrum Health-Helen Devos Hospital, Grand Rapids, MI
[10]Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA
[11]Brown Center for the Study of Children at Risk, Women and Infants Hospital, Providence, RI
[12]Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, RI
[13]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA

3.1 BACKGROUND

Bronchopulmonary dysplasia (BPD), the most common neonatal morbidity among very preterm infants, is a respiratory disorder in which infant lung development is disrupted by mechanisms that include oxidative stress and inflammation, necessitating treatment with supplemental oxygen, assisted ventilation, and/or pharmacological interventions, such as anti-inflammatory corticosteroids, to improve gas exchange. However, these interventions can result in oxidative stress and perturb the hypothalamic-pituitary-adrenal (HPA) axis, an innate human stress response system, in patients with BPD increasing the physical and physiological stress. Infants born very preterm (<30 weeks' gestational age) are at a particularly high risk for developing poor neuromotor, cognitive, and behavioral outcomes that can persist through adulthood [67-78]. Additionally, very preterm infants with BPD are at increased risk for respiratory conditions and neurodevelopmental impairments [8; 79]. Very preterm infants also have longer stays in the neonatal intensive care unit (NICU) and more hospital encounters following NICU discharge compared with infants without BPD [80]. Despite the public health burden of preterm birth and BPD, there is a dearth of tools that can be easily used at NICU discharge to understand the ramifications of prematurity, and existing clinical indicators are limited in their ability to predict longer-term developmental impairments. Thus, there is a need to better understand the links between acute neonatal morbidities, such as BPD, and long-term negative health outcomes.

Our group has shown that neonatal neurobehavioral responses [81] and an increasing number of neonatal morbidities, including BPD [11; 82], are associated with differences in DNAm throughout the genome, specifically, at *NR3C1* which encodes a primary glucocorticoid (GC) receptor [83]. Another study observed a relationship between BPD and an epigenetic marker of biological aging [84]. Additionally, several epigenome-wide studies of peripheral tissues collected at birth, such as cord blood and placenta, have identified differential DNAm associated with BPD[85-87]. Thus, early life exposures and morbidities, including BPD, appear to leave a fingerprint on the early life epigenome, even in peripheral tissues. These epigenetic associations may provide information on important developmental systems whose epigenetic patterns contribute to BPD, and/or have been affected by these early exposures and experiences.

Environmental, biological, and psychosocial stressors during critical windows of development have been associated with changes in DNAm and linked to poorer health and disease [17-19]. Epigenetic alterations to the stress response systems may play a critical role in explaining some of these long-term effects. For example, a cross-tissue polyepigenetic GC score was developed by Provençal et al. with the goal of creating a tool to study how early life stress may impact psychiatric outcomes later in life, and in subsequent analyses the GC score was associated with exogenous GC exposure, maternal stress, and a higher risk of children's negative mental and behavioral outcomes[19]. In humans, exposure to dexamethasone and hydrocortisone, GC agonists, alter DNAm in peripheral tissues, such as blood [88; 89] and buccal cells [90]. Similarly, in hippocampal progenitor cells, dexamethasone induces persistent changes in DNAm and gene expression when the exposure occurs during cellular proliferation and differentiation [19], which happens between gestational weeks 10 and 25 in humans. Additionally, DNAm in human buccal tissue was found to be associated with high-dose GC exposure, specifically in the *FKBP5* and *NR3C1* genes, which are vital for HPA axis function [90].

Among neonates born very preterm, infants with BPD are among those with the greatest exposure to exogenous GC [91], suggesting that BPD may be associated with altered DNAm at stress response genes. We also hypothesized that epigenetic differences in the stress responses system would be detectable in peripheral tissues like buccal cells, since stress response is a systemic process and can impact the activity of cells throughout the body[92]. However, the epigenetic effects of BPD and steroid treatment on HPA axis genes in infants born preterm has not been well studied, although evidence suggests that the early environment can permanently influence the genome through epigenetic mechanisms and modify endocrine function, metabolism, and behavior of offspring [93; 94].

This exploratory and hypothesis-generating study aimed to demonstrate whether an innovative epigenetic biomarker for early life endogenous and exogenous GC exposure or CpG-specific methylation of HPA axis genes were associated with BPD severity or with antenatal steroid treatment in very preterm infants. Exogenous steroid treatment is an exposure that could directly interact with the HPA axis system, while BPD itself exerts physiological stress changes that potentially triggers the HPA system and impacts the

methylome of infants born very preterm. We hypothesized that the polyepigenetic GC score and methylation levels of 8 HPA axis genes would be altered in association with antenatal steroid exposure and increasing BPD severity.

3.2 METHODS

3.2.1 Characterization of the Study Population

The goal of Environmental influences on Child Health Outcomes (ECHO) program is to investigate the effects of environmental exposures on child health and development[95-97]. The two ECHO cohorts involved in this study are prospective studies of early life influences on child health outcomes among infants born very preterm for which epigenetics data were available.

Demographics of the Neonatal Neurobehavioral Outcomes in Very Preterm Infants (ECHO NOVI) Cohort

A total of 538 infants in the ECHO NOVI study had DNAm, BPD, antenatal steroid, and post-menstrual age (PMA) data available. In this cohort, gestational age (GA) at birth ranged from 22 to 29.86 weeks (mean=26.99) and PMA ranged from 32.14 to 51.43 weeks (mean=39.15 weeks) (Table 7). ECHO NOVI is a diverse cohort based on maternal self-reported race and ethnicity, where 52% of infants were White (n=279); 23% were Black (n=121); 9.5% were Other Race (n=51); 7.6% were Asian (n=41); 7.1% were Hawaiian/Pacific Islander (n=38); >1% were Native American (n<5); and <1% did not report race (n<5). Overall, 22% of infants were of Hispanic ethnicity. Approximately 21% of infants were outborn (N=112), with 74 of these infants having a BPD diagnosis (66%). In this cohort, 7% of infants were classified as having fetal growth restriction, based on birthweight-for-gestational-age percentile. Over 51% of infants had BPD, with 130 (24.2%), 122 (22.7%), and 25 (4.2%) classified as mild, moderate, and severe BPD, respectively. Almost 89% of the infants' mothers were treated with antenatal steroids during pregnancy (N=477); 15% of infants were treated with steroids for BPD (N=81); and 12% of infants were exposed to both antenatal steroids and steroids for BPD (N=67).

Demographics of the Extremely Low Gestational Age Newborn (ELGAN ECHO) Cohort

DNAm, GA, and antenatal steroid data were available for 365 infants in the ELGAN ECHO cohort (Table 7). The average GA of this cohort was 26.07 weeks (range=23.00-27.86 weeks). Approximately 63% of infants were White (n=230); 27% were Black (n=97); 4.3% were Other Race (n=16); >2% were Asian (n < 5); >1% were Native American (n < 5); 3% were Multiracial (n=11); and >1% did not report race. In this cohort, 10% of infants were of Hispanic ethnicity (n=35). All infants were born in the hospital where NICU care was provided (no outborn). Overall, 5% of infants were classified with fetal growth restriction. Approximately 54% of infants had BPD, with 142 (38.9%), 31 (8.5%), and 23 (6.3%) classified as mild, moderate, and severe BPD, respectively. Over 90% of the infants' mothers were treated with antenatal steroids during pregnancy (N=331).

3.2.2 Description of Study Samples

This study was implemented using data from two cohorts of very preterm infants who were participating in the ECHO Program: the ECHO NOVI study and the ELGAN ECHO study. The timeline of sample collection for each cohort is shown in Figure 3. The ECHO NOVI is explained in detail in section 2.2.1. In the ECHO NOVI cohort, BPD was defined as requiring supplemental oxygen at or before 36 weeks PMA and was classified into three levels of severity: mild (oxygen alone or high flow nasal cannula without supplemental oxygen), moderate (high flow nasal cannula with supplemental oxygen or continuous positive airway pressure or nasal intermittent mandatory ventilation), and severe (mechanical ventilation or high frequency ventilation through an endotracheal tube) [98].

The ELGAN ECHO study [32] includes 1506 neonates born between 2002-2004 at <28 weeks' gestation in 14 U.S. hospitals. A total of 1222 infants survived to discharge from neonatal intensive care, and 1198 survived to 2 years. Approximately 1190 are presumed alive and currently range in age from 18 to 20 years. Prenatal demographic, socioeconomic, physical health, pregnancy complication, and neonatal medical complication data were collected from review of medical records during NICU stays and maternal interviews around the time of birth of the ELGAN study participant. Neonatal dried blood spots were collected during the first several days of life.

Inclusion/Exclusion Criteria for this study

Among the participants in ECHO NOVI and ELGAN ECHO, those with missing DNAm, BPD, and antenatal steroid data, and age at sample collection (GA and/or PMA) were excluded from this study. In the ECHO NOVI cohort, 4 of the 542 participants were excluded due to missing antenatal steroid data; therefore, 538 infants were included in this study. In the ELGAN ECHO cohort, 857 of the 1222 infants were excluded, primarily due to missing DNAm, leaving 365 infants included in the study.

3.2.3 DNA Methylation (DNAm) Measurement, Quality Control, and Preprocessing

Both cohorts used their pre-developed pre-processing and quality control pipelines that aligned with their previously published work. To assess the methylation profiles of the ECHO NOVI study cohort, we analyzed the DNAm data from buccal tissue at NICU discharge as measured using the Illumina MethylationEPIC microarray. This data has undergone established quality control and processing pipelines mentioned in section 2.2.4. After these exclusions, a total of 706,278 probes were available from 538 ECHO NOVI samples.

The ELGAN ECHO cohort used similar quality control procedures, such as the exclusion of blood spot samples with a detection p-value >0.01 for greater than 10% of samples, the removal of sex mismatches, and functional normalization using *minfi* [99] and *ShinyMethyl* [100] R packages. Final beta values were generated following batch effect correction using PCA checks to identify batches and *ComBat* [101] to remove these effects. After exclusions, a total of 786,267 probes were available from 365 ELGAN samples.

3.2.4 Data Harmonization

The definition of BPD severity (4-level factor) was harmonized between the ECHO NOVI and ELGAN ECHO cohorts, with a 4-level factor for BPD severity: no BPD (no oxygen requirement), mild BPD (supplemental oxygen or high-flow nasal cannula), moderate BPD (supplemental oxygen via high-flow nasal cannula, nasal continuous positive airway pressure or nasal positive airway pressure ventilation), and severe BPD (mechanical ventilation via endotracheal tube) at ≥36 weeks PMA [102; 103].

GA was calculated in weeks, and PMA at NICU discharge was defined as the number of days from the last day of the mother's menstrual cycle to the day of discharge from the NICU, also calculated in weeks. Outborn status [35], maternal smoking during pregnancy, and steroid variables were binary "yes" versus "no." Fetal growth restriction was also binary with "yes" defined as a birthweight for GA more than 2 standard deviations below the population-specific growth curve. Antenatal steroids were given to mothers who were at risk of preterm delivery to accelerate fetal organ maturation and the variable was defined as a binary "yes" or "no."

3.2.5 GC Score Calculation

The neonatal GC score was calculated by multiplying the DNAm beta values by the 24 CpG weights from the cross-tissue polyepigenetic GC score established by Provençal et al. [19] and summing those weighted methylation levels for each infant in both cohorts. The weights represent the coefficients from an elastic net regression model with lower GC score indicating greater GC exposure[19]. Additionally, the GC score was approximately normally distributed, but with a minimal right skew. The estimated polyepigenetic GC scores from the neonatal DNAm data in the ECHO NOVI cohort yielded scores that ranged from -1.78 to -1.10 with a median score of -1.61. While the GC scores from the neonatal data in the ELGAN ECHO cohort ranged from -0.375 to -2.09 with a median score of -1.19.

3.2.6 HPA Axis Genes

Genes previously associated with the HPA axis were identified via literature review [104]. The HPA system involves complex interactions across many genes. We focused on a subset of genes that have been previously studied in relation to prospective outcomes, and we identified 8 genes for this analysis: *FKBP5, NR3C1*, *NR3C2*, *CRH*, *CRHR1*, *CRHR2, POMC,* and *HSP90AA1.* To capture CpGs annotated to these genes, the UCSC Genome Browser (hg19) [105] was used to find regions +/- 2500 bp beyond the start and stop of each gene. The regions for each gene are as follows: *FKBP5* (chr6: 35538862-35698897), *NR3C1* (chr5: 142654996-142817577), *NR3C2* (chr4: 148076262-148447198), *CRH* (chr8: 67086112-67093346), *CRHR1* (chr17: 45781820-45838328), *CRHR2* (chr7: 30649442-30684956), *POMC* (chr2: 25381222-25394059), and *HSP90AA1* (chr14: 102544575-102608586). A total of 344 CpGs were

annotated to these genes. Due to differences in quality control and processing steps between the two cohorts, the ECHO NOVI cohort captured 317 of these CpGs while the ELGAN ECHO cohort captured 294 CpGs.

3.2.7 Data Analysis

For both cohorts, DNAm was profiled using the Illumina MethylationEPIC BeadArray, which measures over 850,000 CpG loci at a single nucleotide resolution. After data cleaning and harmonization, our analysis included 538 infants from the ECHO NOVI cohort and 365 infants from the ELGAN ECHO cohort with complete data. Since DNA was profiled using two different tissues (neonatal blood spot samples and buccal cells) and two different sample collection time points (near birth or at NICU discharge), we performed and interpreted all analyses separately within the ECHO NOVI and ELGAN ECHO cohorts while adjusting for the same covariates consistent with the sample collection timeline and relevant exposures (Figure 3). We used complete cases for children with DNAm and phenotype data (BPD and steroid treatment). We used directed acyclic graphs to identify potential confounders (Supplementary Figure S8). First, we tested whether the rank-normalized neonatal GC scores were associated with BPD and increasing BPD severity while adjusting for potential confounders, including weeks of PMA at NICU discharge , GA at birth, sex, study site, outborn status, fetal growth restriction, maternal age, maternal smoking during pregnancy, cell type proportions (estimated from reference methylomes) [34; 106-109], and maternal antenatal steroid exposure. Buccal tissue from the ECHO NOVI cohort consisted primarily of epithelial cells (>90%), and the blood spot samples from the ELGAN ECHO cohort consisted of CD8 T cells, CD4 T cells, natural killer cells, B cells, monocytes, granulocytes, and nucleated RBCs. Next, we explored whether BPD severity and steroid exposure were associated with differential methylation of the 8 HPA genes of interest while adjusting for the same confounders. Lastly, we stratified the data for each cohort by sex to investigate whether there were sex-specific associations between HPA DNAm, BPD severity, and steroid treatment. All statistical analyses were performed in the R statistical environment, and all regression models were fit with generalized estimating equations. We generated coefficients, confidence intervals, and p-values. We used a

significance threshold of α=0.05 to assess statistical significance for the GC score models, whereas we used a 5% FDR threshold to account for multiple testing in the CpG-specific analyses of HPA genes; all tests were two-sided.

3.3 RESULTS

3.3.1 GC Score and BPD

After regressing the GC score on BPD and BPD severity while adjusting for antenatal steroids and steroids for BPD, we observed that the GC score decreased with increasing BPD severity in the ECHO NOVI cohort. After adjusting for additional confounders, the inverse relationship between BPD severity and GC score was not statistically significant in this cohort (Figure 4A). The proportion of epithelial cells was strongly associated with GC score and likely confounded our initial findings in the ECHO NOVI cohort. In the ELGAN ECHO cohort, BPD severity was not associated with the GC score in any of the models, although GC score estimates decreased with increasing BPD severity (Figure 4B), similar to what we observed in the ECHO NOVI cohort. Cell proportion for both monocytes ($p=4.97x10^{-13}$) and granulocytes (p=0.0001) and GA (p=0.02) were associated with the GC score in the ELGAN ECHO cohort. Overall, while we found weak evidence that GC score decreases with increasing BPD severity, the strong correlations between cell composition and GC score appeared to explain much of this relationship.

3.3.2 HPA Axis Genes and BPD Severity

Of the 706,323 CpGs captured in the ECHO NOVI cohort, 317 CpGs were annotated to the 8 HPA genes (Supplementary Table 25). BPD severity was not significantly associated with differential methylation of the HPA genes after accounting for multiple testing. However, we did observe nominally significant associations (raw p-value <0.05): mild BPD (11 CpGs), moderate BPD (27 CpGs), and severe BPD (2 CpGs) (Supplementary Table 26). Interestingly, we observed significant associations between antenatal steroid exposure and the methylation of 10 CpGs within 5 HPA genes (*FKBP5*, *NR3C1*, *NR3C2*, *CRHR1*, *POMC*, and *HSP90AA1*) (false discovery rate [FDR] <0.05) (Table 8). The most significant

differentially methylated CpGs were cg06798479 (*HSP90AA1;* coef=0.003 p-value=$2.16x10^{-5}$), cg05857597 (*NR3C1;* coef=-0.003; p-value=$7.8x10^{-5}$), and cg04281268 (*HSP90AA1;* coef=0.001; p-value=$8.59x10^{-5}$). One of the top-most significant CpGs associated with antenatal steroid exposure, cg20598211 annotated to *NR3C1* (coef=-0.044; p-value=$4.85x10^{-4}$), was also found to be nominally significant with mild BPD (coef=-0.025; p-value=0.019) and moderate BPD (coef=-0.038; p-value=0.005).

In the ELGAN cohort, 294 CpGs were annotated to HPA genes (Supplementary Table 27). As was found in the ECHO NOVI cohort, BPD severity was not significantly associated with the methylation of the 8 HPA genes at the FDR threshold, although several CpGs did yield raw p-values <0.05 (Supplementary Table 28). Similarly, antenatal steroid exposure was not significantly associated with HPA gene methylation, although 10 CpGs produced nominally significant raw p-values <0.05 (Supplementary Table 29).

3.3.3 Sex-Specific HPA DNAm

Studies have shown that antenatal steroids affect males and females differently [110]; therefore, we conducted a secondary analysis that was stratified by newborn sex. After sex-stratification, we observed sex-specific associations between antenatal steroid treatment and altered DNAm of HPA genes in the ECHO NOVI cohort. Males were observed to have differential methylation at 10 CpGs annotated to 4 HPA genes: *CRHR1*, *HSP90AA1*, *NR3C1*, and *NR3C2* (Table 9). Additionally, 6 CpGs were uniquely altered in males, with cg09775582 (*CRHR1*; coef = 0.005; p-value=$7.55x10^{-5}$) being the most significantly associated with antenatal steroid treatment. The other 4 CpGs were also observed in the unstratified model, with cg03396557 *(CRHR1*; coef = 0.006; p-value = $6.28x10^{-5}$) and cg04281268 (*HSP90AA1*; coef=0.002; p-value=$1.31x10^{-4}$) being the most strongly associated with antenatal steroid treatment in males. We conducted a Spearman correlation analysis of the sex-specific regression coefficients and observed that males and females had a positive correlation for the differential methylation that was associated with antenatal steroid treatment (rho=0.32; p-value=$4.12x10^{-9}$). Similarly, after comparing males and females in the ECHO NOVI cohort, regression coefficients were correlated for

moderate BPD (rho=0.21; p-value=$2x10^{-4}$) and severe BPD (rho=0.125; p-value=0.025). These findings suggest that the methylomes of both males and females tend to be impacted similarly by antenatal steroids and moderate to severe BPD, but males may experience greater differential methylation at specific CpGs.

In the ELGAN ECHO cohort, we observed significant sex-specific associations with severe BPD (FDR <5%) at two CpGs: cg26981809 (*FKBP5*, coef=-0.0200; p-value=$2.49x10^{-6}$) and cg02757179 (*POMC*; coef=0.0791; p-value=$3.43x10^{-5}$) (Table 10). However, the Spearman correlation analysis did not identify significant correlations between the sex-specific regression coefficients of BPD severity or antenatal steroids between males and females in this cohort.

3.4 DISCUSSION

Our study focused on BPD- and steroid-associated perturbations to the stress response system using an established polyepigenetic GC score and CpG-specific DNA methylation of 8 HPA genes. We examined whether the GC score was associated with BPD severity (in the ELGAN ECHO cohort), whether this score was responsive to BPD severity (in the ECHO NOVI cohort), and whether BPD and/or antenatal steroids were associated with the differential methylation of known HPA genes.

The polyepigenetic GC score was not associated with BPD severity, neither in the ELGAN ECHO cohort or the ECHO NOVI cohort. Tissue specificity and time of collection are likely crucial when utilizing the polyepigenetic GC score [19], as the differences between our samples and those used to train the GC score likely impacted its utility in the current study. The original GC score was calculated based on overlapping epigenetic responses to GC in human hippocampal progenitor cell lines and adult blood samples then it was tested in newborn cord blood, whereas our DNA samples were collected from buccal tissue and neonatal blood spot from infants born very and extremely preterm. Our minimally adjusted models appeared to reveal associations between increasing BPD severity and lower GC scores, which would have aligned with our expectations as higher GC exposure was associated with a lower GC score in the original investigations of the score. However, the adjustment for cellular proportions and age metrics substantially attenuated these initial findings. Thus, age and cell type may be strong confounders of these

relationships. These two factors should be considered when utilizing this polyepigenetic GC score. Buccal tissue primarily consists of epithelial cells with very small proportions of immune cells and fibroblasts, whereas blood consists of several cell types, including nucleated red blood cells (nRBCs) and white blood cells (WBCs). Thus, our samples in the ECHO NOVI cohort (mostly epithelial) and in the ELGAN ECHO cohort (nRBCs and WBCs) potentially provide different methylome profiles from those described by Suarez et al. [111] and Provençal et al. [19]. Lastly, the demographics of the participants in the Provençal et al. study were less racially diverse than our study, and thus the polyepigenetic GC score may not be as generalizable to our population [19].

Several studies have found that adversity during pregnancy is associated with altered HPA function in offspring [112; 113], and a combination of epidemiologic and experimental studies have recently shown that synthetic corticosteroids induce changes in DNAm [88] and has lasting effects [19]. Altered programming of the HPA axis has been implicated as a potential mechanism linking early life stress exposure to neurodevelopmental impairments. BPD is a substantial physiologic stressor in infants born very preterm, and its treatments could constitute additional stress. Contrary to our hypothesis, BPD severity did not show a strong impact on the methylome of 8 HPA genes. However, we did observe differential DNAm at two CpGs among males that developed severe BPD in the ELGAN cohort. As hypothesized, antenatal steroid exposure was associated with differential methylation at some CpG sites, but only in buccal tissue, suggesting that this association was tissue- or age-specific. These findings suggest that CpG methylation in HPA genes in some tissues are responsive to prenatal steroid exposure while other tissues are not. *FKBP5*, *CRHR1*, *HSP90AA1, NR3C1*, *NR3C2*, and *POMC* were the most significantly associated with antenatal steroid exposure in buccal tissue (FDR <0.05). *FKBP5* is a co-chaperone in the GC receptor complex and is often used as a proxy for HPA axis function [114]. *CRHR1* is the receptor of corticotropin-releasing hormone (CRH), which stimulates the secretion of adrenocorticotropin hormone (ACTH) when bound [115]. *HSP90AA1* is a molecular chaperone protein that regulates hormone signaling in response to stress [116]. *NR3C1* is a glucocorticoid receptor that when bound by cortisol negatively regulates CRH and ACTH production [117]. *NR3C2* is a mineralocorticoid receptor

that also influences cortisol secretion [118]. *POMC* precedes ACTH and endorphin production in the pituitary and is responsible for cortisol secretion from the adrenal cortex [119]. Alterations to these genes have been associated with panic disorder development, depression, and anxiety [104].

We also examined whether there were sex-specific associations between antenatal steroid exposure and HPA gene methylation, since prior work, including some from our group, has shown that males and females are impacted differently by early life exposures, including those introduced by the external environment, from maternal behavior, and from an adverse intrauterine environment [120-126]. CpGs within *CRHR1*, *HSP90AA1*, *NR3C1*, and *NR3C2* were significantly associated with antenatal steroid exposure in buccal tissue from the ECHO NOVI cohort; with 6 CpGs uniquely observed in the sex-stratified analysis for males only and 4 CpGs appearing in the full analysis. Thus, it is possible that some individual CpGs exhibit sex-specific responses to antenatal steroids, but overall, we observed relatively similar findings between males and females when we compared the regression coefficients from both models.

Although we did not observe significant association between BPD severity and the GC score or a strong significant association with DNAm of HPA genes, this does not rule out the possibility that BPD influences the methylome of other genes and biological systems. In previous work, our group observed that out of 4 neonatal morbidities (BPD, brain injury, retinopathy, and serious infection), BPD had the strongest impact on the neonatal methylome of preterm infants [11; 82]. Additionally, Martin et al. found that BPD was associated with more adverse neurodevelopmental and behavioral outcomes in NOVI infants at 2 years adjusted age [102]. While this project specifically focused on potential impacts within the HPA system, it is certainly possible that other genes or biological pathways are differentially methylated among infants diagnosed with BPD. Additionally, observational studies[127] and clinical trials[128], suggest that adrenal insufficiency, which might decrease cumulative secretion of adrenal glucocorticoids, is associated with a higher risk of BPD. It is plausible that the relationship between adrenal insufficiency and BPD may have confounded our analyses, if this condition limited endogenous GC secretion among our neonates with BPD. Such confounding would have biased our results towards the null. These

limitations highlight the complexity of studying this medical condition as well as antenatal treatments, and all of its associated factors. Despite these limitations, our findings provide valuable insights into some epigenetic features that appear to not be responsive to BPD and its treatments, while some individual genes within the HPA axis warrant further research. Future studies should use untargeted epigenome-wide approaches to explore which genes and systems are most differentially methylated after BPD diagnosis and treatment, and to identify the impact of BPD severity on methylation.

While our findings are interesting and align with our hypothesis that steroid treatment and BPD may impact the methylation of some HPA axis genes, these findings should be considered along with the limitations of this study [129]. First, a limitation of this study is that the steroid variables used in this analysis were limited to a binary "yes" or "no," which does not tell us about timing, dose, or steroid type. Therefore, our findings are hypothesis-generating to further explore the impact of steroids on HPA methylation in cohorts with richer data on prenatal steroid exposure. Another limitation of this study is that there were differences in the timing and tissue type for DNA methylation samples. Thus, we could not perform a pooled analysis that would have improved the statistical power to detect small effects. Additionally, because we studied epigenetic associations in buccal cells, our study is unable to explore the DNAm landscape of the lung tissue that is directly impacted by BPD, and where the epigenetic responses and indicators may be the strongest. . Overall, there was limited evidence of association between BPD severity and HPA axis methylation in infants born very preterm, and there is some evidence of HPA methylation differences with antenatal steroid exposure, but these findings require follow-up studies for confirmation and further exploration.

3.5 CONCLUSIONS

In this study, the GC score was not associated with BPD, and BPD severity did not strongly influence the methylome at CpGs within the stress response system. BPD and its treatments do not appear to be associated with the methylation of HPA genes in buccal tissue at NICU discharge; however, blood spot samples collected shortly after birth from males with severe BPD were found to have differential

methylation within two HPA genes: *FKBP5* and *POMC*. Antenatal steroids may have a lasting impact on the methylome of neonatal buccal cells at CpGs within *FKBP5*, *CRHR1*, *HSP90AA1, NR3C1*, *NR3C2*, and *POMC.* These findings were tissue-specific, as no differential methylation was observed in neonatal blood spots. Overall, the epigenome of the HPA system appears to be modestly impacted by antenatal steroid exposure and not BPD severity. Despite not observing a strong association between BPD severity and the methylation of HPA genes, methylation patterns of other biological pathways may still be impacted by BPD and its severity. Additionally, future studies should explore the effects of antenatal steroid exposure on the methylome in similar cohorts with richer prenatal and postnatal steroid exposure data and examine whether any identified epigenetic responses are risk factors for developmental outcomes.

3.6 AIM 2 TABLES

Table 7. Distribution of demographic characteristics, neonatal morbidities, and maternal/fetal characteristics of the ECHO NOVI and ELGAN ECHO cohort populations

Demographics by cohort	NOVI (N=538)	ELGAN (N=365)
Gestational age (weeks)	26.99 (1.92)	26.09 (1.92)
PMA* (weeks)	39.15 (3.37)	-
Infant Race (%)		
White	279 (51.9)	230 (63.4)
Black	121 (22.5)	97 (26.7)
Other Race	51 (9.5)	36 (9.9)
Asian	41 (7.6)	n<5
Hawaiian/Pacific Islander	38 (7.1)	-
Native American	n<5	n<5
Not Reported	n<5	n<5
Infant Ethnicity (% Hispanic)	114 (21.2)	35 (9.6)
Outborn (%)	112 (20.8)	365 (100)
Fetal growth restriction (%)	39 (7.2)	19 (5.2)
Male (%)	298 (55.4)	194 (53.2)
Maternal age (years)	29.1 (6.37)	29.48 (6.62)
Bronchopulmonary dysplasia (BPD) (%)		
No BPD	261 (48.5)	169 (46.3)
Mild BPD	130 (24.2)	142 (38.9)
Moderate BPD	122 (22.7)	31 (8.5)
Severe BPD	25 (4.6)	23 (6.3)
Education <high school/GED (%)	73 (13.6)	136 (37.3)
Lowest socioeconomic status (SES) (%)	43 (8.0)	43 (8.2)

*Post-menstrual age is defined as the time from conception to sample collection.
GED, general educational development.
Low SES was defined as Hollingshead Level V.

Table 8. CpGs significantly associated with antenatal steroid exposure in the ECHO NOVI cohort after adjustment for potential confounders*

CpG site	Gene	Coefficients	P-value	FDR <10%
cg06798479	*HSP90AA1*	0.0032	2.16E-05	0.0068
cg05857597	*NR3C2*	-0.0037	7.80E-05	0.0091
cg04281268	*HSP90AA1*	0.0018	8.59E-05	0.0091
cg03396557	*CRHR1*	0.0048	2.83E-04	0.0224
cg20598211	*NR3C1*	-0.0444	4.86E-04	0.0299
cg26495008	*FKBP5*	-0.0045	5.87E-04	0.0299
cg11845071	*FKBP5*	0.0026	6.61E-04	0.0299
cg19176661	*NR3C1*	-0.0181	8.97E-04	0.0324
cg24718866	*POMC*	0.0037	9.77E-04	0.0324
cg21410754	*HSP90AA1*	0.0141	1.02E-03	0.0324

*Coefficients are the slope of the relationship after adjustments for potential confounders: post-menstrual age (PMA), gestational age (GA) at birth, sex, study site, outborn status, fetal growth restriction, maternal age, maternal smoking during pregnancy, and cell type proportions (epithelial cells).
FDR, false discovery rate.

Table 9. Sex-specific associations between HPA axis CpGs and antenatal steroid exposure in ECHO NOVI males (N=298) after adjustment for potential confounders*

CpG Site	Gene	Coefficients	P-value	FDR < 10%
cg03396557	*CRHR1*	0.0062	6.28E05	0.0116
cg09775582	*CRHR1*	0.0054	7.55E-05	0.0116
cg04281268	*HSP90AA1*	0.0024	1.31E-04	0.0116
cg26720913	*NR3C1*	-0.0409	1.46E-04	0.0116
cg06798479	*HSP90AA1*	0.0037	2.47E-04	0.0157
cg06320026	*HSP90AA1*	0.0045	3.05E-04	0.0161
cg05857597	*NR3C2*	-0.0049	4.83E-04	0.0194
cg22491627	*NR3C2*	-0.0176	4.89E-04	0.0194
cg04414295	*HSP90AA1*	0.0059	5.75E-04	0.0202
cg24641427	*NR3C2*	-0.0308	1.09E-03	0.0344

*Coefficients are the slope of the relationship after adjustments for potential confounders: gestational age (GA) at birth, sex, study site, outborn status, fetal growth restriction, maternal age, maternal smoking during pregnancy, and cell type proportions (epithelial cells).
FDR, false discovery rate; HPA, hypothalamic-pituitary-adrenal.

Table 10: Sex-specific associations between HPA-axis CpGs and antenatal steroid exposure in ELGAN ECHO males (N=194) after adjustment for potential confounders*.

CpG site	Gene	Coefficients	P-value	FDR <10%
cg26981809	*FKBP5*	-0.0200	2.49E-06	0.0007
cg02757179	*POMC*	0.0791	3.43E-05	0.0050

*Coefficients are the slope of the relationship after adjustments for potential confounders: gestational age (GA) at birth, sex, study site, outborn status, fetal growth restriction, maternal age, maternal smoking during pregnancy, and cell type proportions (CD8 T cells, CD4 T cells, natural killer cells, B cells, monocytes, granulocytes, and nRBCs).
FDR, false discovery rate; HPA, hypothalamic-pituitary-adrenal.

3.7 AIM 2 FIGURES

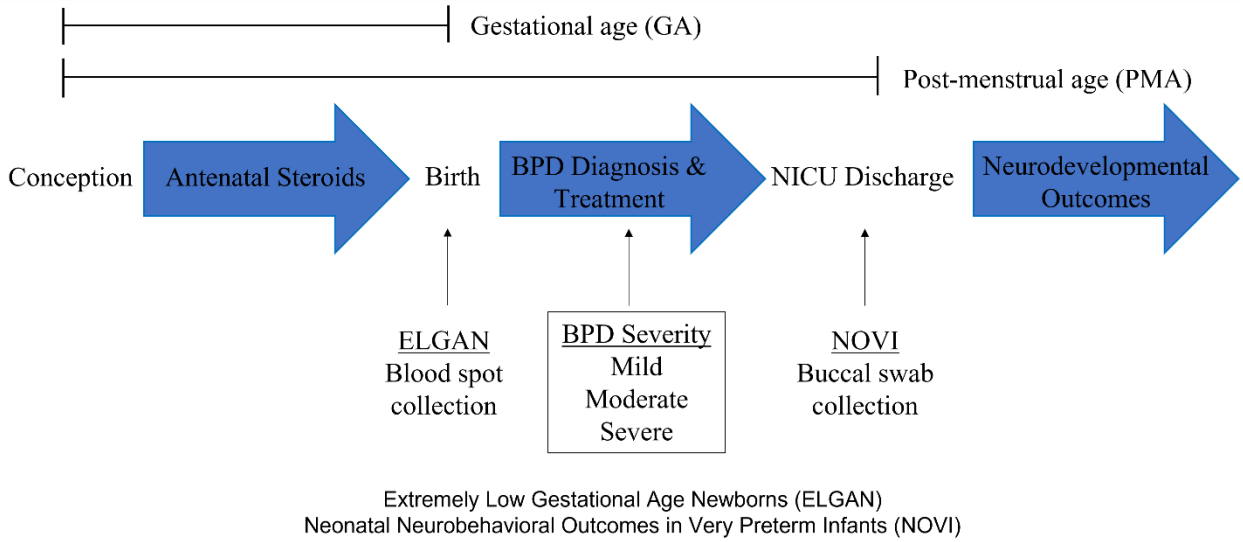


Figure 3. Timeline of ECHO cohort sample collection. ELGAN blood spot samples were collected at birth (gestational age) before BPD diagnosis and treatment. NOVI buccal swabs were collected at NICU discharge (post-menstrual age) after BPD diagnosis and treatment.

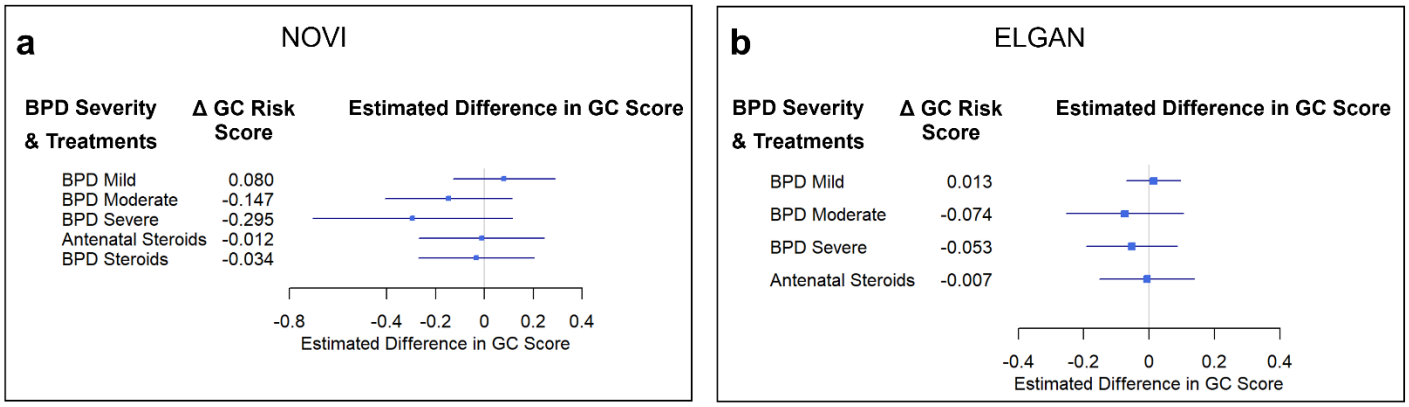
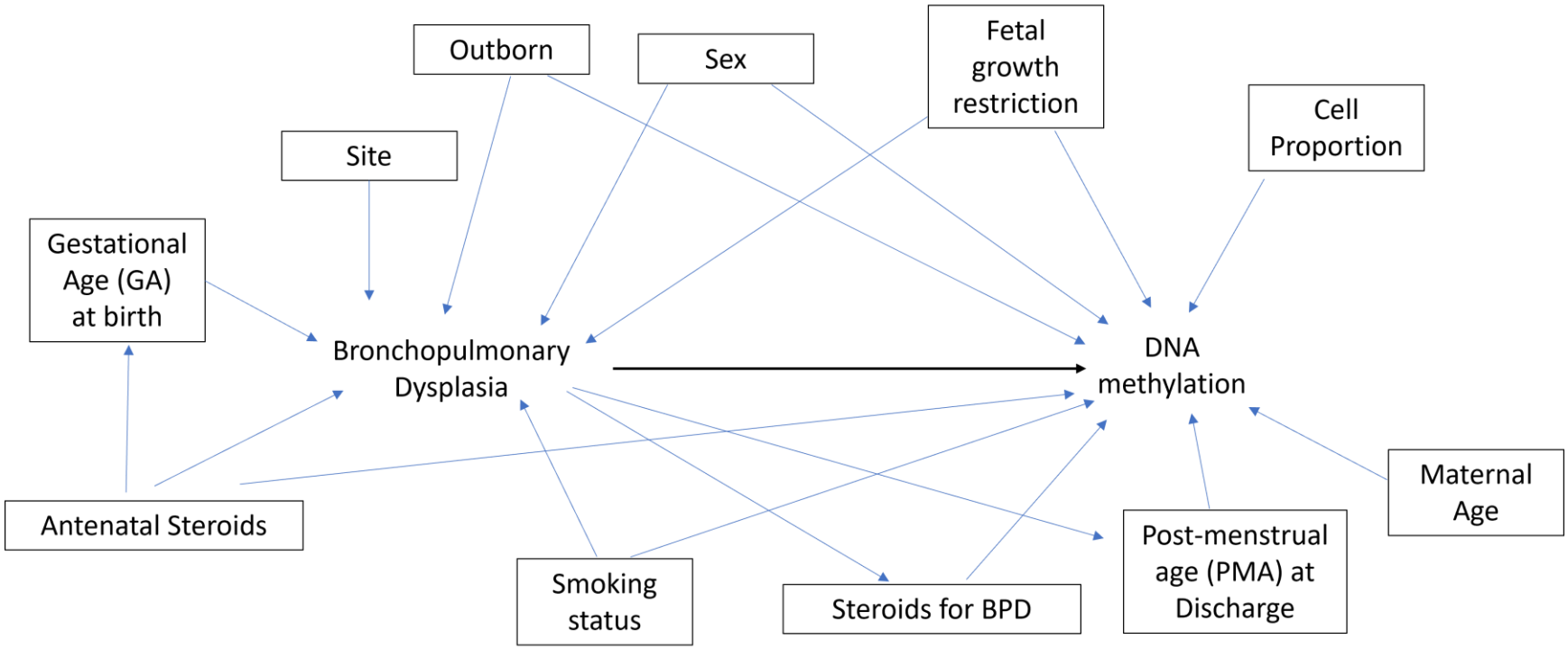


Figure 4. The established polyepigenetic GC score was not a significant measure of a) risk in the ECHO NOVI cohort or b) a predictor of BPD severity in the ELGAN ECHO cohort. No significant association was observed between GC score and BPD severity when adjusting for cell proportion, fetal growth restriction, maternal age, antenatal steroids, and/or steroids for BPD. However, GC score decreased as BPD severity increased, which is indicative of exposure to increased glucocorticoids (endogenous and exogeneous).

3.8 AIM 2 SUPPLEMENTARY TABLES AND FIGURES

All supplementary tables can be found in Supplementary_Tables_Chapter3.xlsx file.



Supplementary Figure S8: Directed acyclic graph to identify potential confounders

CHAPTER 4: DNA methylation as a mediator between neonatal morbidities and neurodevelopmental outcome in infants born very preterm

Kenyaita M. Hodge[1], Marie Camerota[2,3], Sarina Abrishamcar[4], Brian S. Carter[5], Amber A. Burt[1], Jennifer Check[6], Karen N. Conneely[7], Jennifer Helderman[6], Julie A. Hofheimer[8], Anke Hüls[1,4,9], Elisabeth C. McGowan[10], Charles R. Neal[11], Steven L. Pastyrnak[12], Lynne M. Smith[13], Sheri A. DellaGrotta[3], Lynne M. Dansereau[3], T. Michael O'Shea[8], Carmen J. Marsit[1,4], Barry M. Lester[2,3,10] and Todd M. Everson[1,4*]

[1]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA
[2]Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, RI
[3]Brown Center for the Study of Children at Risk, Women and Infants Hospital, Providence, RI
[4]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA
[5]Department of Pediatrics-Neonatology, Children's Mercy Hospital, Kansas City, MO
[6]Department of Pediatrics, Wake Forest School of Medicine, Winston-Salem, NC
[7]Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA
[8]Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC
[9]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA
[10]Department of Pediatrics, Warren Alpert Medical School of Brown University and Women and Infants Hospital, Providence, RI
[11]Department of Pediatrics, University of Hawaii John A. Burns School of Medicine, Honolulu, HI
[12]Department of Pediatrics, Spectrum Health-Helen Devos Hospital, Grand Rapids, MI
[13]Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA

Corresponding Author: Todd M. Everson, PhD, Associate Professor, Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322; **Phone**: 404-727-8136; **Email**: todd.m.everson@emory.edu

4.1 BACKGROUND

Infants born very preterm (< 30 weeks of gestation) are at heightened risk of acute medical complications and long-term developmental impairments. Bronchopulmonary dysplasia (BPD), the most common morbidity among infants born very preterm (VPT), has a significant impact on neurodevelopmental outcomes. Martin et al observed that VPT infants with moderate to severe BPD were less attentive and more likely to have impaired language skills and motor function at 24 months[102]. They also observed that infants with moderate to severe BPD were at increased risk of being diagnosed with cerebral palsy (CP) and it is known that CP is associated with motor function, speech, and language impairment[130; 131]. Similarly, Oluwole et al found that infants diagnosed with BPD were at increased odds of having impaired cognitive development, and motor function[132]. Other neonatal morbidities associated with impaired neurodevelopment were identified in a systematic review, concluding that preterm infants with serious brain injury (SBI), specifically intraventricular hemorrhage and white matter injury, exhibited two to five-fold increased risk of neurodevelopmental impairment and CP diagnosis compared to infants without brain injury[133]. The odds increased significantly with severity of brain injury. Similarly, a meta-analysis of retinopathy of prematurity (ROP) and neurodevelopmental outcomes found that preterm infants with ROP had an increased risk of cognitive impairment, CP, behavioral problems, and overall neurodevelopmental impairment[134]. Most recently, a meta-analysis of preterm infants with necrotizing enterocolitis (NEC), described an increased risk of attention deficits, CP, and cognitive, language, and motor delays[135]. In addition to these individual conditions infants born preterm, particularly those born VPT, are more likely to have multiple medical morbidities[11] and the cumulative number of morbidities has been associated with an incremental increase in the risk of developmental impairments[10].

The study of epigenetics via DNA methylation (DNAm) could provide insights into how neonatal morbidities are linked to subsequent developmental outcomes since epigenetic changes capture information about neonatal exposures, health and development[136]. Our group has previously shown that these same neonatal morbidities (BPD, ROP, SBI, and NEC, and sepsis) may disrupt underlying biological processes through altered epigenetic regulation, with BPD having the strongest effect on

neonatal DNAm[11]. We found that 125 CpGs were differentially methylated among infants born VPT, and that the magnitude of the response became stronger with increasing numbers of morbidities. In the same cohort, McGowan et al found that at 24 months of age, infants with multiple neonatal morbidities performed significantly lower on the Bayley Scales of Infant and Toddler Development (BSID-III), a widely used assessment that measures cognitive development, language skills, and motor function in children[137]. Yet, it is not clear whether the observed differential DNA methylation associated with neonatal morbidities plays a significant role in the neurodevelopmental outcomes of children born VPT.

Mediation analysis is a useful tool to identify mechanisms or interrelationships between exposures (i.e., neonatal morbidities) and outcomes (i.e., neurodevelopment) by disentangling indirect (mediating) effects from direct effects of the exposure on the outcome[138]. The role of DNAm as a possible mediator between various exposures and outcomes is still an emerging field of study; however, there are promising methods that can accommodate high-dimensional epigenomic data. High-dimensional mediation (HDM) is useful in the context of DNA methylation because it identifies the total mediating effect of a group of potential mediators (e.g., a group of CpG sites) in explaining the relationship between exposure and outcome. There are a multitude of HDM methods that can be utilized in epigenetics research, which are thoroughly reviewed and discussed by Clark-Boucher et al (2023). In this review, the high dimensional mediation analysis (HDMA) method by Goa et al (2019) was identified as one of the most robust techniques for identifying CpG-specific mediators while also estimating the combined indirect effects[139; 140]. The HDMA method is a penalized regression approach in which the outcome model is fit with a desparsified LASSO penalty, which first screens the potential mediators being tested.

In this current study, we hypothesize that neonatal DNA methylation acts as a mediator between neonatal morbidities and neurodevelopmental outcomes at 24 months of age among children born VPT. First, we conduct HDMA analysis on the 125 CpGs that we previously described, to identify a set of CpGs that may act as mediators. Next, we examined whether individual CpGs in the set identified via HDMA remained significant as individual mediators using causal mediation analysis.

4.2 METHODS

4.2.1 Study Population

The NOVI Study was conducted at 9 university affiliated NICUs in Providence, RI, Grand Rapids, MI, Kansas City, MO, Honolulu, HI, Winston Salem, NC, and Torrance and Long Beach, CA from April 2014 through June 2016. These NICUs were also Vermont Oxford Network (VON) participants. Eligibility was determined based on the following inclusion criteria: 1) birth at <30 weeks gestational age (GA); 2) parental ability to read and speak English or Spanish and 3) residence within 3 hours of the NICU and follow-up clinic. Exclusion criteria included maternal age <18 years, maternal cognitive impairment, maternal death, infants with major congenital anomalies, including central nervous system, cardiovascular, gastrointestinal, genitourinary, chromosomal, and nonspecific anomalies, and NICU death. Parents of eligible infants were invited to participate in the study when survival to discharge was determined to be likely by the attending neonatologist. Overall, 704 eligible infants were enrolled. Written informed consent was obtained for 93% of infants for epigenomic screening and buccal cells swabs were successfully obtained from 624 infants. Ultimately, 420 samples were included in this analysis from infants for whom DNAm data, information on neonatal morbidities, Bayley Scales of Infant and Toddler Development, 3rd Ed (BSID-III) scores at 24-months, and covariate data were available.

4.2.3 Measures

Maternal and Neonatal Data Collection: Maternal interviews were performed to collect demographic information such as age, race, ethnicity, partner status, and educational attainment, while the Hollingshead Index was used to assess socioeconomic status (SES) based on maternal report on education and occupation. We defined low SES as Hollingshead level V[141]. Infant medical records were reviewed to collect birthweight, gestational age, length of NICU stay, whether the newborn was outborn, and diagnoses of neonatal morbidities comprising a previously validated risk index[10]. This risk index includes a count of the following severe neonatal morbidities collected by standardized neonatal medical record reviews per VON criteria[142]: BPD, ROP, sepsis or NEC, and SBI[98] as defined in our prior publications utilizing this risk index[11; 143].

Neonatal DNA methylation: As previously described, DNAm profiles of buccal tissue collected at NICU discharge were measured using the Illumina MethylationEPIC array. Quality control and data processing excluded samples with >5% probes yielding detection p-values > 1.0E-5, mismatch between reported and predicted sex, or incomplete phenotype data[11; 81; 144]. Functional normalization and beta-mixture quantile (BMIQ) normalization were performed[40], and then we excluded probes located on the X and Y chromosomes, those with single nucleotide polymorphisms (SNP) within the binding region, those able to cross-hybridize to other regions of the genome[41], or those with low variability[42]. These analyses specifically use the 125 CpGs previously observed to be associated with neonatal morbidities[11]. Each CpG has a different mean, range, and variance of DNAm levels. To allow for effect sizes to be comparable across all CpGs sites, we standardized all CpGs by their interquartile range (IQR). This rescaling means that all model coefficients refer to the expected change in outcome associated with a change from the 25th to 75th percentile of DNAm.

Neurodevelopment assessments (24 months corrected age): Neurodevelopment was assessed at 24 months using the Bayley Scales of Infant and Toddler Development, 3rd Ed (BSID-III)[145] which was administered by a trained professional. Composite scores for cognitive, motor and language domains are normed with a population mean of 100 and standard deviation of 15.

The presence or absence of CP at 24 months was determined using the NICHD NRN Neurological Exam and CP severity was assessed using the Gross Motor Function Classification System[146]. CP is defined as abnormality of tone, reflexes, coordination, and movement, and a delay in motor milestones with a disorder of at least 1 motor function.

4.2.2 Statistical Analysis

Description of Confounding variables

We used directed acyclic graphs (DAGs) to identify potential confounders of exposure-mediator, mediator-outcome, and exposure-outcome associations As described in Everson et al[11] and Hodge et al[144], we adjusted for potential confounders of infant sex, GA, fetal growth restriction, study site, array sample plate, low socioeconomic status (SES), cellular heterogeneity[34] and outborn status[35] (Figure 5). Outborn

refers to infants that were born in a hospital without subspecialty providers of neonatal intensive care and were transferred, almost always on the day of birth, to a tertiary center for subspecialty care. Given that outborn status increases the risk of neonatal morbidities and neurobehavioral deficits and is associated with longer stays in the NICU[11; 36; 37], we controlled for this variable in all models. We adjusted for sex[38] given its association with DNAm and neurodevelopmental outcomes, and since males have higher risk of poorer outcomes compared to females[147]. Cellular heterogeneity, the proportion of epithelial, fibroblast, and immune cells, were estimated in our buccal samples using reference methylomes[34]. As in our previous studies, we controlled for the proportion of epithelial cells given that they make up the majority of our samples and are strongly inversely related to proportion of immune cells[11]. e adjusted for fetal growth restriction due its strong association with long-term negative neurodevelopmental outcomes[148-150]. Additionally, we performed sensitivity analysis to examine the effect of adjusting for diagnosis of cerebral palsy (CP) by 24 months of age, when BSID-III assessments were performed, since CP may provide a competing indirect pathway between neonatal morbidity and neurodevelopmental outcome.

4.2.3 High Dimensional Mediation Analysis

All statistical analyses were performed in R version 4.1.0. Our group previously found that 125 DNAm sites were associated with serious neonatal morbidities in the NOVI study[11]. We utilized the *hdmed*[139] R package to test whether a subset of these 125 CpG sites act as mediators between neonatal morbidities (exposure) and BSID-III performance at 24-months (outcome). The *hdmed* incorporates a suite of high dimensional mediation analysis methods within a single package. HDMA was chosen due to its robustness as well as its ability to identify CpG-specific mediated effects, while also estimating the total, direct, and indirect effects for a group of CpGs[139; 140]. In Model 1, the cumulative morbidity risk score[10] was defined as a 4-level numeric (0-3) variable which represented a count of the following neonatal morbidities: BPD, ROP, serious infection (defined as sepsis or NEC), and SBI. Only one neonate experienced all four morbidities and was grouped with the infants who experienced three morbidities. Similar to our previous work, Model 1 adjusted for infant sex, GA, fetal growth restriction, outborn

status, study site, sample plate, low SES, and cell type proportions. In Model S the sensitivity analyses, CP status at 24 months was included as a covariate due to strong relationships between CP diagnosis, some neonatal morbidities, and performance on the BSID-III. In all models the BSID-III outcomes were retained as continuous composite scores.

4.2.4 Causal Mediation

The CpGs that were identified as mediators via the HDM models were then analyzed individually using the generalized linear models R function '*glm*'. function. The total effect (TE) was first calculated using linear regression adjusting for the same confounders as the HDM models (BSID-III ~ multiple neonatal morbidities + DNAm at individual CpGs + confounders). Similarly, the indirect effect or mediated effect (IE) was calculated individually for each of the CpGs (DNAm ~ multiple neonatal morbidities + confounders). The *mediation*[151-154] R package was then used to calculate the proportion mediated (PM) and confidence intervals (CI) for each of the significant CpG sites using the results from the total and mediated linear regression models. Lastly, the correlations between the identified mediators were calculated and used to generate a heatmap using the *gplots*[155] R package function *heatmap.2*.

4.3 RESULTS

4.3.1 Characteristics of the study population

There were 420 infants in the NOVI study for whom DNAm data, BSID-III at 24-months, and covariate data were available. In this subset, GA at birth ranged from 22.0 to 29.9 weeks (mean = 27.0 weeks) and corrected GA at buccal swab collection ranged from 32.1 to 51.4 weeks (mean = 39.0 weeks; Table 11). There were 335 singletons and 85 (20.0%) multiple gestation infants. Of the 420 infants, 85 were outborn (20.0%). The majority of infants were diagnosed with at least one neonatal morbidity including BPD (51.0%), SBI (11.0%), NEC/SEP (19.3%), or severe ROP (6.2%). There were 95 (22.6%) infants with multiple neonatal morbidities. Approximately 16% of NOVI infants received a diagnosis of cerebral palsy (CP) by 24 months of age.

The average BSID-III scores in our sample were 91.9 (SD = 15.18) for cognitive, 88.61 (SD= 16.18) for language, and 92.08 (SD = 14.7) for motor. The number of infants with either one or multiple impairments by BSID-III at 24 months (score < 85) are illustrated in Figure 6. The most prevalent impairments were among language scores, with 161 (38%) infants scoring <85 (1SD below the standardization sample mean); 68 (16%) demonstrated only language impairments. A total of 103 (24.5%) infants demonstrated impaired cognitive scores, and 84 (20%) had impaired motor scores. Fifty-six (13%) infants had impaired cognitive, language, and motor scores. We also examined differences in BSID-III scores for those with CP in NOVI, and adjusted for CP in secondary analyses to identify whether or not CP was a competing mediating pathway. Infants with CP at 24 months had significantly lower cognitive (estimate = -11.9, $p = 3.3x10^{-9}$), language (estimate = -11.7, $p = 4.6x10^{-8}$) and motor scores (estimate = -17.0, $p < 2.2x10^{-16}$) compared to those without CP. Twenty-three (35%) of infants with CP had impaired cognitive, language, and motor scores.

4.3.2 Model 1: HDM of the Cumulative Morbidity Risk Score and Neonatal DNAm on BSID-III at 24 months

We first used HDMA via the *hdmed*[139] R package to identify sets of CpGs, among the 125 candidates, as potential mediators of BSID-III scores. We identified nine CpGs as potential mediators of the effect of neonatal morbidities on language scores, which had the strongest indirect effect (IE), or mediated effect, out of three BSID-III assessments. A one-IQR change in methylation at these nine CpGs was associated with a -2.0 point decrement in language score (IE) out of the -3.7 total effect (TE), which equates to a 0.54 or 54% proportion mediated (PM; Figure 7a). At four of eight CpGs the cumulative neonatal morbidities were associated with lower DNAm (cg09372094, cg11300147, cg12233751, and cg12457231) whereas the other five CpGs showed higher methylation (cg07846767, cg14783303, cg17689473, cg19192146, and cg23224501) with increasing morbidities. Four CpGs (cg04416326, cg07846767, cg14783303, and cg20749212) were identified as mediators of cognitive scores (IE = -1.6, DE = -2.9, TE = -4.5, PM = 0.36) and all five CpGs exhibited higher DNAm with increasing morbidity

(Figure 7b). Lastly, the association between cumulative morbidities and motor scores was found to be modestly mediated (IE = -0.6, DE= -5.6, TE = -6.2, PM = 0.10) by four CpGs with lower DNAm (cg07866909, cg09372094, cg12132046, and cg17481912) and two CpG with higher DNAm (cg14783303 and cg23224501) in association with increasing morbidities (Figure 7c). In total, 14 distinct CpGs were identified as mediators of at least one of the BSID-III scores (Table 12).

4.3.3 Sensitivity Analysis: The Effect of Cerebral Palsy Adjustment on HDM Analysis of Model 1

When adjusting for CP status at 24 months, cognitive and language were observed to have the strongest IE. Eight CpGs were identified as mediators of language (cg05191942, cg07846767, cg11300147, cg12233751, cg12457231, cg14783303, cg17689473 and cg23224501); seven of which were observed in Model 1 with very similar magnitudes of effect (IE = -1.7, DE = -0.7, TE = -2.4, PM = 0.71) (Figure 8a). One CpG (cg05191942) was newly identified as a potential mediator between cumulative morbidities and language scores when adjusting for CP status at 24 months. Four CpGs (cg04416326, cg07846767, cg14783303, and cg18787224) were identified as mediators of cognitive scores, with three CpGs having similar magnitudes of effect in Model 1 (IE = -2.5, DE = -1.01, TE = -3.52, PM = 0.71) (Figure 8b). One CpG (cg18787224; annotated to *LINC-PINT*) was newly identified as a potential mediator between cumulative morbidities and cognitive scores. Interestingly, adjusting for CP status at 24 months strengthened the proportion mediated by DNAm between cumulative morbidities and language scores (71% vs. 54%) as well as cognitive scores (73% vs. 36%). Lastly, DNAm was found to have a modest mediation effect for motor scores (IE = -0.7, DE = -3.6, TE = -4.3, PM = 0.16) with 7 CpGs acting as mediators: cg07866909, cg09372094, cg12132046, cg14783303, cg17481912, cg23224501, and cg23942665 (Figure 8c). One CpG (cg23942665; annotated to *GLRA1*) was a newly identified potential mediator between cumulative morbidities and motor scores. In total, 15 CpGs were identified as mediators of at least one of the BSID-III assessments when adjusting for CP status at 24 months, and 12 of these 15 were similarly found to be mediators in Model 1 (Table 13). Like Model 1,

cg07846767 *(FGFR1OP),* cg14783303 (*SMYD3*) and cg23224501 were found to be mediators of more than one BSID-III score.

Overall, 17 CpGs were identified as mediators between neonatal morbidities and at least one BSID-III assessment at 24 months (Table 14). Three CpGs were mediators between neonatal morbidities and cognitive scores across both HDM models: cg14783303 (*SMYD3*), cg07846767 (*FGFR1OP*), and cg04416326 (*BLNK*). Seven CpGs were mediators of language scores across both HDM models: cg07846767 (*FGFR1OP*), cg11300147 (*CD163L1*), cg12233751, cg12457231, cg14783303 (*SMYD3*), cg17689473 (*UNC5CL*) and cg23224501. Lastly, six CpGs were mediators between neonatal morbidities and motor scores in both HDM models: cg07866909, cg09372094 (*TMEM245*), cg12132046 (*ILKAP*), cg14783303 (*SMYD3*), cg17481912 (*LXN;GFM1*), cg23224501, and cg23942665 (*GLRA1*).

4.3.4 Individual Mediation Analysis of Identified HDM Mediators of BSID-III at 24 months

We next aimed to understand whether the CpGs that we identified as mediators via HDMA would also be considered mediators when analyzed one at a time via the causal mediation framework. Of the 17 CpGs identified as significant mediators via HDM approaches, eleven were found to be independent mediators of at least one BSID-III score (Figure 9). Four CpGs previously identified by HDMA were individual mediators of cognitive scores: cg14783303, cg07846767, cg04416326 and cg18787224. Five CpGs were individual mediators of language scores: cg14783303, cg07846767, cg17689473, cg09372094, and cg05191942. Finally, two CpGs previously identified by HDM were individual mediators of motor scores: cg14783303 and cg09372094. Overall, cg14783303 and cg09372094 were found to be significant individual mediators of performance on all three BSID-III scores at 24 months, while cg07846767 was a mediator of cognitive and language scores at 24 months. After sensitivity analysis adjusting for CP status at 24 months, only cg14783303 remained a mediator of all three BSID-III scores (Figure 10). Individual mediation analysis confirmed that CpGs found to be group mediators by HDM were indeed mediators of BSID-III at 24 months.

Lastly, we aimed to understand if the CpGs that we identified as mediators were conveying similar or unique information, so we explored the correlation between the 17 CpGs identified as mediators of BSID-III performance (Figure 11). Overall, while we did observe several significant positive and negative pairwise correlations among the 17 potential mediating CpGs, very few of these correlation coefficients were strong (> .40).

4.4 DISCUSSION

We identified 17 distinct CpGs that significantly mediate the effect of neonatal morbidities on neurodevelopmental outcome among 420 infants born VPT. Two CpGs (cg14783303 in *SMYD3* and cg09372094 in *TMEM245*) were found to be individual mediators of cognitive, language, and motor skills at 24 months, with cg14783303 (*SMYD3*) exhibiting the strongest and most consistent mediating effects across all analyses. The strongest mediated effect of DNAm was observed between the neonatal morbidities and language scores (54%) and cognitive scores (36%). Additionally, the mediating effect of neonatal DNAm was substantially increased when controlling for CP status at 24 months as a covariate (cognitive:73%; language: 71%) suggesting that CP acts may be an alternative mediator , but not one that explains the associations observed with DNA methylation.

In our group's prior work, we showed that the number of neonatal morbidities that infants experienced was significantly associated with their DNAm at 125 CpG sites[11], and that two or more neonatal morbidities was associated with greater risk for neurodevelopmental impairment as measured by the BSID-III at 24 months[137]. Now we also demonstrate that DNAm may partially explain some of the links between these morbidities and neurodevelopmental outcomes in late infancy and early childhood. While our study is the first to examine the specific question of whether DNAm may mediate the effects of neonatal morbidity on neurodevelopmental outcomes, others have successfully utilized these approaches in similar settings. For instance, Camerota et al used HDM and found that DNAm at 309 CpG sites mediated the association between cumulative perinatal risk and cognitive development at 36 months of age in the NOVI population sample[156]. Additionally, other cohorts have used HDM to characterize how

DNA methylation may mediate the effects of environmental exposures on neurodevelopment[157; 158]. Taken together, these collective findings demonstrate that mediation is a useful tool for identifying intermediaries that may be amenable to early targeted interventions, as well as illuminating some of the biological processes that link these early life exposures with subsequent health outcomes among young children.

4.4.1 CpG Based HDM

Our HDM models identified 17 differentially methylated CpGs as significant mediators between the morbidity risk scores and cognitive, language, and/or motor scores. There was an overlap in significant mediators between the two HDM models. Three CpGs were identified as mediators of more than one BSID assessment and thus may be broadly important for development: cg14783303 (*SMYD3*), cg07846767 (*FGFR1OP*), and cg09372094 (*TMEM245*). Some of these CpGs are within genes known to be associated with neurodevelopment and immune system function. *SMYD3* (SET and MYND Domain Containing 3) encodes a histone methyltransferase which functions with RNA polymerase II. In mouse models, inhibition of Smyd3 rescues NMDA receptor (primary excitatory neurotransmitter in humans) and cognitive deficits of Alzheimer's disease[159]. Epigenetic dysregulation of *SMYD3* and the upregulation of SMYD3 protein has also been found to increase risk of various cancers and it plays a role in skeletal muscle development, heart development, and bone and cartilage differentiation[160-162]. *FGFR1OP* (fibroblast growth factor receptor 1) is a centrosomal protein required for anchoring microtubules to subcellular structures, which is required for hippocampal growth by promoting the proliferation of hippocampal stem cells and progenitor cells during development[163]. Alterations in gene expression of *FGFR1OP* may also be associated with myeloproliferative disorder[164] and autoimmune disorders such as Crohn's disease[165], Graves' disease[166], and vitiligo[167; 168]. *TMEM245* encodes a transmembrane protein that is associated with cognitive function and Alzheimer's disease according to GWAS[169-173].

Individual mediation of the HDM identified mediators showed that nine out of 17 distinct CpGs were individual mediators of at least one BSID-III assessment. The sensitivity analysis adjusting for CP

status at 24 months for the individual mediation models provided similar magnitudes of the estimated IE and the direction of the estimate remained consistent (Figure 6). Since the proportion mediated increased after CP adjustment, this suggests that our observed epigenetic mediating pathways are independent of the CP-mediating pathway. Overall, our findings suggests that having multiple morbidities has a stronger effect on the prospective performance on the BSID-III which is consistent with the findings of McGowan et al[137].

4.4.2 Strengths and Limitations

A major strength of this study is the sample size available to address our research question. This study was well powered to analyze the mediation of 125 CpGs and to adjust for the necessary potential confounders. Additionally, this study is one of few that explores the mediation effect of DNAm between neonatal morbidities and prospective neurodevelopmental outcomes at 24 months in a cohort of infants born very preterm, who are at increased risk of developmental impairments. A unique feature of this study is that we utilized HDM along with individual mediation to identify and filter out non-mediators. Lastly, we had a priori knowledge of the CpGs significantly associated with neonatal morbidities giving confidence that the HDM models were appropriate. Future research within this longitudinal cohort will explore the changes in DNAm over time to identify persistent mediators of neurodevelopmental outcomes.

Limitations of this study include tissue-related questions, since DNAm patterns are known to be cell and tissue-specific and this study focused on DNAm data from buccal tissue. Target tissues for the brain, lung, gut, and blood-borne morbidities and outcomes that we studied are likely not buccal cells, but rather are those found in other critical organ systems. Importantly however, the collection of buccal tissue is a more suitable, non-invasive option for a cohort of infants born VPT. Additionally, our group previously demonstrated that neonatal morbidities were associated with neonatal buccal DNAm and with neurodevelopmental outcomes within this cohort[11; 137]. This study utilized those findings as a foundation for examining DNAm as a mediator of neurodevelopmental outcomes through neonatal morbidities.

While mediation can be employed as a causal inference method, our observational study design precludes assumptions about causality. The mediators that we identify may be on the pathway from exposures to outcomes, but are more likely serving as indicators of the internal biological processes that link neonatal morbidity to neurodevelopmental functioning over time. Additionally, while the HDMA approach is robust it does not provide confidence intervals for the TE and PM. However, pairing HDMA with a traditional mediation approach exploring the individual effect of each potential mediator, provided the individual confidence intervals necessary to support the HDMA findings. Lastly, while we carefully considered the role of confounding, the relationships between neonatal morbidity and prospective health are complex, and residual confounding may still play a role in some of the observed associations. Despite these limitations, our findings contribute important and novel information about the earliest biological intermediaries with potential to improve trajectories from neonatal risks to subsequent health and well-being among children born very preterm.

4.4.3 Conclusion

We found that neonatal DNAm may partially mediate the associations between neonatal morbidities and neurodevelopment outcomes of infants born VPT. These novel findings provide evidence that epigenetic markers of neonatal morbidity may also play an important role in neurodevelopment.

4.5 AIM 3 TABLES

Table 11: Distribution of demographic characteristics, neonatal morbidities and maternal/fetal characteristics of the NOVI study population.

Sample Characteristics (N= 420)	
	Overall
GA (weeks (SD))	27.07 (1.89)
PMA (weeks (SD))	39.05 (3.32)
Infant Race (%)	
Asian	31 (7.4)
Black	96 (22.9)
Hawaiian/Pacific Islander	22 (5.2)
White	230 (54.8)
Other	39 (9.0)
Hispanic (%)	88 (20.7)
Outborn (%)	85 (20.2)
Male (%)	229 (54.5)
Maternal age (years (SD))	29.23 (6.32)
Education < HS/GED (%)	56 (13.3)
Lowest SES (%)	34 (8.1)
Cerebral Palsy (%)	65 (15.8)
Neonatal Medical Morbidities (%)	
Bronchopulmonary Dysplasia	214 (50.7)
Serious Infection (NEC/SEP)	81 (19.3)
Serious Brain Injury (SBI)	46 (11.0)
Retinopathy (ROP)	26 (6.2)
Multiple Neonatal Medical Morbidities (%)	95 (22.6)
Fetal growth restriction (%)	35 (8.3)
Bayley Scores	
Cognitive Composite (SD)	91.90 (15.18)
Language Composite (SD)	88.61 (16.18)
Motor Composite (SD)	92.08 (14.70)
Proportion of Epithelial (SD)	99.0 (0.02)

Table 12: Significant mediators (CpG IDs) between the cumulative morbidity score of infants born very preterm (VPT) and their prospective BSID-III assessments at 24 months of age identified by HDMA while adjusting for sex, gestational age, fetal growth restriction, socioeconomic status, outborn, study site, sample plate and epithelial proportion (Model 1).

Cognitive Scores							
CpG ID	Gene	β_a	(p-value)	β_b	(p-value)	$\beta_a * \beta_b$	(p-value)
cg04416326	*BLNK*	0.202	7.99E-06	-3.043	1.54E-02	-0.616	1.54E-02
cg07846767	*FGFR1OP*	0.214	7.12E-06	-2.850	9.08E-03	-0.609	9.08E-03
cg14783303	*SMYD3*	0.248	1.96E-07	-3.303	2.65E-03	-0.820	2.65E-03
cg20749212		0.169	1.99E-04	2.810	3.06E-02	0.474	3.06E-02
Language Scores							
CpG ID	Gene	β_a	(p-value)	β_b	(p-value)	$\beta_a * \beta_b$	(p-value)
cg07846767	*FGFR1OP*	0.214	7.12E-06	-2.57	3.28E-02	-0.550	3.28E-02
cg09372094	*TMEM245*	-0.159	2.33E-04	2.82	4.52E-02	-0.449	4.52E-02
cg11300147	*CD163L1*	-0.111	1.37E-02	-2.83	4.54E-02	0.314	4.54E-02
cg12233751		-0.161	1.60E-02	-1.87	3.93E-02	0.302	3.93E-02
cg12457231		-0.116	4.08E-03	3.62	1.67E-02	-0.420	1.67E-02
cg14783303	*SMYD3*	0.248	1.96E-07	-3.11	1.03E-02	-0.772	1.03E-02
cg17689473	*UNC5CL*	0.170	1.08E-04	-2.80	2.06E-02	-0.476	2.06E-02
cg19192146	*MINK1*	0.183	2.28E-03	-1.80	4.70E-02	-0.328	4.70E-02
cg23224501		0.146	2.19E-03	2.60	4.73E-02	0.378	4.73E-02
Motor Scores							
CpG ID	Gene	β_a	(p-value)	β_b	(p-value)	$\beta_a * \beta_b$	(p-value)
cg07866909		-0.196	1.87E-04	-2.811	9.37E-03	0.552	9.37E-03
cg09372094	*TMEM245*	-0.159	2.33E-04	3.256	7.01E-03	-0.518	7.01E-03
cg12132046	*ILKAP*	-0.139	1.14E-03	2.714	1.42E-02	-0.378	1.42E-02
cg14783303	*SMYD3*	0.248	1.96E-07	-3.667	4.25E-04	-0.910	4.25E-04
cg17481912	*LXN*	-0.221	1.99E-04	-1.586	4.63E-02	0.351	4.63E-02
cg23224501		0.146	2.19E-03	2.261	4.43E-02	0.329	4.43E-02

βa = Effect estimates for increase in the morbidity risk score on DNAm levels; βb = Effect estimates for DNAm association with BSID-III assessment summary scores; $\beta a * \beta b$ = effect estimates for the mediated pathway from cumulative morbidity, to DNAm differences, to BSID-III scores at 24 months

Table 13: Significant mediators (CpG IDs) between the cumulative morbidity score of infants born very preterm (VPT) and their prospective BSID-III assessments at 24 months of age identified by HDMA sensitivity analysis while adjusting for sex, gestational age, fetal growth restriction, socioeconomic status, cerebral palsy status at 24 months, outborn, study site, sample plate and epithelial proportion (Model S).

Cognitive Development							
CpG ID	Gene	β_a	(p-value)	β_b	(p-value)	$\beta_a * \beta_b$	(p-value)
cg04416326	*BLNK*	0.196	3.56E-05	-2.927	1.78E-02	-0.573	1.78E-02
cg07846767	*FGFR1OP*	0.208	2.89E-05	-2.888	7.18E-03	-0.601	7.18E-03
cg14783303	*SMYD3*	0.256	2.97E-07	-3.200	3.07E-03	-0.819	3.07E-03
cg18787224	*LINC-PINT*	-0.153	4.42E-04	2.466	4.89E-02	-0.378	4.89E-02
Language Skills							
CpG ID	Gene	β_a	(p-value)	β_b	(p-value)	$\beta_a * \beta_b$	(p-value)
cg05191942		0.246	2.43E-06	-2.485	4.57E-02	-0.611	4.57E-02
cg07846767	*FGFR1OP*	0.208	2.89E-05	-2.621	2.77E-02	-0.546	2.77E-02
cg11300147	*CD163L1*	-0.108	2.22E-02	-2.880	3.97E-02	0.310	3.97E-02
cg12233751		-0.169	1.60E-02	-1.859	3.87E-02	0.314	3.87E-02
cg12457231		-0.107	1.14E-02	3.470	2.02E-02	-0.371	2.02E-02
cg14783303	*SMYD3*	0.256	2.97E-07	-3.020	1.17E-02	-0.773	1.17E-02
cg17689473	*UNC5CL*	0.159	5.50E-04	-2.732	2.20E-02	-0.433	2.20E-02
cg23224501		0.139	5.19E-03	2.632	4.20E-02	0.366	4.20E-02
Motor Function							
CpG ID	Gene	β_a	(p-value)	β_b	(p-value)	$\beta_a * \beta_b$	(p-value)
cg07866909		-0.211	1.27E-04	-2.662	1.01E-02	0.562	1.01E-02
cg09372094	*TMEM245*	-0.136	2.50E-03	2.681	2.04E-02	-0.365	2.04E-02
cg12132046	*ILKAP*	-0.145	1.19E-03	2.889	6.38E-03	-0.420	6.38E-03
cg14783303	*SMYD3*	0.256	2.97E-07	-3.460	5.12E-04	-0.886	5.12E-04
cg17481912	*LXN*	-0.225	3.04E-04	-1.570	3.92E-02	0.353	3.92E-02
cg23224501		0.139	5.19E-03	2.350	2.89E-02	0.327	2.89E-02
cg23942665	*GLRA1*	-0.141	6.11E-03	2.174	4.96E-02	-0.306	4.96E-02

βa = Effect estimates for increase in the morbidity risk score on DNAm levels; βb = Effect estimates for DNAm association with BSID-III assessment summary scores; $\beta a * \beta b$ = effect estimates for the mediated pathway from cumulative morbidity, to DNAm differences, to BSID-III scores at 24 months

Table 14: CpGs identified as mediators of at least one prospective BSID-III assessment at 24-months by HDMA Models 1 and the sensitivity analysis.

CpG ID	Chromosome	Position	Gene	Region	Model
cg04416326	chr10	97994880	*BLNK*	Body	1, S
cg05191942	chr1	22439320			S
cg07846767	chr6	167443722	*FGFR1OP*	Body	1, S
cg07866909	chr2	19849829			1, S
cg09372094	chr9	111880232	*TMEM245*	Body	1, S
cg11300147	chr12	7596703	*CD163L1*	1stExon	1, S
cg12132046	chr2	239111040	*ILKAP*	Body	1, S
cg12233751	chr8	39417353			1, S
cg12457231	chr10	85896102			1, S
cg14783303	chr1	246230074	*SMYD3*	Body	1, S
cg17481912	chr3	158390821	*LXN;GFM1*	TSS1500;Body	1, S
cg17689473	chr6	40995202	*UNC5CL*	3'UTR	1, S
cg18787224	chr7	130645353	*LINC-PINT*	Body	S
cg19192146	chr17	4740214	*MINK1*	Body	1
cg20749212	chr11	31262762	*DCDC1*		1
cg23224501	chr15	82406316			1, S
cg23942665	chr5	151250531	*GLRA1*	Body	S

4.6 AIM 3 FIGURES

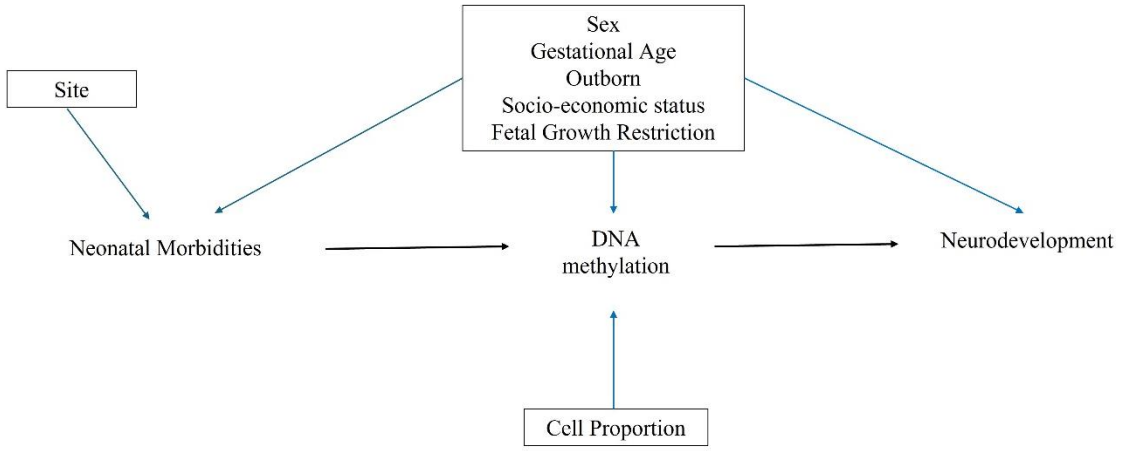


Figure 5: Directed acyclic graphs (DAGs) of neonatal morbidities (exposure) and neonatal DNAm (mediator), DNAm (mediator) and neurodevelopment (outcome), and neonatal morbidities (exposure) and neurodevelopment associations (outcome).

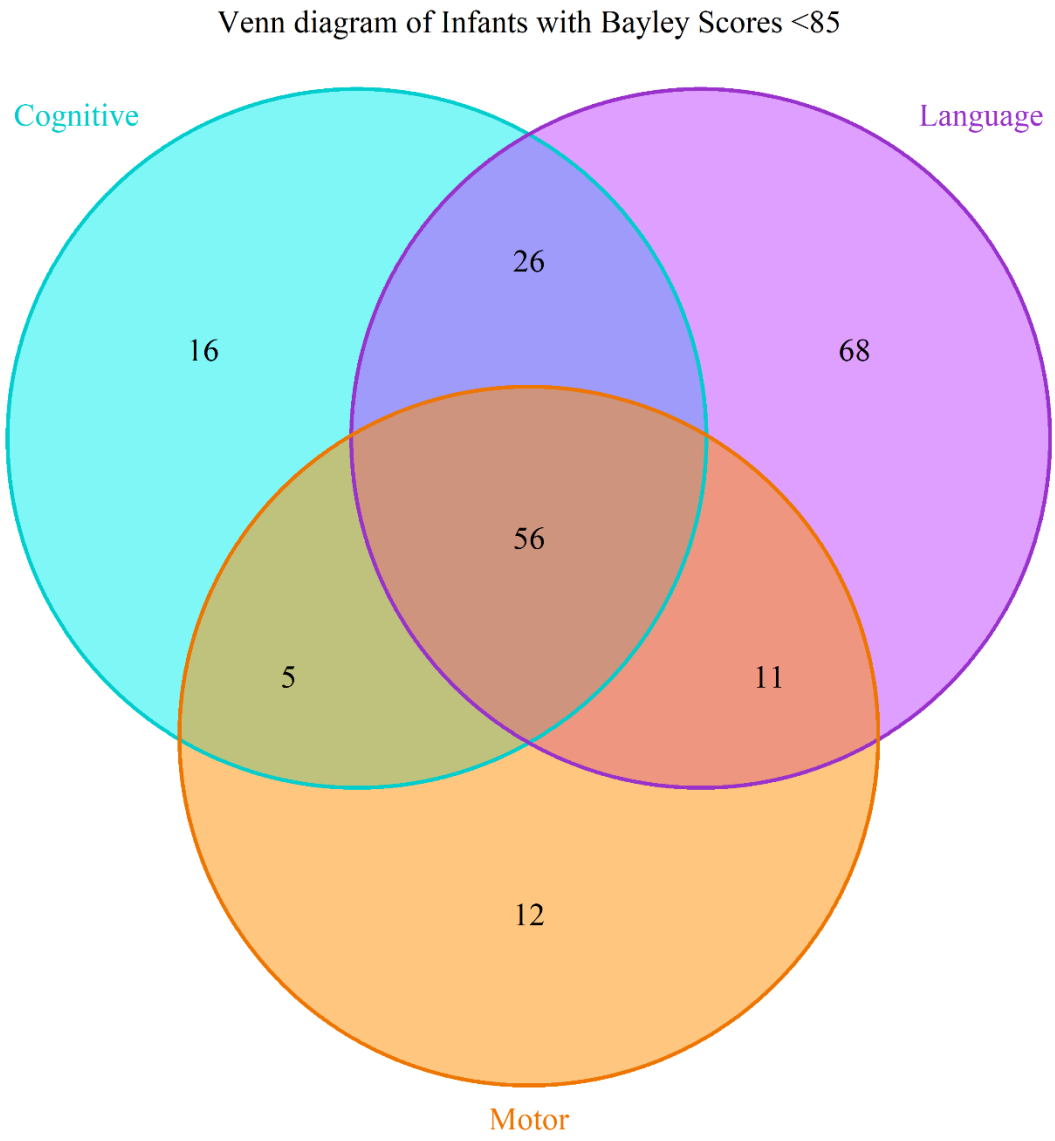


Figure 6: Venn diagram of impaired performance (score <85) on cognitive development, language skills, and/or motor function at 24 months.

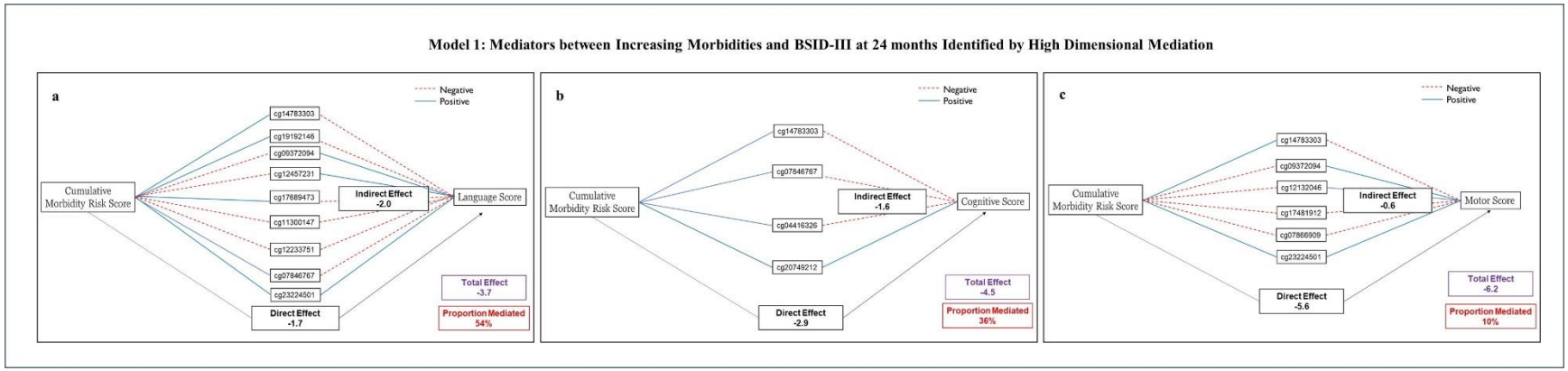


Figure 7: High dimensional mediation for the association between cumulative neonatal morbidity risk score (exposure), DNAm (mediator) and BSID-III performance at 24 months (outcomes) using HDMA via *hdmed* for CpGs previously identified as significantly associated with neonatal morbidities (Model 1). Mediators significantly associated with **a**) cognitive development scores, **b**) language skills scores, and **c**) motor function scores along with the indirect (mediated) effect (IE), direct effect (DE), total effect (TE), and proportion mediated (PM). Model 1 adjusts for sex, gestational age at birth, study site, sample plate, fetal growth restriction, socioeconomic status, outborn status, and cell proportions of epithelial cells.

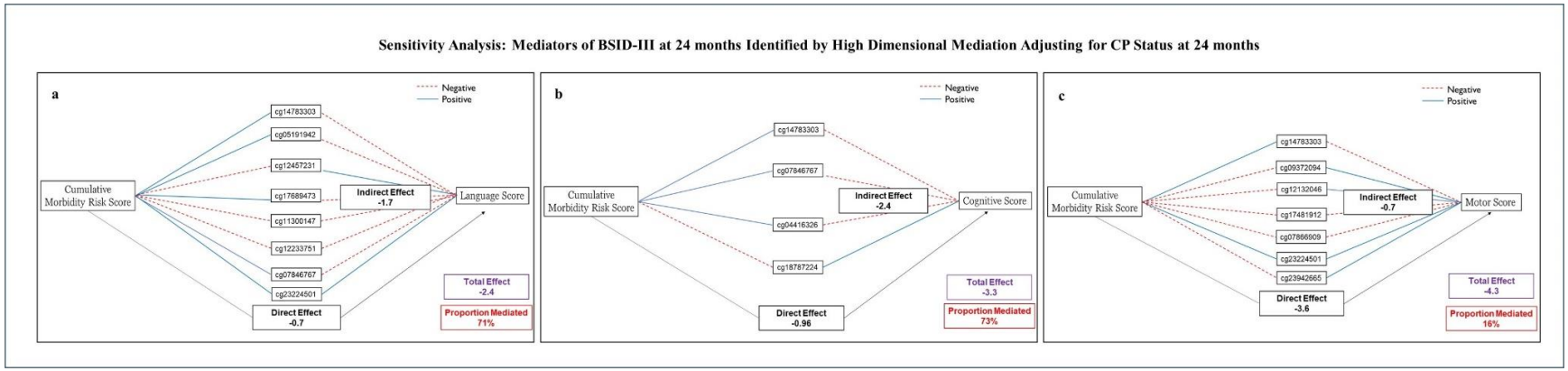


Figure 8: High dimensional mediation sensitivity analysis for the association between the cumulative neonatal morbidity risk score (exposure), DNAm (mediator) and BSID-III performance at 24 months (outcomes) when adjusting for the effect of cerebral palsy status at 24 months using HDMA via *hdmed* for CpGs previously identified as significantly associated with neonatal morbidities (Model S). Mediators significantly associated with a) cognitive development scores, b) language skills scores, and c) motor function scores along with the indirect (mediated) effect (IE), direct effect (DE), total effect (TE), and proportion mediated (PM). The solid blue line indicated a positive relationship, and the dotted red line indicates a negative association. Model S adjusts for cerebral palsy status at 24 months, sex, gestational age at birth, study site, sample plate, fetal growth restriction, socioeconomic status, outborn status, cerebral palsy status at 24 months, and cell proportions of epithelial cells.

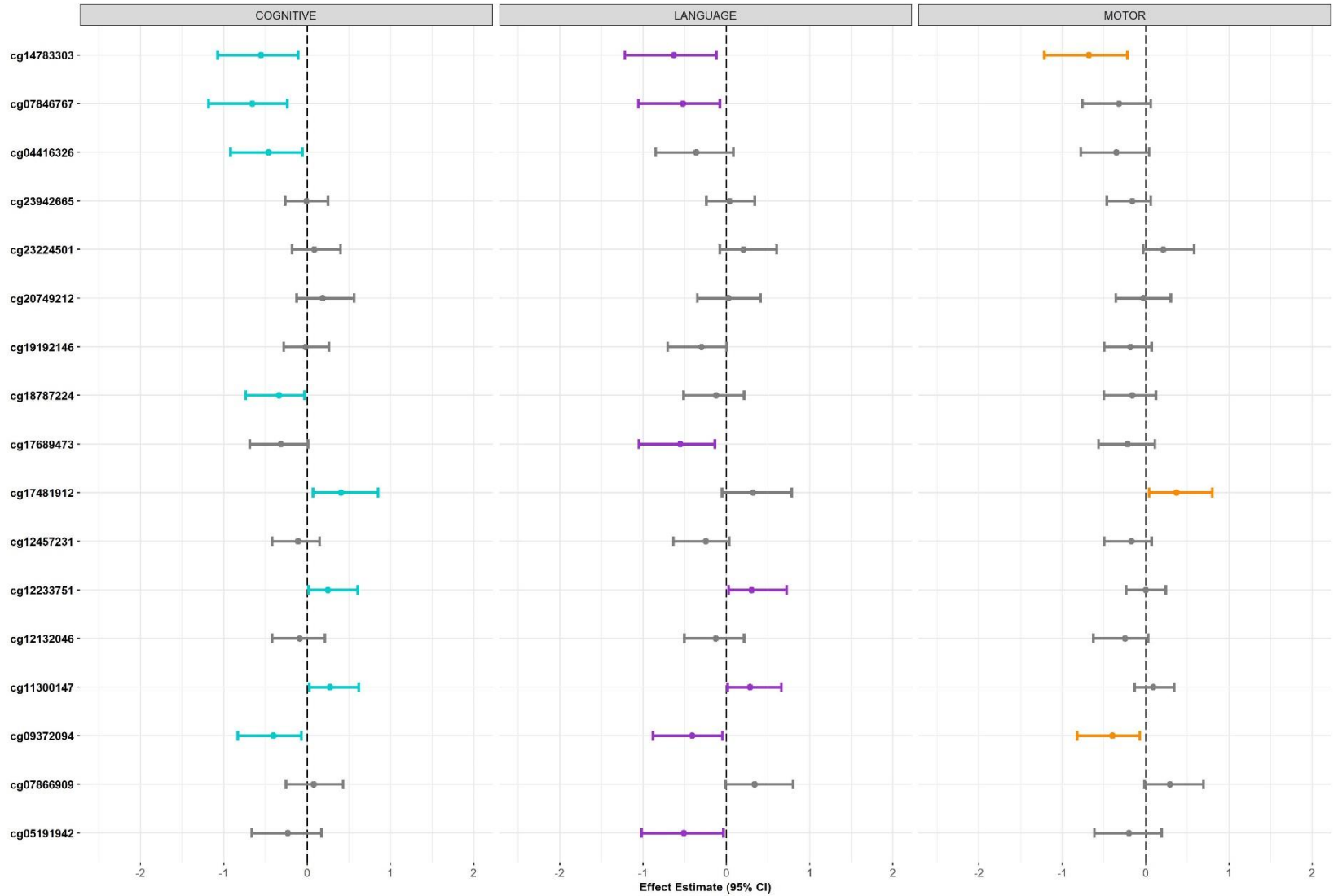


Figure 9: Individual mediation of the CpGs identified as mediators between cumulative neonatal morbidities and cognitive development, language skills, and motor function by HDM (Model 1). Bars in color denote significance threshold of $p < 0.05$. Model adjusts for sex, gestational age at birth, study site, fetal growth restriction, socioeconomic status, outborn status, and cell proportion of epithelial cells.

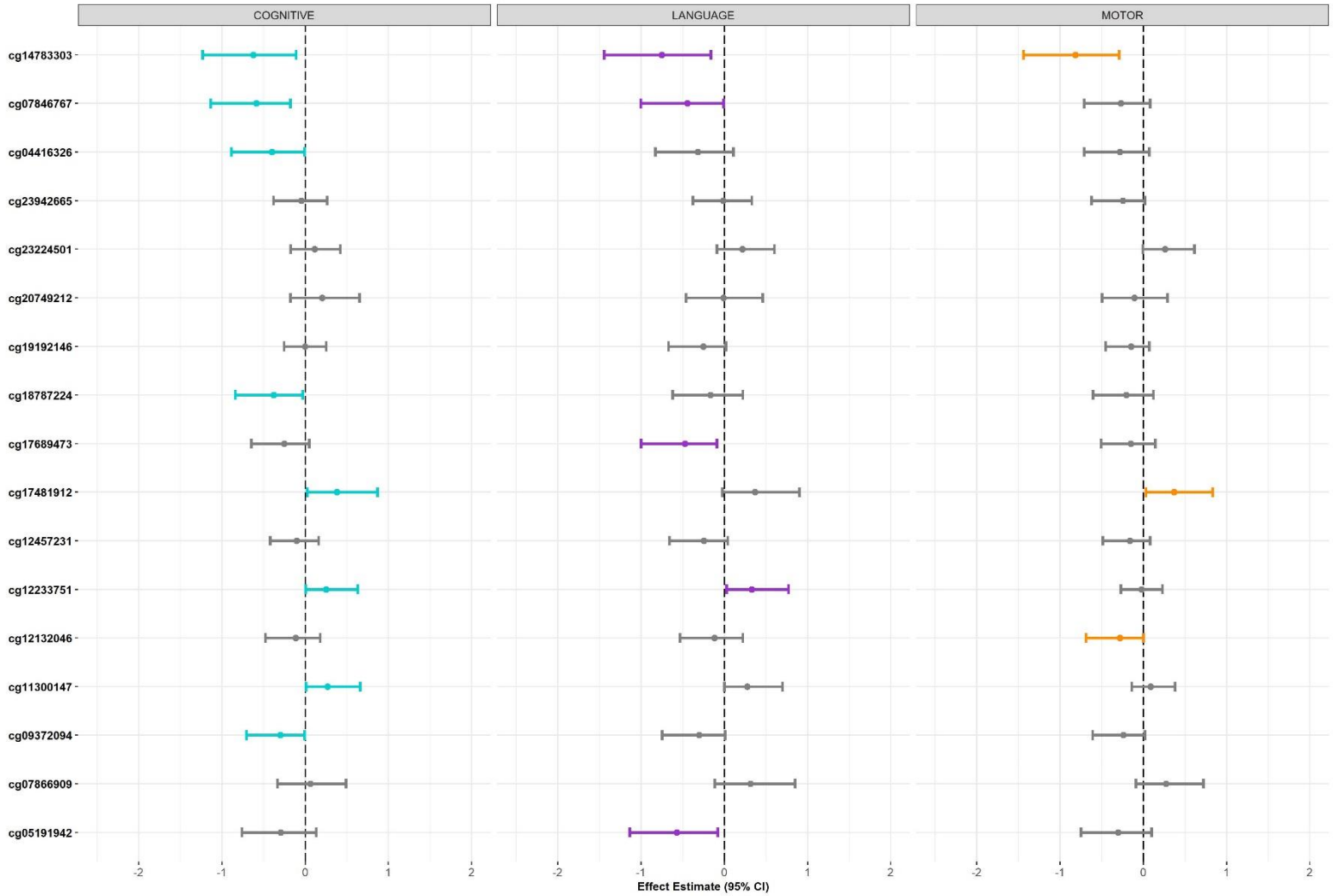


Figure 10: Individual mediation of the CpGs identified as mediators between cumulative neonatal morbidities and cognitive development, language skills, and motor function in the HDM sensitivity analysis adjusting for CP status at 24 months. Bars in color denote significance threshold of $p < 0.05$. Model S adjusts for sex, gestational age at birth, study site, fetal growth restriction, socioeconomic status, outborn status, cerebral palsy status at 24 months, and cell proportion of epithelial cells.

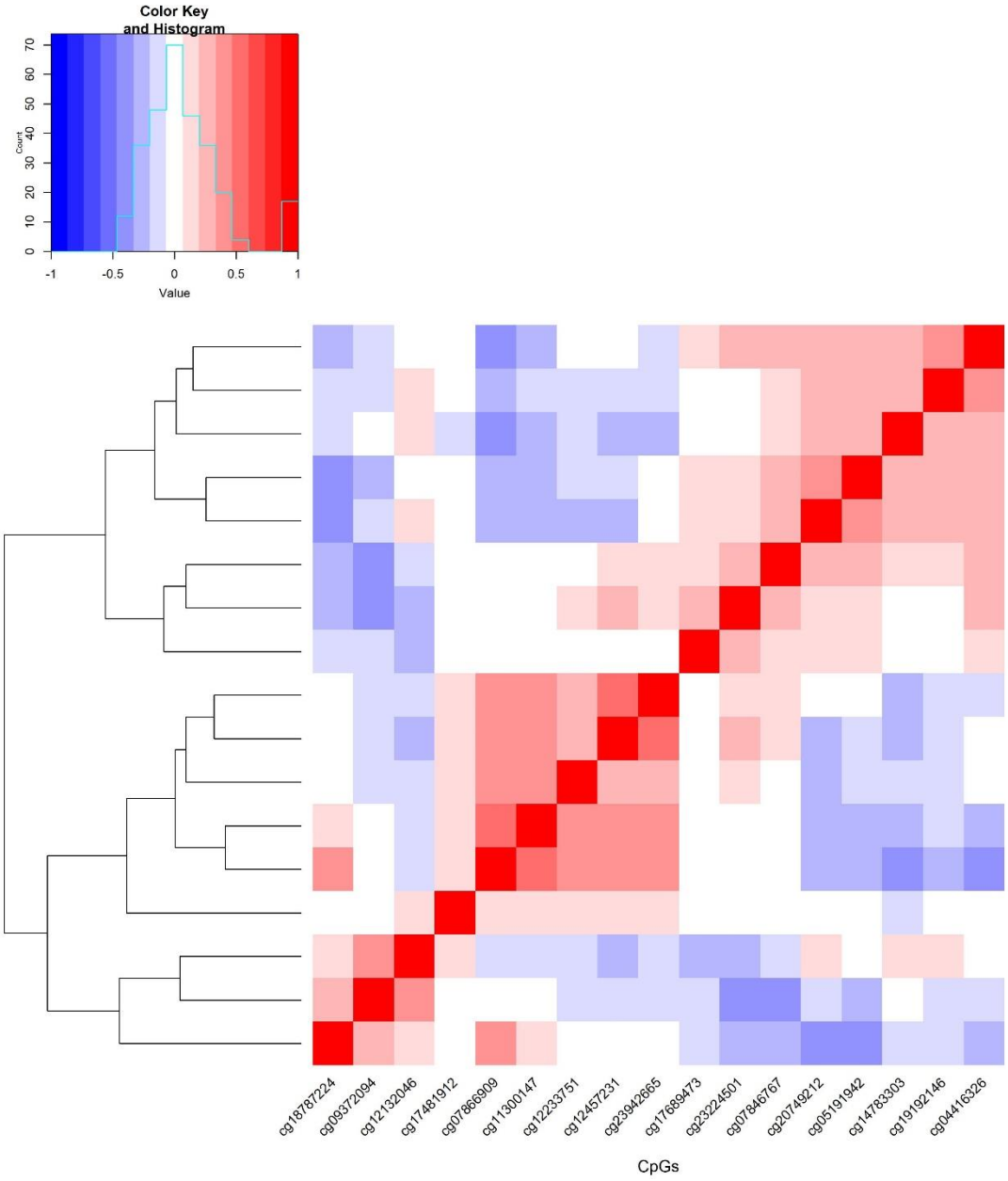


Figure 11: Heatmap of the DNAm of the 17 CpGs identified as mediators of cognitive development, language skills, and motor function at 24 months from both HDM models. Red indicates a positive correlation, and blue indicates a negative correlation. The dendrogram denotes the closely correlated DNAm of the CpGs.

CHAPTER 5: Summary of Aims

Epigenetics, particularly DNAm, is increasingly recognized as a crucial factor in understanding the biological mechanisms underlying preterm birth and its long-term health consequences. Research has demonstrated that preterm birth and its morbidities can lead to lasting changes in the epigenome, which may influence developmental outcomes and overall health. This dissertation focused on investigating the epigenetic changes associated with neonatal age and neonatal morbidities, and how these morbidities alter DNAm that potentially mediates neurodevelopmental outcomes in infants born VPT.

Epigenetic Association of Neonatal Age

Aim 1 identified significant age-associated differences in the neonatal buccal tissue epigenome of VPT infants. Both GA and PMA were associated with widespread differential methylation across the genome. GA-associated DNAm changes reflect the impact of the timing of birth, while PMA-associated changes indicate how the methylome evolves with development from conception to NICU discharge. Key genes such as *IRX4*, *SLC7A5*, and *PDE9A*, which are involved in cardiac and neural development, showed significant differential methylation associated with GA. These findings suggest that early life aging in preterm born neonates exhibits unique and evolving epigenetic signatures that may be related to critical developmental process, including cardiac and neurodevelopmental health.

Impact of Bronchopulmonary Dysplasia (BPD) and Steroid Exposure

Aim 2 examined the impact of BPD severity and antenatal steroid exposure on the epigenome of HPA genes. We also utilized a polyepigenetic GC score, to comprehensively examine potential impacts on the epigenetic machinery of the stress response system. While the GC score was not significantly associated with BPD severity, specific differential DNAm patterns were observed in HPA axis genes such as *FKBP5*, *CRHR1, HSP90AA1*, *NR3C1*, *NR3C2*, and *POMC* in response to antenatal steroids. These changes were particularly evident in buccal tissue, highlighting tissue-specific epigenetic responses. Additionally, sex-specific associations were found, indicating that males and females may have different

epigenetic responses to prenatal steroid exposure. The lack of significant association between the GC score and BPD severity suggests that the timing of sample collection and tissue specificity are important considerations in epigenetic research. However, the observed DNAm changes in response to antenatal steroids suggest lasting epigenetic effects of prenatal exposures, which could influence long-term health outcomes.

Epigenetic Mediation of Neurodevelopmental Outcomes

Aim 3 focused on 125 CpG that were previously identified as associated with neonatal morbidities, to test whether these may act as mediators on the pathway towards poorer neurodevelopment. We identified 17 distinct CpGs that partially mediated the effect of neonatal morbidities on neurodevelopmental outcomes at 24 months, as measured by the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III). The strongest mediated effects were observed in language skills and cognitive development, and we found that three CpGs mediated performance across multiple BSID-III assessments: cg14783303 (*SMYD3*), cg09372094 (*TMEM245*), and cg07846767 (*FGFR1OP*). This study is the first to examine neonatal DNAm as a mediator between neonatal morbidities and neurodevelopmental outcomes in VPT infants, providing evidence that early epigenetic responses may play a crucial role in shaping neurodevelopment. These findings highlight the potential of DNAm as a predictor of neurodevelopmental outcomes and highlights the importance of early interventions targeting epigenetic changes to improve long-term health and developmental trajectories in preterm infants.

Strengths and Limitations

A key strength across all aims is the use of a large sample size and a well-characterized cohort of very preterm infants, which significantly strengthens the study's power to detect meaningful associations and mediating effects in an at-risk population. The availability of detailed phenotypic and epigenetic data enhanced the study's ability to adjust for potential confounders, ensuring the robustness and reliability of the findings. This rich dataset allowed for comprehensive analyses, providing valuable insights into the

biological mechanisms underlying preterm birth outcomes. The focus on clinically relevant outcomes, such as neurodevelopmental performance and the impact of neonatal morbidities, directly addresses important health issues faced by preterm infants, increasing the potential for translating research findings into clinical practice and interventions. Additionally, the methodological approaches provide robust and reliable results, enhancing the study's validity.

However, there are limitations to consider. DNAm patterns are known to be tissue-specific, and this study focused on DNAm data from buccal tissue. While buccal tissue is a suitable and non-invasive option for studying infants, it may not fully capture the epigenetic changes occurring in the brain or lungs, which are the target tissues of the morbidities and outcomes that we studied. Despite this limitation, previous research has demonstrated associations between neonatal morbidities, buccal DNAm, and neurodevelopmental outcomes, supporting the relevance of our findings. The exploratory nature of the epigenetic age EWAS likely resulted in some false positive findings, and we could not identify a suitable external publicly available data set to try and replicate our results. However, we did compare our findings to several published studies of age-related changes in childhood and show that there is substantial overlap with our results. Our results provide a vast resource for future hypothesis driven research and highlight the importance of appropriately accounting for age-differences when studying neonatal epigenetics. In our study of BPD, most of our findings were related to antenatal steroids rather than BPD itself. The binary exposure variables, antenatal steroid treatment (yes/no), provide some evidence of association with DNAm of HPA axis genes but more descriptive variables including steroid type, dosage, and duration are better suited to characterize such relationships in future studies. Lastly, we performed mediation analysis which is a causal inference approach. While our study did robustly account for confounding, and our design establishes the appropriate temporal relationships between the exposure, the mediators, and the outcomes, we cannot claim causal relationships in our interpretations. Despite our best efforts, residual confounding may still play a role in the observed relationships, and our focus on buccal cell DNAm reduces the likelihood that these molecular responses or mechanistic links between the exposures and outcomes. However, this study provides compelling evidence that some of the DNAm responses to

neonatal morbidities are in turn predictive of neurodevelopmental outcomes in VPT infants. These CpGs that we identified as mediators may behave as surrogate markers of what is happening in other important tissues, but we could not examine that here. Additionally, the majority of preterm infants that experience these morbidities do not have neurodevelopmental impairments. Our findings may help to distinguish which infants that do experience these morbidities are at heightened risk. The findings emphasize the need for further research to validate these results and explore the long-term impact of early epigenetic changes on health and development.

Conclusions

The findings from this dissertation highlight the profound relationships between epigenetic modifications and the health and development of infants born VPT. By investigating the age-associated epigenetic changes, the effects of antenatal steroids and BPD on the epigenome, and the role of DNAm as a mediator of neurodevelopmental outcomes, we have illuminated critical pathways that could influence long-term health outcomes. Future research should leverage the longitudinal data provided by the NOVI study and similar cohorts with other tissue types to explore the persistence of these epigenetic changes over time and their implications for health and development outcomes throughout childhood. Together, these aims not only enhance our understanding of the biological mechanisms underlying preterm birth outcomes but also pave the way for developing targeted interventions to improve the health and neurodevelopment of this vulnerable population.

REFERENCES

1. Cruickshank, M.N., Oshlack, A., Theda, C., Davis, P.G., Martino, D., Sheehan, P., Dai, Y., Saffery, R., Doyle, L.W., and Craig, J.M. (2013). Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. Genome Medicine 5, 96.
2. Kashima, K., Kawai, T., Nishimura, R., Shiwa, Y., Urayama, K.Y., Kamura, H., Takeda, K., Aoto, S., Ito, A., Matsubara, K., et al. (2021). Identification of epigenetic memory candidates associated with gestational age at birth through analysis of methylome and transcriptional data. Scientific Reports 11, 3381.
3. Malin, K.J., Gondwe, K.W., Fial, A.V., Moore, R., Conley, Y., White-Traut, R., and Griffith, T. (2023). Scoping Review of Early Toxic Stress and Epigenetic Alterations in the Neonatal Intensive Care Unit. Nurs Res 72, 218-228.
4. Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 13, 97-109.
5. Monk, C., Spicer, J., and Champagne, F.A. (2012). Linking prenatal maternal adversity to developmental outcomes in infants: the role of epigenetic pathways. Dev Psychopathol 24, 1361-1376.
6. Moore, L.D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. Neuropsychopharmacology 38, 23-38.
7. Alisch, R.S., Barwick, B.G., Chopra, P., Myrick, L.K., Satten, G.A., Conneely, K.N., and Warren, S.T. (2012). Age-associated DNA methylation in pediatric populations. Genome Res 22, 623-632.
8. Thébaud, B., Goss, K.N., Laughon, M., Whitsett, J.A., Abman, S.H., Steinhorn, R.H., Aschner, J.L., Davis, P.G., McGrath-Morrow, S.A., and Soll, R.F. (2019). Bronchopulmonary dysplasia. Nature reviews Disease primers 5, 1-23.
9. Wang, X., and Shi, Y. (2023). An updated review on systemic glucocorticoids in the prevention and treatment of bronchopulmonary dysplasia. Pediatric Discovery 1, e6.
10. Bassler, D., Stoll, B.J., Schmidt, B., Asztalos, E.V., Roberts, R.S., Robertson, C.M., Sauve, R.S., and Trial of Indomethacin Prophylaxis in Preterms, I. (2009). Using a count of neonatal morbidities to predict poor outcome in extremely low birth weight infants: added role of neonatal infection. Pediatrics 123, 313-318.
11. Everson, T.M., O'Shea, T.M., Burt, A., Hermetz, K., Carter, B.S., Helderman, J., Hofheimer, J.A., McGowan, E.C., Neal, C.R., Pastyrnak, S.L., et al. (2020). Serious neonatal morbidities are associated with differences in DNA methylation among very preterm infants. Clinical Epigenetics 12, 151.
12. Stiles, J., and Jernigan, T.L. (2010). The basics of brain development. Neuropsychol Rev 20, 327-348.
13. Gilmore, J.H., Knickmeyer, R.C., and Gao, W. (2018). Imaging structural and functional brain development in early childhood. Nature Reviews Neuroscience 19, 123-137.
14. Smith, L.J., McKay, K.O., van Asperen, P.P., Selvadurai, H., and Fitzgerald, D.A. (2010). Normal Development of the Lung and Premature Birth. Paediatric Respiratory Reviews 11, 135-142.
15. Hernandez, D.G., Nalls, M.A., Gibbs, J.R., Arepalli, S., van der Brug, M., Chong, S., Moore, M., Longo, D.L., Cookson, M.R., and Traynor, B.J. (2011). Distinct DNA methylation changes highly correlated with chronological age in the human brain. Human molecular genetics 20, 1164-1172.
16. Wikenius, E., Moe, V., Smith, L., Heiervang, E.R., and Berglund, A. (2019). DNA methylation changes in infants between 6 and 52 weeks. Scientific reports 9, 1-12.
17. Cao-Lei, L., Laplante, D.P., and King, S. (2016). Prenatal Maternal Stress and Epigenetics: Review of the Human Research. Current Molecular Biology Reports 2, 16-25.

18. Parets, S.E., Conneely, K.N., Kilaru, V., Fortunato, S.J., Syed, T.A., Saade, G., Smith, A.K., and Menon, R. (2013). Fetal DNA Methylation Associates with Early Spontaneous Preterm Birth and Gestational Age. PLoS One 8, e67489.
19. Provençal, N., Arloth, J., Cattaneo, A., Anacker, C., Cattane, N., Wiechmann, T., Röh, S., Ködel, M., Klengel, T., Czamara, D., et al. (2020). Glucocorticoid exposure during hippocampal neurogenesis primes future stress response by inducing changes in DNA methylation. Proceedings of the National Academy of Sciences 117, 23280.
20. Breton, C.V., Marsit, C.J., Faustman, E., Nadeau, K., Goodrich, J.M., Dolinoy, D.C., Herbstman, J., Holland, N., LaSalle, J.M., Schmidt, R., et al. (2017). Small-Magnitude Effect Sizes in Epigenetic End Points are Important in Children's Environmental Health Studies: The Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group. Environ Health Perspect 125, 511-526.
21. McEwen Lisa, M., O'Donnell Kieran, J., McGill Megan, G., Edgar Rachel, D., Jones Meaghan, J., MacIsaac Julia, L., Lin David Tse, S., Ramadori, K., Morin, A., Gladish, N., et al. (2020). The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. Proceedings of the National Academy of Sciences 117, 23329-23335.
22. Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome biology 14, 1-20.
23. Graw, S., Camerota, M., Carter, B.S., Helderman, J., Hofheimer, J.A., McGowan, E.C., Neal, C.R., Pastyrnak, S.L., Smith, L.M., and DellaGrotta, S.A. (2021). NEOage clocks-Epigenetic clocks to estimate post-menstrual and postnatal age in preterm infants. Aging (Albany NY) 13, 23527.
24. Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., Lawn, J.E., Cousens, S., Mathers, C., and Black, R.E. (2016). Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet 388, 3027-3035.
25. UNICEF, W., World Bank, UN-DESA Population Division. (2018). Levels and trends in child mortality: report 2017. In. (
26. Kramer, B.W., Niklas, V., and Abman, S. (2022). Bronchopulmonary Dysplasia and Impaired Neurodevelopment-What May Be the Missing Link? Am J Perinatol 39, S14-S17.
27. Majnemer, A., Riley, P., Shevell, M., Birnbaum, R., Greenstone, H., and Coates, A.L. (2000). Severe bronchopulmonary dysplasia increases risk for later neurological and motor sequelae in preterm survivors. Dev Med Child Neurol 42, 53-60.
28. Piyasena, C., Cartier, J., Provençal, N., Wiechmann, T., Khulan, B., Sunderesan, R., Menon, G., Seckl, J.R., Reynolds, R.M., Binder, E.B., et al. (2016). Dynamic Changes in DNA Methylation Occur during the First Year of Life in Preterm Infants. Frontiers in Endocrinology 7, 158.
29. Schuster, J., Uzun, A., Stablia, J., Schorl, C., Mori, M., and Padbury, J.F. (2019). Effect of prematurity on genome wide methylation in the placenta. BMC Medical Genetics 20, 116.
30. Wang, X.-M., Tian, F.-Y., Fan, L.-J., Xie, C.-B., Niu, Z.-Z., and Chen, W.-Q. (2019). Comparison of DNA methylation profiles associated with spontaneous preterm birth in placenta and cord blood. BMC Medical Genomics 12, 1.
31. Braun, P.R., Han, S., Hing, B., Nagahama, Y., Gaul, L.N., Heinzman, J.T., Grossbach, A.J., Close, L., Dlouhy, B.J., Howard, M.A., 3rd, et al. (2019). Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. Transl Psychiatry 9, 47.
32. O'Shea, T.M., Allred, E.N., Dammann, O., Hirtz, D., Kuban, K.C., Paneth, N., Leviton, A., and Investigators, E.s. (2009). The ELGAN study of the brain and related disorders in extremely low gestational age newborns. Early Hum Dev 85, 719-725.
33. McElrath, T.F., Hecht, J.L., Dammann, O., Boggess, K., Onderdonk, A., Markenson, G., Harper, M., Delpapa, E., Allred, E.N., Leviton, A., et al. (2008). Pregnancy disorders that lead to delivery

before the 28th week of gestation: an epidemiologic approach to classification. Am J Epidemiol 168, 980-989.
34. Zheng, S.C., Webster, A.P., Dong, D., Feber, A., Graham, D.G., Sullivan, R., Jevons, S., Lovat, L.B., Beck, S., Widschwendter, M., et al. (2018). A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. Epigenomics 10, 925-940.
35. Natarajan, G., Pappas, A., Shankaran, S., Laptook, A.R., Walsh, M., McDonald, S.A., Ehrenkranz, R.A., Tyson, J.E., Goldberg, R.N., Bara, R., et al. (2012). Effect of inborn vs. outborn delivery on neurodevelopmental outcomes in infants with hypoxic-ischemic encephalopathy: secondary analyses of the NICHD whole-body cooling trial. Pediatr Res 72, 414-419.
36. Ambalavanan, N., Van Meurs, K.P., Perritt, R., Carlo, W.A., Ehrenkranz, R.A., Stevenson, D.K., Lemons, J.A., Poole, W.K., Higgins, R.D., and for the Nichd Neonatal Research Network, B.M.D. (2008). Predictors of death or bronchopulmonary dysplasia in preterm infants with respiratory failure. Journal of Perinatology 28, 420-426.
37. Redpath, S., Shah, P.S., Moore, G.P., Yang, J., Toye, J., Perreault, T., Lee, K.-S., Canadian Neonatal Transport, N., and Canadian Neonatal Network, I. (2020). Do transport factors increase the risk of severe brain injury in outborn infants <33 weeks gestational age? Journal of Perinatology 40, 385-393.
38. Singmann, P., Shem-Tov, D., Wahl, S., Grallert, H., Fiorito, G., Shin, S.Y., Schramm, K., Wolf, P., Kunze, S., Baran, Y., et al. (2015). Characterization of whole-genome autosomal differences of DNA methylation between men and women. Epigenetics Chromatin 8, 43.
39. Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet 3, 1724-1735.
40. Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 29, 189-196.
41. Pidsley, R., Zotenko, E., Peters, T.J., Lawrence, M.G., Risbridger, G.P., Molloy, P., Van Djik, S., Muhlhausler, B., Stirzaker, C., and Clark, S.J. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 17, 208.
42. Logue, M.W., Smith, A.K., Wolf, E.J., Maniates, H., Stone, A., Schichman, S.A., McGlinchey, R.E., Milberg, W., and Miller, M.W. (2017). The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. Epigenomics 9, 1363-1371.
43. Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11, 587.
44. Xie, C., Leung, Y.K., Chen, A., Long, D.X., Hoyo, C., and Ho, S.M. (2019). Differential methylation values in differential methylation analysis. Bioinformatics 35, 1094-1097.
45. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28, 882-883.
46. Carey, V.J., Lumley, T., and Ripley, B.D. (2019). gee: Generalized Estimation Equation Solver.
47. Suderman, M., Staley, J.R., French, R., Arathimos, R., Simpkin, A., and Tilling, K. (2018). dmrff: identifying differentially methylated regions efficiently with power and control. bioRxiv.
48. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb) 2, 100141.
49. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology 16, 284-287.

50. Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. Protein Sci 28, 1947-1951.
51. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res 51, D587-D592.
52. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28, 27-30.
53. Ren, X., and Kuan, P.F. (2019). methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. Bioinformatics 35, 1958-1959.
54. Breeze, C.E. (2021). Cell type-specific signal analysis in EWAS. bioRxiv, 2021.2005.2021.445209.
55. Breeze, C.E., Paul, D.S., van Dongen, J., Butcher, L.M., Ambrose, J.C., Barrett, J.E., Lowe, R., Rakyan, V.K., Iotchkova, V., Frontini, M., et al. (2016). eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. Cell Rep 17, 2137-2150.
56. Breeze, C.E., Reynolds, A.P., van Dongen, J., Dunham, I., Lazar, J., Neph, S., Vierstra, J., Bourque, G., Teschendorff, A.E., Stamatoyannopoulos, J.A., et al. (2019). eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. Bioinformatics 35, 4767-4769.
57. Wheater, E.N.W., Galdi, P., McCartney, D.L., Blesa, M., Sullivan, G., Stoye, D.Q., Lamb, G., Sparrow, S., Murphy, L., Wrobel, N., et al. (2021). DNA methylation and brain dysmaturation in preterm infants. medRxiv, 2021.2004.2008.21255064.
58. Greer, C., Troughton, R.W., Adamson, P.D., and Harris, S.L. (2022). Preterm birth and cardiac function in adulthood. Heart 108, 172-177.
59. Sparrow, S., Manning, J.R., Cartier, J., Anblagan, D., Bastin, M.E., Piyasena, C., Pataky, R., Moore, E.J., Semple, S.I., Wilkinson, A.G., et al. (2016). Epigenomic profiling of preterm infants reveals DNA methylation differences at sites associated with neural function. Transl Psychiatry 6, e716.
60. Scarborough, J., Mattei, D., Dorner-Ciossek, C., Sand, M., Arban, R., Rosenbrock, H., Richetto, J., and Meyer, U. (2021). Symptomatic and preventive effects of the novel phosphodiesterase-9 inhibitor BI 409306 in an immune-mediated model of neurodevelopmental disorders. Neuropsychopharmacology 46, 1526-1534.
61. Harms, J.F., Menniti, F.S., and Schmidt, C.J. (2019). Phosphodiesterase 9A in Brain Regulates cGMP Signaling Independent of Nitric-Oxide. Front Neurosci 13, 837.
62. Benton, M.C., Sutherland, H.G., Macartney-Coxson, D., Haupt, L.M., Lea, R.A., and Griffiths, L.R. (2017). Methylome-wide association study of whole blood DNA in the Norfolk Island isolate identifies robust loci associated with age. Aging (Albany NY) 9, 753-768.
63. Shi, L., Jiang, F., Ouyang, F., Zhang, J., Wang, Z., and Shen, X. (2018). DNA methylation markers in combination with skeletal and dental ages to improve age estimation in children. Forensic Sci Int Genet 33, 1-9.
64. Safran, M., Rosen, N., Twik, M., BarShir, R., Stein, T.I., Dahary, D., Fishilevich, S., and Lancet, D. (2021). The GeneCards Suite. In Practical Guide to Life Science Databases, I. Abugessaisa and T. Kasukawa, eds. (Singapore, Springer Nature Singapore), pp 27-56.
65. Cameron, V.A., Jones, G.T., Horwood, L.J., Pilbrow, A.P., Martin, J., Frampton, C., Ip, W.T., Troughton, R.W., Greer, C., Yang, J., et al. (2023). DNA methylation patterns at birth predict health outcomes in young adults born very low birthweight. Clin Epigenetics 15, 47.
66. Parets, S.E., Bedient, C.E., Menon, R., and Smith, A.K. (2014). Preterm birth and its long-term effects: methylation to mechanisms. Biology (Basel) 3, 498-513.
67. Aarnoudse-Moens, C.S., Weisglas-Kuperus, N., van Goudoever, J.B., and Oosterlaan, J. (2009). Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. Pediatrics 124, 717-728.
68. Hack, M., Taylor, H.G., Drotar, D., Schluchter, M., Cartar, L., Wilson-Costello, D., Klein, N., Friedman, H., Mercuri-Minich, N., and Morrow, M. (2005). Poor predictive validity of the Bayley Scales of

Infant Development for cognitive function of extremely low birth weight children at school age. Pediatrics 116, 333-341.
69. Vohr, B.R., Wright, L.L., Poole, W.K., and McDonald, S.A. (2005). Neurodevelopmental outcomes of extremely low birth weight infants <32 weeks' gestation between 1993 and 1998. Pediatrics 116, 635-643.
70. Aylward, G.P. (2005). Neurodevelopmental outcomes of infants born prematurely. J Dev Behav Pediatr 26, 427-440.
71. Grunau, R.E., Whitfield, M.F., and Davis, C. (2002). Pattern of learning disabilities in children with extremely low birth weight and broadly average intelligence. Arch Pediatr Adolesc Med 156, 615-620.
72. Hack, M., Youngstrom, E.A., Cartar, L., Schluchter, M., Taylor, H.G., Flannery, D., Klein, N., and Borawski, E. (2004). Behavioral outcomes and evidence of psychopathology among very low birth weight infants at age 20 years. Pediatrics 114, 932-940.
73. Hack, M., Taylor, H.G., Schluchter, M., Andreias, L., Drotar, D., and Klein, N. (2009). Behavioral outcomes of extremely low birth weight children at age 8 years. J Dev Behav Pediatr 30, 122-130.
74. Hille, E.T., Dorrepaal, C., Perenboom, R., Gravenhorst, J.B., Brand, R., Verloove-Vanhorick, S.P., and Dutch, P.-C.S.G. (2008). Social lifestyle, risk-taking behavior, and psychopathology in young adults born very preterm or with a very low birthweight. J Pediatr 152, 793-800, 800 e791-794.
75. Taylor, H.G., Klein, N., and Hack, M. (2000). School-age consequences of birth weight less than 750 g: a review and update. Dev Neuropsychol 17, 289-321.
76. Stephens, B.E., and Vohr, B.R. (2009). Neurodevelopmental outcome of the premature infant. Pediatr Clin North Am 56, 631-646, Table of Contents.
77. Allen, M.C. (2008). Neurodevelopmental outcomes of preterm infants. Curr Opin Neurol 21, 123-128.
78. Schmidt, B., Asztalos, E.V., Roberts, R.S., Robertson, C.M., Sauve, R.S., Whitfield, M.F., and Trial of Indomethacin Prophylaxis in Preterms, I. (2003). Impact of bronchopulmonary dysplasia, brain injury, and severe retinopathy on the outcome of extremely low-birth-weight infants at 18 months: results from the trial of indomethacin prophylaxis in preterms. JAMA 289, 1124-1129.
79. O'Reilly, M., Sozo, F., and Harding, R. (2013). Impact of preterm birth and bronchopulmonary dysplasia on the developing lung: long-term consequences for respiratory health. Clin Exp Pharmacol Physiol 40, 765-773.
80. Mowitz, M.E., Ayyagari, R., Gao, W., Zhao, J., Mangili, A., and Sarda, S.P. (2019). Health care burden of bronchopulmonary dysplasia among extremely preterm infants. Frontiers in pediatrics 7, 510.
81. Everson, T.M., Marsit, C.J., Michael O'Shea, T., Burt, A., Hermetz, K., Carter, B.S., Helderman, J., Hofheimer, J.A., McGowan, E.C., Neal, C.R., et al. (2019). Epigenome-wide Analysis Identifies Genes and Pathways Linked to Neurobehavioral Variation in Preterm Infants. Sci Rep 9, 6322.
82. Paniagua, U., Lester, B.M., Marsit, C.J., Camerota, M., Carter, B.S., Check, J.F., Helderman, J., Hofheimer, J.A., McGowan, E.C., Neal, C.R., et al. (2023). Epigenetic age acceleration, neonatal morbidities, and neurobehavioral profiles in infants born very preterm. Epigenetics 18, 2280738.
83. Giarraputo, J., DeLoach, J., Padbury, J., Uzun, A., Marsit, C., Hawes, K., and Lester, B. (2017). Medical morbidities and DNA methylation of NR3C1 in preterm infants. Pediatric Research 81, 68-74.
84. Knight, A.K., Smith, A.K., Conneely, K.N., Dalach, P., Loke, Y.J., Cheong, J.L., Davis, P.G., Craig, J.M., Doyle, L.W., and Theda, C. (2018). Relationship between Epigenetic Maturity and Respiratory Morbidity in Preterm Infants. The Journal of Pediatrics 198, 168-173.e162.
85. Cho, H.Y., Wang, X., Campbell, M.R., Panduri, V., Coviello, S., Caballero, M.T., Bennett, B.D., Kleeberger, S.R., Polack, F.P., Ofman, G., et al. (2023). Prospective epigenome and transcriptome analyses of cord and peripheral blood from preterm infants at risk of bronchopulmonary dysplasia. Sci Rep 13, 12262.

86. Jackson, W.M., Santos, H.P., Jr., Hartwell, H.J., Gower, W.A., Chhabra, D., Hagood, J.S., Laughon, M.M., Payton, A., Smeester, L., Roell, K., et al. (2022). Differential placental CpG methylation is associated with chronic lung disease of prematurity. Pediatr Res 91, 1428-1435.
87. Wang, X., Cho, H.Y., Campbell, M.R., Panduri, V., Coviello, S., Caballero, M.T., Sambandan, D., Kleeberger, S.R., Polack, F.P., Ofman, G., et al. (2022). Epigenome-wide association study of bronchopulmonary dysplasia in preterm infants: results from the discovery-BPD program. Clin Epigenetics 14, 57.
88. Wiechmann, T., Röh, S., Sauer, S., Czamara, D., Arloth, J., Ködel, M., Beintner, M., Knop, L., Menke, A., Binder, E.B., et al. (2019). Identification of dynamic glucocorticoid-induced methylation changes at the FKBP5 locus. Clinical Epigenetics 11, 83.
89. Yang, R., Xu, C., Bierer, L.M., Flory, J.D., Gautam, A., Bader, H.N., Lehrner, A., Makotkine, I., Desarnaud, F., Miller, S.A., et al. (2021). Longitudinal genome-wide methylation study of PTSD treatment using prolonged exposure and hydrocortisone. Transl Psychiatry 11, 398.
90. Braun, P.R., Tanaka-Sahker, M., Chan, A.C., Jellison, S.S., Klisares, M.J., Hing, B.W., Shabbir, Y., Gaul, L.N., Nagahama, Y., and Robles, J. (2019). Genome-wide DNA methylation investigation of glucocorticoid exposure within buccal samples. Psychiatry and clinical neurosciences 73, 323-330.
91. Doyle, L.W. (2021). Postnatal Corticosteroids to Prevent or Treat Bronchopulmonary Dysplasia. Neonatology 118, 244-251.
92. Godoy, L.D., Rossignoli, M.T., Delfino-Pereira, P., Garcia-Cairasco, N., and de Lima Umeoka, E.H. (2018). A Comprehensive Overview on Stress Neurobiology: Basic Concepts and Clinical Implications. Front Behav Neurosci 12, 127.
93. Weaver, I.C.G., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., and Meaney, M.J. (2004). Epigenetic programming by maternal behavior. Nature Neuroscience 7, 847-854.
94. Weaver, I.C.G., Diorio, J., Seckl, J.R., Szyf, M., and Meaney, M.J. (2004). Early Environmental Regulation of Hippocampal Glucocorticoid Receptor Gene Expression: Characterization of Intracellular Mediators and Potential Genomic Target Sites. Annals of the New York Academy of Sciences 1024, 182-212.
95. Blackwell, C.K., Wakschlag, L.S., Gershon, R.C., Cella, D., and with the, E.P.R.O.C. (2018). Measurement framework for the Environmental influences on Child Health Outcomes research program. Curr Opin Pediatr 30, 276-284.
96. Gillman, M.W., and Blaisdell, C.J. (2018). Environmental influences on Child Health Outcomes, a Research Program of the National Institutes of Health. Curr Opin Pediatr 30, 260-262.
97. Jacobson, L.P., Lau, B., Catellier, D., and Parker, C.B. (2018). An Environmental influences on Child Health Outcomes viewpoint of data analysis centers for collaborative study designs. Curr Opin Pediatr 30, 269-275.
98. Hofheimer, J.A., Smith, L.M., McGowan, E.C., O'Shea, T.M., Carter, B.S., Neal, C.R., Helderman, J.B., Pastyrnak, S.L., Soliman, A., Dansereau, L.M., et al. (2020). Psychosocial and medical adversity associated with neonatal neurobehavior in infants born before 30 weeks gestation. Pediatr Res 87, 721-729.
99. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30, 1363-1369.
100. Fortin, J.P., Fertig, E., and Hansen, K. (2014). shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. F1000Res 3, 175.
101. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118-127.

102. Martin, M., Smith, L., Hofheimer, J.A., McGowan, E.C., Shea, T.M., Pastyrnak, S., Carter, B.S., Helderman, J., Check, J., Neal, C., et al. (2022). Bronchopulmonary dysplasia and neurobehavioural outcomes at birth and 2 years in infants born before 30 weeks. Archives of Disease in Childhood - Fetal and Neonatal Edition, fetalneonatal-2021-323405.
103. Jensen, E.A., Dysart, K., Gantz, M.G., McDonald, S., Bamat, N.A., Keszler, M., Kirpalani, H., Laughon, M.M., Poindexter, B.B., Duncan, A.F., et al. (2019). The Diagnosis of Bronchopulmonary Dysplasia in Very Preterm Infants. An Evidence-based Approach. American Journal of Respiratory and Critical Care Medicine 200, 751-759.
104. Zou, Z., Xiang, M., Zhang, Y., Huang, Y., Wang, J., He, Y., Min, W., and Zhou, B. (2022). Associations of DNA methylation of HPA axis-related genes and neuroendocrine abnormalities in panic disorder. Psychoneuroendocrinology 142, 105777.
105. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res 12, 996-1006.
106. Zheng, S.C., Breeze, C.E., Beck, S., Dong, D., Zhu, T., Ma, L., Ye, W., Zhang, G., and Teschendorff, A.E. (2019). EpiDISH web server: Epigenetic Dissection of Intra-Sample-Heterogeneity with online GUI. Bioinformatics.
107. Gervin, K., Salas, L.A., Bakulski, K.M., van Zelm, M.C., Koestler, D.C., Wiencke, J.K., Duijts, L., Moll, H.A., Kelsey, K.T., Kobor, M.S., et al. (2019). Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood DNA methylation data. Clin Epigenetics 11, 125.
108. Koestler, D.C., Jones, M.J., Usset, J., Christensen, B.C., Butler, R.A., Kobor, M.S., Wiencke, J.K., and Kelsey, K.T. (2016). Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). BMC Bioinformatics 17, 120.
109. Salas, L.A., Koestler, D.C., Butler, R.A., Hansen, H.M., Wiencke, J.K., Kelsey, K.T., and Christensen, B.C. (2018). An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol 19, 64.
110. Ramos-Navarro, C., Sanchez-Luna, M., Zeballos-Sarrato, S., and Pescador-Chamorro, I. (2022). Antenatal corticosteroids and the influence of sex on morbidity and mortality of preterm infants. J Matern Fetal Neonatal Med 35, 3438-3445.
111. Suarez, A., Lahti, J., Lahti-Pulkkinen, M., Girchenko, P., Czamara, D., Arloth, J., Malmberg, A.L.K., Hämäläinen, E., Kajantie, E., Laivuori, H., et al. (2020). A polyepigenetic glucocorticoid exposure score at birth and childhood mental and behavioral disorders. Neurobiology of Stress 13, 100275.
112. McGowan, P.O., and Matthews, S.G. (2018). Prenatal Stress, Glucocorticoids, and Developmental Programming of the Stress Response. Endocrinology 159, 69-82.
113. Talge, N.M., Neal, C., Glover, V., Early Stress, T.R., Prevention Science Network, F., Neonatal Experience on, C., and Adolescent Mental, H. (2007). Antenatal maternal stress and long-term effects on child neurodevelopment: how and why? J Child Psychol Psychiatry 48, 245-261.
114. Syed, S.A., and Zannas, A.S. (2021). Chapter 29 - Epigenetics in psychotherapy. In Epigenetics in Psychiatry (Second Edition), J. Peedicayil, D.R. Grayson, andD. Avramopoulos, eds. (Academic Press), pp 701-709.
115. Tsigos, C., and Chrousos, G.P. (2002). Hypothalamic–pituitary–adrenal axis, neuroendocrine factors and stress. Journal of Psychosomatic Research 53, 865-871.
116. Criado-Marrero, M., Rein, T., Binder, E.B., Porter, J.T., Koren, J., 3rd, and Blair, L.J. (2018). Hsp90 and FKBP51: complex regulators of psychiatric diseases. Philos Trans R Soc Lond B Biol Sci 373.
117. Somvanshi, P.R., Mellon, S.H., Yehuda, R., Flory, J.D., Makotkine, I., Bierer, L., Marmar, C., Jett, M., and Doyle, F.J. (2020). Role of enhanced glucocorticoid receptor sensitivity in inflammation in

PTSD: insights from computational model for circadian-neuroendocrine-immune interactions. American Journal of Physiology-Endocrinology and Metabolism 319, E48-E66.
118. Plieger, T., Felten, A., Splittgerber, H., Duke, É., and Reuter, M. (2018). The role of genetic variation in the glucocorticoid receptor (NR3C1) and mineralocorticoid receptor (NR3C2) in the association between cortisol response and cognition under acute stress. Psychoneuroendocrinology 87, 173-180.
119. Raffin-Sanson, M.L., de Keyzer, Y., and Bertagna, X. (2003). Proopiomelanocortin, a polypeptide precursor with multiple functions: from physiology to pathological conditions. European Journal of Endocrinology Eur J Endocrinol 149, 79-90.
120. DiPietro, J.A., and Voegtline, K.M. (2017). The gestational foundation of sex differences in development and vulnerability. Neuroscience 342, 4-20.
121. De Coster, S., van Leeuwen, D.M., Jennen, D.G., Koppen, G., Den Hond, E., Nelen, V., Schoeters, G., Baeyens, W., van Delft, J.H., Kleinjans, J.C., et al. (2013). Gender-specific transcriptomic response to environmental exposure in Flemish adults. Environ Mol Mutagen 54, 574-588.
122. Gabory, A., Roseboom, T.J., Moore, T., Moore, L.G., and Junien, C. (2013). Placental contribution to the origins of sexual dimorphism in health and diseases: sex chromosomes and epigenetics. Biol Sex Differ 4, 5.
123. McCarthy, N.S., Melton, P.E., Cadby, G., Yazar, S., Franchina, M., Moses, E.K., Mackey, D.A., and Hewitt, A.W. (2014). Meta-analysis of human methylation data for evidence of sex-specific autosomal patterns. BMC Genomics 15, 981.
124. Rosenfeld, C.S. (2015). Sex-Specific Placental Responses in Fetal Development. Endocrinology 156, 3422-3434.
125. Martin, E., Smeester, L., Bommarito, P.A., Grace, M.R., Boggess, K., Kuban, K., Karagas, M.R., Marsit, C.J., O'Shea, T.M., and Fry, R.C. (2017). Sexual epigenetic dimorphism in the human placenta: implications for susceptibility during the prenatal period. Epigenomics 9, 267-278.
126. Clark, J., Martin, E., Bulka, C.M., Smeester, L., Santos, H.P., O'Shea, T.M., and Fry, R.C. (2019). Associations between placental CpG methylation of metastable epialleles and childhood body mass index across ages one, two and ten in the Extremely Low Gestational Age Newborns (ELGAN) cohort. Epigenetics 14, 1102-1111.
127. Watterberg, K.L., Scott, S.M., and Naeye, R.L. (1997). Chorioamnionitis, cortisol, and acute lung disease in very low birth weight infants. Pediatrics 99, E6.
128. Baud, O., and Watterberg, K.L. (2019). Prophylactic postnatal corticosteroids: Early hydrocortisone. Semin Fetal Neonatal Med 24, 202-206.
129. Meakin, C.J., Martin, E.M., Santos, H.P., Jr., Mokrova, I., Kuban, K., O'Shea, T.M., Joseph, R.M., Smeester, L., and Fry, R.C. (2018). Placental CpG methylation of HPA-axis genes is associated with cognitive impairment at age 10 among children born extremely preterm. Horm Behav 101, 29-35.
130. Hustad, K.C., Allison, K., McFadd, E., and Riehle, K. (2014). Speech and language development in 2-year-old children with cerebral palsy. Dev Neurorehabil 17, 167-175.
131. Mei, C., Reilly, S., Bickerton, M., Mensah, F., Turner, S., Kumaranayagam, D., Pennington, L., Reddihough, D., and Morgan, A.T. (2020). Speech in children with cerebral palsy. Dev Med Child Neurol 62, 1374-1382.
132. Oluwole, I., Tan, J.B.C., DeSouza, S., Hutchinson, M., Leigh, R.M., Cha, M., Rodriguez, A., Hou, G., Rao, S.S., Narang, A., et al. (2023). The association between bronchopulmonary dysplasia grade and risks of adverse neurodevelopmental outcomes among preterm infants born at less than 30 weeks of gestation. J Matern Fetal Neonatal Med 36, 2167074.

133. Rees, P., Callan, C., Chadda, K.R., Vaal, M., Diviney, J., Sabti, S., Harnden, F., Gardiner, J., Battersby, C., Gale, C., et al. (2022). Preterm Brain Injury and Neurodevelopmental Outcomes: A Meta-analysis. Pediatrics 150, e2022057442.
134. Diggikar, S., Gurumoorthy, P., Trif, P., Mudura, D., Nagesh, N.K., Galis, R., Vinekar, A., and Kramer, B.W. (2023). Retinopathy of prematurity and neurodevelopmental outcomes in preterm infants: A systematic review and meta-analysis. Front Pediatr 11, 1055813.
135. Wang, Y., Liu, S., Lu, M., Huang, T., and Huang, L. (2024). Neurodevelopmental outcomes of preterm with necrotizing enterocolitis: a systematic review and meta-analysis. Eur J Pediatr.
136. Colwell, M.L., Townsel, C., Petroff, R.L., Goodrich, J.M., and Dolinoy, D.C. (2023). Epigenetics and the Exposome: DNA Methylation as a Proxy for Health Impacts of Prenatal Environmental Exposures. Exposome 3.
137. McGowan, E.C., Hofheimer, J.A., O'Shea, T.M., Kilbride, H., Carter, B.S., Check, J., Helderman, J., Neal, C.R., Pastyrnak, S., Smith, L.M., et al. (2022). Analysis of Neonatal Neurobehavior and Developmental Outcomes Among Preterm Infants. JAMA Network Open 5, e2222249-e2222249.
138. Agler, R., and De Boeck, P. (2017). On the Interpretation and Use of Mediation: Multiple Perspectives on Mediation Analysis. Front Psychol 8, 1984.
139. Clark-Boucher, D., Zhou, X., Du, J., Liu, Y., Needham, B.L., Smith, J.A., and Mukherjee, B. (2023). Methods for mediation analysis with high-dimensional DNA methylation data: Possible choices and comparisons. PLoS Genet 19, e1011022.
140. Gao, Y., Yang, H., Fang, R., Zhang, Y., Goode, E.L., and Cui, Y. (2019). Testing Mediation Effects in High-Dimensional Epigenetic Studies. Front Genet 10, 1195.
141. Hollingshead, A. (1975). Four factor index of social status. In., pp 47-55.
142. (2022). Manual of Operations: Part 2 Data Definitions & Infant Data Booklets. (Vermont-Oxford Network.
143. Camerota, M., McGowan, E.C., Carter, B.S., Check, J., Dansereau, L.M., DellaGrotta, S.A., Helderman, J.B., Hofheimer, J.A., Neal, C.R., O'Shea, T.M., et al. (2023). Maternal Prenatal Risk Phenotypes and Neurobehavioral Outcomes among Infants Born Very Preterm. J Pediatr 260, 113521.
144. Hodge, K.M., Zhabotynsky, V., Burt, A.A., Carter, B.S., Fry, R.C., Helderman, J., Hofheimer, J.A., McGowan, E.C., Neal, C.R., Pastyrnak, S.L., et al. (2024). Epigenetic associations in HPA axis genes related to bronchopulmonary dysplasia and antenatal steroids. Pediatr Res.
145. Bayley, N. (2006). Bayley scales of infant and Toddler development - Third Edition.(San Antonio, Texas: Harcourt Assessment).
146. Palisano, R.J.R., P.; Bartlett, D.; Livingston, M.H. (2007). GMFCS – E & R Gross Motor Function Classification System Expanded and Revised.(Hamilton, ON: Institute for Applied Health Sciences, McMaster University).
147. Peacock, J.L., Marston, L., Marlow, N., Calvert, S.A., and Greenough, A. (2012). Neonatal and infant outcome in boys and girls born very prematurely. Pediatr Res 71, 305-310.
148. Morsing, E., Asard, M., Ley, D., Stjernqvist, K., and Marsal, K. (2011). Cognitive function after intrauterine growth restriction and very preterm birth. Pediatrics 127, e874-882.
149. Voßbeck, S., de Camargo, O.K., Grab, D., Bode, H., and Pohlandt, F. (2001). Neonatal and neurodevelopmental outcome in infants born before 30 weeks of gestation with absent or reversed end-diastolic flow velocities in the umbilical artery. European Journal of Pediatrics 160, 128-134.
150. Wu, B.A., Chand, K.K., Bell, A., Miller, S.L., Colditz, P.B., Malhotra, A., and Wixey, J.A. (2024). Effects of fetal growth restriction on the perinatal neurovascular unit and possible treatment targets. Pediatric Research 95, 59-69.
151. Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. Psychol Methods 15, 309-334.

152. Imai, K., Tingley, D., and Yamamoto, T. (2013). Experimental Designs for Identifying Causal Mechanisms. Journal of the Royal Statistical Society Series A: Statistics in Society 176, 5-51.
153. Imai, K., and Yamamoto, T. (2013). Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments. Political Analysis 21, 141-171.
154. Tingley D, Y.T., Hirose K, Keele L, Imai K. (2013). mediation: R Package for Causal Mediation Analysis. In. (
155. Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., and Moeller, S. (2009). gplots: Various R programming tools for plotting data. In R package version. (
156. Camerota, M., Lester, B.M., McGowan, E.C., Carter, B.S., Check, J., Dansereau, L.M., DellaGrotta, S.A., Helderman, J.B., Hofheimer, J.A., Loncar, C.M., et al. (2024). Contributions of prenatal risk factors and neonatal epigenetics to cognitive outcome in children born very preterm. Dev Psychol.
157. Abrishamcar, S., Chen, J., Feil, D., Kilanowski, A., Koen, N., Vanker, A., Wedderburn, C.J., Donald, K.A., Zar, H.J., Stein, D.J., et al. (2022). DNA methylation as a potential mediator of the association between prenatal tobacco and alcohol exposure and child neurodevelopment in a South African birth cohort. Translational Psychiatry 12, 418.
158. Feil, D., Abrishamcar, S., Christensen, G.M., Vanker, A., Koen, N., Kilanowski, A., Hoffman, N., Wedderburn, C.J., Donald, K.A., Kobor, M.S., et al. (2023). DNA methylation as a potential mediator of the association between indoor air pollution and neurodevelopmental delay in a South African birth cohort. Clinical Epigenetics 15, 31.
159. Williams, J.B., Cao, Q., Wang, W., Lee, Y.-H., Qin, L., Zhong, P., Ren, Y., Ma, K., and Yan, Z. (2023). Inhibition of histone methyltransferase Smyd3 rescues NMDAR and cognitive deficits in a tauopathy mouse model. Nature Communications 14, 91.
160. Binh, M.T., Hoan, N.X., Giang, D.P., Tong, H.V., Bock, C.T., Wedemeyer, H., Toan, N.L., Bang, M.H., Kremsner, P.G., Meyer, C.G., et al. (2020). Upregulation of SMYD3 and SMYD3 VNTR 3/3 polymorphism increase the risk of hepatocellular carcinoma. Scientific Reports 10, 2797.
161. Nigam, N., Bernard, B., Kim, S., Burkitt, K., Tsai, D., Robbins, Y., Sievers, C., Clint, A., Bennett, R.L., Tettey, T.T., et al. (2022). Epigenetic silencing by SMYD3 represses tumor intrinsic interferon response in HPV-negative squamous cell carcinoma of the head and neck. bioRxiv, 2022.2011.2004.515098.
162. Rueda-Robles, A., Audano, M., Álvarez-Mercado, A.I., and Rubio-Tomás, T. (2021). Functions of SMYD proteins in biological processes: What do we know? An updated review. Archives of Biochemistry and Biophysics 712, 109040.
163. Ohkubo, Y., Uchida, A.O., Shin, D., Partanen, J., and Vaccarino, F.M. (2004). Fibroblast growth factor receptor 1 is required for the proliferation of hippocampal progenitor cells and for hippocampal growth in mouse. J Neurosci 24, 6057-6069.
164. Popovici, C., Zhang, B., Gregoire, M.J., Jonveaux, P., Lafage-Pochitaloff, M., Birnbaum, D., and Pebusque, M.J. (1999). The t(6;8)(q27;p11) translocation in a stem cell myeloproliferative disorder fuses a novel gene, FOP, to fibroblast growth factor receptor 1. Blood 93, 1381-1389.
165. Yang, S.K., Hong, M., Zhao, W., Jung, Y., Baek, J., Tayebi, N., Kim, K.M., Ye, B.D., Kim, K.J., Park, S.H., et al. (2014). Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. Gut 63, 80-87.
166. Chu, X., Pan, C.M., Zhao, S.X., Liang, J., Gao, G.Q., Zhang, X.M., Yuan, G.Y., Li, C.G., Xue, L.Q., Shen, M., et al. (2011). A genome-wide association study identifies two new risk loci for Graves' disease. Nat Genet 43, 897-901.

167. Quan, C., Ren, Y.Q., Xiang, L.H., Sun, L.D., Xu, A.E., Gao, X.H., Chen, H.D., Pu, X.M., Wu, R.N., Liang, C.Z., et al. (2010). Genome-wide association study for vitiligo identifies susceptibility loci at 6q27 and the MHC. Nat Genet 42, 614-618.
168. Yi, X., Du, L., Hou, S., Li, F., Chen, Y., Kijlstra, A., and Yang, P. (2013). FGFR1OP tagSNP but not CCR6 polymorphisms are associated with Vogt-Koyanagi-Harada syndrome in Chinese Han. PLoS One 8, e69358.
169. Gouveia, C., Gibbons, E., Dehghani, N., Eapen, J., Guerreiro, R., and Bras, J. (2022). Genome-wide association of polygenic risk extremes for Alzheimer's disease in the UK Biobank. Sci Rep 12, 8404.
170. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linner, R., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50, 1112-1121.
171. Demange, P.A., Malanchini, M., Mallard, T.T., Biroli, P., Cox, S.R., Grotzinger, A.D., Tucker-Drob, E.M., Abdellaoui, A., Arseneault, L., van Bergen, E., et al. (2021). Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. Nat Genet 53, 35-44.
172. Williams, C.M., Labouret, G., Wolfram, T., Peyre, H., and Ramus, F. (2023). A General Cognitive Ability Factor for the UK Biobank. Behav Genet 53, 85-100.
173. Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun 9, 2098.

ADDITIONAL INFORMATION

Data availability statement

The raw and processed DNAm data for NOVI are publicly accessible through NCBI Gene Expression Omnibus (GEO) via accession series GSE128821. Select de-identified data from the ECHO Program are available through NICHD's Data and Specimen Hub (DASH). Information on study data not available on DASH, such as some Indigenous datasets, can be found on the ECHO study DASH webpage.

Acknowledgements

The authors wish to thank our ECHO Colleagues; the medical, nursing, and program staff; and the children and families participating in the ECHO cohorts. We also acknowledge the contribution of the following ECHO Program collaborators:

ECHO Components—Coordinating Center: Duke Clinical Research Institute, Durham, North Carolina: Smith PB, Newby LK; Data Analysis Center: Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland: Jacobson LP; Research Triangle Institute, Durham, North Carolina: Catellier DJ; Person-Reported Outcomes Core: Northwestern University, Evanston, Illinois: Gershon R, Cella D.

ECHO Awardees and Cohorts—Baystate Children's Hospital, Springfield, MA: Vaidya R; Beaumont Children's Hospital, Royal Oak, MI: Obeid R; Boston Children's Hospital, Boston, MA: Rollins C; East Carolina University, Brody School of Medicine, Greenville, NC: Bear K; Michigan State University College of Human Medicine, East Lansing, MI: Lenski M; Tufts University School of Medicine, Boston, MA: Singh R; University of Chicago, Chicago, IL: Msall M; University of Massachusetts Chan Medical School, Worcester, MA: Frazier J; Atrium Health Wake Forest Baptist, Winston Salem, NC: Gogcu S; Yale School of Medicine, New Haven, CT: Montgomery A; Boston Medical Center, Boston, MA: Kuban K, Douglass L, Jara H; Boston University, Boston, MA: Joseph R.

Competing Interests

The authors declare that they have no competing interests.

Funding

Research reported in this publication was supported by the Environmental influences on Child Health Outcomes (ECHO) Program, Office of the Director, National Institutes of Health, under Award Numbers U2COD023375 (Coordinating Center), U24OD023382 (Data Analysis Center), U24OD023319 with co-funding from the Office of Behavioral and Social Science Research (PRO Core), UH3OD023347 (Lester and Marsit) and UH3OD023348 (O'Shea and Fry). The Neonatal Neurobehavior and Outcomes in Very Preterm Infants (NOVI) cohort was also supported by R01HD072267, R01HD084515, and P30 ES019776. Dr. Camerota was supported by a career development award (K01MH129510).

Author contributions

TME, CJM, and BML initiated, acquired the funding for, and, along with KMH, designed this investigation. KMH performed the statistical analyses and KMH, VZ, TME, AH, KNC and MC interpreted the results. AAB, BSC, JC, JH, JAH, ECM, CRN, SLP, LMS, SAD, LMD and TMO coordinated data collection. KMH and TME drafted the manuscript. All authors contributed to interpretation of the results and revisions to the manuscript. All authors read and approved the final manuscript.