Distribution Agreement

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

Signature:

_____ _____

Michael Mortillo Date

Functionality of placental 5-hydroxymethylcytosine and associations with birthweight

By

Michael Mortillo
Doctor of Philosophy
Genetics and Molecular Biology

Carmen Marsit, PhD
Advisor

Karen Conneely, PhD
Committee Member

Mike Epstein, PhD
Committee Member

Todd Everson, PhD
Committee Member

Anke Huels, PhD
Committee Member

Accepted:

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

Date

Functionality of placental 5-hydroxymethylcytosine and associations with birthweight

By

Michael Mortillo
MPH, Emory University, 2019
BS, Virginia Tech, 2015

Advisor: Carmen Marsit, 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
2023

Abstract

Functionality of placental 5-hydroxymethylcytosine and associations with birthweight

By Michael Mortillo

Abstract

Genetic and environmental influences during *in utero* development may contribute to the etiology of adverse health outcomes in both early life and adulthood. As the main regulator of early development, the placenta acts as an interface in the exchange of maternal physiological and environmental cues with the fetus. It also functions to ensure the success of the pregnancy and allow for proper fetal growth and development. Together, these roles position the placenta as a unique window to understanding molecular mechanisms underlying these potentially deleterious genetic and environmental impacts. One mechanism that can regulate placental function through control of gene expression is the DNA modification 5-hydroxymethylcytosine (5hmC). 5hmC is an epigenetic modification found on the cytosine base of a cytosine-phosphate-guanine (CpG) dinucleotide. It is formed through oxidation of 5-methylcytosine (5mC), an epigenetic mark known to play a complicated role in gene expression control. Though evidence has suggested 5hmC to be a transient intermediate on the demethylation pathway, it may also play a functional role similar to that of 5mC. In this study, placentae from the Rhode Island Child Health Study (RICHS) were selected for CpG 5hmC and 5mC profiling, as well as total mRNA sequencing. The RICHS cohort also collected anthropomorphic and medical data from mother-infant pairs, including data related to newborn birthweight, an early life predictor of long-term health. Here, we characterize associations between placental 5hmC and gene expression, as well as associations between 5hmC, 5mC, and birthweight. We show that although 5hmC proportions are generally low across placenta, it is positively associated with expression of specific genes based on genomic context. We also identify few relationships between 5hmC and birthweight, but show that the characterization of 5mC is not overtly biased by 5hmC content. The characterization of placental 5hmC and its associations with expression suggests a role for this mark in the placenta, with as of yet unknown phenotypic consequence. Future work should consider potentially other roles for 5hmC and explore how environmental and physiologic factors impact this functional mark.

Functionality of placental 5-hydroxymethylcytosine and associations with birthweight

By

Michael Mortillo
MPH, Emory University, 2019
BS, Virginia Tech, 2015

Advisor: Carmen Marsit, 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
2023

Acknowledgements

In the words of The Shawshank Redemption's Brooks Hatlen, "The world went and got itself in a big damn hurry." (For those that have met me, you know I am a massive cinephile and couldn't not start these acknowledgements without a movie quote). Anyway, truer words have never been spoken, in my opinion. As I sit here writing these acknowledgements and reflecting on my time in the GMB program, I recognize that I have done my best not to conform to the constant hustle the world finds itself in now. I knew that when I started a PhD program, it would require dedication, effort, and most importantly, time. Time that I would never get back. Time away from my friends and family. But ultimately, I realized it would be worth the wait. Well today, I can confidently say that the wait is over. PhD, complete.

There are so many people who have supported or encouraged me over the course of these last four years that I can thank, though the limitations of 1" margins prevent me from mentioning them all. Therefore, I'll just land on the greatest hits. First off, my family, composed of my parents, brothers, and a countless number of aunts, uncles, and cousins have all helped me get to where I am today. My parents especially have been huge supporters throughout my life and I thank them every day for instilling a sense of kindness and acceptance in me. Thank you both for never looking at me or treating me differently for the decisions I've made. I would also like to give a shoutout to my grandparents, who I am convinced are the two kindest, most sincere people in the world. Their opinion is what matters most to me.

If there is one silver lining for being in school for 25+ years, it's getting the opportunity to meet such amazing people. One person in particular, my girlfriend Allyson, has been so supportive throughout this process. I am also incredibly grateful that I dated a biostatistician while I worked on a statistics-centric project! Her love, support, and patience while helping me make graphs in ggplot is something I will never take for granted. I would also like to thank my three best friends from college, Ben, Sham, and PJ, for always picking me up on my worst days and helping me succeed. It's important to take time for yourself once in a while, and you three were always there to take my mind off school for a few hours while we talk fantasy football and endlessly quote movies to each other. And finally, I'd like to thank all my graduate school friends, both Masters and PhD folks, who helped me along the way. You guys are amazing and I am so proud of the work you all continue to do. To my GMB cohort buddies, thank you for letting me play in my corner of the sandbox over these past four years. It's meant the world to me.

To my advisor, Dr. Carmen Marsit, thank you so much for allowing me to join the lab and work alongside you. I was in a tough spot when faced with the realization that I needed to switch labs, and you were so supportive and helpful with the transition. I am forever grateful for your mentorship and helping me grow as both a scientist and a person. Also, to Dr. Liz Kennedy, thank you for your patience with me and always telling me what I needed to hear. I couldn't have done this without you.

I'm running out of space now and you're probably falling asleep reading this so I'll wrap it up. Thanks again to everyone I've mentioned and the countless others I couldn't here. I look forward to keeping in touch, and please come visit if you're ever in the Washington, D.C. area!

Table of Contents

Chapter 1 – Introduction

Birthweight outcomes are an early life predictor of chronic disease risk through the developmental origins of health and disease

Birthweight has been identified as a risk indicator not only for infant mortality, but also as a risk factor for developing chronic diseases in adulthood, including high blood pressure, cardiovascular disease, and type 2 diabetes[1-3]. Throughout gestation, humans experience a large amount of developmental plasticity, though this plasticity quickly diminishes following birth[4]. Despite this reduction in plasticity, in utero changes in the physiology and metabolism of the offspring can produce lifelong health outcomes[4] (**Figure 1-1**). Fetal growth trajectories are established early in development and are highly susceptible to in utero conditions, and thus alterations in the fetal environment can lead to a perturbed fetal growth trajectory[5]. While birthweight itself does not directly contribute to the etiology of these chronic diseases in adulthood, it may serve as a marker of underlying mechanisms that are developmentally programming lifelong health outcomes[6, 7]. This relationship was originally highlighted in observational studies of economically disadvantaged areas of England and Wales, where it was shown that these regions experienced significantly higher rates of mortality due to cardiovascular disease compared to wealthier regions[8]. This seemed counterintuitive, as prevalence of cardiovascular disease was expected to be higher in wealthier regions where fatty foods are more abundant and accessible. Further investigation revealed that people of lower birth weight had a higher risk of adult cardiovascular disease[9]. Specifically, in countries like England, Wales, and Finland, lower birth weight was associated with increased risk of coronary heart disease later in life[10-13]. This work resulted in the postulation of the **D**evelopmental **O**rigins of **H**ealth **a**nd

Disease hypothesis (DOHaD), which states that environmental exposures throughout the perinatal period can impact diseases later in life[14, 15].

The placenta is an important mediator of the maternal-fetal environment and can be used to assess the DOHaD hypothesis

The placenta is a highly specialized organ that supports the growth and development of a fetus throughout pregnancy[16-18]. It begins formation following the invasion of the trophectoderm into the extracellular matrix of the uterine wall, and is critical for implantation [16, 19]. The main function of the placenta is to act as an interface between the maternal and fetal systems, and to perform this, it contains two separate circulatory systems: 1) the maternal-placenta system, and 2) the fetal-placenta system[20-22]. Through its complex, microvilli-based system[23, 24], the placenta maintains intrauterine homeostasis by carrying out critical functions including exchange of nutrients, gas, and waste[25-28], along with immunomodulation[29, 30] and endocrine signaling[31, 32] **(Figure 1-2)**. Due to its position as an ephemeral fetal organ that acts as an interface in the exchange of maternal physiological and environmental cues with the developing fetus, the placenta is uniquely positioned to assess the mechanisms that may be involved in the DOHaD hypothesis[33].

Placental function is regulated by the collective responses of maternal decidual cells, trophoblastic cells, and fetal endothelial cells[25, 34]. Therefore, disruption of the maternal environment via stress and other stimuli can have adverse effects on placental structure and function, including integrity of the transplacental barrier, nutrient and oxygen exchange, and placental endocrine action[25, 35, 36]. Studies have shown that these effects act in a largely time-specific framework; perturbations occurring in early pregnancy are more likely to produce prolonged effects on placental function related to trophoblast differentiation and vascular remodeling[37]. These placental function deficiencies are associated with adverse gestational

outcomes, such as fetal growth restriction, which serves as a significant risk indicator for the development of chronic diseases throughout life[38]. However, clinical placental pathologies are fairly uncommon in the general population, and thus they do not entirely account for the prevalence of adverse gestational outcomes. However, it is widely believed that the subtle alterations in placental function induced via a response to changing gestational conditions may in part be responsible for the developmental programming of long-term health outcomes in the offspring[39].

While this relationship between gestational conditions, placental function, and longitudinal health outcomes in offspring is well characterized, the molecular mechanisms involved in this programming remain poorly understood[40, 41]. It has been theorized that mechanisms in the placental epigenome may be altered in response to gestational conditions[42], leading to placental dysregulation and subsequent health outcomes, and investigating these epigenetic mechanisms may provide novel insight into the etiology of these outcomes.

Placental methylation is a potential biomarker for fetal birthweight

DNA methylation is the most common type of epigenetic modification, with the most common type of DNA methylation being 5-methylcytostine (5mC), which occurs at unmethylated cytosines along the sugar-phosphate backbone of the DNA. Though it serves a variety of functions, the main function of methylation is to regulate gene expression. Typically, methylation in a gene promoter leads to a decrease in transcription of the gene, mainly by recruiting proteins involved in gene silencing or by inhibiting transcription factors from binding to the DNA and initiating transcription[43, 44]. Methylation marks are established by the addition of a methyl group to the 5^{th} carbon position of the cytosine via DNA methyltransferases (DNMTs)[45-48]. As a functional epigenetic modification, 5mC plays a critical role in genomic

imprinting[49, 50], X-chromosome inactivation[51, 52], and tissue-specific gene expression[53, 54].

DNA methylation may represent the regulatory links between genetic and environmental influences on birthweight[55, 56]. Thus, identifying birthweight-associated methylation loci can provide clues to detect molecular biomarkers of aberrant fetal growth[56] (**Figure 1-3**). DNA methylation in placenta has been shown to regulate expression and may be one mechanism by which environmental exposures and genetic variation are influencing fetal growth; a recent study assessing the relationship between methylation-regulated expression of imprinted genes and fetal growth showed that enrichment of methyl marks in the imprinted CDKN1C gene was positively associated with birthweight[57]. Furthermore, evidence has shown that aberrant imprinting leads to abnormalities in placental and fetal growth in animal models[58-61]. Additionally, altered placental expression of imprinted genes is associated with fetal growth restriction in humans[62, 63]. Another proposed link between infant growth restriction and epigenetic alterations is the association between methylation and regulation of genes involved in trophoblast migration and invasion[64], as well as the reception of the endometrium to implantation[65].

Placental methylation in other areas outside of candidate gene regions has also been shown to be significantly associated with birthweight; epigenome-wide association studies (EWAS) have identified CpG sites that are significantly associated with birthweight[56, 66, 67], as well as an association with methylation in long interspersed nuclear elements (LINE)-1[68, 69] and global methylation (defined as the total level of 5mC content relative to total cytosine content)[70].

Previous studies that aim to understand the association between placental methylation and fetal growth have been limited, mainly due to the fact that the methylation signal may be

contaminated by hydroxymethylation (which may or may not be present and have a functional impact), and thus these results may be confounded. It is vital that not only do future studies assess the relationship between placental 5hmC and birthweight, but explore the relationship between 5mC, 5hmC, and birthweight. The latter focus will hopefully allow researchers to understand which modification at a given CpG is acting more robustly on birthweight.

5hmC is formed through oxidation of 5mC and is a stable epigenetic modification that regulates expression.

5hmC is a DNA modification formed through oxidation of 5mC by ten-eleven translocation (TET) methylcytosine dioxygenases[71-73]. It is understood that 5hmC plays an intermediary role in the demethylation pathway, as it is subsequently converted to various other intermediaries before returning to an unmethylated cytosine[74-76] (**Figure 1-4**). Though its role on the demethylation pathway is well-characterized, recent evidence has suggested that 5hmC is also a stable epigenetic modification[77, 78] that may play a variety of roles related to the maintenance of pluripotency in embryonic stem cells (ESCs) and tumorigenesis[79-81]. 5hmC has been found to be associated with expression, as it is generally found in the bodies of actively transcribed genes, and TET1 has been observed at the transcription start site (TSS) of genes with CpG-rich promoters that are decorated with the bivalent histone signature of H3K27me3 and H3K4me3[82-86]. Thus, it is assumed the 5hmC may regulate gene expression through modulating the chromatin accessibility of transcriptional machinery, or by inhibiting the binding of repressors to DNA[82]. Despite these findings, the relationship between 5hmC abundance and gene expression is still evolving; it may be acting in a cell type-dependent regulatory network instead of a simple activation or repression of expression[82], as differentially hydroxymethylated regions (DHMRs) have been uncovered in mouse ESCs and neural progenitor cells (NPCs)[87]. Recent evidence also suggests that tissue type is likely a

major modifier for 5hmC content in human genes, as it has been observed that there may be a > 20-fold change in gene body 5hmC between the same gene in different tissues[88].

Currently, our knowledge of the epigenetic mechanisms regulating placental dynamics is limited, though recent studies have begun to reveal how epigenetic marks are involved in the regulation of placental development processes such as cell fate determination[89], syncytialization[90], and extravillous trophoblasts (EVT) migration and invasion[91]. Additionally, epigenetic regulation of the placenta may be playing a pivotal role in the mediation of developmental programming of chronic diseases. Recent studies employing genome-wide DNA methylation arrays have demonstrated that sexually dimorphic patterns in placental DNA methylation profiles may regulate responses to environmental stimuli and disease susceptibility[92, 93]. Specifically, epigenetic processes in the placenta have been shown to mediate relations between prenatal factors and neurodevelopmental outcomes, with these factors including low infant birthweight[94], maternal depression during pregnancy[95], and maternal obesity and exposure to toxicants[96, 97].

Though some studies have characterized the distribution of placental 5hmC and its role in gene expression[78], few have been able to draw an empirical correlation between 5hmC and transcript levels. Additionally, few studies have attempted to elucidate an association between placental 5hmC and fetal growth. Understanding the epigenetic landscape of 5hmC in the placenta and its association with fetal growth is critical to furthering our understanding of the functionality of the placental epigenome, as well as elucidating these underlying mechanisms that may be programming long-term health outcomes.

Dissertation Overview

The placenta is a critical organ in ensuring successful gestational outcomes, and perturbations to its physiology and function during gestation can lead to restricted fetal growth,

which is a risk indicator for longitudinal health outcomes in infants. It is hypothesized that epigenetic mechanisms are the underlying cause of this developmental programming, as placental methylation has been shown to regulate gene expression, which in turn may influence fetal birth outcomes. However, these studies have assessed methylation using bisulfite-converted methylation arrays, which do not discriminate 5mC from 5hmC. Additionally, few studies have drawn an empirical correlation between placental 5hmC and expression, as well as investigating the association between 5hmC and birthweight outcomes. This dissertation investigates the epigenetic landscape of placental 5hmC and its association with gene expression and birthweight outcomes utilizing epidemiological and molecular data gathered from a large mother-infant cohort from the northeastern United States (**Figure 1-5**). The primary aims of this work are:

Aim 1: Characterize the distribution of 5hmC across the placenta and identify areas of 5hmC that may play a direct role in gene expression, using placenta data collected from the Rhode Island Child Health Study (RICHS). Hypothesis: Placental 5hmC is associated with gene expression in the RICHS cohort.

Aim 2: Investigate the association between placental hydroxymethylation and methylation with birthweight outcomes in RICHS. Hypothesis: Placental 5hmC and 5mC are associated with birthweight in RICHS.

These aims are addressed in the following dissertation chapters. Chapter 2 acts as a set-up to Aim 1 by investigating the influence of early-life environmental exposures on placental methylation. The goal of this study was to review recent literature that demonstrated an association between early-life environmental exposures and DNA methylation in the placenta, in the hopes of understanding how placental methylation is regulated. Chapter 3 addresses Aim 1 by investigating the epigenetic landscape of 5hmC and how it associates with placental gene

expression. The goal of this study was to develop a better understanding of the distribution and functional relevance of 5hmC in the placenta. Chapter 4 addresses Aim 2 by investigating the association between placental hydroxymethylation and methylation on birthweight. The goal of this study was to develop an understanding of how placental 5hmC relates to 5mC in regards to infant birthweight. Lastly, Chapter 5 provides a summation of the research conclusion and discusses future directions.

Figures

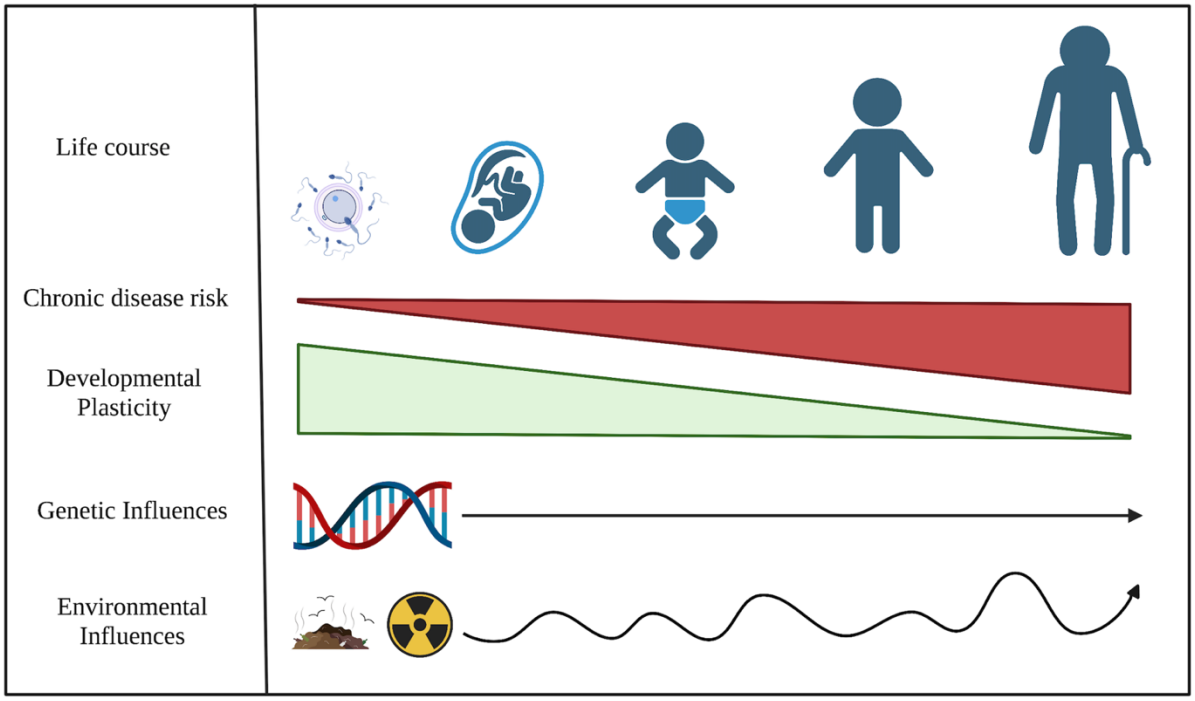


Figure 1-1: Developmental programming of chronic disease. Throughout gestation, plasticity levels are high. This plasticity diminishes following birth, though in utero exposures can increase risk of chronic diseases. Thus, developmental plasticity and chronic disease risk are inversely proportional throughout an individual's lifespan. Genetic influences of developmental programming remain constant throughout lifespan, while environmental influences fluctuate. Created with www.BioRender.com

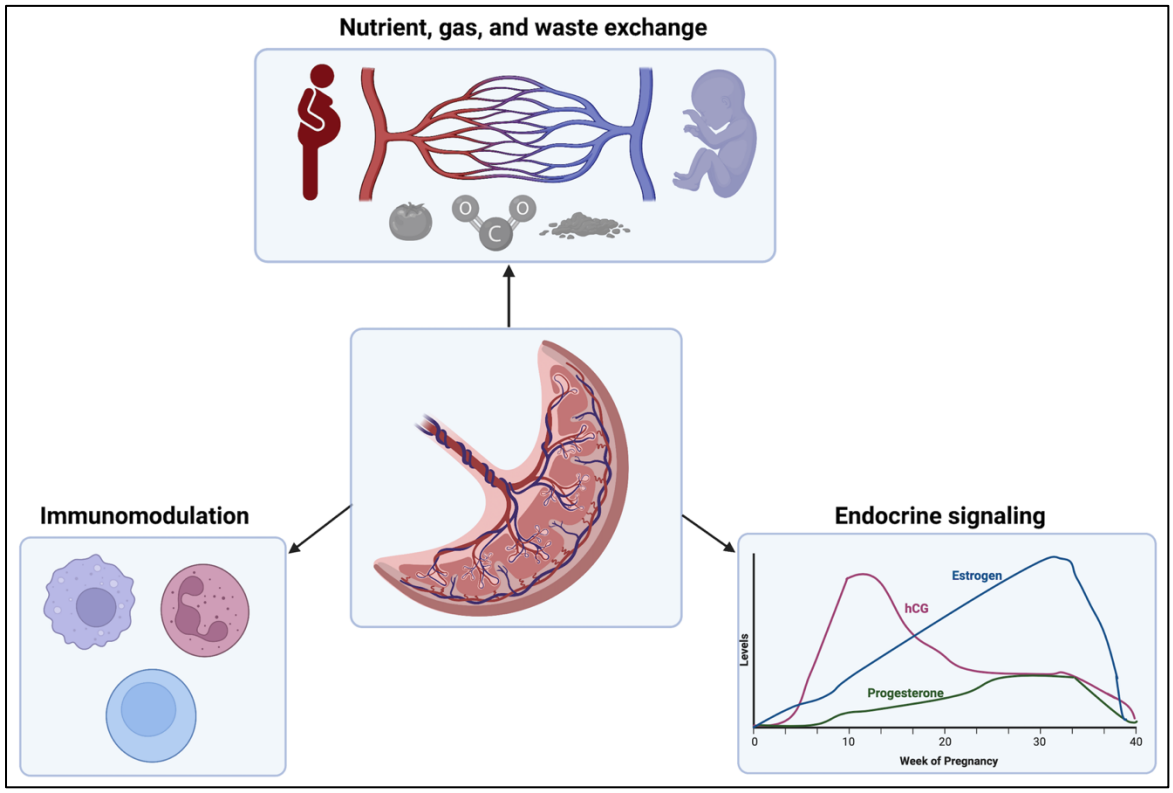


Figure 1-2: Overview of the critical functions of the placenta. The placenta is responsible for a myriad of processes designed to support and sustain gestational growth and development, including 1) exchange of nutrients, gas, and waste between the mother and fetus, 2) immunomodulation, which protects the fetus from harmful pathogens, and 3) endocrine signaling, where it makes a number of hormones to support the developing fetus. Created with www.BioRender.com

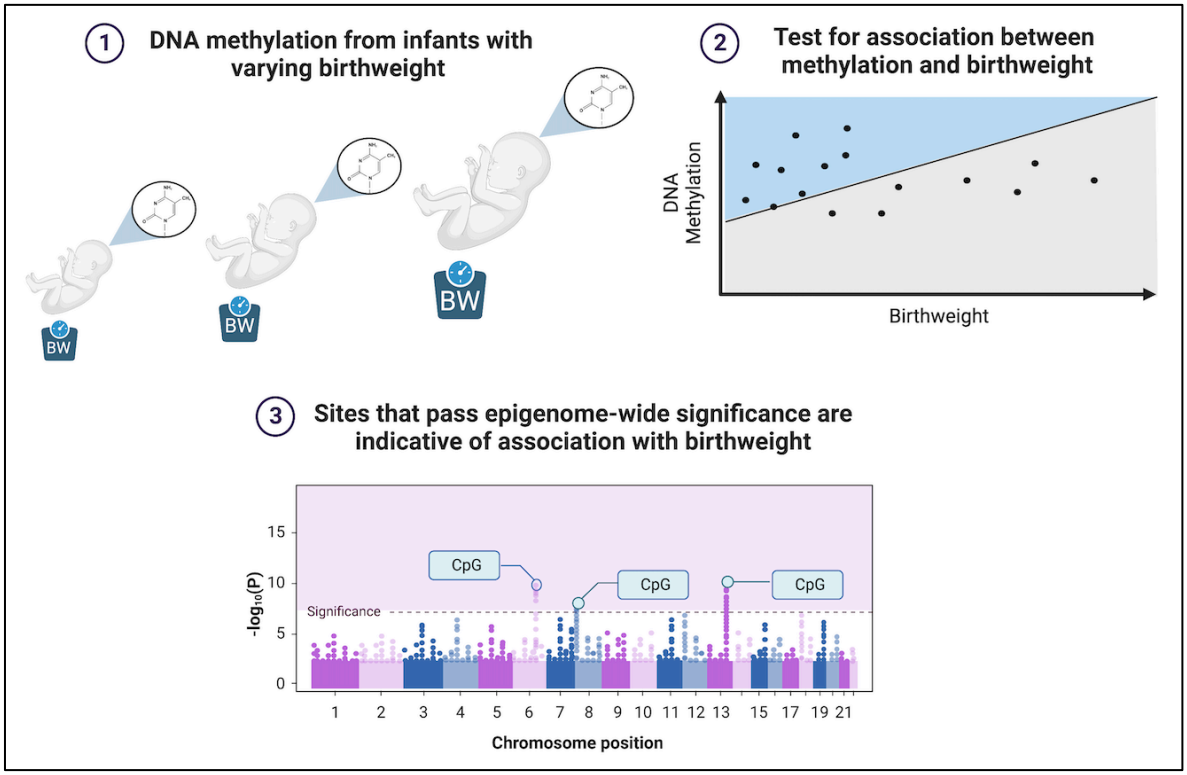


Figure 1-3: Association between DNA methylation and infant birthweight. Methylation is assayed across infants with varying birthweight levels, and an epigenome-wide association study (EWAS) is performed to elucidate an association. EWAS results are then queried to identify CpG sites where methylation is associated with birthweight. Created with www.BioRender.com

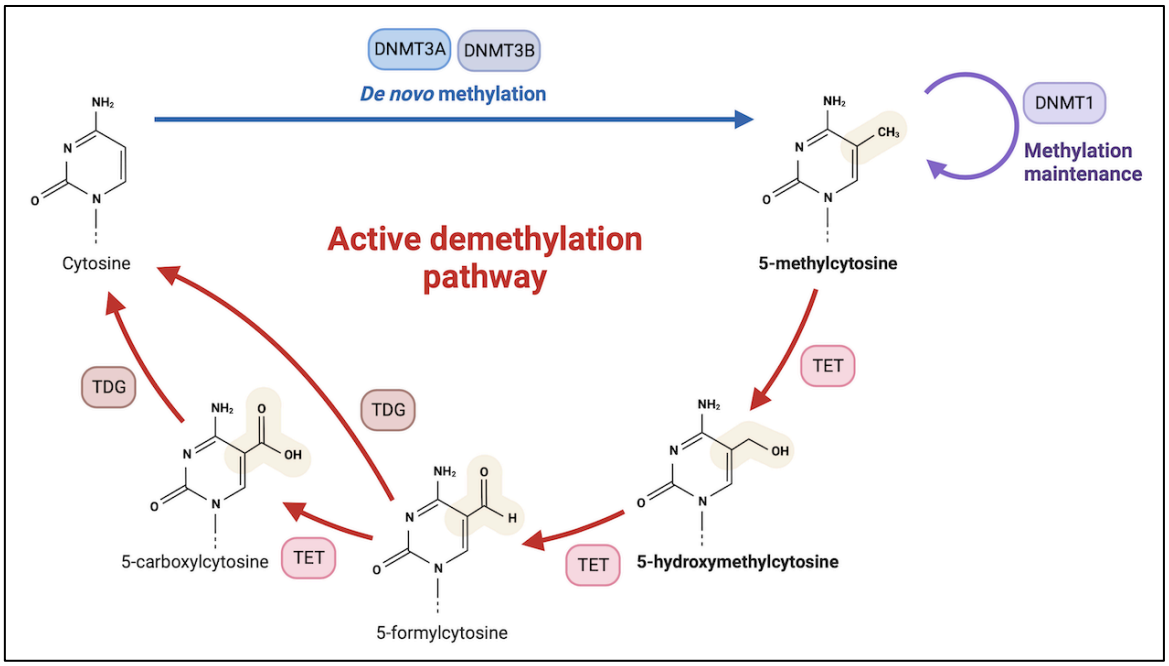


Figure 1-4: Overview of the active demethylation pathway. The pathway begins with 5-methylcytosine (5mC), which has been created through the addition of a methyl group to the 5th carbon position on an unmethylated cytosine via DNA methyltransferases (DNMTs), being oxidized to 5-hydroxymethylcytosine (5hmC) via Ten-Eleven-Translocation (TET) proteins. These TET proteins generate additional intermediates in 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). 5fC and 5caC can be further excised by thymine-DNA glycosylase (TDG) to restore the unmethylated cytosine. Created with www.BioRender.com

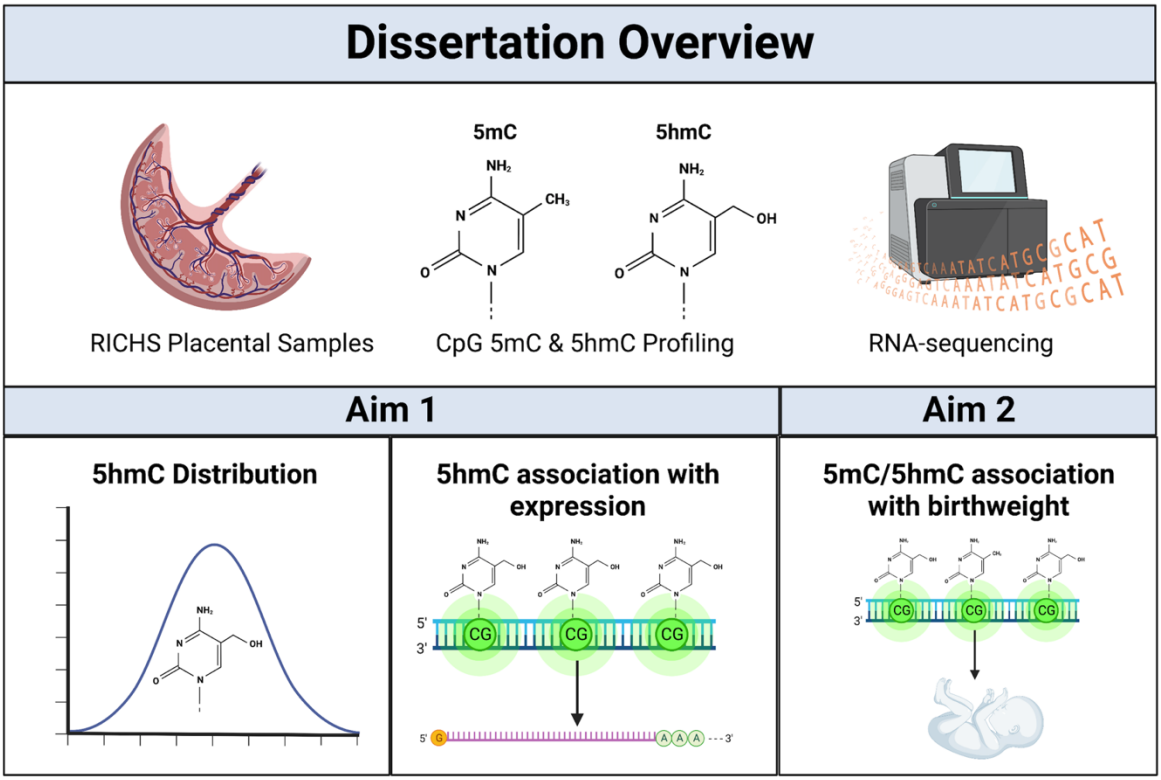


Figure 1-5: Dissertation Overview. Using placental samples collected from the RICHS cohort, we have performed CpG 5mC and 5hmC profiling, along with RNA-sequencing. In **Aim 1**, we plan to characterize the distribution of 5hmC across the placenta, and then investigate the association between 5hmC and expression. In **Aim 2**, we plan to investigate the association between 5mC and 5hmC to birthweight. Created with www.BioRender.com

References

1. Alexander, B.T., J. Henry Dasinger, and S. Intapad, *Effect of low birth weight on women's health.* Clinical therapeutics, 2014. **36**(12): p. 1913-1923.2.
2. Belbasis, L., et al., Birth weight in relation to health and disease in later life: an umbrella review of systematic reviews and meta-analyses. BMC Med, 2016. **14**(1): p. 147.
3. Lester, B.M. and C.J. Marsit, *Epigenetic mechanisms in the placenta related to infant neurodevelopment.* Epigenomics, 2018. **10**(3): p. 321-333.
4. Barker, D.J.P., *Sir Richard Doll Lecture. Developmental origins of chronic disease.* Public Health, 2012. **126**(3): p. 185-189.
5. Barker, E.D., et al., *The role of growth trajectories in classifying fetal growth restriction.* Obstet Gynecol, 2013. **122**(2 Pt 1): p. 248-254.
6. Li, X., et al., *Association between ambient fine particulate matter and preterm birth or term low birth weight: An updated systematic review and meta-analysis.* Environ Pollut, 2017. **227**: p. 596-605.
7. Wang, S., et al., *Changing trends of birth weight with maternal age: a cross-sectional study in Xi'an city of Northwestern China.* BMC Pregnancy Childbirth, 2020. **20**(1): p. 744.
8. Barker, D.J. and C. Osmond, *Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales.* Lancet, 1986. **1**(8489): p. 1077-81.
9. Osmond, C. and D.J. Barker, *Fetal, infant, and childhood growth are predictors of coronary heart disease, diabetes, and hypertension in adult men and women.* Environ Health Perspect, 2000. **108 Suppl 3**(Suppl 3): p. 545-53.
10. Martyn, C.N., D.J. Barker, and C. Osmond, *Mothers' pelvic size, fetal growth, and death from stroke and coronary heart disease in men in the UK.* Lancet, 1996. **348**(9037): p. 1264-8.
11. Osmond, C., et al., *Early growth and death from cardiovascular disease in women.* BMJ, 1993. **307**(6918): p. 1519-24.
12. Forsen, T., et al., *Mother's weight in pregnancy and coronary heart disease in a cohort of Finnish men: follow up study.* BMJ, 1997. **315**(7112): p. 837-40.
13. Frankel, S., et al., *Birthweight, body-mass index in middle age, and incident coronary heart disease.* Lancet, 1996. **348**(9040): p. 1478-80.
14. Barker, D.J., *The origins of the developmental origins theory.* J Intern Med, 2007. **261**(5): p. 412-7.
15. Lapehn, S. and A.G. Paquette, *The Placental Epigenome as a Molecular Link Between Prenatal Exposures and Fetal Health Outcomes Through the DOHaD Hypothesis.* Curr Environ Health Rep, 2022. **9**(3): p. 490-501.
16. Gude, N.M., et al., *Growth and function of the normal human placenta.* Thromb Res, 2004. **114**(5-6): p. 397-407.
17. Maltepe, E. and S.J. Fisher, *Placenta: the forgotten organ.* Annu Rev Cell Dev Biol, 2015. **31**: p. 523-52.
18. Woods, L., V. Perez-Garcia, and M. Hemberger, *Regulation of Placental Development and Its Impact on Fetal Growth-New Insights From Mouse Models.* Front Endocrinol (Lausanne), 2018. **9**: p. 570.
19. Turco, M.Y. and A. Moffett, *Development of the human placenta.* Development, 2019. **146**(22).
20. Wang, Y. and S. Zhao, in *Vascular Biology of the Placenta*. 2010: San Rafael (CA).

21. Goldstein, J.A., et al., *Maternal-Fetal Inflammation in the Placenta and the Developmental Origins of Health and Disease.* Front Immunol, 2020. **11**: p. 531543.
22. Mortillo, M. and C.J. Marsit, *Select Early-Life Environmental Exposures and DNA Methylation in the Placenta.* Curr Environ Health Rep, 2023. **10**(1): p. 22-34.
23. Burton, G.J. and A.L. Fowden, *The placenta: a multifaceted, transient organ.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1663): p. 20140066.
24. Cross, J.C., et al., *Genes, development and evolution of the placenta.* Placenta, 2003. **24**(2-3): p. 123-30.
25. Bronson, S.L. and T.L. Bale, *The Placenta as a Mediator of Stress Effects on Neurodevelopmental Reprogramming.* Neuropsychopharmacology, 2016. **41**(1): p. 207-18.
26. Herrick, E.J. and B. Bordoni, *Embryology, Placenta*, in *StatPearls*. 2022: Treasure Island (FL).
27. Punshon, T., et al., *Placental metal concentrations in relation to placental growth, efficiency and birth weight.* Environ Int, 2019. **126**: p. 533-542.
28. Gaccioli, F. and S. Lager, *Placental Nutrient Transport and Intrauterine Growth Restriction.* Front Physiol, 2016. **7**: p. 40.
29. Kreis, N.N., et al., *A Message from the Human Placenta: Structural and Immunomodulatory Defense against SARS-CoV-2.* Cells, 2020. **9**(8).
30. Kshirsagar, S.K., et al., *Immunomodulatory molecules are released from the first trimester and term placenta via exosomes.* Placenta, 2012. **33**(12): p. 982-90.
31. Rutherford, J.N., *Fetal signaling through placental structure and endocrine function: illustrations and implications from a nonhuman primate model.* Am J Hum Biol, 2009. **21**(6): p. 745-53.
32. Stern, C., et al., *Placental Endocrine Activity: Adaptation and Disruption of Maternal Glucose Metabolism in Pregnancy and the Influence of Fetal Sex.* Int J Mol Sci, 2021. **22**(23).
33. Tarrade, A., et al., *Placental contribution to nutritional programming of health and diseases: epigenetics and sexual dimorphism.* J Exp Biol, 2015. **218**(Pt 1): p. 50-8.
34. Fowden, A.L., et al., *The placenta and intrauterine programming.* J Neuroendocrinol, 2008. **20**(4): p. 439-50.
35. Jansson, T. and T.L. Powell, *Role of the placenta in fetal programming: underlying mechanisms and potential interventional approaches.* Clin Sci (Lond), 2007. **113**(1): p. 1-13.
36. Myatt, L., *Placental adaptive responses and fetal programming.* J Physiol, 2006. **572**(Pt 1): p. 25-30.
37. Watson, E.D. and J.C. Cross, *Development of structures and transport functions in the mouse placenta.* Physiology (Bethesda), 2005. **20**: p. 180-93.
38. Audette, M.C. and J.C. Kingdom, *Screening for fetal growth restriction and placental insufficiency.* Semin Fetal Neonatal Med, 2018. **23**(2): p. 119-125.
39. Nugent, B.M. and T.L. Bale, *The omniscient placenta: Metabolic and epigenetic regulation of fetal programming.* Front Neuroendocrinol, 2015. **39**: p. 28-37.
40. Leiva, A., et al., *Maternal hypercholesterolemia in pregnancy associates with umbilical vein endothelial dysfunction: role of endothelial nitric oxide synthase and arginase II.* Arterioscler Thromb Vasc Biol, 2013. **33**(10): p. 2444-53.

41. Sobrevia, L., et al., *Review: Differential placental macrovascular and microvascular endothelial dysfunction in gestational diabetes.* Placenta, 2011. **32 Suppl 2**: p. S159-64.
42. Bhattacharya, A., et al., *Placental genomics mediates genetic associations with complex health traits and disease.* Nat Commun, 2022. **13**(1): p. 706.
43. Kass, S.U., D. Pruss, and A.P. Wolffe, *How does DNA methylation repress transcription?* Trends in Genetics, 1997. **13**(11): p. 444-449.
44. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function.* Neuropsychopharmacology, 2013. **38**(1): p. 23-38.
45. Breiling, A. and F. Lyko, *Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond.* Epigenetics Chromatin, 2015. **8**: p. 24.
46. Kumar, S., V. Chinnusamy, and T. Mohapatra, *Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond.* Front Genet, 2018. **9**: p. 640.
47. Taryma-Lesniak, O., K.E. Sokolowska, and T.K. Wojdacz, *Short history of 5-methylcytosine: from discovery to clinical applications.* J Clin Pathol, 2021. **74**(11): p. 692-696.
48. Wang, L., G. Ren, and B. Lin, *Expression of 5-methylcytosine regulators is highly associated with the clinical phenotypes of prostate cancer and DNMTs expression predicts biochemical recurrence.* Cancer Med, 2021. **10**(16): p. 5681-5695.
49. Turpin, M. and G. Salbert, *5-methylcytosine turnover: Mechanisms and therapeutic implications in cancer.* Front Mol Biosci, 2022. **9**: p. 976862.
50. Bartolomei, M.S. and S.M. Tilghman, *Genomic imprinting in mammals.* Annu Rev Genet, 1997. **31**: p. 493-525.
51. Panning, B. and R. Jaenisch, *RNA and the epigenetic regulation of X chromosome inactivation.* Cell, 1998. **93**(3): p. 305-8.
52. Kubiura, M., et al., *Chromosome-wide regulation of euchromatin-specific 5mC to 5hmC conversion in mouse ES cells and female human somatic cells.* Chromosome Res, 2012. **20**(7): p. 837-48.
53. Ponnaluri, V.K., et al., *Association of 5-hydroxymethylation and 5-methylation of DNA cytosine with tissue-specific gene expression.* Epigenetics, 2017. **12**(2): p. 123-138.
54. Nagase, H. and S. Ghosh, *Epigenetics: differential DNA methylation in mammalian somatic tissues.* FEBS J, 2008. **275**(8): p. 1617-23.
55. Smith, Z.D. and A. Meissner, *DNA methylation: roles in mammalian development.* Nat Rev Genet, 2013. **14**(3): p. 204-20.
56. Tekola-Ayele, F., et al., *DNA methylation loci in placenta associated with birthweight and expression of genes relevant for early development and adult diseases.* Clin Epigenetics, 2020. **12**(1): p. 78.
57. Piyasena, C., et al., *Placental 5-methylcytosine and 5-hydroxymethylcytosine patterns associate with size at birth.* Epigenetics, 2015. **10**(8): p. 692-7.
58. Constancia, M., et al., *Placental-specific IGF-II is a major modulator of placental and fetal growth.* Nature, 2002. **417**(6892): p. 945-8.
59. DeChiara, T.M., A. Efstratiadis, and E.J. Robertson, *A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting.* Nature, 1990. **345**(6270): p. 78-80.
60. Tunster, S.J., A.B. Jensen, and R.M. John, *Imprinted genes in mouse placental development and the regulation of fetal energy stores.* Reproduction, 2013. **145**(5): p. R117-37.

61. Tunster, S.J., B. Tycko, and R.M. John, *The imprinted Phlda2 gene regulates extraembryonic energy stores.* Mol Cell Biol, 2010. **30**(1): p. 295-306.
62. Guo, L., et al., *Altered gene expression and methylation of the human chromosome 11 imprinted region in small for gestational age (SGA) placentae.* Dev Biol, 2008. **320**(1): p. 79-91.
63. Moore, G.E., et al., *The role and interaction of imprinted genes in human fetal growth.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1663): p. 20140074.
64. Rahnama, F., et al., *Epigenetic regulation of human trophoblastic cell migration and invasion.* Endocrinology, 2006. **147**(11): p. 5275-83.
65. Rahnama, F., et al., *Epigenetic regulation of E-cadherin controls endometrial receptivity.* Endocrinology, 2009. **150**(3): p. 1466-72.
66. Banister, C.E., et al., *Infant growth restriction is associated with distinct patterns of DNA methylation in human placentas.* Epigenetics, 2011. **6**(7): p. 920-7.
67. Chen, P.Y., et al., *Prenatal Growth Patterns and Birthweight Are Associated With Differential DNA Methylation and Gene Expression of Cardiometabolic Risk Genes in Human Placentas: A Discovery-Based Approach.* Reprod Sci, 2018. **25**(4): p. 523-539.
68. Bourque, D.K., et al., *Decreased placental methylation at the H19/IGF2 imprinting control region is associated with normotensive intrauterine growth restriction but not preeclampsia.* Placenta, 2010. **31**(3): p. 197-202.
69. Michels, K.B., H.R. Harris, and L. Barault, *Birthweight, maternal weight trajectories and global DNA methylation of LINE-1 repetitive elements.* PLoS One, 2011. **6**(9): p. e25254.
70. Dwi Putra, S.E., et al., *Being Born Large for Gestational Age is Associated with Increased Global Placental DNA Methylation.* Sci Rep, 2020. **10**(1): p. 927.
71. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* Science, 2009. **324**(5929): p. 930-5.
72. Ito, S., et al., *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.* Nature, 2010. **466**(7310): p. 1129-33.
73. Huang, Y., et al., *Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells.* Proc Natl Acad Sci U S A, 2014. **111**(4): p. 1361-6.
74. Klug, M., et al., *5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes.* Genome Biol, 2013. **14**(5): p. R46.
75. Stoyanova, E., et al., *5-Hydroxymethylcytosine-mediated active demethylation is required for mammalian neuronal differentiation and function.* Elife, 2021. **10**.
76. Yamaguchi, S., et al., *Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming.* Cell Res, 2013. **23**(3): p. 329-39.
77. Bachman, M., et al., *5-Hydroxymethylcytosine is a predominantly stable DNA modification.* Nat Chem, 2014. **6**(12): p. 1049-55.
78. Green, B.B., et al., *Hydroxymethylation is uniquely distributed within term placenta, and is associated with gene expression.* FASEB J, 2016. **30**(8): p. 2874-84.
79. Dawlaty, M.M., et al., *Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development.* Cell Stem Cell, 2011. **9**(2): p. 166-75.
80. Lian, C.G., et al., *Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma.* Cell, 2012. **150**(6): p. 1135-46.

81. Thomson, J.P., et al., *Loss of Tet1-Associated 5-Hydroxymethylcytosine Is Concomitant with Aberrant Promoter Hypermethylation in Liver Cancer.* Cancer Res, 2016. **76**(10): p. 3097-108.
82. Shi, D.-Q., et al., *New Insights into 5hmC DNA Modification: Generation, Distribution and Function.* Frontiers in Genetics, 2017. **8**.
83. Branco, M.R., G. Ficz, and W. Reik, *Uncovering the role of 5-hydroxymethylcytosine in the epigenome.* Nat Rev Genet, 2011. **13**(1): p. 7-13.
84. Ficz, G., et al., *Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation.* Nature, 2011. **473**(7347): p. 398-402.
85. Szulwach, K.E., et al., *Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells.* PLoS Genet, 2011. **7**(6): p. e1002154.
86. Pastor, W.A., et al., *Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells.* Nature, 2011. **473**(7347): p. 394-7.
87. Tan, L., et al., *Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method.* Nucleic Acids Res, 2013. **41**(7): p. e84.
88. Nestor, C.E., et al., *Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes.* Genome Res, 2012. **22**(3): p. 467-77.
89. Wu, B., et al., *Epigenetic regulation of Elf5 is associated with epithelial-mesenchymal transition in urothelial cancer.* PLoS One, 2015. **10**(1): p. e0117510.
90. Novakovic, B., et al., *Increased methylation and decreased expression of homeobox genes TLX1, HOXA10 and DLX5 in human placenta are associated with trophoblast differentiation.* Sci Rep, 2017. **7**(1): p. 4523.
91. Gamage, T., et al., *The role of DNA methylation in human trophoblast differentiation.* Epigenetics, 2018. **13**(12): p. 1154-1173.
92. Martin, E., et al., *Sexual epigenetic dimorphism in the human placenta: implications for susceptibility during the prenatal period.* Epigenomics, 2017. **9**(3): p. 267-278.
93. Barouki, R., et al., *Epigenetics as a mechanism linking developmental exposures to long-term toxicity.* Environ Int, 2018. **114**: p. 77-86.
94. Marsit, C.J., et al., *Placental 11-beta hydroxysteroid dehydrogenase methylation is associated with newborn growth and a measure of neurobehavioral outcome.* PLoS One, 2012. **7**(3): p. e33794.
95. Conradt, E., et al., *The roles of DNA methylation of NR3C1 and 11beta-HSD2 and exposure to maternal mood disorder in utero on newborn neurobehavior.* Epigenetics, 2013. **8**(12): p. 1321-9.
96. Maccani, J.Z., et al., *Placental DNA Methylation Related to Both Infant Toenail Mercury and Adverse Neurobehavioral Outcomes.* Environ Health Perspect, 2015. **123**(7): p. 723-9.
97. Stroud, L.R., et al., *Prenatal Major Depressive Disorder, Placenta Glucocorticoid and Serotonergic Signaling, and Infant Cortisol Response.* Psychosom Med, 2016. **78**(9): p. 979-990.

Chapter 2 - Select early-life environmental exposures and DNA methylation in the placenta

Michael Mortillo[1], Carmen J. Marsit[1]

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

Abstract

Purpose of review
To summarize recent literature relating early-life environmental exposures on DNA methylation in the placenta, to identify how variation in placental methylation is regulated in an exposure-specific manner and to encourage additional work in this area.

Recent findings
Multiple studies have evaluated associations between prenatal environmental exposures and placental methylation in both gene-specific and epigenome-wide frameworks. Specific exposures lead to unique variability in methylation, and cross-exposure assessments have uncovered certain genes that demonstrate consistency in differential placental methylation. Exposure studies that assess methylation effects in a trimester-specific approach tend to find larger effects during 1st trimester exposure. Earlier studies have more targeted gene-specific approaches to methylation, while later studies have shifted towards epigenome-wide, array-based approaches. Studies focusing on exposures such as air pollution, maternal smoking, environmental contaminants, and trace metals appear to be more abundant, while studies of socioeconomic adversity and circadian disruption are scarce but demonstrate remarkable effects.

Summary
Understanding the impacts of early-life environmental exposures on placental methylation is critical to establishing the link between the maternal environment, epigenetic variation, and long-term health. Future studies into this field should incorporate repeated measures of exposure throughout pregnancy, in order to determine the critical windows in which placental methylation is most heavily affected. Additionally, the use of methylation-based scores and sequencing technology could provide important insights into epigenetic gestational age and uncovering more genomic regions where methylation is affected. Studies examining the impact of other exposures on methylation, including pesticides, alcohol, and other chemicals are also warranted.

Introduction

DNA methylation is the best-characterized and most stable epigenetic modification, influencing gene expression through the disruption of transcription factor binding, chromatin structure, and subsequent gene silencing [1-3]. DNA methylation typically involves methylation of the 5th carbon position at a cytosine residue within a CpG dinucleotide (CpG), resulting in 5-methylcytosine (5mC). There are ~28 million CpGs in the genome, most of which are methylated [4]. However, CpGs located in "CpG islands" (regions of high CpG density and commonly found in gene promoters) tend to be unmethylated, as methylation in these regions generally represses transcription of the gene [5]. DNA methylation is primarily measured in this locus/gene-specific way, though other forms of measuring methylation do exist. One such example is building a methylation profile across several unique sites in the genome, through a genome-wide analysis [6]. Global methylation is another form of methylation profiling, and is defined as the total level of 5mC content in a sample relative to total cytosine content [6]. Though these types of methylation are generally studied in nuclear DNA, other types of DNA have proven to be useful tools for understanding methylation. Altered methylation of mitochondrial DNA (mtDNA) which is located outside the nucleus and plays an important role in cell life and death [7], has been implicated in a number of human diseases including cancer [8], and cardiovascular disease (CVD) [9].

DNA methylation can be influenced by both genetic and environmental factors, and recent studies have provided concrete evidence of a link between methylation and certain environmental exposures, including tobacco smoke [10-12], air pollution [13-15], toxic metals [16-18], and chemical compounds [19-21]. Exposure to these contaminants, particularly early in life, is associated with an increased later-life disease risk [22-24] (**Fig. 2-1**). Much of this literature linking the maternal environment, epigenetic variation, and developmental

programming has focused on DNA methylation in cord blood, due to its availability in birth cohort studies and its utility as a surrogate marker of target offspring tissue or as a target itself of environmental impacts on the developing immune system [25, 26].

The placenta, though, may also be a highly relevant target tissue for the environment during in utero development. This ephemeral organ acts as a regulator of the intrauterine environment and physiologic interface between the mother and developing fetus. It is the first organ to develop and plays an important role throughout pregnancy, coordinating nutrient, gas, and waste exchange, as well as acting as an immunologic and endocrine organ [27-29]. An overview of placental structure and its role in function is described in **Figure 2-2.** Placental development is essential for proper fetal development, and can be influenced by the maternal environment [30, 31]. Exogenous exposures including environmental contaminants, pharmaceuticals, and psychosocial factors, as well as endogenous characteristics including maternal metabolic state contribute to that maternal environment and subsequently influence fetal development and potentially lifelong offspring health [32-34] through interactions with or effects on the placenta [35, 36]. Those environmental impacts to the placenta can be reflected in its molecular landscape, including changes to gene and protein expression and the upstream mechanisms which control these cellular products, in particularly DNA methylation, which exhibits a unique profile in the placental genome [37].

In this review, we summarize the existing literature relating early-life environmental exposures on DNA methylation in the placenta, to identify how variation in placental methylation is regulated in an exposure-specific manner and to encourage additional work in this area.

Methods

In December 2021, we searched PubMed for literature concerning placental DNA methylation and early-life environmental exposures. We utilized general keywords “placenta”, “methylation”, “environment”, “exposures”, as well as specific exposure keywords “smoking”, “pollution”, “metals”, “chemicals”, “circadian”, and “socioeconomic.” From our queries, we included studies published in English within the last 5 years that represented new research on placental methylation in relation to in-utero environmental exposures and were either open access or accessible through a typical University library. We excluded studies that were systematic review articles, non-human studies, and studies where environmental exposures’ impacts were assessed in-vitro. If multiple papers assessed the same environmental exposure and methylation type in nested, overlapping sample sources, only the paper with the largest sample size was retained. Papers that met inclusion criteria but were later found to have a scope not matching inclusion criteria were also excluded. See **Figure 2-3** for flow chart of literature search strategy. For papers meeting inclusion criteria, relevant data, including paper title, authors, year and place of publication, cohort size and source, exposure and specific type of DNA methylation assessed, research question, methods, and key findings, was extracted and maintained in a local database.

Results

A total of 108 studies were initially retrieved for screening and assessed for possible inclusion. After exclusion of non-pertinent articles, 28 studies met final inclusion criteria for assessing placental DNA methylation in relation to an environmental exposure; 9 (32.1%) studied air pollution, 8 (28.6%) studied maternal smoking, 4 (14.3%) studied environmental chemicals, 4 (14.3%) studied trace metals, 2 (7.1%) studied socioeconomic status (SES), and 1

(3.6%) studied circadian disruption **(Fig. 2-3).** Results from all studies are summarized in **Table 2-1.**

Air pollution

From our review, exposure to air pollution was the most highly-studied environmental exposure in relation to placental DNA methylation. One study assessed global placental DNA methylation based on $PM_{2.5}$ exposure during different during different time windows of pregnancy [38], observing an overall decrease in methylation as exposure to $PM_{2.5}$ increased during whole pregnancy, particularly with exposure during the 1st trimester and specifically the early first trimester when implantation occurs. A separate study assessed methylation of 7 CpGs in the promoter of leptin, a hormone that plays a functional role in embryo implantation, intrauterine development, and fetal growth during pregnancy [39]. The authors observed decreased methylation across all 7 sites among mothers with increased $PM_{2.5}$ exposure in the 2nd trimester [40].

A 2018 study assessed placental methylation among genes within the circadian clock pathway in response to $PM_{2.5}$ exposure [41]. The authors observed that increased $PM_{2.5}$ exposure during 3rd trimester led to increased DNA methylation in the promoters of NPAS2, CRY1, PER2, and PER3, while an inverse association was seen between 1st trimester exposure and CLOCK methylation. Another study examined exposure to $PM_{2.5}$, black carbon, and NO_2 on promoter methylation of tumor suppressor and DNA repair genes, including genes on the nucleotide excision repair (NER) and base excision repair (BER) pathways [42]. Increased $PM_{2.5}$ exposure throughout pregnancy was found to be positively associated with promoter methylation of repair genes APEX1, OGG1, and ERCC4, as well as tumor-suppressor gene p53. Increased black carbon exposure was also positively associated with methylation in APEX and ERCC4. Interestingly, pollution exposure during the 1st and 2nd trimesters of pregnancy mainly affected

methylation of tumor suppressor genes, whereas later pregnancy exposure affected genes of the BER pathway. These gene and trimester-specific trends were observed in additional studies, including one that studied $PM_{2.5}$ exposure on promoter methylation among 5 candidate placental genes and its role as a mediator of the exposure's impact on fetal growth [43]. The authors noted that in the case of IGF2, a growth hormone that plays a crucial role in fetal development [44], increased $PM_{2.5}$ exposure during 2nd or 3rd trimester and entire pregnancy was associated with decreases in promoter methylation, while increases in exposure across the same windows resulted in increased promoter methylation of BID, an apoptosis regulator that has been shown to be susceptible to oxidative stress and immune response induced by environmental risk factors [45, 46]. A mediation analysis also showed that $PM_{2.5}$ exposure might influence fetal growth through BID methylation.

Several studies used qualitative measures of air pollution exposure, grouping their participants based on exposure levels. This included a 2018 study done in Tehran, Iran, assessing impact of $PM_{2.5}$ and PM_{10} exposure on global placental DNA methylation [14]. Positive correlations were observed between $PM_{2.5}$/PM_{10} exposure in 1st trimester and methylation of all participants in "polluted" and "non polluted" groups. Stronger correlations were also seen in the "polluted" group compared to "non-polluted" group. Another study assessed prenatal $PM_{2.5}$ exposure on placental methylation and how these changes modulate vitamin D deficiency and atopic dermatitis in offspring [47]. They observed significant hypomethylation in the promoter of the AHRR gene, as well as decreased expression of AHRR targets (AHR), in mothers with high $PM_{2.5}$ exposure, low cord blood vitamin D, and offspring atopic dermatitis compared to other groups. AHR is a transcription factor that responds to chemicals regulating expression of genes with toxic or protective effects [48]. Placental hypomethylation of AHRR could suppress

expression of AHR, thereby decreasing AHR signaling and disrupting immune response, which could increase risk of atopic dermatitis in offspring.

Several studies also investigated genomic regions that exhibit differential methylation patterns in response to varying levels of pollution exposure. One study assessed NO_2 and PM_{10} exposure on global methylation in Alu and LINE-1 repetitive elements, as well as in specific CpG-sites [49]. The authors identified 27 differentially methylated regions (DMRs) associated with air pollutants, including 4 located in CD81, DAXX, NOTCH3, and P2RX4 genes, all of which have been implicated in preeclampsia phenotypes [50-53]. A similar trend was recently reported in a study looking at NO_2 and O_3 exposure impact on methylation across genomic locations, including CpG islands, shores, shelves, and open seas [54]. The authors noted location-specific variability in methylation, but more importantly they identified 5 hypomethylated DMRs in placenta mapped to genes ZNF442, PTPRH, SLC25A44, F11R, and STK38. Several of these genes are involved in immune and inflammatory processes, and these processes have been implicated as biological targets of air pollutant exposure [55, 56].

Studies of air pollution are often unique in that they can consider various time windows of exposure in pregnancy, something that is generally not seen in studies requiring biomarkers of exposure, due to the burden and cost of sampling multiple times throughout pregnancy [57]. Several genes exhibiting variable methylation in response to air pollution exposure have been previously implicated in environmental exposure-induced changes to immune response and oxidative stress [43, 47], making them good candidates for future studies on how changes in their methylation levels may impact these processes. Finally, a few studies assessed global methylation in the form of LINE-1 and Alu non-coding, repetitive elements, but such studies of

global patterns of methylation are becoming more rare, replaced with studies of gene-specific and genome-wide methylation, made possible with array-based technologies and sequencing.

Maternal smoking

There is also an abundance of studies focusing on maternal smoking either prior to or during pregnancy, and placental methylation. In one study [58], investigators aimed to assess the effect of prenatal exposure to smoking on methylation of mtDNA and in the promoter of the CYP1A1 gene, which is involved in detoxification and may be activated by constituents of tobacco smoke [59]. The authors noted that direction of effect varies based on the type of DNA (genomic vs. mtDNA); current smokers had neonates with lower CpG-specific methylation at CYP1A1 compared to non-smokers, but higher mtDNA methylation at specific loci (specifically the MT-RNR1 gene). Another study that also assessed CYP1A1 promoter methylation found no association with prenatal smoking exposure, though it is worth nothing that this study only sampled placenta from pre-term births [60]. This may denote that early-pregnancy smoking is not sufficient to elicit CYP1A1 methylation changes. Likewise, a recent study [61] that also assessed placental mtDNA methylation found significantly higher D-loop methylation in smokers compared to non-smokers, but no difference in methylation of LDLR2 between groups. The D-loop and LDLR2 both lie on the displacement loop of mtDNA, with D-loop on the heavy chain and LDLR2 on the light chain [62, 63]. These findings suggest that prenatal exposure to smoking may have a differential impact on the displacement loop of placental mtDNA.

Recently, a large epigenome-wide association study (EWAS) aimed to understand if placental methylation would mediate the established association of prenatal smoking and lower infant birthweight [64]. Authors found 153 DMRs between smokers and non-smokers, with increases and decreases based on exposure. Interestingly, methylation of 7 CpGs had a mediating effect of lower birthweight in smokers. A separate study looking at impacts of Vitamin C

supplementation on smoking-associated changes in placental methylation found differential methylation between those groups and suggested Vitamin C could be a potential intervention [65].

One study investigated maternal smoking effects on tissue-specific impacts in the placenta [66]. Authors observed significant decreases in methylation on the fetal side of the placenta among those who smoked throughout pregnancy, and, to a lesser extent, quit during pregnancy compared to non-smokers. Methylation levels on maternal sides of the placenta displayed less significant changes between groups, suggesting that smoking-induced alterations reflect smoking through pregnancy rather than long-term smoking history, under the assumption that maternal tissue would already have the smoking associated differences. along the lines of understanding the "biological memory" of smoking on methylation [67], a recent EWAS found that ~ 75% of DMRs showed "reversible" alterations of DNA methylation present in placentas of current smokers, and 26 of these DMRs were also present in placentas of former smokers (which were not exposed directly to cigarette smoke), suggesting an "epigenetic memory" of placentas exposed to smoking prior to pregnancy. These DMRs also contained "enhancer-like" epigenetic marks enriched for H3K4me1 and H3K27ac, suggesting these placenta enhancer regions are particularly sensitive to tobacco smoke [68].

A recent EWAS performed by the Pregnancy and Childhood Epigenetics (PACE) consortium [69] represents the largest study of smoking and placental methylation, and serves as a complement to their prior work linking smoking and cord blood methylation [11]. This analysis identified over 400 differentially methylated CpG sites, with very few overlapping with those identified in cord blood, and many exhibiting effect estimates substantially larger than those observed in blood. Additionally, many of these CpGs were associated with smoking-related birth

outcomes, including preterm birth and small birthweight, lending support to the critical role that the placenta plays in mediating the impact of adverse environmental exposures on newborn health.

Environmental chemicals

Environmental chemicals including persistent organic pollutants (POPs), bisphenol A (BPA), and polybrominated diphenyl ethers (PBDEs), also emerged as widely-studied prenatal exposure related to placental DNA methylation. Effects on placental methylation varied in both methylation-specific and chemical-specific contexts. A recent study conducted in the Shanghai-Minhang Birth Cohort Study (S-MBCS) assessed prenatal BPA exposure on CpG-specific methylation in the placenta [70]. Authors noted a trend towards hypermethylation of CpGs in the high BPA-exposed group. This result aligns with a previous study in mouse placentas that showed increased BPA exposure led to increased DNA methylation (and subsequently reduced expression) of WNT-2, a gene involved in cell proliferation and differentiation during embryogenesis [71, 72]. A similar association was seen in an EWAS evaluating prenatal exposure to 9 synthetic phenols on methylation [73]. Researchers identified 596 phenol-associated CpG sites, with > 90% of sites showing a positive association between urinary phenol concentration and methylation. A similar association was seen in 97% of DMRs (n = 180), with many of these DMRs related to exposure to triclosan, which has been shown in mice to have adverse effects on placental development and nutrient transport [74, 75]. Triclosan-associated DMRs were also shown to overlap with imprinted genes, suggesting triclosan could impact these important drivers of fetal development.

Examining imprinted placental genes was further investigated in a couple of studies. The first [76] was a 2018 study looking at exposure to POPs on IGF2 and H19, expressed from the paternal and maternal allele, respectively, on chromosome 11. IGF2 lies upstream of H19, and

when it is down-regulated, expression of H19 increases. Decreased IGF2 levels have been shown to impair fetal growth [77]. Researchers observed significant hypermethylation of CpGs in IGF2 and hypomethylation in H19 in response to higher serum POP concentrations. Increased methylation in IGF2 may lead to decreased expression, and since IGF2 is known to be a major regulator of placental and fetal growth [78], this could severely impair fetal growth. The second study assessed methylation of IGF2 and HSD11B2, a non-imprinted gene, in response to in-utero exposures to PBDEs among a cohort of children with fetal growth restriction (FGR) [79]. In contrast to the 2018 study, authors observed hypomethylation at IGF2-associated CpGs in response to increased cord blood concentrations of BDE-17-190, a PBDE congener. However, they noted a positive association between that congener and methylation of HSD11B2. HSD11B2 converts maternal cortisol to cortisone [80], and hypermethylation of its promoter has been shown to reduce placental expression [81], leading to increased cortisol being able to enter fetal circulation. High levels can affect fetal growth [82], which may explain the prevalence of FGR in these infants.

Overall, the reviewed chemical-methylation association studies showed variability in methylation in a chemical-and-gene specific framework. Specifically, BPA and phenol exposures appeared to show a more positive association with placental methylation, while POP and PBDE exposures tend to show more variability, with opposite effect directions based on specific genes. One trend to note was that we did not find any recent placental methylation studies looking at exposure to specific pesticides, although there are a number of such studies greater than five years ago.

Trace metals

Our group has been at the forefront of understanding the effects of prenatal trace metal exposures on placental methylation with a number of studies conducted by our group within the

Rhode Island Child Health Study (RICHS) or New Hampshire Birth Cohort Study (NHBCS) which have extensive data on placental DNA methylation. One study [83] looked at concentrations of various neurotoxic metals and their impact on methylation of NR3C1, a glucocorticoid receptor that has been linked to neurobehavioral outcomes at birth [84, 85], and is involved in the development of a child's hypothalamic-pituitary-adrenal (HPA) axis [86]. We found that higher levels of Arsenic (As), Cadmium (Cd), Lead (Pb), Manganese (Mn), and Mercury (Hg), and lower levels of Zinc (Zn) were associated with increased methylation of NR3C1. We noted that this metal-induced hypermethylation may reduce expression of NR3C1, thereby affect the child's developing HPA axis, which may increase cognitive and neurodevelopmental risk later in life.

Cadmium (Cd) is a unique metal to examine in the placenta, given that it can be sequestered in the placenta and not pass into fetal circulation, and so may elicit its toxicity within the placenta itself [87]. In a study of Cd-associated placental methylation [88], we found that increased Cd concentrations in the placenta was associated with differential methylation at 17 CpGs. Additionally, these Cd-associated CpGs may play a functional role in birth outcomes as methylation at specific CpGs was associated with increased expression of genes such as TNFAIP2 and ACOT7. Higher expression of these genes is associated with lower birthweight in our cohorts. In another study [89], we identified that increasing placental selenium was associated with increased methylation of a CpG in the GFI1 gene, and that methylation of that gene was associated with greater muscle tone in the infants. In an EWAS [90] on placental copper (Cu) and DNA methylation, we identified Cu-associated differentially methylated sites and regions, including the antioxidant GSTP1 gene, and the ZNF197 transcription factor, which has as transcriptional targets a number of Cu metabolism genes [91]. These studies of metals

suggest that DNA methylation likely plays a role in the impacts of these metals on various fetal and newborn health outcomes and may be involved in the regulation of those metals in the placenta.

Socioeconomic adversity

It is also important to consider how demographic, social, and structural factors, which contribute to exposure to adverse environmental agents, also impact placental DNA methylation and may act to modify the effects of the exposures. In one study conducted in the Extremely Low Gestational Age Newborns (ELGAN) [92] cohort, an adversity risk score was developed, based on four indicators: maternal education, marital status, eligibility for public health insurance, and supplemental nutrition assistance. Higher scores indicate greater socioeconomic adversity. Investigators identified 33 CpGs with methylation associated with at least one indicator, with 19 (58%) hypermethylated and 14 (42%) hypomethylated. Additionally, 15 (45%) of these were associated with the summative risk score, with placentas from female infants showing more robust differential methylation than male placentas. Thus, epigenomic effects may be linked to embedding of adversity, potentially effecting long term health outcomes. However, these effects may be attributed to this cohort consisting only of pre-term infants, and therefore these findings may not be generalizable to placentas of term pregnancies.

Our group had previously assessed adversity on placental methylation in RICHS [93]. Similar to the ELGAN study, we developed a cumulative risk score for adversity. We tested the association of this exposure on HSD11B2 methylation, and found that infants whose mothers experienced the greatest levels of adversity during pregnancy had the lowest extent of placental HSD11B2 methylation, particularly among males. Low maternal education, prenatal tobacco use, and higher cumulative risk scores were associated with significant HSD11B2 hypomethylation, with a one-unit increase in risk score correlating with a 2.3% decrease in methylation.

In general, there is an overall lack of literature on the impact of adversity on placental methylation, and additional work in this area may provide important insights on the mechanisms underlying the impacts of adversity on health across generations. Thus, additional studies about adversity and incorporating measures of adversity with other environmental exposures is warranted.

Circadian Disruption (CD)

The last environmental exposure we sought literature on was disruptions in circadian rhythm (CD). Our team was one of the first (and to date, only few) groups to examine CD on placental methylation, which we assessed based on night-shift work in the RICHS cohort [94]. We observed differential methylation at 298 CpG sites in night shift workers, with an average methylation decrease of 1.7% compared to non-night shift workers. We hypothesized that this could be due to increased transcription factor (TF) binding to DNA, leading to chromatin changes causing hypomethylation [95]. Additionally, CLOCK, a core component of the circadian clock, acts as a histone acetyltransferase [96], and thus CD could be impacting the epigenetic activity of CLOCK, affecting chromatin state and accessibility. The 298 probes were also found to be associated with traits such as psoriasis, lupus, type-1 diabetes (T1D), and multiple sclerosis (MS). This finding is in concert with a growing literature of the impacts of CD on various health outcomes [97]. Thus, our study shows that CD is impacting placental epigenetics and may also play a role in the development of diseases, though additional studies on this exposure's effect on placental methylation are required.

Conclusions and Future Perspectives

In this review, we have outlined the current state of evidence pertaining to early-life environmental exposures and their impact on placental DNA methylation. We have examined 6

well-studied exposure categories but recognize there were a number of other exposures, including pesticides, alcohol, and other chemicals, that we did not include in this review. From what we have summarized, we can identify a few prevailing themes: **1)** Specific exposures lead to unique variability in methylation, though cross-exposure assessment shows certain genes demonstrating consistency in differential methylation across exposures; **2)** Exposure studies that have looked at trimester-specific exposures' impact on methylation patterns tend to find effects that are most striking during the 1st trimester; **3)** Earlier studies have more targeted gene-specific approaches to methylation or assessed repetitive elements such as LINE-1 and Alu, while later studies are epigenome-wide, array-based.

One challenge in the study of DNA methylation in the placenta has been limitations in the ability to control for cellular heterogeneity in genome-wide studies. Until recently, many studies made use of reference-free methods to address this issue given a lack of reference data. This limitation has been recently overcome, though, with the publication of a reference panel and method, through R package *planet*, to estimate the cellular composition from array-based DNA methylation data [98]. Additionally, since methylation and exposure assessment are often observed coincidentally at birth, it is difficult to elucidate any form of temporality or causation. Incorporating repeated measures of exposure throughout pregnancy could improve this issue and allow for a better understanding of the critical windows during which exposure impacts placental methylation.

As for additional directions of placental epigenetic research, the use of DNA methylation-based scores is likely to become more prevalent as various risk scores are developed. For example, a placental epigenetic clock has been developed for estimating epigenetic gestational age from placental methylation levels [99]. Like other epigenetic clocks,

the deviation between actual gestational age and the estimated age can be used as an outcome and future studies that aim to assess epigenetic gestational age could consider this approach. Also, as sequencing technology continues to improve, it is likely there will be more studies utilizing such an approach in the placenta, and this could provide important new insights to genomic regions that thus far are being missed using array-based and targeted approaches. Finally, as we begin to uncover specific genes that have identified epigenetic alterations related to exposures, there will be opportunities for the development of more robust biomarkers, leading to a better understanding of how environmental exposures work.

The placenta clearly plays a critical role in fetal development and newborn health, and is increasingly being recognized as a critical organ in the developmental origins of long-term health. As cohorts and studies with placental data mature, there will be incredible opportunities to look empirically at these relationships and we encourage ongoing efforts to establish the links between the developmental environment, the placenta, and long-term health.

Statements and Declarations

Funding
This work was supported by the National Institutes of Health (NIH-NIGMS T32GM008490, NIH-NIEHS R24ES028507, NIH-NIEHS R01ES025145, NIH-NIEHS P30 ES019776).

Conflicts of Interest
Michael Mortillo declares that he has no conflict of interest.

Human and animal rights informed consent
This article does not contain any studies with human or animal subjects performed by any of the authors.

Tables

1st author, year	Study cohort & size	Exposure	Assay	Direction of effect on DNAme
Air Pollution Studies				
Janssen et. al, 2013 [38]	ENVIRONAGE (n = 240)	$PM_{2.5}$	Bisulfite sequencing	↓ global DNAme (1st trimester, whole pregnancy, implantation)
Saenen et. al, 2017 [40]	ENVIRONAGE (n = 361)	$PM_{2.5}$	Bisulfite sequencing	↓ LEP (2nd trimester)
Nawrot et. al, 2018 [41]	ENVIRONAGE (n = 407)	$PM_{2.5}$	Bisulfite sequencing	↑ NPAS2, CRY1, PER2, PER3 (3rd trimester) ↓ PER1, CLOCK (3rd trimester)
Neven et. al, 2018 [42]	ENVIRONAGE (n = 463)	$PM_{2.5}$, black carbon, NO_2	Bisulfite sequencing	↑ APEX1, OGG1, ERCC4, p53 ($PM_{2.5}$) ↓ DAPK1 ($PM_{2.5}$) ↑ APEX1, ERCC4 (black carbon)
Zhao et. al, 2021 [43]	Shanghai MCPC (n = 287)	$PM_{2.5}$	Infinium 450K	↑ IGF2, ↓ BID (2nd & 3rd trimester, whole pregnancy) ↑ FOX3N (2nd trimester)

Maghbooli et. al, 2018 [14]	• Prenatal care clinics in Tehran, Iran (n = 92)
	  ○ Polluted = 48
	  ○ Non-polluted = 44
		$PM_{2.5}$, PM_{10}	HPLC	↑ global DNAme ($PM_{2.5}$ & PM_{10}, 1st trimester)
				↑ global DNAme ($PM_{2.5}$, 3rd trimester, polluted)

Yang et. al, 2020 [47]	• COCOA (n = 1,180)
	• Grouped analysis:
	  ○ n = 6 for each group (8 groups)
	  ▪ High/low exposure groups,
		$PM_{2.5}$	Infinium 450K	↓ AHRR, DPP10, HLA-DRB1 (high $PM_{2.5}$, low CB VD, and AD group)

	high/low CB VD within those, AD within those			
Abraham et. al, 2018 [49]	EDEN (n = 668)	NO_2, PM_{10}	Infinium 450K	↓ ADORA2B (NO_2, 1st & 2nd trimester, whole pregnancy) ↑ PXT1, KCTD20 (NO_2, 2nd trimester & whole pregnancy) ↓ TUBGCP, ↑ TGM6, ADCK5 (PM_{10}, month before birth) ↓ LINE-1, Alu (PM_{10}, 1st trimester)
Ladd-Acosta et. al, 2019 [54]	EARLI (n = 133)	NO_2, O_3	Infinium 450K	↓ All probes, CpG islands (NO_2) ↑ shores, islands, ↓ shelves (O_3) ↓ ZNF442, PTPRH, SLC25A44, F11R (NO_2), STK38 (O_3)
Maternal Smoking Studies				
Janssen et. al, 2017 [58]	• ENVIRONAGE (n = 382) o Smokers (n = 62) o Past-smokers (n = 65) o Non-smokers (n = 255)	Cigarette smoke	Bisulfite sequencing	↑ MT-RNR1 ↓ CYP1A1
Fa et. al, 2017 [60]	• Women seeking legal abortions in Denmark (n = 39) o Smoking-exposed (n = 17)	Cigarette smoke	Bisulfite sequencing	↑ AHRR
Vos et. al, 2021 [61]	• ENVIRONAGE (n = 60) o Smokers (n = 20)	Cigarette smoke	Bisulfite sequencing	↑ D-loop

Cardenas et. al, 2019 [64]	• Gen3G (n = 441) o Smokers (n = 38) o Non-smokers (n = 403)	Cigarette smoke	Infinium 450K	↑ MDS2, PBX1, CYP1A2, VPBRP, CD28, CDK6 ↓ WBP1L
Shorey-Kendrick et. al, 2021 [65]	• VCSIP (n = 96) o Smokers (n = 72) ▪ VCS (n = 37) ▪ Placebo (n = 35) o Non-smokers (n = 24)	Cigarette smoke	Infinium 450K	↑ DIP2C, APOH/PRKCA (VCS)
Van Otterdijk et. al, 2017 [66]	• HEBC (n = 120) o Smoked during entire pregnancy (n = 27) o Quit smoking while pregnant (n = 32) o Non-smokers (n = 61)	Cigarette smoke	Bisulfite sequencing	↓ AHRR, CYP1A1 (smoked throughout pregnancy) ↓ GFI1 (quit while pregnant)
Rousseaux et. al, 2020 [67]	• EDEN (n = 568) o Current smokers (n = 117) o Former smokers (n = 70) o Non-smokers (n = 381)	Cigarette smoke	Infinium 450K	↓ LINE-1, ↑ DMRs (current smokers)

Everson et. al, 2021 [69]	• PACE (n = 1,700) ○ Any MSDP (n = 344) ○ Sustained MSDP (n = 163) ○ No MSDP (n = 1,193)	Cigarette smoke	Infinium 450K	↓ CpG DNAme (Any/sustained MSDP)
Chemical Studies				
Song et. al, 2021 [70]	• S-MBCS (n = 146) ○ Low BPA (n = 108) ○ High BPA (n = 38)	BPA	Infinium 450K	↑ CpG DNAme, HLA-DRB6 (high BPA)
Jedynak et. al, 2021 [73]	EDEN (n = 202)	Phenols	Infinium 450K	↑ DMPs, DMRs
Kim et. al, 2018 [76]	CHECK (n = 109)	POPs	Bisulfite sequencing	↓ LINE-1, H19, ↑ IGF2
Zhao et. al, 2019 [79]	• Wenzhou Birth Cohort (n = 249) ○ FGR cases (n = 124) ○ Controls (n = 125)	PBDEs	Bisulfite sequencing	↓ IGF2, ↑ HSD11B2
Trace Metal Studies				
Appleton et. al, 2017 [83]	RICHS (n = 222)	As, Cd, Pb, Mn, Hg, Zn	Bisulfite sequencing	↑ NR3C1 (higher As, Cd, Pb, Mn, Hg, lower Zn)
Everson et. al, 2018 [88]	• RICHS (n = 141) • NHBCS (n = 343)	Cd	Infinium 450K	↑ CpG DNAme

Tian et. al, 2020 [89]	• RICHS (n = 141) • NHBCS (n = 343)	Se	Infinium 450K	↑ GFI1, CAPN9, SKIDA1, ↓ ZNF496, TBC1D5
Kennedy et. al, 2020 [90]	• RICHS (n = 141) • NHBCS (n = 306)	Cu	Infinium 450K	↓ Enhancers, active TSS
Socioeconomic Adversity Studies				
Santos et. al, 2019 [92]	ELGAN (n = 426)	SES	Infinium 450K	↑ ↓ DMPs
Appleton et. al, 2013 [93]	RICHS (n = 444)	SES	Bisulfite sequencing	↓ HSD11B2
CD Studies				
Clarkson-Townsend et. al, 2019 [94]	• RICHS (n = 237) o Night shift (n = 53) o No night shift (n = 184)	CD	Infinium 450K	↓ DMPs, NAV1, MXRA8, GABRG1, PRDM16, WNT5A, FOXG1 (night shift) ↑ TDO2, ADAMTSL3, DLX2, SERPINA1

Table 2-1: **Summary of recent publications focusing on various environmental exposures in relation to placental DNA methylation.**

Abbreviations: ENVIRONAGE = Environmental Influence on Early Aging; $PM_{2.5}$ = particulate matter ≤ 2.5 μm; DNAme = DNA methylation; NO_2 = nitrogen dioxide; Shanghai MCPC = Shanghai Maternal-Child Pairs Cohort; PM_{10} = particulate matter ≤ 10 μm; HPLC = high performance liquid chromatography; COCOA = Cohort for Childhood Origin of Asthma and Allergic Diseases; CB = cord blood; VD = vitamin D; AD = atopic dermatitis; EARLI = Early Autism Risk Longitudinal Investigation; O_3 = ozone; CpG shores = CpG probe ± 2 kb from CpG island; CpG shelves = CpG probe ± 2-4 kb from island; D-loop = displacement loop; Gen3G = Genetics of Glucose regulation in Gestation and Growth; VCSIP = Vitamin C to decrease the effects of smoking in pregnancy on infant lung function; VCS = vitamin C supplement; HEBC = Harvard Epigenetic Birth Cohort; DMRs = differentially methylated regions; PACE = Pregnancy and Childhood Epigenetics; MSDP = maternal smoking during pregnancy; S-MBCS = Shanghai-Minhang Birth Cohort Study; BPA = bisphenol A; DMPs = differentially methylated probes; CHECK = Children's Health and Environmental Chemicals in Korea; POPs = persistent organic pollutants; FGR = fetal growth retardation; PBDEs = polybrominated diphenyl ethers; RICHS = Rhode Island Child Health Study; As = arsenic; Cd = cadmium; Pb = lead; Mn = manganese; Hg = mercury; Zn = zinc; NHBCS = New Hampshire Birth Cohort Study; Se = selenium; Cu = copper; TSS = transcription start site; ELGAN = Extremely Low Gestational Age Newborns; SES = socioeconomic status; CD = circadian disruption

Figures

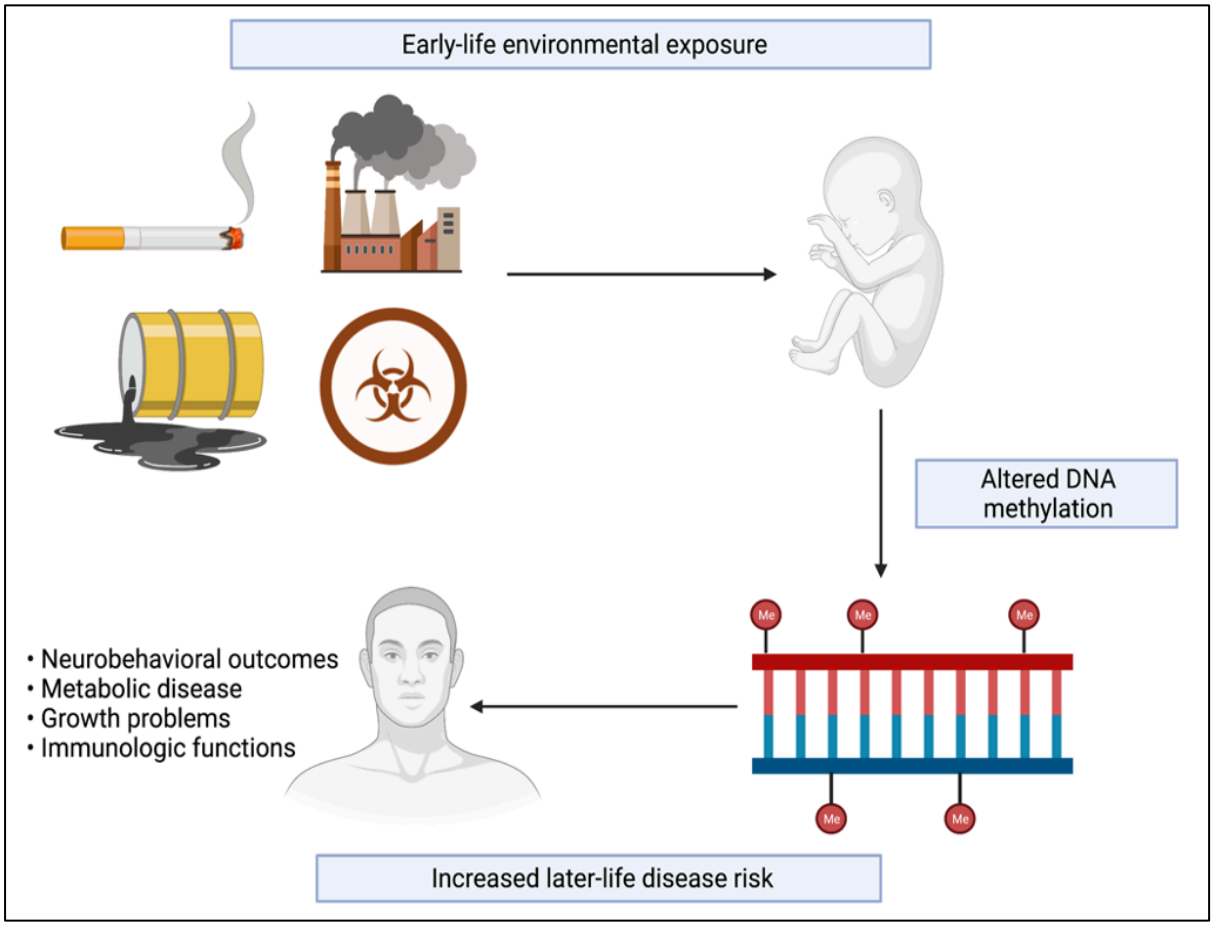


Figure 2-1: Early life environmental exposures and their impact on genome-wide DNA methylation and disease risk.

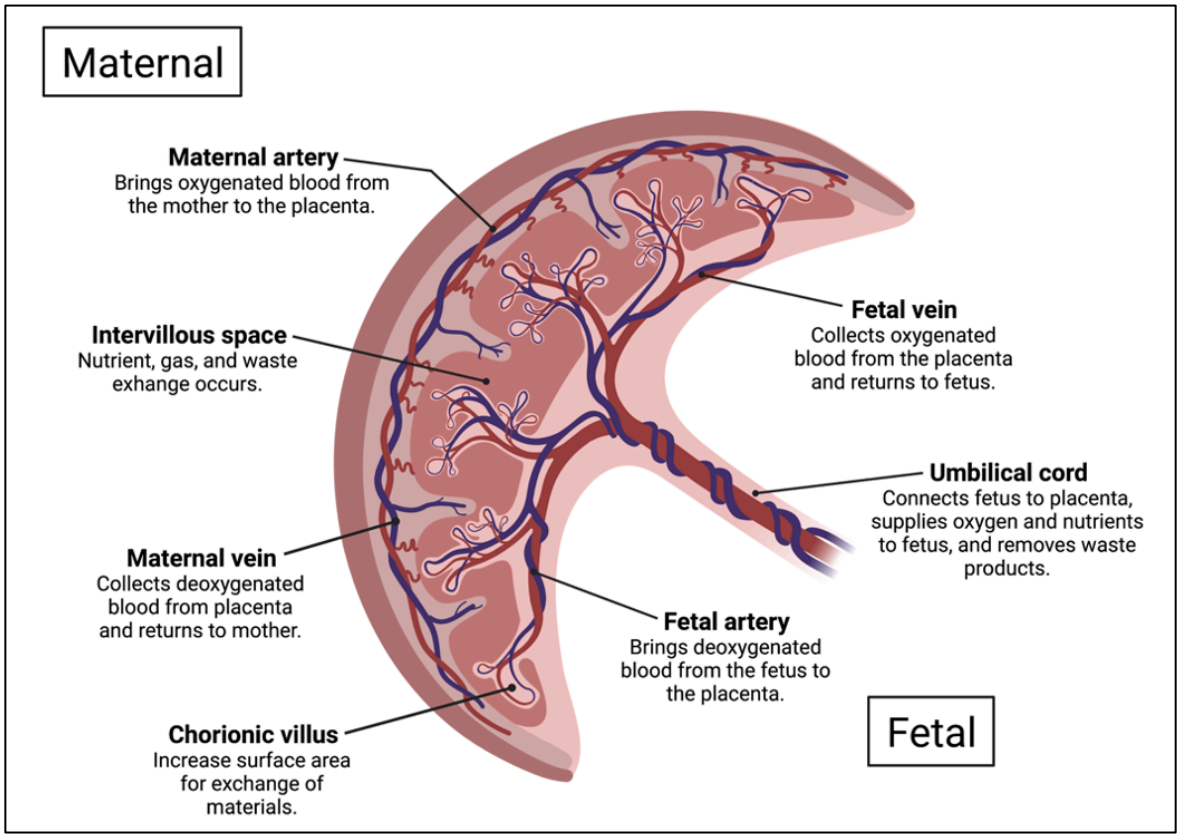


Figure 2-2: Human placental structure and function.

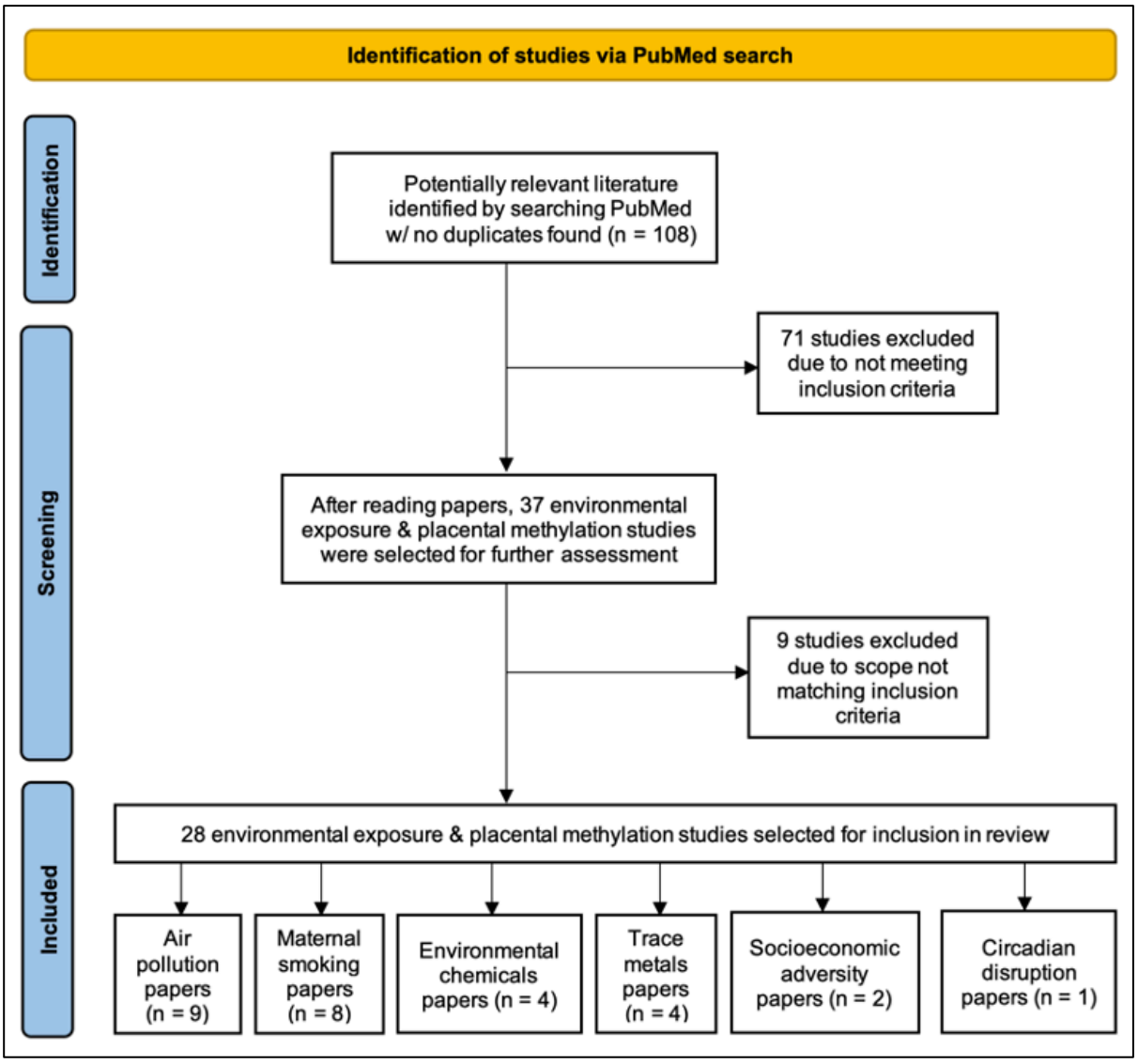


Figure 2-3: Flow chart illustrating paper selection process for conducting literature search.

References

1. Newell-Price, J., A.J. Clark, and P. King, *DNA methylation and silencing of gene expression.* Trends Endocrinol Metab, 2000. **11**(4): p. 142-8.
2. Hannon, E., et al., *Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins.* PLoS Genet, 2018. **14**(8): p. e1007544.
3. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.
4. Robinson, W.P. and E.M. Price, *The human placental methylome.* Cold Spring Harb Perspect Med, 2015. **5**(5): p. a023044.
5. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.* Nat Genet, 2007. **39**(4): p. 457-66.
6. Vryer, R. and R. Saffery, *What's in a name? Context-dependent significance of 'global' methylation measures in human health and disease.* Clin Epigenetics, 2017. **9**: p. 2.
7. Stoccoro, A. and F. Coppede, *Mitochondrial DNA Methylation and Human Diseases.* Int J Mol Sci, 2021. **22**(9).
8. Dong, Z., L. Pu, and H. Cui, *Mitoepigenetics and Its Emerging Roles in Cancer.* Front Cell Dev Biol, 2020. **8**: p. 4.
9. Baccarelli, A.A. and H.M. Byun, *Platelet mitochondrial DNA methylation: a potential new marker of cardiovascular disease.* Clin Epigenetics, 2015. **7**: p. 44.
10. Tsaprouni, L.G., et al., *Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation.* Epigenetics, 2014. **9**(10): p. 1382-96.
11. Joubert, B.R., et al., *DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis.* Am J Hum Genet, 2016. **98**(4): p. 680-96.
12. Elliott, H.R., et al., *Differences in smoking associated DNA methylation patterns in South Asians and Europeans.* Clin Epigenetics, 2014. **6**(1): p. 4.
13. Gruzieva, O., et al., *Epigenome-Wide Meta-Analysis of Methylation in Children Related to Prenatal NO2 Air Pollution Exposure.* Environ Health Perspect, 2017. **125**(1): p. 104-110.
14. Maghbooli, Z., et al., *Air pollution during pregnancy and placental adaptation in the levels of global DNA methylation.* PLoS One, 2018. **13**(7): p. e0199772.
15. Mostafavi, N., et al., *Associations Between Genome-wide Gene Expression and Ambient Nitrogen Oxides.* Epidemiology, 2017. **28**(3): p. 320-328.
16. Goodrich, J.M., et al., *Mercury biomarkers and DNA methylation among Michigan dental professionals.* Environ Mol Mutagen, 2013. **54**(3): p. 195-203.
17. Tajuddin, S.M., et al., *Genetic and non-genetic predictors of LINE-1 methylation in leukocyte DNA.* Environ Health Perspect, 2013. **121**(6): p. 650-6.
18. Tellez-Plaza, M., et al., *Association of global DNA methylation and global DNA hydroxymethylation with metals and other exposures in human blood DNA samples.* Environ Health Perspect, 2014. **122**(9): p. 946-54.
19. Hanna, C.W., et al., *DNA methylation changes in whole blood is associated with exposure to the environmental contaminants, mercury, lead, cadmium and bisphenol A, in women undergoing ovarian stimulation for IVF.* Hum Reprod, 2012. **27**(5): p. 1401-10.

20. Itoh, H., et al., *Association between serum organochlorines and global methylation level of leukocyte DNA among Japanese women: a cross-sectional study.* Sci Total Environ, 2014. **490**: p. 603-9.
21. Watkins, D.J., et al., *Associations between serum perfluoroalkyl acids and LINE-1 DNA methylation.* Environ Int, 2014. **63**: p. 71-6.
22. Clark, N.A., et al., *Effect of early life exposure to air pollution on development of childhood asthma.* Environ Health Perspect, 2010. **118**(2): p. 284-90.
23. Lindoso, L., et al., *The Effect of Early-Life Environmental Exposures on Disease Phenotype and Clinical Course of Crohn's Disease in Children.* Am J Gastroenterol, 2018. **113**(10): p. 1524-1529.
24. Pesce, G., et al., *Foetal exposure to heavy metals and risk of atopic diseases in early childhood.* Pediatr Allergy Immunol, 2021. **32**(2): p. 242-250.
25. Lin, X., et al., *Choice of surrogate tissue influences neonatal EWAS findings.* BMC Med, 2017. **15**(1): p. 211.
26. Lura, M.P., et al., *Response of cord blood cells to environmental, hereditary and perinatal factors: A prospective birth cohort study.* PLoS One, 2018. **13**(7): p. e0200236.
27. Fisher, S.J., *The placenta dilemma.* Semin Reprod Med, 2000. **18**(3): p. 321-6.
28. Maltepe, E. and S.J. Fisher, *Placenta: the forgotten organ.* Annu Rev Cell Dev Biol, 2015. **31**: p. 523-52.
29. Nugent, B.M. and T.L. Bale, *The omniscient placenta: Metabolic and epigenetic regulation of fetal programming.* Front Neuroendocrinol, 2015. **39**: p. 28-37.
30. Barker, D.J. and K.L. Thornburg, *Placental programming of chronic diseases, cancer and lifespan: a review.* Placenta, 2013. **34**(10): p. 841-5.
31. Kusuyama, J., et al., *Effects of maternal and paternal exercise on offspring metabolism.* Nat Metab, 2020. **2**(9): p. 858-872.
32. Mitro, S.D., T. Johnson, and A.R. Zota, *Cumulative Chemical Exposures During Pregnancy and Early Development.* Curr Environ Health Rep, 2015. **2**(4): p. 367-78.
33. Robinson, J.F., et al., *Differences in cytochrome p450 enzyme expression and activity in fetal and adult tissues.* Placenta, 2020. **100**: p. 35-44.
34. Sachdeva, P., B.G. Patel, and B.K. Patel, *Drug use in pregnancy; a point to ponder!* Indian J Pharm Sci, 2009. **71**(1): p. 1-7.
35. Myatt, L., *Placental adaptive responses and fetal programming.* J Physiol, 2006. **572**(Pt 1): p. 25-30.
36. Saenen, N.D., et al., *Air pollution-induced placental alterations: an interplay of oxidative stress, epigenetics, and the aging phenotype?* Clin Epigenetics, 2019. **11**(1): p. 124.
37. Christensen, B.C., et al., *Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context.* PLoS Genet, 2009. **5**(8): p. e1000602.
38. Janssen, B.G., et al., *Placental DNA hypomethylation in association with particulate air pollution in early life.* Part Fibre Toxicol, 2013. **10**: p. 22.
39. Sagawa, N., et al., *Possible role of placental leptin in pregnancy: a review.* Endocrine, 2002. **19**(1): p. 65-71.
40. Saenen, N.D., et al., *Lower Placental Leptin Promoter Methylation in Association with Fine Particulate Matter Air Pollution during Pregnancy and Placental Nitrosative Stress at Birth in the ENVIRONAGE Cohort.* Environ Health Perspect, 2017. **125**(2): p. 262-268.

41. Nawrot, T.S., et al., *Placental circadian pathway methylation and in utero exposure to fine particle air pollution.* Environ Int, 2018. **114**: p. 231-241.
42. Neven, K.Y., et al., *Placental promoter methylation of DNA repair genes and prenatal exposure to particulate air pollution: an ENVIRONAGE cohort study.* Lancet Planet Health, 2018. **2**(4): p. e174-e183.
43. Zhao, Y., et al., *Prenatal fine particulate matter exposure, placental DNA methylation changes, and fetal growth.* Environ Int, 2021. **147**: p. 106313.
44. White, V., et al., *IGF2 stimulates fetal growth in a sex- and organ-dependent manner.* Pediatr Res, 2018. **83**(1-1): p. 183-189.
45. Garcia-Perez, C., et al., *Bid-induced mitochondrial membrane permeabilization waves propagated by local reactive oxygen species (ROS) signaling.* Proc Natl Acad Sci U S A, 2012. **109**(12): p. 4497-502.
46. Wang, P., et al., *Phosphorylation of the proapoptotic BH3-only protein bid primes mitochondria for apoptosis during mitotic arrest.* Cell Rep, 2014. **7**(3): p. 661-71.
47. Yang, S.I., et al., *Prenatal PM2.5 exposure and vitamin D-associated early persistent atopic dermatitis via placental methylation.* Ann Allergy Asthma Immunol, 2020. **125**(6): p. 665-673 e1.
48. Hao, N. and M.L. Whitelaw, *The emerging roles of AhR in physiology and immunity.* Biochem Pharmacol, 2013. **86**(5): p. 561-70.
49. Abraham, E., et al., *Pregnancy exposure to atmospheric pollution and meteorological conditions and placental DNA methylation.* Environ Int, 2018. **118**: p. 334-347.
50. Ding, H., et al., *Upregulation of CD81 in trophoblasts induces an imbalance of Treg/Th17 cells by promoting IL-6 expression in preeclampsia.* Cell Mol Immunol, 2019. **16**(1): p. 302-312.
51. Novakovic, B., et al., *Variable DAXX gene methylation is a common feature of placental trophoblast differentiation, preeclampsia, and response to hypoxia.* FASEB J, 2017. **31**(6): p. 2380-2392.
52. Roberts, V.H., et al., *Post-Translational Modifications of the P2X(4) purinergic receptor subtype in the human placenta are altered in preeclampsia.* Placenta, 2007. **28**(4): p. 270-7.
53. Zhao, W.X., et al., *Expression of notch family proteins in placentas from patients with early-onset severe preeclampsia.* Reprod Sci, 2014. **21**(6): p. 716-23.
54. Ladd-Acosta, C., et al., *Epigenetic marks of prenatal air pollution exposure found in multiple tissues relevant for child health.* Environ Int, 2019. **126**: p. 363-376.
55. Johannesson, S., et al., *Urban air pollution and effects on biomarkers of systemic inflammation and coagulation: a panel study in healthy adults.* Inhal Toxicol, 2014. **26**(2): p. 84-94.
56. Ruckerl, R., et al., *Associations between ambient air pollution and blood markers of inflammation and coagulation/fibrinolysis in susceptible populations.* Environ Int, 2014. **70**: p. 32-49.
57. Strauss, W.J., et al., *Improving cost-effectiveness of epidemiological studies via designed missingness strategies.* Stat Med, 2010. **29**(13): p. 1377-87.
58. Janssen, B.G., et al., *Placental mitochondrial DNA and CYP1A1 gene methylation as molecular signatures for tobacco smoke exposure in pregnant women and the relevance for birth weight.* J Transl Med, 2017. **15**(1): p. 5.

59. Ezzeldin, N., et al., *Genetic polymorphisms of human cytochrome P450 CYP1A1 in an Egyptian population and tobacco-induced lung cancer.* Genes Environ, 2017. **39**: p. 7.
60. Fa, S., et al., *Changes in first trimester fetal CYP1A1 and AHRR DNA methylation and mRNA expression in response to exposure to maternal cigarette smoking.* Environ Toxicol Pharmacol, 2018. **57**: p. 19-27.
61. Vos, S., et al., *Mitochondrial DNA methylation in placental tissue: a proof of concept study by means of prenatal environmental stressors.* Epigenetics, 2021. **16**(2): p. 121-131.
62. Sharma, H., et al., *Mutations in the mitochondrial DNA D-loop region are frequent in cervical cancer.* Cancer Cell Int, 2005. **5**: p. 34.
63. van Loon, N.M., D. Lindholm, and N. Zelcer, *The E3 ubiquitin ligase inducible degrader of the LDL receptor/myosin light chain interacting protein in health and disease.* Curr Opin Lipidol, 2019. **30**(3): p. 192-197.
64. Cardenas, A., et al., *Mediation by Placental DNA Methylation of the Association of Prenatal Maternal Smoking and Birth Weight.* Am J Epidemiol, 2019. **188**(11): p. 1878-1886.
65. Shorey-Kendrick, L.E., et al., *Impact of vitamin C supplementation on placental DNA methylation changes related to maternal smoking: association with gene expression and respiratory outcomes.* Clin Epigenetics, 2021. **13**(1): p. 177.
66. van Otterdijk, S.D., A.M. Binder, and K.B. Michels, *Locus-specific DNA methylation in the placenta is associated with levels of pro-inflammatory proteins in cord blood and they are both independently affected by maternal smoking during pregnancy.* Epigenetics, 2017. **12**(10): p. 875-885.
67. Rousseaux, S., et al., *Immediate and durable effects of maternal tobacco consumption alter placental DNA methylation in enhancer and imprinted gene-containing regions.* BMC Med, 2020. **18**(1): p. 306.
68. Dolinoy, D.C., J.R. Weidman, and R.L. Jirtle, *Epigenetic gene regulation: linking early developmental environment to adult disease.* Reprod Toxicol, 2007. **23**(3): p. 297-307.
69. Everson, T.M., et al., *Placental DNA methylation signatures of maternal smoking during pregnancy and potential impacts on fetal growth.* Nat Commun, 2021. **12**(1): p. 5095.
70. Song, X., et al., *Differential methylation of genes in the human placenta associated with bisphenol A exposure.* Environ Res, 2021. **200**: p. 111389.
71. Steinhart, Z. and S. Angers, *Wnt signaling in development and tissue homeostasis.* Development, 2018. **145**(11).
72. Ye, Y., et al., *Bisphenol A exposure alters placentation and causes preeclampsia-like features in pregnant mice involved in reprogramming of DNA methylation of WNT2.* FASEB J, 2019. **33**(2): p. 2732-2742.
73. Jedynak, P., et al., *Pregnancy exposure to synthetic phenols and placental DNA methylation - An epigenome-wide association study in male infants from the EDEN cohort.* Environ Pollut, 2021. **290**: p. 118024.
74. Cao, X., et al., *Exposure of pregnant mice to triclosan impairs placental development and nutrient transport.* Sci Rep, 2017. **7**: p. 44803.
75. Wang, X., et al., *Triclosan causes spontaneous abortion accompanied by decline of estrogen sulfotransferase activity in humans and mice.* Sci Rep, 2015. **5**: p. 18252.

76. Kim, S., et al., *Prenatal exposure to persistent organic pollutants and methylation of LINE-1 and imprinted genes in placenta: A CHECK cohort study.* Environ Int, 2018. **119**: p. 398-406.
77. St-Pierre, J., et al., *IGF2 DNA methylation is a modulator of newborn's fetal growth and development.* Epigenetics, 2012. **7**(10): p. 1125-32.
78. Constancia, M., et al., *Placental-specific IGF-II is a major modulator of placental and fetal growth.* Nature, 2002. **417**(6892): p. 945-8.
79. Zhao, Y., et al., *Associations between in utero exposure to polybrominated diphenyl ethers, pathophysiological state of fetal growth and placental DNA methylation changes.* Environ Int, 2019. **133**(Pt B): p. 105255.
80. Bro-Rasmussen, F., O. Buus, and D. Trolle, *Ratio cortisone/cortisol in mother and infant at birth.* Acta Endocrinol (Copenh), 1962. **40**: p. 579-83.
81. Alikhani-Koopaei, R., et al., *Epigenetic regulation of 11 beta-hydroxysteroid dehydrogenase type 2 expression.* J Clin Invest, 2004. **114**(8): p. 1146-57.
82. Zhao, Y., et al., *Site-specific methylation of placental HSD11B2 gene promoter is related to intrauterine growth restriction.* Eur J Hum Genet, 2014. **22**(6): p. 734-40.
83. Appleton, A.A., et al., *Prenatal exposure to neurotoxic metals is associated with increased placental glucocorticoid receptor DNA methylation.* Epigenetics, 2017. **12**(8): p. 607-615.
84. Bromer, C., et al., *Genetic and epigenetic variation of the glucocorticoid receptor (NR3C1) in placenta and infant neurobehavior.* Dev Psychobiol, 2013. **55**(7): p. 673-83.
85. Monk, C., et al., *Distress During Pregnancy: Epigenetic Regulation of Placenta Glucocorticoid-Related Genes and Fetal Neurobehavior.* Am J Psychiatry, 2016. **173**(7): p. 705-13.
86. Seckl, J.R., *Glucocorticoids, feto-placental 11 beta-hydroxysteroid dehydrogenase type 2, and the early life origins of adult disease.* Steroids, 1997. **62**(1): p. 89-94.
87. Gundacker, C. and M. Hengstschlager, *The role of the placenta in fetal exposure to heavy metals.* Wien Med Wochenschr, 2012. **162**(9-10): p. 201-6.
88. Everson, T.M., et al., *Cadmium-Associated Differential Methylation throughout the Placental Genome: Epigenome-Wide Association Study of Two U.S. Birth Cohorts.* Environ Health Perspect, 2018. **126**(1): p. 017010.
89. Tian, F.Y., et al., *Selenium-associated DNA methylation modifications in placenta and neurobehavioral development of newborns: An epigenome-wide study of two U.S. birth cohorts.* Environ Int, 2020. **137**: p. 105508.
90. Kennedy, E., et al., *Copper associates with differential methylation in placentae from two US birth cohorts.* Epigenetics, 2020. **15**(3): p. 215-230.
91. Le, A., et al., *Characterization of timed changes in hepatic copper concentrations, methionine metabolism, gene expression, and global DNA methylation in the Jackson toxic milk mouse model of Wilson disease.* Int J Mol Sci, 2014. **15**(5): p. 8004-23.
92. Santos, H.P., Jr., et al., *Epigenome-wide DNA methylation in placentas from preterm infants: association with maternal socioeconomic status.* Epigenetics, 2019. **14**(8): p. 751-765.
93. Appleton, A.A., et al., *Patterning in placental 11-B hydroxysteroid dehydrogenase methylation according to prenatal socioeconomic adversity.* PLoS One, 2013. **8**(9): p. e74691.

94. Clarkson-Townsend, D.A., et al., *Maternal circadian disruption is associated with variation in placental DNA methylation.* PLoS One, 2019. **14**(4): p. e0215745.
95. Martin, E.M. and R.C. Fry, *A cross-study analysis of prenatal exposures to environmental contaminants and the epigenome: support for stress-responsive transcription factor occupancy as a mediator of gene-specific CpG methylation patterning.* Environ Epigenet, 2016. **2**(1).
96. Doi, M., J. Hirayama, and P. Sassone-Corsi, *Circadian regulator CLOCK is a histone acetyltransferase.* Cell, 2006. **125**(3): p. 497-508.
97. Lyons, A.B., et al., *Circadian Rhythm and the Skin: A Review of the Literature.* J Clin Aesthet Dermatol, 2019. **12**(9): p. 42-45.
98. Yuan, V., et al., *Accurate ethnicity prediction from placental DNA methylation data.* Epigenetics Chromatin, 2019. **12**(1): p. 51.
99. Lee, Y., et al., *Placental epigenetic clocks: estimating gestational age using placental DNA methylation levels.* Aging (Albany NY), 2019. **11**(12): p. 4238-4253.

Chapter 3 - Epigenetic landscape of 5-hydroxymethylcytosine and associations with gene expression in placenta

Michael Mortillo[1], Elizabeth M. Kennedy[1], Karen E. Hermetz[1], Amber A. Burt[1], Carmen J. Marsit[1]

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

Adapted from an original manuscript currently under review at *Epigenetics*

Abstract

5-hydroxymethylcystosine (5hmC) is produced through enzymatic oxidation of 5-methylcytosine (5mC). 5hmC is an intermediary in the DNA demethylation pathway, though recent evidence suggests that it acts as a functional epigenetic modification. The landscape of placental 5hmC and its role in gene expression is poorly understood. We aimed to characterize the distribution of 5hmC across the placenta and identify genomic regions of 5hmC that play a direct role in gene expression. Using 5hmC signals at CG dinucleotide (CpG) sites along with RNA-sequencing (RNA-seq) data, we assessed 5hmC distribution in the placenta and evaluated the association between 5hmC and gene expression using an expression quantitative trait hydroxymethylation (eQTHM) analysis. We identified ~47,000 loci with consistently elevated (systematic) 5hmC levels. We observed significant depletion ($p < 0.0001$) of systematic 5hmC at CpG islands (CGI), and enrichment ($p < 0.0001$) in "open sea" regions (CpG > 4 Mb from CGI). 5hmC levels were highest and lowest among CpGs in enhancers and active transcription start sites (TSS), respectively ($p < 0.05$). eQTHM analysis produced 499 significant (empirical-$p < 0.05$) cis CpG-gene pairs, with 75.4% of eQTHMs demonstrating a positive correlation between 5hmC and expression. We observed significant enrichment and depletion of eQTHMs in enhancers and active TSS, respectively ($p < 0.05$ for both). Finally, we identified 107 differentially hydroxymethylated regions (DHMRs, $p < 0.05$) across 100 genes. Our study provides novel insight into placental distribution of 5hmC, and sheds light on the functional capacity of this epigenetic modification.

Introduction

The placenta is an organ that acts as the interface between the fetus and its mother[1]. Its crucial roles include nutrient transfer, gas exchange, waste removal, immune protection, and various neuroendocrine functions[2-4]. Disruptions in placental processes can lead to pregnancy complications including preeclampsia, inflammation, preterm delivery and fetal growth restriction, with subsequent implications for the health of both the infant and the mother[5-7].

5-methylcytosine (5mC) is formed through DNA methyltransferases (DNMTs) adding a methyl group to the 5th carbon position of an unmethylated cytosine[8]. It is one of the most abundant DNA modifications in the human genome and is involved in a number of epigenetic functions, including modulation of transcription factor binding to regulatory regions[9]. When found in the promoter regions of genes, it generally leads to repression in transcription[10]. 5-hydroxymethylcytosine (5hmC) is produced through enzymatic oxidation of 5mC by ten-eleven translocation (TET) methylcytosine dioxygenases[11, 12]. Though it is most commonly believed that 5hmC acts as an intermediary in the DNA demethylation pathway, there is also evidence to suggest that 5hmC is stable and could act as a stand-alone epigenetic modification[13]. Within the gene body, 5hmC is correlated with increased transcription, depending on cell and tissue type[14, 15]. In the placenta, 5hmC is sparse[16], and has been observed to be enriched at imprinted loci[17], although when evaluated by Piyasena et al.[18] at imprinting control regions (ICRs) of the genes IGF2, H19, and CDKN1C, was not associated with gene expression. Our group has previously leveraged publicly-available placental gene expression data to demonstrate that 5hmC is positively associated with transcription in actively transcribed genes, but did not assess 5hmC and gene expression in the same samples[19].

In this study, we aim to characterize 5hmC distribution across the placenta and identify specific areas of 5hmC that relate to gene expression through the use of expression quantitative

trait hydroxymethylation (eQTHM). We believe that drawing a direction correlation between 5hmC and expression will shed light on the functionality and distribution of placental 5hmC and provide a framework for future studies of the placental epigenome.

Materials and Methods

Study population

The study was conducted in placenta samples collected from participants enrolled in the Rhode Island Child Health Study (RICHS). This hospital-based cohort (n = 840) recruited women 18-40 years old, with no history of psychological disorders and in good physical health, and who delivered infants from healthy, non-pathologic pregnancies at term (≥ 37 gestational weeks). Mother-infant pairs were recruited between September 1, 2009 and July 31, 2014 from the Women and Infants Hospital of Rhode Island (WIH, Providence, RI). The cohort was established to examine the relationship between molecular features of the placenta and birthweight and was oversampled for infants born small for gestational age (SGA, <10% 2013 Fenton Growth Curve) and large for gestational age (LGA, >90% 2013 Fenton Growth Curve), each matched on gender, gestational age, and maternal age to infants born appropriate for gestational age (AGA, 10-90% 2013 Fenton Growth Curve)[20].

Participant demographic data was collected from interviewer-administered questionnaires and clinical data were abstracted from a structured review of medical records. The current study focuses on a subset of the enrolled participants with available placental 5hmC data (n = 227).

Placental sample collection

Placental parenchyma was obtained within two hours of delivery, taken from the fetal side of the placenta, 2 cm from the umbilical cord insertion site. All samples were free of maternal decidua. Samples were placed in RNALater at 4ºC. At least 72 hours later, samples

were removed from RNALater, snap-frozen in liquid nitrogen, pulverized to homogenize the samples, and stored at -80^{o}C until extraction.

CpG methylation and hydroxymethylation profiling

Hydroxymethylation profiling was performed as previously described[19]. Briefly, DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Germantown, MD, USA) following manufacturer's protocol, quantified with the Qubit Fluorometer (Thermo Fisher Scientific Life Sciences), and stored at −80°C. Bisulfite (BS) and oxidative bisulfite (oxBS) conversion were performed on placenta-derived DNA using the TrueMethyl oxBS Module (NuGen, Redwood City, CA, USA), following manufacturer's optimized protocol of 500 ng gDNA input for downstream analysis using the Infinium HD Methylation EPIC Bead Chip Array (Illumina, San Deigo, CA, USA).

Cross-reactive CpG probes[21], probes that failed detection p-value ($p > 0.01$) in > 1 sample, and probes overlapping single nucleotide polymorphisms (SNPs) were removed from the analysis. Three samples were removed due to failing sex quality control or failing detection p-value (p-value > 0.01 in >2% of probes). After quality control, 706,435 CpGs were available for normalization. Normalization of background correction and dye bias of raw signals from each of the BS and oxBS-converted samples was done using the Noob procedure, followed by normalization of probe-type bias using SWAN, both of which are available in the R/Bioconductor package minfi (version 1.24.0; *https://www.bioconductor.org*). Estimation of 5hmC beta values (proportion of CpGs across all cells in a given sample that are hydroxymethylated) for each CpG on the array was performed using OxyBS (version 1.5)[22]. Potential confounding due to array chip was removed using the ComBat function in R/bioconductor package sva[23]. Beta values were logit-transformed (M-values) to better approximate a normal distribution[24]. Finally, we limited our analysis to autosomal probes

only, to avoid any confounding due to sex-specific effects. The final filtered, normalized dataset contained 689,815 CpGs.

RNA-sequencing, Quality Control and Read Filtering

RNA-sequencing was performed as previously described[25]. Briefly, Total RNA was isolated from homogenized placental tissue, stored, and quantified. RNA was then converted to cDNA, and transcriptome-wide 50 bp single-end RNA sequencing was conducted using the HiSeq 2500 platform (Illumina, San Diego, CA)[26]. Samples were run in three sequencing batches, with 10% of the samples run in triplicate within each batch.

The raw RNA sequencing data (fastq files) were assessed for quality control, including read length and GC content. Reads that passed quality control were mapped to the human reference genome (hg19) in a splice-aware manner, with common SNPs in the reference genome masked prior to alignment. Genes with counts per million <1 in greater than 30 samples were considered unexpressed and removed. Read counts were adjusted for GC content, followed by TMM correction for library size differences across samples. The data were then $\log_2$-transformed to account for the mean-variance relationship. Following assessments of Pearson correlations in gene expression among the triplicate samples, duplicated repeat samples were removed from the analysis. The data were adjusted to remove batch effect due to flow cell, using the ComBat function in R/bioconductor package sva[23]. We then removed poorly-defined transcripts (transcripts containing the phrases “LOC”, “orf”, “KIAA”, “NCRNA”, and “MIR”), as well as genes on sex chromosomes. The final filtered, normalized dataset contained reads mapped to 12,744 genes. For our 5hmC-expression analysis (see section below, “*eQTHM Identification*”), we used samples that had both placental 5hmC *and* gene expression data available (n = 197). To assess 5hmC enrichment across genes with varying expression levels, we grouped our genes into

expression quartiles (0 – 25%, 26% – 50%, 51% – 75%, 76% – 100%) based on mean log_2TMM expression counts across all samples.

Annotation

We annotated CpGs using the R/bioconductor package "IlluminaHumanMethylationEPICanno.ilm10b2.hg19"[27]. This includes annotating each CpG based on its location relative to the nearest gene, as well as annotating CpGs that fall within a CpG islands (CGI) interval. The available gene compartments from the EPIC array package are as follows: 1) 5' untranslated region (5' UTR), 2) TSS 200 (1-200 base pairs (bp) upstream of the TSS), 3) TSS 1500 (201-1500 bp upstream of TSS), 4) 1st exon, 5) gene bodies, 6) exon boundaries, and 7) 3' untranslated region (3' UTR). CpGs that fell in either the TSS200 or TSS1500 intervals were combined into one "TSS" interval. CpGs that fell into gene bodies, 1st exons, or exon boundaries were combined into one "gene body" class. For the CGI regions, CpGs were annotated to a CGI "shore" if they were within a 2 kb region flanking a CGI, a CGI "shelf" if they were within a 2 kb region flanking a CGI shore, or an "open sea" if they were not within a shore, shelf, or CGI. CpGs were also annotated to chromatin-based genomic categories using ChromHMM[28], derived from the Roadmap Epigenomics Consortium[29] and applied to fetal placenta cells. We combined enhancers and genic enhancers into one "enhancers" class, strong and weak transcription into a "transcribed" class, bivalent enhancers and bivalent/poised transcription start sites (TSS) into "poised TSS/enhancers" class, and finally polycomb repressed and weak polycomb repressed into a single "polycomb repressed" class.

In order to relate the location of each CpG to its associated gene for eQTHM analysis, gene TSS and transcription termination site (TTS) positions were obtained from the switchDbTss track of the UCSC table browser (hg19/GRCh37) [30]. 5' and 3'UTR positions were obtained from the NCBI RefSeq track of the table browser. When multiple TSS, TTS, 5'UTR, and 3'UTR

positions were annotated for one gene, we used the most 5' upstream and most 3' downstream position. To annotate gene body coordinates, we selected the most 5' upstream start position of the 1st exon and the most 3' downstream end position of the last exon.

When mapping the distance of the CpG to its target eQTHM gene, CpGs upstream of the TSS were calculated as negative bp to the TSS, while CpGs downstream of the TTS were calculated as positive bp. For CpGs that fall within their target gene, their distance is represented as the proportion of the way through the gene (to account for variability in gene length).

Systematic 5hmC distribution

As in prior work[19], CpGs at which 5hmC proportion ≥ 0.10 in at least 50% of samples were defined as regions of systematic 5hmC (n = 113). Distribution of systematic 5hmC in relation to nearest genes and CGI feature was assessed using odds ratios (ORs) and 95% confidence intervals (CIs) derived from a Fisher's exact test. This allowed us to compare the proportions of loci demonstrating systematic 5hmC within a gene/CGI compartment against loci not demonstrating systematic 5hmC.

eQTHM Identification

eQTHM is an extension of expression quantitative trait methylation (eQTM), a method used to identify specific positions in the genome where the proportion of methylated CpGs at one locus is associated with transcript abundance for a given gene[31]. In the current study, we assess how 5hmC proportion at a given CpG in the placental epigenome associates with the abundance of a gene transcript. To this end, we have conducted a cis eQTHM analysis using the Matrix eQTL R package[32]. Matrix eQTL fits a multivariate linear regression model: $y = b_0 + b_1HM + Ta + e$, where y is the normalized transcript counts for each gene, HM is 5hmC M-values at each CpG, b_0, b_1, and a are regression coefficients, and T represents covariates. We identified covariates as any variable with significant univariate association ($p < 0.05$) with any of

the top 5 principal components (PCs) for the 5hmC data and any of the top 5 PCs of the transcriptomic data. From this, we identified infant sex, birthweight percentile, and estimated proportions of syncytiotrophoblasts (STBs, estimated using R/Bioconductor package planet[33]) as being significantly associated with both pairs of PCs, and included them as covariates in our model. With respect to the transcript under investigation, we restricted our analyses to only CpGs that were 1) in the gene, 2) up to 1 Mb upstream of the transcription start site (TSS), and 3) up to 1 Mb downstream of the transcription termination site (TTS). We then implemented an empirical p-value approach (see section below) to find eQTHMs among all cis CpGs. Among eQTHMs, we conducted a Pearson's correlation between 5hmC at the CpG and transcript level of the target gene, according to the Matrix eQTL guidelines[32].

Empirical p-value threshold. CpGs can be mapped to multiple genes that fall within a 1 Mb window, so we employed an empirical p-value threshold to ensure that genes paired to a higher number of CpGs do not have a greater chance of being part of an eQTHM. We defined a significance threshold using the procedure previously described for eQTL analyses in the Genotype-Tissue Expression (GTEx) project[34], and subsequently for expression quantitative trait methylation (eQTM) analysis by Ruiz-Arenas et al.[31]. Briefly, we performed the eQTHM analysis described above on data with shuffled sample identifiers, collecting the minimum p-value obtained for each gene in the permutation. We obtained 1,000 permuted p-values for each gene. We modeled the null distribution with the permuted *P*-values using a beta distribution, generating the parameters with maximum likelihood estimation[35]. Using the beta distributions, we estimated empirical p-values for the minimum observed (non-permuted) p-values obtained for each gene. False discovery rates (FDR) were calculated for gene-wise empirical p-values using the Benjamini-Hochberg method[36]. Genes with FDR less than 5% were considered

significant. Finally, we defined an empirical p-value threshold as the empirical p-value of the gene closest to the 5% FDR threshold (i.e. the highest empirical p-value among genes significant at 5% FDR). Among all cis CpG-gene pair associations, if the nominal p-value for the CpG was less than the empirical threshold *and* the corresponding gene had an FDR < 0.05, it was deemed an eQTHM.

Differentially-hydroxymethylated regions (DHMRs)

We also conducted a DHMR analysis, where we aimed to find contiguous local genomic regions with differential hydroxymethylation associated with gene expression, considering these regions may be more biologically relevant than individual eQTHMs. To perform this, we utilized the Comb-p method[37], which identifies regions enriched for low p-values (using Matrix eQTL p-values). Comb-p corrects p-values for auto-correlation with neighboring CpGs (within 1 kb) using the Stouffer-Liptak method, and then builds DHMRs for a specific gene based on CpGs with corrected p-values falling below a threshold. We specified a corrected p-value threshold of 1e-4. Once a DHMR is identified, a regional p-value is generated and adjusted for multiple testing using the Sidak correction. DHMRs with an adjusted regional p-value < 0.05, *and* containing at least three CpGs were considered significant. This process was conducted for every gene in our eQTHM analysis.

Enrichment tests

Fisher's exact tests were used to test for enrichment of systematic 5hmC within gene compartments and CGI regions, as well as enrichment of systematic CpGs among CpGs in the eQTHMs and DHMRs. Fisher's exact tests were also used to test for enrichment of positively and negatively-correlated eQTHMs in target gene compartments, CGI regions, and ChromHMM states, as well as enrichment of DHMR CpGs across ChromHMM states. One-way repeated measures analysis of variance (ANOVA) was used to test for differences in 5hmC proportions

within gene compartments of genes with varying expression levels, as well as to assess 5hmC differences across ChromHMM compartments.

All analyses were conducted using R version 4.1.1.

Results

Sample cohort

This study analyzed data from 227 mother-infant pairs from the RICHS cohort, with the mother and infant demographics displayed in **Table 3-1**. The sample consisted mainly of white mothers (77.1% of samples), with a mean age of 30.9 years. There was a nearly equal distribution of male and female infants (51.1% vs. 48.9%, respectively), with a mean gestational age of 39.4 weeks, and by study design, the sample was over-represented by infants born SGA (14.5%) and LGA (30.8%).

Placental 5hmC distribution

Placental 5hmC proportions were notably low with very little variation. Mean 5hmC across all 689,815 autosomal CpGs ranged from 0 – 56%, with the 227 samples having a grand mean of 2.98% (**Fig. 3-1**). Among all CpGs, 46,921 (6.8%) were deemed systematic (5hmC proportion ≥ 0.10 in at least 50% of samples), with a mean 5hmC of 15.57% (**Fig. 3-1**).

Across gene compartments, we observed significant enrichment of systematic 5hmC in gene bodies (OR = 1.26; 95% CI = 1.23, 1.28; $p < 0.0001$), 3’ UTRs (OR = 1.14; 95% CI = 1.10, 1.19; $p < 0.0001$), and 5’ UTRs (OR = 1.10; 95% CI = 1.07, 1.13; $p < 0.0001$). Significant depletion of systematic 5hmC was observed in TSS (OR = 0.66; 95% CI = 0.64, 0.68; $p < 0.0001$) (**Fig. 3-1**). Among CGI regions, we observed a significant enrichment of systematic 5hmC at "open sea” regions (OR = 2.09; 95% CI = 2.05, 2.13; $p < 0.0001$), and depletion in CGIs (OR = 0.09; 95% CI = 0.08, 0.10; $p < 0.0001$) (**Fig. 3-1**).

Across ChromHMM states, mean 5hmC was most abundant among enhancer CpGs (mean = 5.2%), flanking transcribed (mean = 5.1%), and transcribed regions (mean = 4.2%). 5hmC was least abundant among CpGs found in active TSS, flanking poised TSS/enhancers, and flanking active TSS (means = 0.39%, 1.1%, and 2.0%, respectively) (**Fig. 3-2**). Significant differences in 5hmC proportions were found across all ChromHMM states ($p < 0.05$).

We found that 5hmC proportions varied by both genic region and gene expression level. Although proportions of 5hmC were largely dependent on the location of the CpG site relative to the gene, hydroxymethylation proportion generally increased with transcript abundance. 5hmC among CpGs lying in the gene body or 3' UTR increased as transcript abundance increased from transcriptionally silent (represented by the first quartile of transcript abundances) to active genes (fourth quartile) (**Fig. 3-3**). Among CpGs in the 5' UTR, 5hmC proportion was highest in transcriptionally silent genes, then decreased as transcription increased, and finally increased slightly in the most actively transcribed genes. 5hmC was negatively correlated with transcript in TSS; as transcript abundance increased, 5hmC proportion decreased. Within all gene compartments, 5hmC proportion differed across expression quartiles ($p < 0.05$, for all) (**Fig. 3-3**). Thus, we observed significant differences in 5hmC proportions across expression quartiles, indicating a potential association between 5hmC and expression.

eQTHM analysis

We identified 499 eQTHMs (**Table S3-1**), consisting of 473 unique CpGs and 284 unique genes. Among the 473 unique CpGs, 165 (34.9%) were deemed systematic CpGs from our placental 5hmC distribution results, representing a significant enrichment (OR = 6.79; 95% CI = 5.60, 8.21; $p < 0.0001$). We also observed an overall positive correlation between 5hmC and expression, with 75.4% of significant eQTHMs demonstrating a positive correlation with cis gene transcript abundance (**Table 3-2**).

Among eQTHMs positively correlated with transcript abundance (positively correlated eQTHMs), we observed a depletion (OR = 0.60; 95% CI = 0.39, 0.93; $p < 0.05$) of CpGs that fall within their target gene (**Fig. 3-4**), along with an enrichment (OR = 2.89; 95% CI = 1.84, 4.57; $p < 0.001$) in "open sea" regions of the genome (**Fig. 3-4**). Among eQTHMs negatively correlated with transcript abundance (negatively correlated eQTHMs), we observed an enrichment (OR = 8.93; 95% CI = 3.64, 24.16; $p < 0.0001$) in the TSS regions (**Fig. 3-4**) and in CGIs (OR = 19.45; 95% CI = 8.49, 50.44; $p < 0.0001$) (**Fig. 3-4**).

When plotting the distance between the eQTHMs and their associated gene, we observed a larger proportion of eQTHMs lying close to the TSS of the associated gene (**Fig. 3-4**). As the distance from the TSS increases in the upstream direction we observed a smaller proportion of eQTHMs in these areas. This pattern of increased proportions at the 5' end of the genes and decreased proportions further away from the 5' end was also observed across all cis eQTHM-gene pairs (**Fig. 3-4**).

We found a significant enrichment of eQTHMs among enhancers CpGs (OR = 2.29; 95% CI = 1.85, 2.80; $p < 0.01$) **(Fig. 3-4),** with particular enrichment of positively correlated eQTHMs in these regions (OR = 2.01; 95% CI = 1.17, 3.58; $p < 0.01$) **(Fig. 3-4).** Conversely, we observed an overall depletion of eQTHMs among CpGs in active TSS regions (**Fig. 3-4**), and among eQTHMs that did reside there, there was an overall enrichment of negatively correlated eQTHMs (OR = 28.36; 95% CI = 12.56, 72.85; $p < 0.0001$) **(Fig. 3-4).** Likewise, there was an overall depletion of eQTHMs in regions flanking active TSS (**Fig. 3-4**), though amongst eQTHMs in these regions, we observed an enrichment (OR = 3.90; 95% CI = 1.76, 8.72; $p < 0.001$) of negatively correlated eQTHMs **(Fig. 3-4)**. We also found a strong enrichment of

positively correlated eQTHMs in quiescent regions (OR = 4.61; 95% CI = 2.14, 11.37; p < 0.0001) **(Fig. 3-4).**

DHMR analysis

We identified 107 significant (≥ 3 CpGs in DHMR and regional p-value < 0.05) DHMRs across 100 genes (**Table S3-2**) and encompassing 479 unique CpGs. Among these CpGs, 85 (17.7%) were considered to have systematic 5hmC, representing a significant enrichment (OR = 2.96; 95% CI = 2.31, 3.75; p < 0.0001). Additionally, 121 of these CpGs (25.3%) were also eQTHMs.

Among the 107 significant DHMRs, 42 (39.3%) fell within their target gene, with most (n=36, 33.6%) in the gene body (**Fig. 3-5**). Additionally, 47 DHMRs (43.9%) were in "open sea" regions and 30 (28.0%) in CGIs (**Fig. 3-5**). We found an enrichment of DHMR CpGs compared to CpGs outside of DHMRs in active TSS regions (OR = 2.14; 95% CI = 1.71, 2.67; p < 0.0001), enhancers (OR = 1.65; 95% CI = 1.32, 2.06; p < 0.0001), regions flanking active TSS (OR = 2.69; 95% CI = 2.08, 3.45; p < 0.0001), and poised TSS or enhancers (OR = 2.19; 95% CI = 1.58, 2.96; p < 0.0001) (**Fig. 3-5**). We observed a depletion of DHMR CpGs in polycomb repressed regions (OR = 0.16; 95% CI = 0.08, 0.27; p < 0.0001) and quiescent regions (OR = 0.29; 95% CI = 0.20, 0.40; p < 0.0001) (**Fig. 3-5**).

The DHMR with the most significant association with expression is on chromosome 19 in the body of B3GNT3 (**Fig. 3-5**). This DHMR also overlaps with a poised TSS/enhancer region. We assessed the 5hmC proportion across the eight CpGs found in this DHMR, stratified by the B3GNT3 expression quartile across all subjects, and observed a negative correlation between 5hmC proportion and B3GNT3 transcript abundance across the DMHR (**Fig. 3-5**). Interestingly, we found 29% of all DHMRs overlapped with an enhancer region. We also assessed whether DHMRs tend to be more associated with active transcripts than inactive

transcripts, and found that the proportions were evenly distributed across transcript abundance quartiles (28%, 26.2%, 28%, and 17.8% of DHMRs were associated with transcripts in the first, second, third, and fourth quartile of transcript abundance, respectively). Among the 65 DHMRs that fell outside their target gene, 51.6% and 48.4% were located in the upstream and downstream direction, respectively.

Discussion

In this study, we have achieved two distinct goals: 1) we have described the genomic landscape of hydroxymethylation in a healthy subset of infant placental tissue samples, and 2) we have identified the extent to which 5hmC is associated with placental gene expression. These findings will provide important reference data for future studies of 5hmC in placental tissue and offer insight into the roles this modification plays.

Although 5hmC proportions were relatively low across all placental samples, we found ~47,000 loci that met the criteria for demonstrating systematic placental hydroxymethylation (**Fig. 3-1**). Hydroxymethylation has been described as both an intermediate in the demethylation pathway[11], as well as a stable epigenetic mark[38]. By classifying 5hmC as systematic based on our criteria, we aim to distinguish 5hmC that fall within each of these categories. The criteria were established due to the notion that hydroxymethylation present as a result of cells undergoing active demethylation would likely be observed in only a limited number of samples, while stable, functional 5hmC regions would more likely be seen in a greater proportion of samples[19].

Our findings regarding the enrichment or depletion of systematic 5hmC at functionally relevant genomic regions (**Fig. 3-1**) are in strong concordance with results from previous placental studies. Green et al.[19] utilized placental methylation data from the 450 HumanMethylationBead Chip (450k) Array, and also observed a depletion of systematic 5hmC

in areas like the TSS and particularly, CGIs. CGIs are often associated with promoters and have shown 5hmC depletion in brain tissue[39]. Similarly, Mora Hernandez et al.[17] profiled placental 5hmC using the 450k array, and in agreement with our results, found an enrichment of 5hmC in “open seas”, along with a depletion in CGIs. Our 5hmC distribution results are also consistent with previous studies in brain samples; one study[40] utilized 450k data to quantify genome-wide patterns of 5hmC in the cerebellum and also observed significant depletion in CGIs, as well as enrichment in “shores”, “shelves”, and areas outside of CG-rich regions (“open seas”). Another study by Spiers et al.[41] characterizing 5hmC and 5mC across human fetal brain tissue found about ¼ of the autosomal CpGs assayed were characterized by non-detectable 5hmC in all brain samples, with these sites enriched in CGIs and other regulatory regions including TSS. Among CpGs that contained detectable 5hmC, the authors noted these sites were enriched in “shores”, “shelves”, and gene bodies. Finally, this relationship has been replicated in a 5hmC study[42] of breast tissue, with the authors utilizing 450k methylation data to identify genomic loci containing elevated 5hmC. The authors observed 5hmC enrichment in “open seas” and “shores”, as well as a depletion in CGIs. Our 5hmC distribution findings, along with those from other studies of placental, brain, and breast tissue, further highlight the paradoxical relationship of 5hmC abundance in CpG-poor regions and depletion in CpG-rich regions. Numerous mechanistic models have been proposed for how 5hmC is regulated in these areas, with one possible mechanism being that in CpG-rich regions, the CpGs are already methylated, and these methylated CpGs can recruit methyl-CpG binding domain (MBD) proteins to establish a constitutive heterochromatin state, thereby making TET1 inaccessible to the hypermethylated sites and preventing the conversion of 5mC to 5hmC[43]. Future studies of the placental 5hmC

could examine specific chromatin marks and associated chromatin modifiers and binding proteins to understand these relationships.

Despite our 5hmC distribution results demonstrating agreement with previous findings in placenta, brain, and breast tissue, a portion of our findings are in contrast with those found in embryonic stem cells (ESCs). Specifically, studies in mouse ESCs have shown *increased* 5hmC at CGIs associated with bivalent (TET1/Polycomb repression complex 2 (PRC2)-cobound) promoters[44]. As TET1 is capable of binding to PRC2-repressed development regulators[45], it is possible that in mouse ESCs, TET1 may have a higher binding affinity to PRC2, thereby allowing for more oxidation of 5mC to 5hmC at CGIs in the bivalent promoters[45, 46]. Additionally, mouse ESCs could have a greater proportion of CGIs bound by PRC2, in comparison to placental samples[47]. As TET1 binds to these areas, this in turn could manifest as an enrichment of 5hmC. As we observed a *depletion* of 5hmC in polycomb repressed regions (**Fig. 3-2**), our findings are in direct contrast to those of ESCs. However, it is worth noting that our 5hmC distribution results are consistent with those from other differentiated cell types; a recent study[48] characterizing 5hmC in trophoblast stem cells (TSCs) within mouse placenta showed a profound lack of TET1 peaks overlapping with bivalent or polycomb repressed regions, leading to a subsequent reduction of 5hmC in these areas.

We observed that higher proportions of 5hmC, particularly in the gene body, are associated with more highly expressed genes (**Fig. 3-3**). Previous studies have also shown 5hmC to be enriched in the gene body of active genes in mouse cerebellum[39] and that oxidation from 5mC to 5hmC prevents the binding of transcriptionally repressive MBD proteins[49]. Thus, it is possible that the observed enrichment of 5hmC in the gene body of actively transcribed genes

could be due to the absence of MBD proteins in those areas, though future studies comprised of MBD-binding marks are needed to confirm this association.

Our study also sheds light on how 5hmC may be involved in regulating expression through regulatory elements, specifically enhancers. We observed an enrichment of 5hmC marks in enhancer regions (**Fig. 3-2**), along with enrichment of eQTHMs in these regions (**Fig. 3-4**), with the majority of these eQTHMs being positively associated with expression (**Fig. 3-4**). Enrichment of 5hmC in these regions is in agreement with previous studies, which have shown that 5hmC tends to accumulate at poised and active enhancers labeled with H3K4me1, H3K18ac, and H3K27ac[50, 51]. The trend we observed between 5hmC and expression in these regions could be due to the aforementioned increased variability of 5hmC in enhancer elements, and recent studies stress the importance of altered methylation patterns at enhancers as a critical component to variation in gene expression[52, 53]. It is also possible that 5hmC at enhancers may be involved in changing the transcriptional landscape of placental tissue, which in turn aids in differentiation. The results from our DHMR analysis further support the notion that 5hmC may be regulating expression through enhancers; 29% of significant DHMRs overlapped with an enhancer region, as we observed an enrichment of DHMR CpGs in these regions (**Fig. 3-5**). As enhancers are known to regulate transcription of both distal and proximal genes[54, 55], it is possible that 5hmC across contiguous CpGs in enhancers is working in concert to further promote transcription of the associated gene.

Our eQTHM findings are also noteworthy in that we observed an enrichment of systematic 5hmC among CpGs in eQTHMs (**Table 3-2**). This is especially significant given the small number of systematic 5hmC loci across all CpGs on the array (6.8% of all CpGs were deemed systematic). As these sites demonstrate consistently higher proportion of 5hmC *and* are

significantly associated with gene expression, this provides a candidate list of loci with potentially functional 5hmC. Future studies should evaluate the functional relevance of these sites in the larger placental epigenome.

The most robust DHMR we identified was in the body of the *B3GNT3* gene (**Fig. 3-5**). Expression of *B3GNT3* is negatively correlated with 5hmC in the DHMR, meaning higher expression of *B3GNT3* associates with lower 5hmC proportions (**Fig. 3-5**). B3GNT3 is a transmembrane protein that has been found to be associated with immune cell infiltration and activation of the NF-kB pathway in gynecologic cancers[56]. A previous study[57] looking at how mutations in IDH2 and TET2 cells modulates tumorigenesis in angioimmunoblastic T cell lymphoma (AITL) also found a negative correlation between CpG hypermethylation in a differentially methylated region (DMR) of *B3GNT3* and its corresponding expression levels.

To our knowledge, this study presents the most comprehensive description of the empirical relationship between placental 5hmC and gene expression through eQTHM analyses using a large sample size (n=227) and the more comprehensive array (Illumina MethylationEpic array) than prior studies. Many of our findings are consistent with 5hmC-expression research in various other tissues and models[17, 19, 39, 40, 42]. We demonstrated that 5hmC is associated, for the most part positively, with gene expression in the placenta. Unlike previous placental epigenetic studies, we were able to leverage paired sample RNA-sequencing data to gain a better understanding of the relationship between 5hmC and expression. We were also very stringent in our control for multiple testing of the eQTHM analysis; by employing an empirical p-value threshold, we ensured that genes paired to a higher number of CpGs did not have a stronger chance of being a part of an eQTHM. We also utilized a DHMR analysis as an additional way to demonstrate the 5hmC and gene expression associations, which adds an additional level of rigor

and confidence in the reported results. This study also benefits from the recruitment of a large cohort of placental samples; while previous placental epigenetics studies were generally conducted in a small cohort of samples, the RICHS cohort represents one of the largest cohorts of placental samples ever assembled, ensuring a well-powered study necessary to evaluate empirical effects. Finally, we employed a reference-based approach when estimating cell-type proportions in the placental samples, thereby limiting confounding due to inter-sample variability in cell compositions.

The results from this study are robust and relevant within the broader area of placental health, though these findings should be interpreted within the context of this study's limitations. This is an observational study in which placental hydroxymethylation and expression were both measured in placenta at term. Therefore, we cannot conclude that our results are representative of the relationship between placental 5hmC and expression throughout development. Hydroxymethylation and expression being collected at the same time point also limits our ability to infer a causal relationship between the two. We also were unable to assess distal CpGs in our eQTHM analysis, mainly due to a limited sample size. Our DHMR analysis utilized eQTHM results with raw p-values $< 1 \times 10^{-4}$, and thus there may be some false-positives in our DHMR hits. Finally, the RICHS cohort consists predominantly of healthy, white mothers and their infants from the New England region of the United States, and thus these findings may not be as generalizable to more diverse populations.

In summary, this study presents an important step in understanding the distribution and functional relevance of 5hmC in the placenta. Although additional studies are needed for a more complete understanding of placental hydroxymethylation, these findings serve as a good starting point for investigators looking to understand the role of 5hmC in placental epigenetics.

Statements and Declarations

Data Availability

Raw data were generated at Emory University. Derived data supporting the findings of this study are available through the Gene Expression Omnibus (GEO, accession number: GSE144129).

Funding

This work was supported by grants from the National Institutes of Health, National Institute of Environmental Health Sciences (P30ES 019776, U24 ES028507, R21 ES028226) and Eunice Kennedy Shriver National Institute of Child Health and Development (R01 HD108310).

Conflicts of Interest

The authors report there are no competing interests to declare.

Ethical Standards

All enrolled participants provided written informed consent and study protocols were approved by the institutional review boards (IRB) at WIH and Emory University.

Tables

	RICHS (n = 227)
Maternal characteristics	
Age in years (mean, SD)	30.9 (4.9)
Educational attainment (n, %)	
High school or less	35 (15.4)
Post-high school or junior college	111 (48.9)
College	81 (35.7)
Self-reported race/ethnicity (n, %)	
Asian	10 (4.4)
Black	12 (5.3)
Indian	2 (0.88)
More than one race	4 (1.8)
Unknown/not reported	24 (10.6)
White	175 (77.1)
Infant characteristics	
Age in weeks (mean, SD)	39.4 (0.9)
Sex (n, %)	
Male	116 (51.1)
Female	111 (48.9)
Birthweight in grams (mean, SD)	3,556.4 (662.5)
Birthweight category (n, %)[a]	
AGA	124 (54.6)
LGA	70 (30.8)
SGA	33 (14.5)

Table 3-1: RICHS participant demographics

[a]Infants born with birthweight percentile ≤ 10% (small for gestational age [SGA]), 10-90% (appropriate for gestational age [AGA]), and ≥ 90% (large for gestational age [LGA])

	All cis CpG-gene pairs (n = 19,546,660)
Pairs below empirical-p threshold (n, %)	499 (0.003)
# unique CpGs	473
# unique genes	284
Unique CpGs that overlap with systematic CpGs (n, %)	165 (34.9)
Correlation sign (n, %)	
+	376 (75.4)
-	123 (24.6)
Estimate (mean, SD)	0.29 (0.46)
r (mean, SD)	0.20 (0.36)

Table 3-2. eQTHM summary results

Abbreviations: r = correlation coefficient

Figures

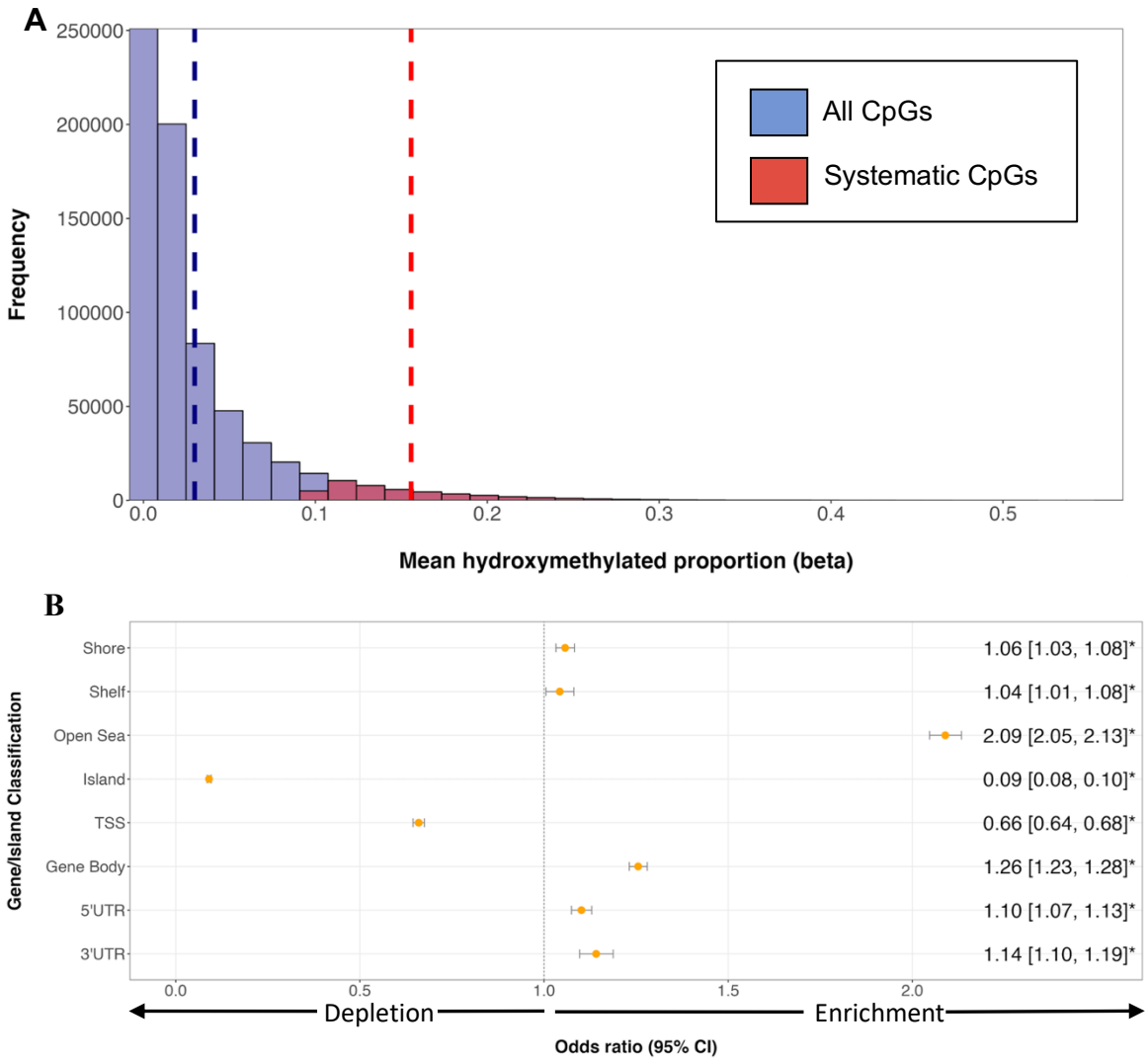


Figure 3-1: Distribution of systematic 5hmC across the placental epigenome. Systematic CpGs were defined as loci with 5hmC proportion > 0.10 in at least 50% (n = 113) of samples. A total of 689,815 autosomal CpGs were assayed. Among those, 46,921 (6.8%) were considered systematic, with the remaining 642,894 (93.2%) deemed non-systematic. **A)** Distribution of all CpGs sites (blue) and systematic sites (red) on EPIC array. 5hmC proportions display a strong right skew, with samples having a mean of 2.98% (indicated by the vertical dashed blue line). Systematic CpGs had a 5hmC mean of 15.57% across samples (indicated by vertical dashed red line). **B**) Distribution of systematic 5hmC by gene and CGI compartments. ORs and 95% CIs were determined by Fisher's exact test, with ORs marked by asterisks defined as significant ($p < 0.05$). ORs above 1.0 indicate enrichment for systematic 5hmC in comparison to other location classifiers, and ORs below 1.0 indicate depletion. CpGs associated with >1 gene class may be counted twice. CGI shores define loci < 2 kb from CGI, shelves are loci 2-4 kb from CGI, and open seas are loci > 4 kb from CGI.

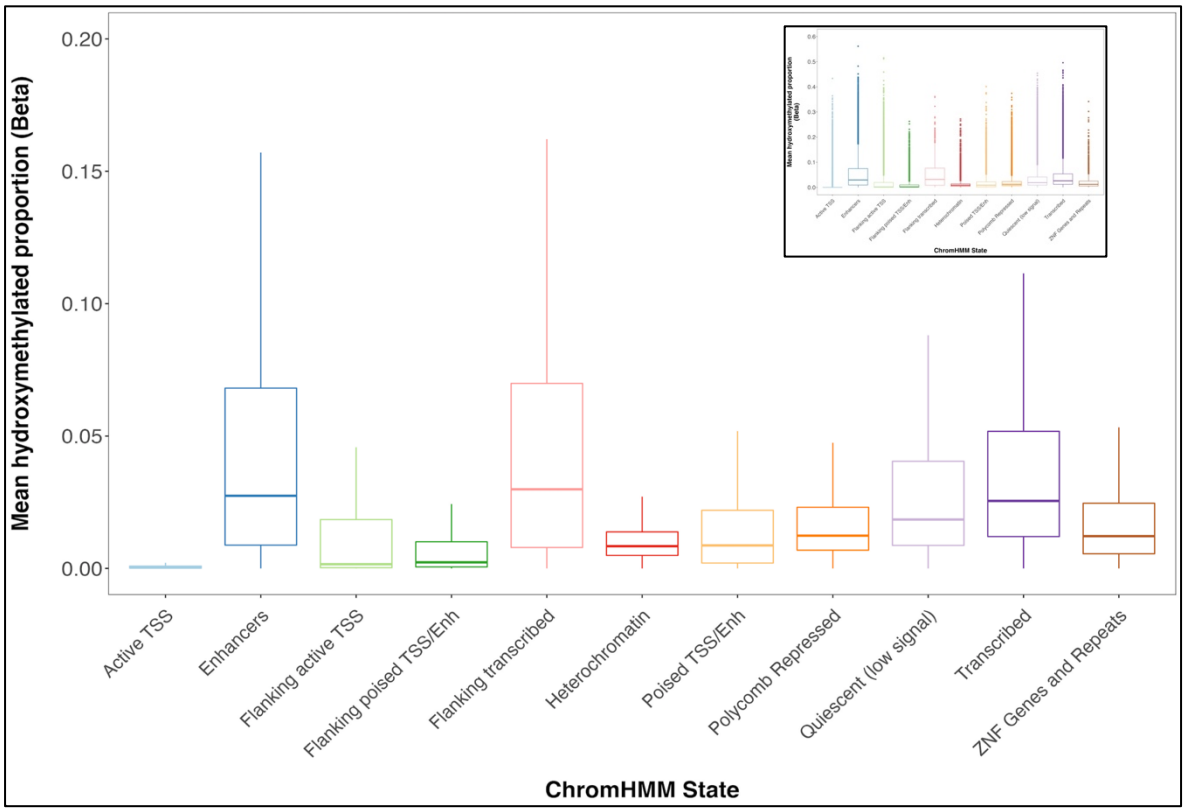


Figure 3-2: Placental 5hmC distribution across ChromHMM states. Box plots denote distribution of 5hmC, with boxes encompassing 25th to 75th percentile (with the length of the box representing the interquartile range (IQR), defined as the difference between the 25th and 75th percentiles), the median denoted as line within box, and the upper and lower whiskers marking the maximum and minimum values no further than 1.5 x IQR, respectively. Outliers were suppressed to improve visualization of differences. Inset plot represents the plot with outliers included. One-way repeated measures ANOVA revealed significant differences in 5hmC levels across states ($p < 0.05$).

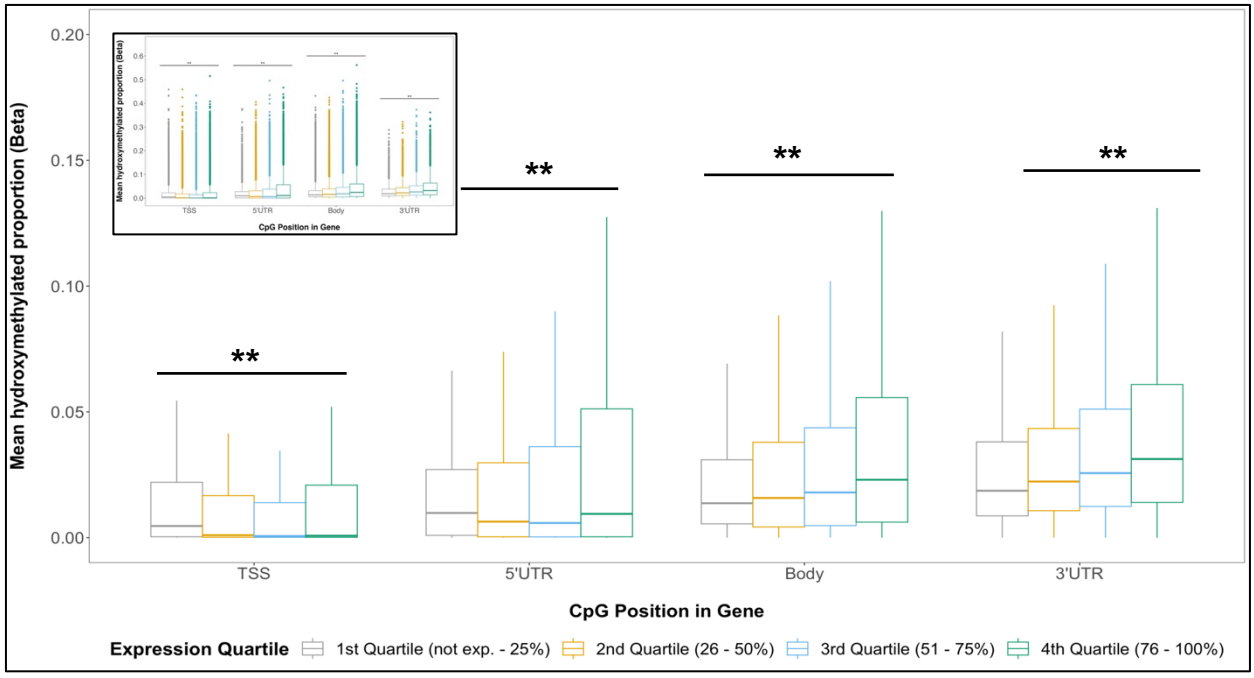
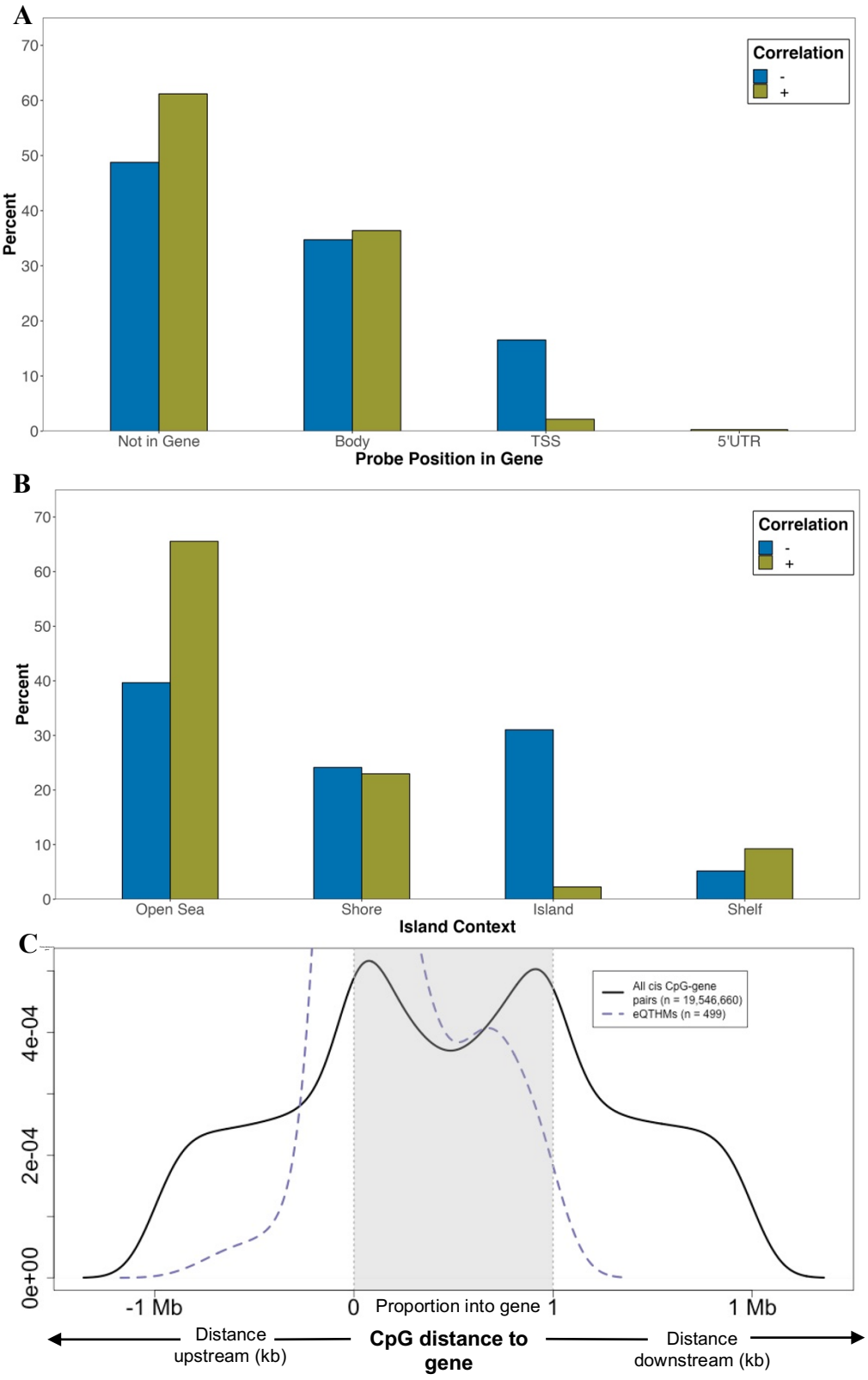


Figure 3-3: Placental 5hmC distribution across gene compartments of genes with varying expression levels. Genes were grouped into expression quartiles based on mean transcript levels across all subjects. CpG probes were mapped to compartment of nearest gene from EPIC array annotation package. Box plots denote distribution of 5hmC, with boxes encompassing 25th to 75th percentile (with the length of the box representing the IQR, defined as the difference between the 25th and 75th percentiles), the median denoted as line within box, and the upper and lower whiskers marking the maximum and minimum values no further than 1.5 x IQR, respectively. Outliers were suppressed to improve visualization of differences. Inset plot represents the plot with outliers included. Asterisks mark significant differences in 5hmC levels across expression quartiles within each gene compartment (ANOVA $p < 0.05$).


A
Correlation
-
+
Percent
70
60
50
40
30
20
10
0
Not in Gene
Body
TSS
5'UTR
Probe Position in Gene
B
Correlation
-
+
Percent
70
60
50
40
30
20
10
0
Open Sea
Shore
Island
Shelf
Island Context
C
All cis CpG-gene pairs (n = 19,546,660)
eQTHMs (n = 499)
4e-04
2e-04
0e+00
-1 Mb
0
Proportion into gene
1
1 Mb
Distance upstream (kb)
CpG distance to gene
Distance downstream (kb)

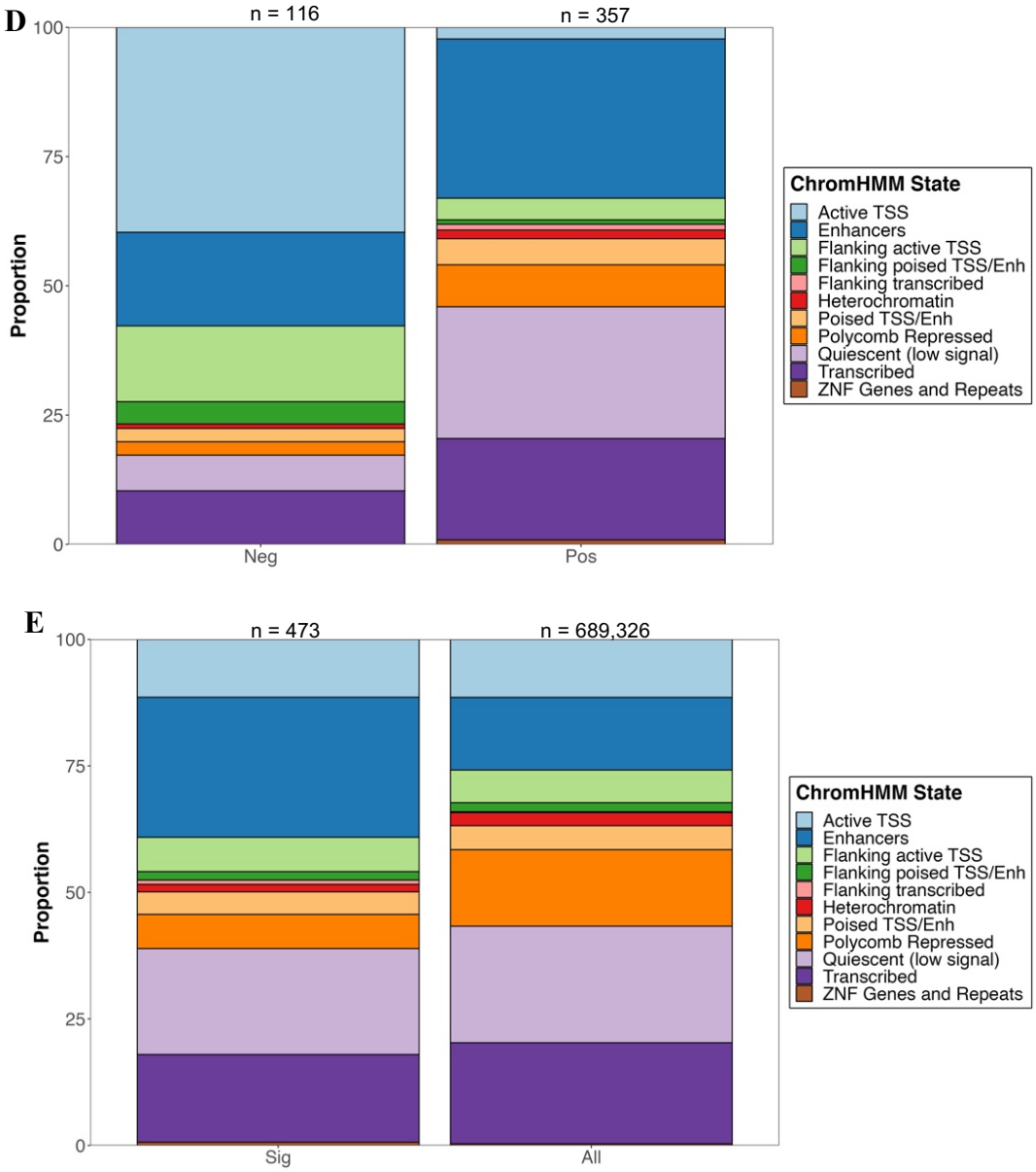


Figure 3-4: Characterization and distribution of genomic location of eQTHM signals. **A)** Percentage of positively and negatively correlated eQTHMs across gene compartments of target gene. **B)** Percentage of positively and negatively correlated eQTHMs across CGI compartments. Fisher's exact tests were used to test for enrichment of positively and negatively-correlated eQTHMs across gene compartments (A) and CGI regions (B). **C)** CpG distance from TSS/TTS of target eQTHM gene. TSS and TTS are represented by 0, 1 on x-axis, respectively. CpGs lying within gene have distance represented by proportion through gene. Shown are significant eQTHMs (dashed purple line) and all cis CpG-gene pairs from Matrix eQTL (solid black line). **D and E)** Proportion of CpGs within ChromHMM states. Shown are negatively (Neg) and positively (Pos) associated CpGs (D), along with all significant CpGs (Sig) and all CpGs on EPIC array (E). Fisher's exact tests were used to test for enrichment of eQTHMs across ChromHMM states. Numbers on top of each bar represent # of CpGs in that group.

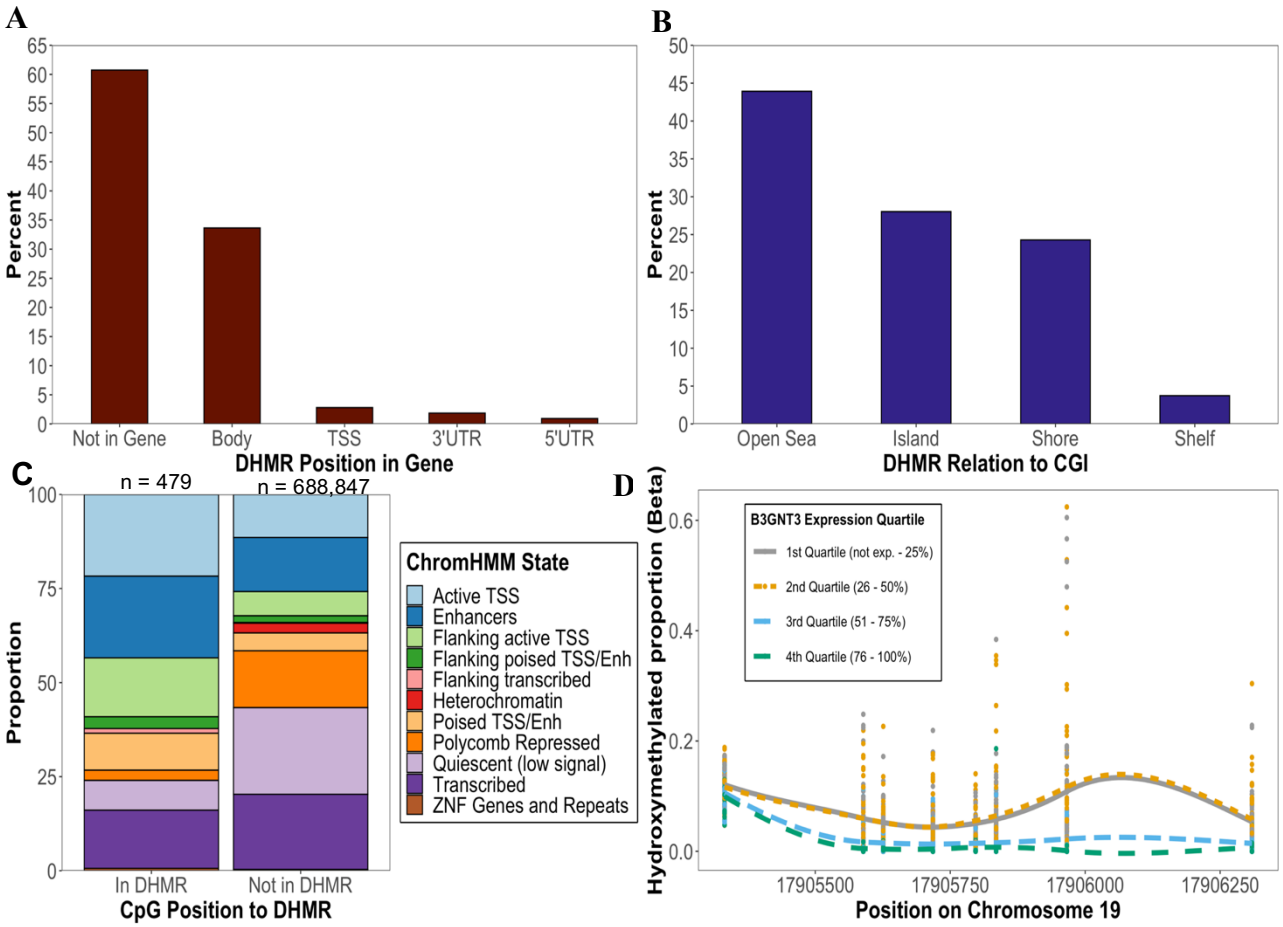


Figure 3-5: Characterization and distribution of transcription-associated DHMRs. Percentage of all significant DHMRs (n = 107) across **A)** genic compartments and **B)** CGI regions. **C)** Proportion of CpGs across ChromHMM states among CpGs in and out of DHMRs. Fisher's exact tests were used to test for enrichment of CpGs across ChromHMM states. Numbers on top of each bar represent # of CpGs in that group. **D)** Most significant DHMR at B3GNT3 gene. Hydroxymethylated proportions among subjects at each of the 8 CpG probes (represented by vertical lines of stacked points) in the DHMR are shown. Data is stratified by B3GNT3 $\log_2$TMM expression quartiles among subjects. CpG site positions are displayed along the x-axis. Lines represent local regression model using the locally estimated scatterplot smoothing (LOESS) method.

Supplementary Tables

Table S3-1: Significant CpG-gene pairs from eQTHM analysis

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg03515999	PSMD5	12.17	1.34E-25	2.62E-18	0.30	0.66	0.44
cg16289896	ATP1A4	11.72	2.98E-24	2.91E-17	0.78	0.65	0.42
cg04677839	SLC6A15	11.50	1.35E-23	8.83E-17	1.54	0.64	0.41
cg15407162	TOB2P1	10.85	1.14E-21	5.59E-15	0.55	0.62	0.38
cg16289896	CASQ1	10.62	5.45E-21	2.13E-14	0.75	0.61	0.37
cg27213577	STON2	10.45	1.71E-20	5.02E-14	0.78	0.60	0.36
cg16389864	B3GNT3	-10.44	1.80E-20	5.02E-14	-0.08	-0.60	0.36
cg26251192	HEATR4	10.40	2.39E-20	5.84E-14	0.23	0.60	0.36
cg04627461	STBD1	10.21	8.46E-20	1.84E-13	0.79	0.59	0.35
cg05772903	DSEL	10.00	3.25E-19	6.34E-13	0.31	0.59	0.34
cg14112503	SMUG1	-9.86	8.22E-19	1.46E-12	-0.07	-0.58	0.34
cg01183122	KCNE1	9.84	9.74E-19	1.59E-12	1.47	0.58	0.34
cg10814298	DYRK1B	9.75	1.73E-18	2.60E-12	1.18	0.58	0.33
cg14298503	ATP2C2	9.67	2.93E-18	4.09E-12	1.00	0.57	0.33
cg10334928	STON2	9.64	3.67E-18	4.78E-12	0.91	0.57	0.33
cg20809402	TOB2P1	9.56	6.25E-18	7.63E-12	1.76	0.57	0.32
cg01733572	DNAAF1	9.38	1.98E-17	2.23E-11	0.18	0.56	0.32
cg01970114	CUL7	-9.37	2.05E-17	2.23E-11	-0.04	-0.56	0.32
cg10878397	STBD1	9.33	2.72E-17	2.80E-11	1.10	0.56	0.31
cg04517722	B3GNT3	-9.28	3.79E-17	3.71E-11	-0.09	-0.56	0.31
cg04940991	STON2	9.18	6.96E-17	6.48E-11	0.90	0.55	0.31
cg21694531	SLC30A8	9.11	1.12E-16	9.91E-11	0.18	0.55	0.30
cg06514344	B3GNT3	-9.10	1.23E-16	1.01E-10	-0.08	-0.55	0.30
cg08105571	ATP1A4	9.09	1.24E-16	1.01E-10	2.26	0.55	0.30
cg21558508	SAR1B	-8.95	3.04E-16	2.37E-10	-0.06	-0.54	0.30
cg16024904	B3GNT3	-8.95	3.16E-16	2.37E-10	-0.09	-0.54	0.30
cg03847020	ARFGAP3	8.87	5.34E-16	3.86E-10	0.32	0.54	0.29
cg15292565	SMUG1	8.86	5.67E-16	3.95E-10	0.26	0.54	0.29
cg21220670	ETV7	8.83	6.72E-16	4.53E-10	0.86	0.54	0.29
cg09228454	SMUG1	8.81	7.55E-16	4.92E-10	0.24	0.54	0.29
cg02657438	STON2	8.65	2.05E-15	1.29E-09	0.94	0.53	0.28
cg08183066	DDX19B	-8.47	6.64E-15	4.06E-09	-0.04	-0.52	0.27
cg21546132	PLA2G2F	8.39	1.09E-14	6.43E-09	0.57	0.52	0.27
cg25703541	GSTT1	-8.38	1.12E-14	6.47E-09	-0.22	-0.52	0.27

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg22424213	SGSM1	8.35	1.33E-14	7.34E-09	1.03	0.52	0.27
cg08105571	CASQ1	8.35	1.35E-14	7.34E-09	2.19	0.52	0.27
cg07612367	CUL7	-8.32	1.67E-14	8.81E-09	-0.04	-0.52	0.27
cg02912161	CUL7	-8.31	1.80E-14	9.26E-09	-0.04	-0.52	0.27
cg25148230	CDCP1	8.28	2.06E-14	1.03E-08	0.99	0.51	0.26
cg19778253	SCCPDH	8.27	2.22E-14	1.08E-08	0.75	0.51	0.26
cg04645556	TMX4	8.21	3.24E-14	1.55E-08	0.21	0.51	0.26
cg18955493	SGSM1	8.20	3.48E-14	1.62E-08	0.72	0.51	0.26
cg25519284	TUSC3	-8.15	4.63E-14	2.10E-08	-0.07	-0.51	0.26
cg17385905	TRPV5	-8.14	5.13E-14	2.28E-08	-0.99	-0.51	0.26
cg04905092	NARS2	8.11	5.92E-14	2.57E-08	0.39	0.51	0.26
cg17879066	PGM1	8.10	6.25E-14	2.65E-08	0.58	0.51	0.26
cg09153565	AQP3	-8.07	7.72E-14	3.21E-08	-0.46	-0.50	0.25
cg04139874	PDGFD	7.96	1.53E-13	6.22E-08	0.37	0.50	0.25
cg24483411	IL36RN	7.95	1.59E-13	6.33E-08	1.16	0.50	0.25
cg26251192	ACOT1	7.94	1.68E-13	6.55E-08	0.14	0.50	0.25
cg16744531	B3GNT3	-7.87	2.63E-13	1.01E-07	-0.10	-0.49	0.24
cg18334211	PRKAG2	-7.86	2.76E-13	1.04E-07	-0.07	-0.49	0.24
cg22131825	CUL7	-7.86	2.81E-13	1.04E-07	-0.03	-0.49	0.24
cg05196621	THNSL2	7.83	3.27E-13	1.18E-07	0.66	0.49	0.24
cg02216667	NOS1AP	7.80	3.85E-13	1.37E-07	0.83	0.49	0.24
cg07203561	P2RX7	-7.73	5.99E-13	2.09E-07	-0.37	-0.49	0.24
cg19768360	ST3GAL3	-7.72	6.19E-13	2.12E-07	-0.57	-0.49	0.24
cg21211480	TMEM106A	-7.71	6.81E-13	2.30E-07	-0.03	-0.49	0.24
cg06967124	FLT4	7.68	8.23E-13	2.73E-07	1.02	0.49	0.24
cg01358620	HEATR5A	-7.66	9.08E-13	2.96E-07	-0.35	-0.48	0.23
cg08526264	AQP11	-7.64	1.04E-12	3.33E-07	-0.15	-0.48	0.23
cg00571421	CUL7	-7.62	1.14E-12	3.61E-07	-0.04	-0.48	0.23
cg13054523	METRNL	7.59	1.35E-12	4.18E-07	0.52	0.48	0.23
cg08986950	ERAP1	7.58	1.50E-12	4.59E-07	0.02	0.48	0.23
cg11058900	CUL7	-7.50	2.39E-12	7.04E-07	-0.03	-0.48	0.23
cg11794635	RAB6B	7.50	2.40E-12	7.04E-07	0.50	0.48	0.23
cg15407162	ZKSCAN4	7.49	2.47E-12	7.04E-07	0.09	0.48	0.23
cg00648558	GJB5	7.49	2.47E-12	7.04E-07	0.43	0.48	0.23
cg17998522	PLA2R1	7.49	2.48E-12	7.04E-07	1.05	0.48	0.23
cg22593405	HOXB2	7.46	3.02E-12	8.42E-07	0.42	0.47	0.23
cg00832635	ZNF208	7.45	3.08E-12	8.49E-07	1.02	0.47	0.23
cg12022967	ATP8A2	7.39	4.34E-12	1.18E-06	0.12	0.47	0.22

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg01040129	TOB2P1	7.37	5.14E-12	1.38E-06	0.15	0.47	0.22
cg22247664	GSTM4	-7.35	5.49E-12	1.45E-06	-0.61	-0.47	0.22
cg02323744	ATP2C2	7.33	6.34E-12	1.65E-06	0.19	0.47	0.22
cg15127088	WNT7A	7.27	8.95E-12	2.30E-06	1.35	0.47	0.22
cg05237429	CDCP1	7.27	9.18E-12	2.33E-06	0.94	0.47	0.22
cg07054978	BTBD11	7.23	1.11E-11	2.78E-06	0.90	0.46	0.22
cg26121234	WNT7A	7.21	1.26E-11	3.13E-06	1.38	0.46	0.21
cg26335760	RAB6B	7.20	1.34E-11	3.27E-06	0.98	0.46	0.21
cg25485560	MAP3K5	-7.19	1.45E-11	3.51E-06	-0.29	-0.46	0.21
cg27540823	ERCC6	-7.17	1.56E-11	3.73E-06	-0.18	-0.46	0.21
cg06243115	DSEL	7.17	1.60E-11	3.78E-06	0.25	0.46	0.21
cg10490670	PIP5K1C	7.16	1.74E-11	4.05E-06	0.64	0.46	0.21
cg09000199	CDCP1	7.15	1.78E-11	4.10E-06	0.40	0.46	0.21
cg17125727	THSD7A	7.14	1.89E-11	4.30E-06	0.79	0.46	0.21
cg18685394	CCK	7.14	1.94E-11	4.33E-06	1.26	0.46	0.21
cg27563138	SLCO4A1	7.14	1.95E-11	4.33E-06	1.48	0.46	0.21
cg10325053	TMX4	7.10	2.34E-11	5.14E-06	0.45	0.46	0.21
cg20847471	B3GNT3	-7.09	2.48E-11	5.39E-06	-0.09	-0.46	0.21
cg21968849	DDX19B	-7.09	2.52E-11	5.42E-06	-0.04	-0.46	0.21
cg26466773	DNAAF1	7.08	2.61E-11	5.54E-06	1.88	0.46	0.21
cg05672540	TMX4	7.08	2.64E-11	5.54E-06	0.20	0.46	0.21
cg04688366	THSD7A	7.08	2.72E-11	5.65E-06	0.15	0.46	0.21
cg20809402	ZKSCAN4	7.06	3.02E-11	6.21E-06	0.28	0.45	0.21
cg02359409	PEX6	-7.06	3.08E-11	6.28E-06	-0.05	-0.45	0.21
cg14815361	PDLIM4	7.02	3.75E-11	7.55E-06	0.99	0.45	0.21
cg08279665	ATP1A4	-7.00	4.11E-11	8.20E-06	-0.11	-0.45	0.20
cg03438754	SGSM1	7.00	4.16E-11	8.21E-06	0.18	0.45	0.20
cg03316864	B3GNT3	-7.00	4.34E-11	8.48E-06	-0.07	-0.45	0.20
cg12294310	DSEL	6.98	4.70E-11	9.09E-06	0.08	0.45	0.20
cg10840704	TACSTD2	6.97	4.99E-11	9.57E-06	0.55	0.45	0.20
cg01743658	HEATR5A	6.96	5.18E-11	9.80E-06	1.14	0.45	0.20
cg10325053	PLCB1	6.96	5.22E-11	9.80E-06	0.56	0.45	0.20
cg21174128	PLCB1	6.93	6.14E-11	1.14E-05	0.09	0.45	0.20
cg23459486	PLCB1	6.93	6.42E-11	1.18E-05	0.43	0.45	0.20
cg12647020	ARFGAP3	-6.91	6.99E-11	1.28E-05	-0.04	-0.45	0.20
cg24252262	CDH12	6.89	7.69E-11	1.39E-05	0.28	0.45	0.20
cg06383401	IL36RN	6.88	8.25E-11	1.48E-05	0.39	0.45	0.20
cg17846098	DNAAF1	-6.88	8.46E-11	1.50E-05	-1.04	-0.45	0.20

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg07303975	ARMS2	-6.86	9.39E-11	1.65E-05	-0.64	-0.44	0.20
cg14343062	ERCC6	-6.84	1.04E-10	1.80E-05	-0.32	-0.44	0.20
cg14909555	TMEM65	6.84	1.05E-10	1.80E-05	0.54	0.44	0.20
cg25042239	CUL7	-6.84	1.05E-10	1.80E-05	-0.03	-0.44	0.20
cg00364287	P2RX7	6.83	1.09E-10	1.85E-05	0.72	0.44	0.20
cg03236032	DSEL	6.83	1.12E-10	1.89E-05	0.11	0.44	0.20
cg01750053	DPYD	6.82	1.17E-10	1.96E-05	0.15	0.44	0.20
cg08472247	DDX19B	-6.81	1.22E-10	2.02E-05	-0.04	-0.44	0.20
cg08377924	TMEM106A	-6.75	1.76E-10	2.89E-05	-0.04	-0.44	0.19
cg18061573	PGM1	6.67	2.65E-10	4.32E-05	0.56	0.43	0.19
cg12556325	LRRC61	6.66	2.83E-10	4.58E-05	0.46	0.43	0.19
cg02028568	SGSM1	6.63	3.37E-10	5.40E-05	0.17	0.43	0.19
cg10034572	PLA2R1	6.62	3.58E-10	5.69E-05	0.56	0.43	0.19
cg03318428	GRAMD4	6.60	3.95E-10	6.23E-05	0.38	0.43	0.19
cg11011131	ARMS2	6.59	4.08E-10	6.38E-05	1.58	0.43	0.19
cg27254924	LY6K	6.59	4.24E-10	6.58E-05	1.04	0.43	0.19
cg18336854	TUSC3	6.59	4.28E-10	6.59E-05	0.62	0.43	0.19
cg06288340	HPGD	6.58	4.47E-10	6.82E-05	0.84	0.43	0.18
cg06755741	P2RX7	-6.56	4.97E-10	7.53E-05	-0.61	-0.43	0.18
cg06289802	SLC9A3R1	6.54	5.56E-10	8.35E-05	0.44	0.43	0.18
cg03992114	ATP11A	6.53	5.77E-10	8.61E-05	0.47	0.43	0.18
cg11545871	IP6K3	6.52	5.99E-10	8.87E-05	0.11	0.43	0.18
cg06061081	AQP11	6.51	6.45E-10	9.43E-05	0.12	0.43	0.18
cg18569135	GAS7	6.51	6.51E-10	9.43E-05	0.26	0.43	0.18
cg11704068	SLCO4A1	6.51	6.51E-10	9.43E-05	0.93	0.43	0.18
cg24014795	STON2	6.51	6.63E-10	9.52E-05	0.41	0.43	0.18
cg17762770	PGM1	6.50	6.96E-10	9.92E-05	0.60	0.43	0.18
cg11012194	PSMD5	6.50	7.01E-10	9.93E-05	0.17	0.43	0.18
cg18894440	CRISPLD1	6.48	7.77E-10	0.00011	0.21	0.42	0.18
cg03698089	PLCB1	6.47	7.95E-10	0.00011	0.11	0.42	0.18
cg17507371	GHR	6.47	8.05E-10	0.00011	0.49	0.42	0.18
cg24624696	DSEL	6.47	8.07E-10	0.00011	0.13	0.42	0.18
cg26141324	B3GNT3	-6.47	8.25E-10	0.00011	-0.08	-0.42	0.18
cg14936846	DSG2	6.45	8.72E-10	0.00012	0.44	0.42	0.18
cg17192247	MAPRE3	6.44	9.32E-10	0.00013	0.68	0.42	0.18
cg23459486	TMX4	6.44	9.37E-10	0.00013	0.32	0.42	0.18
cg08279665	CASQ1	-6.42	1.07E-09	0.00014	-0.10	-0.42	0.18
cg24835545	KCNC4	6.41	1.10E-09	0.00014	1.02	0.42	0.18

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg27474431	IL36RN	-6.41	1.10E-09	0.00014	-0.60	-0.42	0.18
cg25094921	SLC17A5	6.40	1.15E-09	0.00015	0.41	0.42	0.18
cg15808331	CDCP1	6.40	1.18E-09	0.00015	0.20	0.42	0.18
cg03252829	KCNK17	-6.39	1.23E-09	0.00016	-0.15	-0.42	0.18
cg16935597	TMC6	6.39	1.24E-09	0.00016	0.78	0.42	0.18
cg11463696	DDX19B	-6.37	1.35E-09	0.00017	-0.03	-0.42	0.18
cg15623143	STARD5	-6.37	1.41E-09	0.00018	-0.04	-0.42	0.18
cg09898793	KCNC4	6.36	1.43E-09	0.00018	0.82	0.42	0.17
cg12089439	TUSC3	-6.36	1.47E-09	0.00018	-0.04	-0.42	0.17
cg02408321	PDLIM4	6.35	1.50E-09	0.00018	0.58	0.42	0.17
cg06454698	GRAMD4	6.35	1.50E-09	0.00018	0.34	0.42	0.17
cg22462856	SMUG1	6.35	1.51E-09	0.00018	0.14	0.42	0.17
cg17270927	CFL2	6.35	1.53E-09	0.00019	0.25	0.42	0.17
cg07739179	NAGS	6.35	1.55E-09	0.00019	0.47	0.42	0.17
cg02527190	SLC27A6	6.34	1.62E-09	0.00019	0.31	0.42	0.17
cg06993952	CEP128	6.31	1.93E-09	0.00023	0.05	0.42	0.17
cg07019869	STON2	6.31	1.94E-09	0.00023	0.42	0.42	0.17
cg08089513	FAM3B	6.29	2.14E-09	0.00025	0.97	0.41	0.17
cg11940318	CDCP1	6.29	2.17E-09	0.00025	0.38	0.41	0.17
cg15026277	TMEM106A	-6.27	2.33E-09	0.00027	-0.03	-0.41	0.17
cg01703818	NARS2	6.27	2.37E-09	0.00027	0.34	0.41	0.17
cg27426267	FAM118A	6.27	2.39E-09	0.00027	0.27	0.41	0.17
cg07183362	LPCAT1	6.26	2.51E-09	0.00029	0.51	0.41	0.17
cg08406047	PSMG1	6.26	2.52E-09	0.00029	0.28	0.41	0.17
cg13330363	SPESP1	6.25	2.63E-09	0.00030	1.40	0.41	0.17
cg15059622	PVR	6.24	2.73E-09	0.00031	0.32	0.41	0.17
cg07876788	WDR91	6.24	2.78E-09	0.00031	0.54	0.41	0.17
cg04861640	TOB2P1	-6.23	2.96E-09	0.00033	-0.13	-0.41	0.17
cg02475902	MUC4	-6.21	3.21E-09	0.00035	-1.63	-0.41	0.17
cg16485140	ZNF208	6.21	3.22E-09	0.00035	0.10	0.41	0.17
cg10829629	SNCG	6.20	3.41E-09	0.00037	0.36	0.41	0.17
cg02149965	VARS2	6.20	3.45E-09	0.00037	0.43	0.41	0.17
cg11011131	HTRA1	6.20	3.49E-09	0.00038	0.92	0.41	0.17
cg16712789	TMX4	6.19	3.55E-09	0.00038	0.04	0.41	0.17
cg01660001	HEG1	6.18	3.85E-09	0.00041	0.63	0.41	0.17
cg24940138	TMEM106A	-6.16	4.19E-09	0.00045	-0.03	-0.41	0.17
cg16347629	GJB5	6.11	5.37E-09	0.00056	0.84	0.40	0.16
cg02589828	SAR1B	-6.11	5.41E-09	0.00057	-0.05	-0.40	0.16

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg07606636	PIK3C2B	6.11	5.57E-09	0.00058	0.55	0.40	0.16
cg07722372	SV2B	6.10	5.76E-09	0.00060	0.61	0.40	0.16
cg15573406	SNX25	6.09	5.95E-09	0.00061	0.25	0.40	0.16
cg16253157	AMIGO1	6.09	6.14E-09	0.00063	1.08	0.40	0.16
cg26251585	THSD7A	6.08	6.38E-09	0.00065	0.56	0.40	0.16
cg04041654	NCF2	6.08	6.54E-09	0.00066	0.32	0.40	0.16
cg08526264	PAK1	-6.06	6.99E-09	0.00070	-0.04	-0.40	0.16
cg23663999	PDGFD	6.06	7.14E-09	0.00072	0.43	0.40	0.16
cg15371526	SGSM1	6.06	7.18E-09	0.00072	0.47	0.40	0.16
cg18236464	PAPPA2	6.05	7.50E-09	0.00074	1.02	0.40	0.16
cg01811895	CBLB	-6.05	7.60E-09	0.00075	-0.09	-0.40	0.16
cg08412750	FLT4	6.04	7.76E-09	0.00076	0.64	0.40	0.16
cg05524458	ANKRD33B	6.04	7.80E-09	0.00076	0.78	0.40	0.16
cg01067604	VAMP3	-6.03	8.12E-09	0.00079	-0.03	-0.40	0.16
cg01691763	INPP5B	6.03	8.20E-09	0.00079	0.23	0.40	0.16
cg11339420	DNAAF1	6.03	8.28E-09	0.00080	0.93	0.40	0.16
cg19018267	PHYH	-6.02	8.71E-09	0.00084	-0.12	-0.40	0.16
cg25817165	CNDP2	-6.02	8.76E-09	0.00084	-0.05	-0.40	0.16
cg07005960	METRNL	6.01	9.14E-09	0.00087	0.36	0.40	0.16
cg08150575	CEACAM19	6.01	9.23E-09	0.00087	0.70	0.40	0.16
cg10578851	HPCAL1	6.01	9.36E-09	0.00088	0.85	0.40	0.16
cg26586567	CSMD1	5.99	1.02E-08	0.00096	0.29	0.40	0.16
cg17834632	METRNL	5.98	1.10E-08	0.00102	0.45	0.40	0.16
cg07078958	SH3BP5	5.97	1.15E-08	0.00106	0.06	0.40	0.16
cg23105600	LRFN5	5.97	1.15E-08	0.00106	0.30	0.40	0.16
cg19157819	SYNPO2L	5.97	1.15E-08	0.00106	0.54	0.40	0.16
cg01514075	VAMP3	5.97	1.16E-08	0.00106	0.22	0.40	0.16
cg00995241	PLCB1	5.97	1.16E-08	0.00106	0.08	0.40	0.16
cg27356188	TPPP	5.96	1.20E-08	0.00109	0.47	0.40	0.16
cg18291850	PLCB1	-5.95	1.23E-08	0.00111	-0.04	-0.40	0.16
cg21119074	MBOAT1	-5.95	1.24E-08	0.00111	-0.22	-0.40	0.16
cg22362405	KCNE1	5.93	1.38E-08	0.00123	0.53	0.39	0.16
cg15092561	PWP2	5.93	1.41E-08	0.00125	0.10	0.39	0.16
cg02339392	TOB2P1	-5.92	1.43E-08	0.00127	-0.18	-0.39	0.16
cg17484874	STEAP3	-5.92	1.44E-08	0.00127	-0.12	-0.39	0.16
cg24312582	ANXA4	5.91	1.52E-08	0.00134	0.35	0.39	0.15
cg23332989	WDR91	-5.91	1.56E-08	0.00136	-0.51	-0.39	0.15
cg15863052	LYPD5	5.90	1.65E-08	0.0014	0.33	0.39	0.15

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg16277399	ZNF559	5.88	1.76E-08	0.0015	0.06	0.39	0.15
cg22963267	CACNG4	5.88	1.77E-08	0.0015	1.31	0.39	0.15
cg27356188	CEP72	5.88	1.80E-08	0.0015	0.53	0.39	0.15
cg20848186	ALOX5	5.88	1.80E-08	0.0015	0.15	0.39	0.15
cg22131825	MRPL2	-5.87	1.85E-08	0.0016	-0.02	-0.39	0.15
cg18382305	ASXL3	5.87	1.89E-08	0.0016	0.16	0.39	0.15
cg17591198	PDP2	5.87	1.91E-08	0.0016	0.32	0.39	0.15
cg03804409	GABRA4	5.87	1.94E-08	0.0016	0.80	0.39	0.15
cg02557139	EXOC3L4	5.85	2.05E-08	0.0017	0.53	0.39	0.15
cg15112803	RPS28	5.85	2.10E-08	0.0017	0.49	0.39	0.15
cg06114320	HPCAL1	5.83	2.29E-08	0.0019	0.48	0.39	0.15
cg07303975	HTRA1	-5.83	2.30E-08	0.0019	-0.34	-0.39	0.15
cg21832243	TTC3	-5.82	2.44E-08	0.0020	-0.20	-0.39	0.15
cg23088397	MYO1E	5.82	2.47E-08	0.0020	0.20	0.39	0.15
cg15066323	KIF6	5.81	2.52E-08	0.0020	0.53	0.39	0.15
cg15818787	POU6F2	5.81	2.52E-08	0.0020	0.35	0.39	0.15
cg01502373	SGSM1	5.81	2.58E-08	0.0021	0.16	0.39	0.15
cg04524477	TMEM106A	-5.81	2.60E-08	0.0021	-0.05	-0.39	0.15
cg16989380	CLDN16	5.80	2.66E-08	0.0021	0.60	0.39	0.15
cg07881650	CRISPLD1	5.80	2.67E-08	0.0021	0.13	0.39	0.15
cg21744026	PAPPA2	5.80	2.71E-08	0.0022	0.79	0.39	0.15
cg08220966	SNCG	5.79	2.84E-08	0.0022	0.28	0.39	0.15
cg03104820	TMX4	5.79	2.88E-08	0.0023	0.23	0.39	0.15
cg05659314	S100A13	-5.79	2.91E-08	0.0023	-0.13	-0.39	0.15
cg00598758	TBX20	5.78	2.95E-08	0.0023	0.09	0.39	0.15
cg13045913	ZFYVE19	5.78	3.02E-08	0.0024	0.04	0.39	0.15
cg07031334	POLE4	5.77	3.08E-08	0.0024	0.39	0.39	0.15
cg08504942	B3GNT3	5.77	3.09E-08	0.0024	0.39	0.39	0.15
cg12601575	SEC14L5	5.77	3.18E-08	0.0025	0.11	0.39	0.15
cg01775802	RGS6	5.77	3.20E-08	0.0025	0.21	0.39	0.15
cg27383876	PDGFD	5.76	3.37E-08	0.0026	0.19	0.38	0.15
cg14473416	AQP3	-5.76	3.40E-08	0.0026	-0.85	-0.38	0.15
cg08670210	RNF141	5.75	3.54E-08	0.0027	0.49	0.38	0.15
cg14240646	MASTL	-5.75	3.55E-08	0.0027	-0.11	-0.38	0.15
cg16705185	CUL7	5.74	3.64E-08	0.0027	0.03	0.38	0.15
cg27182293	ZNF611	-5.73	3.85E-08	0.0029	-0.03	-0.38	0.15
cg04041654	ARPC5	5.73	3.95E-08	0.0029	0.22	0.38	0.15
cg20972969	IL36RN	5.72	4.03E-08	0.0030	0.98	0.38	0.15

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg13500480	ADCY7	5.71	4.22E-08	0.0031	0.28	0.38	0.15
cg03653088	GAS7	5.70	4.40E-08	0.0032	0.30	0.38	0.15
cg25724460	PDK2	5.69	4.64E-08	0.0034	0.50	0.38	0.15
cg15052791	ADORA1	5.69	4.80E-08	0.0035	0.17	0.38	0.14
cg25940196	NPAS3	5.68	4.99E-08	0.0036	0.46	0.38	0.14
cg10324502	PHYH	5.68	5.06E-08	0.0037	0.25	0.38	0.14
cg19548479	TMEM106A	-5.65	5.64E-08	0.0041	-0.03	-0.38	0.14
cg08525331	GSTM3	-5.64	5.92E-08	0.0043	-0.57	-0.38	0.14
cg07612367	MRPL2	-5.63	6.24E-08	0.0045	-0.02	-0.38	0.14
cg19062298	AP2A2	5.63	6.49E-08	0.0046	0.39	0.38	0.14
cg19205266	SMUG1	5.62	6.53E-08	0.0046	0.10	0.38	0.14
cg22636562	CLDN8	5.62	6.56E-08	0.0046	0.17	0.38	0.14
cg10997251	SAR1B	-5.62	6.61E-08	0.0047	-0.04	-0.38	0.14
cg10507231	PIK3C2B	5.62	6.62E-08	0.0047	0.43	0.38	0.14
cg26143719	C1QTNF6	-5.62	6.69E-08	0.0047	-0.23	-0.38	0.14
cg12454169	LCLAT1	5.61	7.12E-08	0.0050	0.06	0.38	0.14
cg03187338	IP6K3	5.61	7.19E-08	0.0050	0.23	0.38	0.14
cg01970114	MRPL2	-5.60	7.33E-08	0.0051	-0.02	-0.38	0.14
cg08925307	LRRC8E	-5.60	7.50E-08	0.0052	-0.04	-0.38	0.14
cg14460291	CDH8	5.60	7.55E-08	0.0052	0.11	0.38	0.14
cg06659352	LRCH1	5.60	7.56E-08	0.0052	0.31	0.38	0.14
cg27569984	SNX25	5.59	7.66E-08	0.0052	0.19	0.38	0.14
cg10645412	THSD7A	5.59	7.89E-08	0.0054	0.37	0.37	0.14
cg01458219	ERVV-1	5.57	8.66E-08	0.0059	0.15	0.37	0.14
cg04482110	TMEM106A	-5.56	8.86E-08	0.0060	-0.02	-0.37	0.14
cg19811108	TRIM45	5.56	8.97E-08	0.0060	0.55	0.37	0.14
cg13427828	CCK	5.56	9.06E-08	0.0060	0.57	0.37	0.14
cg02265785	PDGFD	5.56	9.15E-08	0.0061	0.06	0.37	0.14
cg08371186	CDCA7L	5.56	9.17E-08	0.0061	0.08	0.37	0.14
cg06851354	SMUG1	5.56	9.21E-08	0.0061	0.10	0.37	0.14
cg12658012	ARPC5	5.55	9.53E-08	0.0063	0.70	0.37	0.14
cg11588907	ASXL3	5.55	9.65E-08	0.0063	0.20	0.37	0.14
cg18371928	FAM184A	5.54	1.00E-07	0.0065	0.78	0.37	0.14
cg25054324	A4GALT	5.53	1.02E-07	0.0066	0.07	0.37	0.14
cg15907473	GPR180	5.53	1.02E-07	0.0066	0.53	0.37	0.14
cg17548735	UNC5CL	5.53	1.05E-07	0.0068	0.70	0.37	0.14
cg17092624	ZNF197	5.53	1.05E-07	0.0068	0.05	0.37	0.14
cg00148729	EVC2	5.53	1.06E-07	0.0068	0.06	0.37	0.14

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg04773602	SLC25A24	-5.52	1.09E-07	0.0070	-0.03	-0.37	0.14
cg17548735	APOBEC2	5.52	1.10E-07	0.0070	0.60	0.37	0.14
cg04645556	PLCB1	5.52	1.11E-07	0.0070	0.20	0.37	0.14
cg08246428	ATP1A4	5.52	1.12E-07	0.0071	0.25	0.37	0.14
cg22459204	ANKRD33B	5.52	1.12E-07	0.0071	0.69	0.37	0.14
cg06575723	RRP7A	-5.51	1.12E-07	0.0071	-0.19	-0.37	0.14
cg07636681	INF2	-5.51	1.13E-07	0.0071	-0.69	-0.37	0.14
cg19563574	METRNL	5.51	1.14E-07	0.0071	0.43	0.37	0.14
cg15729878	STON2	5.51	1.14E-07	0.0071	0.76	0.37	0.14
cg10279470	NOS1AP	5.51	1.16E-07	0.0072	0.40	0.37	0.14
cg16814023	ACTR8	-5.51	1.16E-07	0.0072	-0.37	-0.37	0.14
cg09395732	SLC6A11	5.50	1.21E-07	0.0075	0.94	0.37	0.14
cg25966812	ZFYVE19	5.50	1.22E-07	0.0075	0.61	0.37	0.14
cg05703009	OR52E4	5.49	1.24E-07	0.0076	0.36	0.37	0.14
cg19090871	ECHDC3	-5.49	1.26E-07	0.0077	-0.55	-0.37	0.14
cg15769472	FKBP2	-5.49	1.27E-07	0.0078	-0.05	-0.37	0.14
cg05904364	RHOB	-5.48	1.32E-07	0.0081	-0.30	-0.37	0.14
cg09150320	STARD5	5.48	1.33E-07	0.0081	0.75	0.37	0.14
cg21088514	KCNE1	5.47	1.37E-07	0.0083	1.15	0.37	0.14
cg26293310	SAR1B	-5.47	1.37E-07	0.0083	-0.03	-0.37	0.14
cg26826957	ICA1	5.47	1.43E-07	0.0086	0.05	0.37	0.14
cg23890469	SNCG	5.46	1.46E-07	0.0087	0.63	0.37	0.14
cg07734889	CRISPLD1	5.46	1.49E-07	0.0089	0.28	0.37	0.13
cg07925587	KRT80	5.46	1.49E-07	0.0089	0.52	0.37	0.13
cg02475902	MUC20	-5.46	1.50E-07	0.0089	-0.84	-0.37	0.13
cg23645969	DISP2	5.45	1.54E-07	0.0091	0.17	0.37	0.13
cg22497336	KRT80	5.45	1.58E-07	0.0093	0.21	0.37	0.13
cg20073882	NAA11	-5.44	1.60E-07	0.0094	-0.06	-0.37	0.13
cg07709181	CRISPLD1	5.43	1.73E-07	0.010	0.26	0.37	0.13
cg01146238	RTN4	-5.42	1.74E-07	0.010	-0.12	-0.37	0.13
cg24752487	CYP1B1	5.42	1.76E-07	0.010	1.14	0.37	0.13
cg13588073	PRKAG2	5.42	1.76E-07	0.010	0.27	0.37	0.13
cg07436701	SNCG	5.42	1.79E-07	0.010	0.53	0.37	0.13
cg07165018	STBD1	5.42	1.80E-07	0.010	0.75	0.37	0.13
cg17417378	ZNF205	5.40	1.92E-07	0.011	0.20	0.36	0.13
cg14449309	VPS16	5.40	1.95E-07	0.011	0.09	0.36	0.13
cg13996963	TUSC3	-5.40	1.96E-07	0.011	-0.04	-0.36	0.13
cg04896851	SGSM1	5.40	1.98E-07	0.011	0.28	0.36	0.13

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg02992311	FBXO27	-5.40	1.98E-07	0.011	-0.07	-0.36	0.13
cg21116203	UNC13B	5.40	1.99E-07	0.011	0.18	0.36	0.13
cg26876215	CDON	5.39	2.02E-07	0.011	0.07	0.36	0.13
cg22636562	KRTAP26-1	5.39	2.05E-07	0.011	0.19	0.36	0.13
cg12836958	SNX25	5.39	2.07E-07	0.012	0.13	0.36	0.13
cg00738291	HNRNPA1L2	5.39	2.08E-07	0.012	0.08	0.36	0.13
cg14815361	P4HA2	5.39	2.10E-07	0.012	0.48	0.36	0.13
cg12658012	NCF2	5.38	2.12E-07	0.012	0.93	0.36	0.13
cg12884378	EYA2	5.38	2.21E-07	0.012	0.76	0.36	0.13
cg09355348	PLA2G2F	5.38	2.21E-07	0.012	0.40	0.36	0.13
cg23147597	CEACAM19	5.37	2.26E-07	0.012	0.41	0.36	0.13
cg01682590	WNT7A	5.37	2.32E-07	0.013	0.87	0.36	0.13
cg01288797	GSTT1	-5.36	2.35E-07	0.013	-1.96	-0.36	0.13
cg26792694	FAM118A	5.36	2.35E-07	0.013	0.47	0.36	0.13
cg00773965	VAMP3	-5.36	2.36E-07	0.013	-0.03	-0.36	0.13
cg13845211	ERAP2	-5.36	2.41E-07	0.013	-1.30	-0.36	0.13
cg03147584	TRIM58	5.35	2.45E-07	0.013	0.46	0.36	0.13
cg19539664	ZNF502	5.35	2.47E-07	0.013	0.04	0.36	0.13
cg18763720	DYRK1B	5.35	2.52E-07	0.013	0.05	0.36	0.13
cg09331735	DIP2C	5.35	2.55E-07	0.014	0.45	0.36	0.13
cg03182620	DIP2C	5.35	2.56E-07	0.014	0.42	0.36	0.13
cg08453609	POSTN	5.33	2.76E-07	0.015	0.17	0.36	0.13
cg02019444	ITSN1	-5.33	2.79E-07	0.015	-0.20	-0.36	0.13
cg26217936	COX6C	-5.32	2.88E-07	0.015	-0.22	-0.36	0.13
cg11692488	PDGFD	5.32	2.93E-07	0.015	0.06	0.36	0.13
cg22413023	DSCAM	5.31	2.97E-07	0.016	0.25	0.36	0.13
cg06161915	ARF5	-5.31	2.98E-07	0.016	-0.20	-0.36	0.13
cg11377286	SDCBP	5.31	2.99E-07	0.016	0.29	0.36	0.13
cg08163621	THAP3	-5.31	3.02E-07	0.016	-0.05	-0.36	0.13
cg06543640	FZD10	5.31	3.03E-07	0.016	0.73	0.36	0.13
cg01111341	RAB6B	5.31	3.08E-07	0.016	0.62	0.36	0.13
cg18649503	TOB2P1	5.30	3.15E-07	0.016	0.31	0.36	0.13
cg13644197	METRNL	5.30	3.16E-07	0.016	0.38	0.36	0.13
cg22741977	FBXO32	-5.30	3.18E-07	0.016	-0.14	-0.36	0.13
cg07971797	SGSM1	5.29	3.29E-07	0.017	0.28	0.36	0.13
cg23432370	CLASP1	5.29	3.29E-07	0.017	0.16	0.36	0.13
cg21158502	GCNT4	-5.29	3.32E-07	0.017	-0.20	-0.36	0.13
cg17157956	RRP7A	-5.28	3.46E-07	0.017	-0.21	-0.36	0.13

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg10395101	ANGPT2	-5.27	3.71E-07	0.019	-1.24	-0.36	0.13
cg18362112	ISL1	5.26	3.78E-07	0.019	0.28	0.36	0.13
cg07236562	MTSS1	5.26	3.82E-07	0.019	0.24	0.36	0.13
cg12806347	METRNL	5.26	3.87E-07	0.019	0.79	0.36	0.13
cg22990871	C2CD2	5.26	3.92E-07	0.019	0.34	0.36	0.13
cg08741843	ZNF595	5.25	3.94E-07	0.019	0.05	0.36	0.13
cg05703009	TRIM5	5.25	3.95E-07	0.019	0.07	0.36	0.13
cg06773033	TOE1	5.25	4.04E-07	0.020	0.27	0.36	0.13
cg12061069	RAP1GAP2	5.25	4.06E-07	0.020	0.32	0.36	0.13
cg21999269	LAMB4	5.25	4.07E-07	0.020	0.54	0.35	0.13
cg13927938	PIK3C2B	5.25	4.12E-07	0.020	0.35	0.35	0.13
cg18806980	ATP2C2	5.24	4.20E-07	0.020	0.11	0.35	0.13
cg01183703	ADCY2	5.24	4.20E-07	0.020	0.04	0.35	0.13
cg14751503	OVCH2	5.24	4.21E-07	0.020	1.12	0.35	0.13
cg01784614	CARD6	5.24	4.25E-07	0.020	0.32	0.35	0.13
cg12793733	ATP10D	-5.24	4.27E-07	0.020	-0.14	-0.35	0.13
cg08220278	QSOX1	-5.24	4.28E-07	0.020	-0.20	-0.35	0.13
cg08827804	KCNJ2	5.23	4.52E-07	0.021	0.23	0.35	0.13
cg07435254	SNX25	5.22	4.56E-07	0.022	0.22	0.35	0.13
cg09655116	CHMP4A	-5.22	4.56E-07	0.022	-0.02	-0.35	0.13
cg07816074	SH3TC1	5.22	4.58E-07	0.022	0.23	0.35	0.12
cg05957477	TMX4	5.22	4.62E-07	0.022	0.23	0.35	0.12
cg07277633	ATP2C2	5.22	4.64E-07	0.022	0.28	0.35	0.12
cg22072332	MUC15	5.22	4.65E-07	0.022	0.67	0.35	0.12
cg05940672	HPCAL1	5.21	4.78E-07	0.022	0.74	0.35	0.12
cg12200611	AFF3	5.21	4.80E-07	0.022	0.52	0.35	0.12
cg16537292	ERICH1	5.21	4.82E-07	0.022	0.31	0.35	0.12
cg15274858	MYEF2	5.21	4.86E-07	0.022	0.25	0.35	0.12
cg00766382	N4BP1	5.21	4.96E-07	0.023	0.17	0.35	0.12
cg26880549	LYNX1	5.21	4.97E-07	0.023	0.35	0.35	0.12
cg08150575	CBLC	5.20	5.08E-07	0.023	0.62	0.35	0.12
cg13892059	DDX19B	-5.19	5.35E-07	0.024	-0.03	-0.35	0.12
cg07735013	IL36RN	5.19	5.42E-07	0.024	0.58	0.35	0.12
cg20004389	ANO2	5.18	5.55E-07	0.025	0.57	0.35	0.12
cg26293019	ISL1	5.18	5.60E-07	0.025	0.88	0.35	0.12
cg11708454	CATSPERG	5.18	5.71E-07	0.025	0.06	0.35	0.12
cg05747095	ETV7	-5.17	5.76E-07	0.026	-0.05	-0.35	0.12
cg26446133	CNDP2	-5.17	5.91E-07	0.026	-0.15	-0.35	0.12

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg10057218	ORMDL3	5.16	6.03E-07	0.027	0.53	0.35	0.12
cg08481002	NARS2	5.16	6.10E-07	0.027	0.17	0.35	0.12
cg09489281	NNT	-5.16	6.18E-07	0.027	-0.03	-0.35	0.12
cg07474359	BTBD16	5.15	6.37E-07	0.028	0.90	0.35	0.12
cg21256968	SCARB2	5.15	6.48E-07	0.028	0.53	0.35	0.12
cg18318704	SAR1B	-5.15	6.55E-07	0.028	-0.04	-0.35	0.12
cg14590806	WDR91	-5.15	6.56E-07	0.028	-0.30	-0.35	0.12
cg04422802	ZNF208	5.15	6.56E-07	0.028	0.08	0.35	0.12
cg17589576	SLC6A15	5.14	6.66E-07	0.028	0.26	0.35	0.12
cg21603313	CELF2	5.14	6.80E-07	0.029	0.06	0.35	0.12
cg16125375	PPP2R2B	5.11	7.68E-07	0.032	0.51	0.35	0.12
cg25940485	TRIM45	5.11	7.68E-07	0.032	0.42	0.35	0.12
cg11268590	GLRB	5.11	7.74E-07	0.032	0.62	0.35	0.12
cg21032929	FAM171B	5.11	7.93E-07	0.033	0.69	0.35	0.12
cg02498072	SNX29	5.10	7.97E-07	0.033	0.44	0.35	0.12
cg11316709	ZNF28	5.10	8.05E-07	0.033	0.15	0.35	0.12
cg22990475	FAM98A	5.10	8.10E-07	0.033	0.18	0.35	0.12
cg16280624	RPSA	5.10	8.25E-07	0.034	0.39	0.35	0.12
cg08160623	OR52E4	5.10	8.29E-07	0.034	0.63	0.35	0.12
cg15791719	LRRC8D	5.09	8.35E-07	0.034	0.27	0.35	0.12
cg07321742	TMX4	-5.09	8.37E-07	0.034	-0.03	-0.35	0.12
cg05871892	INPP1	-5.09	8.42E-07	0.034	-0.40	-0.35	0.12
cg14960550	NAALAD2	5.09	8.47E-07	0.034	0.59	0.35	0.12
cg10981598	NNAT	-5.09	8.49E-07	0.034	-0.05	-0.35	0.12
cg18642179	MGMT	5.09	8.53E-07	0.034	0.43	0.35	0.12
cg10624236	TAF5L	5.09	8.56E-07	0.034	0.25	0.35	0.12
cg05672540	PLCB1	5.09	8.58E-07	0.034	0.20	0.35	0.12
cg19071879	AFF3	5.09	8.66E-07	0.034	0.16	0.35	0.12
cg25785733	PPP2R2B	5.08	8.89E-07	0.035	0.80	0.35	0.12
cg12689806	TRIM58	5.08	8.90E-07	0.035	0.07	0.35	0.12
cg27060295	ICA1	5.08	9.00E-07	0.035	0.45	0.34	0.12
cg25063515	CA10	5.07	9.31E-07	0.036	0.08	0.34	0.12
cg14909555	TRMT12	5.07	9.43E-07	0.037	0.22	0.34	0.12
cg04304338	ANKRD33B	5.07	9.53E-07	0.037	0.64	0.34	0.12
cg13136938	HMOX1	5.07	9.57E-07	0.037	0.21	0.34	0.12
cg01287209	FAM171B	5.06	9.68E-07	0.037	0.71	0.34	0.12
cg08392199	LIFR	5.05	1.01E-06	0.038	0.19	0.34	0.12
cg04220914	HEXB	-5.05	1.04E-06	0.039	-0.54	-0.34	0.12

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg01964975	NOS1AP	5.05	1.05E-06	0.039	0.33	0.34	0.12
cg18140087	SPINK5	5.04	1.09E-06	0.040	0.55	0.34	0.12
cg13570585	PLCB1	5.04	1.10E-06	0.040	0.07	0.34	0.12
cg03730490	PEPD	5.03	1.12E-06	0.041	0.04	0.34	0.12
cg04907746	ARAP2	5.03	1.13E-06	0.042	0.20	0.34	0.12
cg02007844	KCNJ8	5.03	1.14E-06	0.042	0.06	0.34	0.12
cg14842970	ASXL3	5.03	1.15E-06	0.042	0.14	0.34	0.12
cg08875710	DMXL1	5.02	1.17E-06	0.043	0.11	0.34	0.12
cg26923577	STARD5	5.02	1.20E-06	0.044	0.40	0.34	0.12
cg18349420	AQP11	5.02	1.20E-06	0.044	0.06	0.34	0.12
cg15689283	PDGFD	5.01	1.23E-06	0.044	0.14	0.34	0.12
cg01159543	CBR1	-5.01	1.25E-06	0.045	-0.05	-0.34	0.12
cg05359525	SLC44A5	5.01	1.26E-06	0.045	0.39	0.34	0.12
cg23858094	ASXL3	5.01	1.26E-06	0.045	0.09	0.34	0.12
cg07528940	PDGFD	5.00	1.28E-06	0.046	0.23	0.34	0.12
cg25339619	KCNJ2	4.99	1.33E-06	0.047	0.04	0.34	0.12
cg03125975	PDGFD	4.99	1.33E-06	0.047	0.08	0.34	0.12
cg01918824	TRIM45	4.99	1.36E-06	0.048	0.53	0.34	0.12
cg20459238	SNCG	4.99	1.37E-06	0.048	0.53	0.34	0.12
cg13919821	ZNF502	4.99	1.37E-06	0.048	0.35	0.34	0.12
cg06765956	LRFN5	4.98	1.42E-06	0.049	0.08	0.34	0.11
cg18452347	RPL23AP53	4.97	1.47E-06	0.050	0.27	0.34	0.11
cg15534578	TRIB1	4.97	1.47E-06	0.050	0.02	0.34	0.11
cg25256669	RAB27A	-4.97	1.51E-06	0.051	-0.04	-0.34	0.11
cg00529567	HEG1	4.95	1.64E-06	0.054	0.45	0.34	0.11
cg23678338	ATP10D	4.94	1.68E-06	0.054	0.28	0.34	0.11
cg05498539	SNCG	4.94	1.71E-06	0.055	0.66	0.34	0.11
cg23098371	CDC7	-4.94	1.72E-06	0.056	-0.17	-0.34	0.11
cg22492435	ABCC2	4.92	1.86E-06	0.059	0.46	0.34	0.11
cg09818397	LRRC8B	4.91	1.96E-06	0.062	0.18	0.33	0.11
cg21856508	CRISPLD1	4.90	2.06E-06	0.064	0.06	0.33	0.11
cg00437019	STON2	4.90	2.08E-06	0.065	0.24	0.33	0.11
cg13944161	CPVL	-4.89	2.10E-06	0.065	-0.06	-0.33	0.11
cg10202113	DPYD	4.89	2.15E-06	0.066	0.41	0.33	0.11
cg09361995	TTPA	4.88	2.24E-06	0.068	0.66	0.33	0.11
cg20379470	STON1	4.88	2.25E-06	0.068	0.18	0.33	0.11
cg15248091	CBLB	4.84	2.65E-06	0.077	0.57	0.33	0.11
cg14145653	INSL6	4.83	2.76E-06	0.079	0.55	0.33	0.11

CpG	Gene	Statistic	p-value	FDR	Beta	r	r2
cg26616283	THSD7A	-4.83	2.80E-06	0.080	-0.06	-0.33	0.11
cg11281200	FGF12	4.83	2.85E-06	0.081	0.10	0.33	0.11
cg15400117	THSD7A	-4.81	3.01E-06	0.084	-0.05	-0.33	0.11
cg16511333	DPYD	4.78	3.54E-06	0.092	0.10	0.33	0.11
cg05946623	CDH8	4.76	3.75E-06	0.096	0.10	0.33	0.11
cg09662906	PDGFD	4.73	4.44E-06	0.11	0.10	0.32	0.10
cg13721814	PEX13	4.68	5.43E-06	0.12	0.02	0.32	0.10
cg05448511	TRIM24	-4.59	7.92E-06	0.15	-0.18	-0.32	0.10
cg03930781	TRIM24	-4.55	9.59E-06	0.17	-0.22	-0.31	0.10

Table S3-2: Significant DHMRs

Gene	Chrom	Start	End	p-value	FDR	Sidak	# probes
B3GNT3	chr19	17905332	17906310	2.74E-32	5.48E-32	0	8
CDCP1	chr3	45181094	45182104	5.01E-27	2.51E-26	0	4
CUL7	chr6	43021181	43022045	3.15E-18	3.15E-18	0	10
DNAAF1	chr16	84209087	84209610	1.03E-19	3.08E-19	0	3
KCNE1	chr21	35883571	35884377	7.09E-19	7.09E-19	0	5
P2RX7	chr12	121570559	121571069	1.07E-18	1.07E-18	0	7
SGSM1	chr22	25170368	25171163	1.21E-17	8.48E-17	0	3
SMUG1	chr12	54577268	54577736	1.96E-27	5.88E-27	0	4
TOB2P1	chr6	28234525	28234681	4.34E-20	1.74E-19	0	4
PSMD5	chr9	123606119	123606676	6.80E-17	6.80E-17	1.11E-16	4
TMEM106A	chr17	41363502	41364122	7.73E-17	7.73E-17	3.33E-16	12
IL36RN	chr2	113825068	113825528	3.40E-16	6.81E-16	4.44E-16	3
NOS1AP	chr1	162346511	162346614	5.07E-16	1.01E-15	4.77E-15	3
ANKRD33B	chr5	10565679	10566783	5.74E-14	5.74E-14	2.40E-14	3
JAK3	chr19	17905332	17905967	1.76E-14	1.76E-14	4.65E-14	7
DDX19B	chr16	70332878	70333125	2.40E-14	2.40E-14	8.84E-14	7
RNF39	chr6	30038720	30039477	9.50E-15	9.50E-15	1.43E-13	21
ERCC6	chr10	50746568	50746730	3.72E-14	3.72E-14	1.46E-13	3
MRPL2	chr6	43021181	43021777	7.52E-14	7.52E-14	1.50E-13	8
IL36RN	chr2	113819714	113820091	1.04E-13	1.04E-13	1.86E-13	5

Gene	Chrom	Start	End	p-value	FDR	Sidak	# probes
TMX4	chr20	8228450	8229346	1.71E-12	7.71E-12	3.31E-13	4
SAR1B	chr5	133967994	133968722	2.64E-13	2.64E-13	3.40E-13	10
VAMP3	chr1	7831168	7832358	1.02E-12	1.02E-12	7.68E-13	7
CHMP4A	chr14	24682303	24682695	3.51E-13	3.51E-13	1.05E-12	5
TOB2P1	chr6	28192457	28192602	2.08E-13	4.16E-13	2.14E-12	3
IER3	chr6	30711138	30711681	9.87E-14	9.87E-14	2.23E-12	12
STON2	chr14	81864623	81865008	5.46E-12	1.27E-11	3.43E-12	4
SNCG	chr10	88717364	88717927	4.15E-12	1.25E-11	4.51E-12	5
ARMS2	chr10	124220091	124220505	9.78E-12	2.93E-11	1.77E-11	4
ABL1	chr9	133709810	133710065	4.76E-12	4.76E-12	1.83E-11	3
CACNG4	chr17	64959947	64960283	4.39E-11	4.39E-11	7.48E-11	4
DYRK1B	chr19	40324226	40325156	6.66E-11	6.66E-11	1.00E-10	12
HTRA1	chr10	124220091	124220505	1.11E-10	3.32E-10	2.04E-10	4
ADCY7	chr16	50300289	50300493	7.25E-11	7.25E-11	3.15E-10	6
S100A13	chr1	153588548	153590244	8.97E-10	8.97E-10	6.15E-10	9
THSD7A	chr7	11871535	11871844	1.02E-09	1.70E-09	9.57E-10	4
ATP1A4	chr1	160159453	160160767	1.57E-09	1.57E-09	1.29E-09	13
PANK1	chr10	91352963	91353163	5.52E-10	5.52E-10	1.33E-09	3
C2	chr6	32055370	32055630	3.41E-11	3.41E-11	1.87E-09	10
AFF3	chr2	100722022	100722208	5.49E-10	5.49E-10	2.20E-09	3
ZNF559	chr19	9488238	9488548	6.89E-10	6.89E-10	2.20E-09	3
TUSC3	chr8	15397637	15398334	1.59E-08	1.59E-08	2.86E-09	7

Gene	Chrom	Start	End	p-value	FDR	Sidak	# probes
DSEL	chr18	65296234	65296540	9.50E-11	1.40E-10	3.03E-09	3
ETV7	chr6	36355542	36356245	2.48E-09	2.48E-09	3.19E-09	11
PIK3C2B	chr1	204459579	204459648	2.10E-10	2.10E-10	3.24E-09	5
ASXL3	chr18	31020973	31021066	1.82E-09	3.65E-09	4.06E-09	3
CATSPERG	chr19	38794514	38794665	5.37E-10	5.37E-10	4.29E-09	5
WDR91	chr7	134854206	134854552	3.10E-09	3.10E-09	5.08E-09	6
ZBTB45	chr19	59024830	59024900	4.92E-10	4.92E-10	5.76E-09	3
SNCG	chr10	88719848	88720026	1.72E-09	2.58E-09	5.92E-09	3
MMP24	chr20	33865797	33866065	1.67E-09	1.67E-09	6.39E-09	3
LYNX1	chr8	143858414	143858637	8.85E-10	1.77E-09	6.63E-09	6
NOP10	chr15	34631062	34631365	4.91E-09	4.91E-09	6.73E-09	3
NOS1AP	chr1	162350943	162351058	9.18E-10	9.18E-10	7.11E-09	4
AQP11	chr11	77299805	77300118	2.96E-09	2.96E-09	7.21E-09	4
TMX4	chr20	8638883	8639259	1.73E-08	2.60E-08	7.97E-09	5
OR52E4	chr11	5841482	5841689	2.13E-09	2.13E-09	8.13E-09	3
CASQ1	chr1	160159453	160160767	1.12E-08	1.12E-08	9.28E-09	13
CBR1	chr21	37442104	37442386	5.73E-09	5.73E-09	1.22E-08	6
CCDC62	chr12	123258736	123259331	7.46E-09	7.46E-09	1.55E-08	10
SLC25A24	chr1	108742232	108742425	1.09E-08	1.09E-08	2.52E-08	4
HPCAL1	chr2	10470114	10470466	9.94E-09	1.99E-08	2.93E-08	5
ERAP2	chr5	96210230	96210395	1.17E-08	1.17E-08	3.05E-08	3
XDH	chr2	31637552	31637850	2.28E-08	2.28E-08	3.28E-08	3

Gene	Chrom	Start	End	p-value	FDR	Sidak	# probes
RABAC1	chr19	23578679	23578687	1.53E-11	1.53E-11	4.12E-08	3
ZFP57	chr6	29631227	29631322	3.96E-10	3.96E-10	4.44E-08	3
PDK2	chr17	48179239	48179460	1.40E-08	1.40E-08	9.36E-08	4
EDEM2	chr20	34242979	34243250	3.03E-08	3.03E-08	1.09E-07	3
HOXB2	chr17	46622012	46622155	1.45E-08	1.45E-08	1.48E-07	4
ZNF208	chr19	22815701	22815854	9.89E-08	1.48E-07	1.54E-07	3
SMUG1	chr12	54558274	54558300	3.12E-09	4.68E-09	1.68E-07	3
SH3BP5	chr3	15311021	15311270	5.54E-08	5.54E-08	1.79E-07	4
CALCOCO1	chr12	53693769	53694019	2.81E-08	2.81E-08	1.82E-07	6
C2CD2	chr21	43374476	43374651	3.13E-08	3.13E-08	1.96E-07	3
TRIM58	chr1	248083146	248083215	2.98E-08	1.19E-07	2.13E-07	3
SLC27A1	chr19	17905589	17905836	4.34E-08	4.34E-08	2.75E-07	5
ERGIC3	chr20	34242979	34243250	8.39E-08	8.39E-08	3.15E-07	3
STARD5	chr15	81588069	81588309	1.47E-07	1.47E-07	3.38E-07	4
SLCO4A1	chr20	61288142	61288352	4.36E-08	4.36E-08	3.59E-07	3
RHOB	chr2	20650578	20650793	1.44E-07	1.44E-07	3.70E-07	3
TRAPPC4	chr11	43918794	43918833	4.07E-10	4.07E-10	3.74E-07	6
MED1	chr17	37608096	37608295	5.54E-08	5.54E-08	4.05E-07	3
ZNF229	chr19	44429318	44429628	1.12E-07	1.12E-07	4.80E-07	3
ZNF100	chr19	863054	863245	6.88E-09	6.88E-09	5.36E-07	4
B4GALT4	chr3	118948906	118949199	3.96E-07	3.96E-07	5.53E-07	4
PDGFD	chr11	103815860	103816170	6.61E-07	1.19E-06	5.69E-07	4

Gene	Chrom	Start	End	p-value	FDR	Sidak	# probes
SLC6A15	chr12	85307131	85307599	2.05E-06	2.05E-06	6.01E-07	6
PRH1-PRR4	chr12	11322362	11322617	3.11E-08	3.11E-08	6.14E-07	3
TRIM58	chr1	248005038	248005247	3.07E-07	4.09E-07	7.24E-07	3
PLCB1	chr20	8229198	8229346	4.10E-07	6.14E-07	7.66E-07	3
CDC42BPA	chr1	226856447	226856528	5.84E-08	6.82E-08	7.94E-07	3
PURA	chr5	139493546	139493681	8.76E-08	8.76E-08	8.10E-07	6
KATNB1	chr16	202482	202566	4.02E-09	4.02E-09	8.79E-07	3
KRT81	chr12	52305261	52305476	2.24E-07	2.24E-07	1.33E-06	5
TTLL4	chr2	219576383	219576540	1.52E-07	1.52E-07	1.53E-06	3
GPNMB	chr7	23285742	23285804	1.98E-07	1.98E-07	1.62E-06	3
FCHO1	chr19	17905589	17905836	3.07E-07	3.07E-07	2.05E-06	5
LY6K	chr8	143782868	143782974	8.28E-09	8.28E-09	2.31E-06	3
PHLDB1	chr11	2019624	2019737	7.52E-09	7.52E-09	2.38E-06	7
P4HA2	chr5	132200490	132200667	6.65E-07	6.65E-07	3.22E-06	5
CCDC124	chr19	17905589	17905836	6.45E-07	6.45E-07	4.48E-06	5
ATF6B	chr6	32055447	32055630	3.47E-07	3.47E-07	2.75E-05	6
CCL28	chr5	43040998	43041153	9.42E-06	9.42E-06	2.82E-05	3
STK19	chr6	30158059	30158091	1.76E-07	1.76E-07	7.95E-05	3
SPN	chr16	30124803	30124905	6.26E-05	6.26E-05	0.00082	3
LAT2	chr7	27209197	27209282	1.41E-05	1.41E-05	0.0028	4
BIRC5	chr17	75446431	75446593	0.0031	0.0031	0.027	5

References

1. Burton, G.J., A.L. Fowden, and K.L. Thornburg, *Placental Origins of Chronic Disease.* Physiol Rev, 2016. **96**(4): p. 1509-65.
2. Gude, N.M., et al., *Growth and function of the normal human placenta.* Thromb Res, 2004. **114**(5-6): p. 397-407.
3. Schoots, M.H., et al., *Oxidative stress in placental pathology.* Placenta, 2018. **69**: p. 153-161.
4. Burton, G.J. and E. Jauniaux, *What is the placenta?* Am J Obstet Gynecol, 2015. **213**(4 Suppl): p. S6 e1, S6-8.
5. Nakashima, A., et al., *Disruption of Placental Homeostasis Leads to Preeclampsia.* Int J Mol Sci, 2020. **21**(9).
6. Sadovsky, Y., V.L. Clifton, and G.J. Burton, *Invigorating placental research through the "Human Placenta Project".* Placenta, 2014. **35**(8): p. 527.
7. Baczkowska, M., et al., *Molecular Changes on Maternal-Fetal Interface in Placental Abruption-A Systematic Review.* Int J Mol Sci, 2021. **22**(12).
8. Breiling, A. and F. Lyko, *Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond.* Epigenetics Chromatin, 2015. **8**: p. 24.
9. Smith, Z.D. and A. Meissner, *DNA methylation: roles in mammalian development.* Nat Rev Genet, 2013. **14**(3): p. 204-20.
10. Baubec, T. and D. Schubeler, *Genomic patterns and context specific interpretation of DNA methylation.* Curr Opin Genet Dev, 2014. **25**: p. 85-92.
11. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* Science, 2009. **324**(5929): p. 930-5.
12. Shi, D.Q., et al., *New Insights into 5hmC DNA Modification: Generation, Distribution and Function.* Front Genet, 2017. **8**: p. 100.
13. Bachman, M., et al., *5-Hydroxymethylcytosine is a predominantly stable DNA modification.* Nat Chem, 2014. **6**(12): p. 1049-55.
14. Nestor, C.E., et al., *Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes.* Genome Res, 2012. **22**(3): p. 467-77.
15. Pfeifer, G.P. and P.E. Szabo, *Gene body profiles of 5-hydroxymethylcytosine: potential origin, function and use as a cancer biomarker.* Epigenomics, 2018. **10**(8): p. 1029-1032.
16. Li, W. and M. Liu, *Distribution of 5-hydroxymethylcytosine in different human tissues.* J Nucleic Acids, 2011. **2011**: p. 870726.
17. Hernandez Mora, J.R., et al., *Profiling of oxBS-450K 5-hydroxymethylcytosine in human placenta and brain reveals enrichment at imprinted loci.* Epigenetics, 2018. **13**(2): p. 182-191.
18. Piyasena, C., et al., *Placental 5-methylcytosine and 5-hydroxymethylcytosine patterns associate with size at birth.* Epigenetics, 2015. **10**(8): p. 692-7.
19. Green, B.B., et al., *Hydroxymethylation is uniquely distributed within term placenta, and is associated with gene expression.* FASEB J, 2016. **30**(8): p. 2874-84.
20. Fenton, T.R. and J.H. Kim, *A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants.* BMC Pediatr, 2013. **13**: p. 59.
21. Pidsley, R., et al., *Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling.* Genome Biol, 2016. **17**(1): p. 208.

22. Houseman, E.A., K.C. Johnson, and B.C. Christensen, *OxyBS: estimation of 5-methylcytosine and 5-hydroxymethylcytosine from tandem-treated oxidative bisulfite and bisulfite DNA.* Bioinformatics, 2016. **32**(16): p. 2505-7.
23. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
24. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.* BMC Bioinformatics, 2010. **11**: p. 587.
25. Deyssenroth, M.A., et al., *Whole-transcriptome analysis delineates the human placenta gene network and its associations with fetal growth.* BMC Genomics, 2017. **18**(1): p. 520.
26. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry.* Nature, 2008. **456**(7218): p. 53-9.
27. Hansen, K., *IlluminaHumanMethylationEPICanno.ilm10b2.hg19: Annotation for Illumina's EPIC methylation arrays.*, in *R package version 0.6.0*. 2016.
28. Ernst, J. and M. Kellis, *Chromatin-state discovery and genome annotation with ChromHMM.* Nat Protoc, 2017. **12**(12): p. 2478-2492.
29. Roadmap Epigenomics, C., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.
30. Karolchik, D., et al., *The UCSC Table Browser data retrieval tool.* Nucleic Acids Res, 2004. **32**(Database issue): p. D493-6.
31. Ruiz-Arenas, C., et al., *Identification of autosomal cis expression quantitative trait methylation (cis eQTMs) in children's blood.* Elife, 2022. **11**.
32. Shabalin, A.A., *Matrix eQTL: ultra fast eQTL analysis via large matrix operations.* Bioinformatics, 2012. **28**(10): p. 1353-8.
33. Yuan, V., *Placental DNA methylation analysis tools*. 2022.
34. Aguet, F., et al., *Genetic effects on gene expression across human tissues.* Nature, 2017. **550**(7675): p. 204-213.
35. Dudbridge, F. and A. Gusnanto, *Estimation of significance thresholds for genomewide association scans.* Genet Epidemiol, 2008. **32**(3): p. 227-34.
36. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* J R Stat Soc Ser B Methodol, 1995(57): p. 289-300.
37. Pedersen, B.S., et al., *Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values.* Bioinformatics, 2012. **28**(22): p. 2986-8.
38. Globisch, D., et al., *Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates.* PLoS One, 2010. **5**(12): p. e15367.
39. Song, C.X., et al., *Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.* Nat Biotechnol, 2011. **29**(1): p. 68-72.
40. Lunnon, K., et al., *Variation in 5-hydroxymethylcytosine across human cortex and cerebellum.* Genome Biol, 2016. **17**: p. 27.
41. Spiers, H., et al., *5-hydroxymethylcytosine is highly dynamic across human fetal brain development.* BMC Genomics, 2017. **18**(1): p. 738.
42. Wilkins, O.M., et al., *Genome-wide characterization of cytosine-specific 5-hydroxymethylation in normal breast tissue.* Epigenetics, 2020. **15**(4): p. 398-418.
43. Xu, Y., et al., *Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells.* Mol Cell, 2011. **42**(4): p. 451-64.

44. Wu, H., et al., *Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells.* Genes Dev, 2011. **25**(7): p. 679-84.
45. Wu, H., et al., *Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells.* Nature, 2011. **473**(7347): p. 389-93.
46. Neri, F., et al., *Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells.* Genome Biol, 2013. **14**(8): p. R91.
47. Ku, M., et al., *Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains.* PLoS Genet, 2008. **4**(10): p. e1000242.
48. Senner, C.E., et al., *TET1 and 5-Hydroxymethylation Preserve the Stem Cell State of Mouse Trophoblast.* Stem Cell Reports, 2020. **15**(6): p. 1301-1316.
49. Valinluck, V., et al., *Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2).* Nucleic Acids Res, 2004. **32**(14): p. 4100-8.
50. Szulwach, K.E., et al., *Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells.* PLoS Genet, 2011. **7**(6): p. e1002154.
51. Szulwach, K.E., et al., *5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging.* Nat Neurosci, 2011. **14**(12): p. 1607-16.
52. Aran, D., S. Sabato, and A. Hellman, *DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes.* Genome Biol, 2013. **14**(3): p. R21.
53. Kennedy, E.M., et al., *An integrated -omics analysis of the epigenetic landscape of gene expression in human blood cells.* BMC Genomics, 2018. **19**(1): p. 476.
54. Arnold, P.R., A.D. Wells, and X.C. Li, *Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate.* Front Cell Dev Biol, 2019. **7**: p. 377.
55. Kim, Y.W., et al., *Chromatin looping and eRNA transcription precede the transcriptional activation of gene in the beta-globin locus.* Biosci Rep, 2015. **35**(2).
56. Xu, J., et al., *Upregulation of B3GNT3 is associated with immune infiltration and activation of NF-kappaB pathway in gynecologic cancers.* J Reprod Immunol, 2022. **152**: p. 103658.
57. Leca, J., et al., *IDH2 and TET2 mutations synergize to modulate T Follicular Helper cell functional interaction with the AITL microenvironment.* Cancer Cell, 2023. **41**(2): p. 323-339 e10.

Chapter 4 - Associations between placental hydroxymethylation and birthweight

Michael Mortillo[1], Elizabeth M. Kennedy[1], Karen E. Hermetz[1], Amber A. Burt[1], Carmen J. Marsit[1]

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

Adapted from an original manuscript currently under review at *Epigenetics*

Abstract

5-hydroxymethylcytosine (5hmC), is formed through oxidation of 5-methylcytosine (5mC), and is found on the cytosine base of cytosine-phosphate-guanine (CpG) dinucleotides. Though it is believed to mainly serve as a transient intermediate in the DNA demethylation pathway, recent evidence suggests that 5hmC may play a functionally relevant role. Here, we have conducted an epigenome-wide association study (EWAS) to assess the association between placenta 5hmC, obtained through parallel bisulfite and oxidative bisulfite modification of DNA and array based assessment, and newborn birthweight using the Rhode Island Child Health Study (RICHS). We also also assess whether correction of 5mC data through the removal of the 5hmC signal impacts the observed results from traditional epigenome-wide studies that rely on BS modification-based 5mC assessment alone. We identified 5hmC at one CpG in the CUBN gene to be significantly associated with birthweight (FDR < 0.05) and demonstrate that expression of that gene was also associated with birthweight. Comparison of EWAS effect estimates showed between 5hmC+5mC and 5mC results indicate strong correlation ($r = 0.77$, $p < 0.0001$). Our study identified little evidence to suggest the 5hmC in the placenta is related to birthweight, and that traditional assessment of 5mC through bisulfite modification alone provides an accurate assessment of the state of CpG specific DNA methylation in the placenta.

Introduction

Low birthweight is associated with increased infant mortality, as well as morbidity later in life[1, 2]. Similarly, high birthweight is associated with an increased risk of adverse health outcomes in developing children and later in life[2, 3]. The placenta is the interface between mother and fetus. It plays a critical role in fetal growth and newborn growth outcomes through its functions of facilitating gas-, waste-, and nutrient-exchange, as well as growth factor and hormone production[4-7]. Placental function can be influenced through genetic variation as well as in response to the environment experienced in pregnancy. Epigenetic variation can be useful in understanding functional variation in the placenta[8-10].

5-methylcytosine (5mC) is one of the most abundant epigenetic modifications in the genome. It represents the addition of a methyl group to the 5th carbon position of an unmethylated cytosine, usually in the context of a cytosine-phosphate-guanine (CpG) dinucleotide[11-13]. As a functional epigenetic mark, 5mC plays a crucial role in genomic imprinting[14, 15], inactivation of the X-chromosome[16, 17], and transposon repression[18, 19]. When localized to gene promoters, 5mC generally leads to transcriptional repression[20-23]. 5-hydroxymethylcytosine (5hmC) is produced via the oxidation of 5mC by ten-eleven translocation (TET) methylcytosine dioxygenases[24-27]. 5hmC is highly enriched in the bodies of transcriptionally active genes, promoters, and enhancers[28-30].Though it is commonly believed that 5hmC acts as a transient intermediate on the demethylation pathway[24, 31, 32], recent evidence suggests that it may also be a stable epigenetic modification[33].

Previous work has shown that DNA methylation patterns in the placenta are significantly associated with infant growth[34], with methylation at a number of individual CpG sites demonstrating a significant association with birthweight[35, 36]. Despite these advancements linking placental methylation, birthweight, and long-term health outcomes[37, 38], there has

been limited assessment of associations between 5hmC in placenta and birthweight[24, 39]. Further, because the most common assay to estimate DNA methylation, bisulfite (BS) conversion, does not discriminate between 5mC and 5hmC at a given cytosine, most existing studies of placental DNA methylation assess combined CpG methylation. In this study, we aim to assess associations between placental 5hmC on birthweight, using parallel assays of total CpG methylation (BS-converted) and 5mC only (oxidative BS-converted) data, to provide a more complete picture of the placenta epigenome and its relationship with birthweight. As a secondary analysis, we will also assess if the correction of 5mC data through the removal of the 5hmC signal impacts the observed results from traditional epigenome-wide studies that are based on bisulfite modification-based 5mC assessment alone.

Methods

Study cohort

The Rhode Island Child Health Study (RICHS) is a mother-infant cohort recruited from the Women & Infants Hospital of Rhode Island (WIH, Providence, RI) between September 2009 and July 2014. Mothers were 18-40 years old, with no history of psychological disorders and in good physical health, who delivered singleton infants free of any significant health complications at term ($\geq$ 37 gestational weeks). The cohort was oversampled for infants born small for gestational age (SGA, <10% 2013 Fenton Growth Curve[40]) and large for gestational age (LGA, >90% 2013 Fenton Growth Curve). Infants who were appropriate for gestational age (AGA, 10-90% 2013 Fenton Growth Curve), and matched for gestational age and maternal age were also enrolled. The current study focuses on a subset of enrolled participants with available placental 5mC and 5hmC data (n = 213).

Study covariates collection

Participant data, including study covariates of infant sex, birthweight, birth length, and gestational age were extracted from medical records. Parity was obtained from responses to an interviewer-administered questionnaire. Birthweight and birth length were Z-score transformed. Parity was recoded as ≥ 1 or 0 based on any documented history of pregnancy (≥ 1 if mother indicated "yes" when asked if they have ever been pregnant before, 0 if they indicated "no").

Placental sample collection

Placental parenchyma was obtained from the fetal side of the placenta, 2 cm from the umbilical cord insertion site, within 2 hours of delivery. All samples were free of maternal decidua. Samples were placed in RNALater at 4ºC for at least 72 hours, then removed, snap frozen in liquid nitrogen, pulverized, and stored at -80ºC until extraction.

CpG methylation and hydroxymethylation profiling and RNA-sequencing of transcripts

Assessment of methylation and hydroxymethylation were performed as previously described[41]. Briefly, BS and oxBS conversion were performed on placenta-derived DNA, following manufacturer's optimized protocol of 500 ng gDNA input for downstream analysis using the Infinium HD Methylation EPIC Bead Chip Array.

Quality control, normalization, and filtering of methylation and hydroxymethylation data were all performed in the R statistical programming language. We excluded cross-reactive probes, probes that failed p-value detection ($p > 0.01$) in > 1 sample, and probes overlapping single nucleotide polymorphisms (SNPs). Three samples were removed due to failing sex quality control or failing detection p-value (p-value > 0.01 in >2% of probes). Normalization of background correction, dye bias, and functional normalization for each of the BS and oxBS-converted samples was performed via the R/Bioconductor package *minfi* (version 1.24.0; *https://www.bioconductor.org*)[42]. Methylation data were exported from minfi as β-values, which represent the proportion of methylated alleles at each CpG. Standardization across

probe types was performed using subset quantile within-array normalization (SWAN)[43]. Estimation of 5hmC and 5mC β-values was performed using OxyBS (version 1.5)[44]. β-values were adjusted for batch effects using the ComBat function in R/Bioconductor package *sva*[45], and then logit-transformed (M-values) to better approximate a normal distribution[46]. Finally, we limited our analysis to autosomal probes only, to limit any confounding due to sex-specific effects. The final filtered, normalized dataset contained 689,815 CpGs.

RNA-sequencing on placenta-derived RNA was performed as previously described[47]. Read counts were corrected for library size differences across samples using the trimmed mean of M-values (TMM) method. The data were then $\log_2$-transformed to account for the mean-variance relationship, resulting in transcript counts expressed as $\log_2$TMM.

Cell type estimations

Proportions of constituent putative cell types in each placental sample were estimated using the reference-based R/Bioconductor package *planet*[48]. From this, we identified six putative constituent cells in placental samples: 1) trophoblasts, 2) stromal cells, 3) Hofbauer cells, 4) endothelial cells, 5) nucleated red blood cells, and 6) synctiotrophoblasts (STBs).

Robust linear modeling of 5hmC and 5mC data (EWAS)

To assess the association between site-specific 5mC, 5hmC, and birthweight, we utilized an epigenome-wide association study (EWAS) design to fit a robust linear model, regressing methylation M-values for each CpG on birthweight Z-score for all samples. This was performed three separate times, one for each type of available methylation data: 1) 5hmC (generated by OxyBS), 2) 5mC (5mC only, generated from oxBS conversion), and 3) 5mC+5hmC (generated from BS conversion). EWAS were performed on samples with available 5hmC, 5mC+5hmC, and 5mC data, as well as available birthweight Z-score and covariate data (n = 213). Covariates were identified through univariate associations ($p < 0.05$) with any of the top 5 principal

components (PCs) for 5mC+5hmC data and birthweight Z-score. Model covariates included infant sex, birth length Z-score, gestational age, parity (coded as 0 = first pregnancy, ≥ 1 = not first pregnancy), and STB proportions. This analysis was performed in R. Models were run using the *MASS* R package[49]. Estimates, standard errors, test statistics, and p-values were generated using the *lmtest* R package[50]. The covariance matrix of the estimated coefficients was supplied using the *sandwich* R package[51, 52]. False discovery rates (FDR) were calculated using the Benjamini-Hochberg procedure[53]. CpGs with an FDR < 0.05 were considered differentially hydroxy/methylated with birthweight.

Statistical tests

Correlations between methylation and hydroxymethylation proportions, as well as between EWAS effect estimates were performed through a Pearson's correlation using the *stats* R package[54]. We employed a multivariate linear regression model to test for associations between 5hmC proportions at any significant CpGs and expression of genes annotated to the CpGs, fitting a model of: ($\log_2$TMM transcript counts) = B_0 + B_1(5hmC proportions) + B_2(gestational age) + e. We employed a similar model to test for associations between transcript counts and birthweight z-score: (Birthweight Z-score) = B_0 + B_1($\log_2$TMM transcript counts) + B_2(gestational age) + e. Gestational age was included as a covariate in both models due to known associations between placental methylation, birthweight, gene expression, and gestational age[55-58].

CpG annotation

To identify associated genes, gene compartments, and CGI regions for the CpGs in our analyses, we annotated CpGs using the R/bioconductor package *IlluminaHumanMethylationEPICanno.ilm10b2.hg19*[59]. The available gene compartments from the EPIC array package are as follows: 1) 5' untranslated region (5' UTR), 2) transcription start

site (TSS) 200 (1-200 base pairs (bp) upstream of the TSS), 3) TSS 1500 (201-1500 bp upstream of TSS), 4) 1st exon, 5) gene bodies, 6) exon boundaries, and 7) 3' untranslated region (3' UTR). CpGs that lie in either the TSS200 or TSS1500 intervals were combined into one "TSS" category. CpGs were annotated to a CGI "shore" if they were within a 2 kb region flanking a CGI, a CGI "shelf" if they were within a 2 kb region flanking a CGI shore, or an "open sea" if they were not within a shore, shelf, or CGI. CpGs were annotated to chromatin-based genomic categories using ChromHMM[60], derived from the Roadmap Epigenomics Consortium[61] and applied to fetal placenta cells.

Results

Sample cohort

This study analyzed data from 213 mother-infant pairs from the RICHS cohort, with the maternal and fetal participant characteristics displayed in **Table 4-1**. The sample consisted mainly of white mothers (77.1% of samples), with a mean age of 30.9 years. There was a nearly equal distribution of male and female infants (51.1% vs. 48.9%, respectively), with a mean gestational age of 39.4 weeks, and by study design, the sample was over-represented by infants born SGA (14.3%) and LGA (30.9%).

Hydroxymethylation and methylation associations with birthweight

Comparison of individual CpG 5hmC and 5mC+5hmC β-values showed very little correlation ($r = 0.34$, $p < 0.0001$; **Fig. 4-1A**), while comparison of 5mC and 5mC+5hmC β-values showed a much stronger concordance ($r = 0.99$, $p < 0.0001$) (**Fig. 4-1B**).

Using an EWAS of 5hmC data, and controlling for infant sex, birth length Z-score, gestational age, parity, and STB proportions, we identified 5hmC at one CpG demonstrating an epigenome-wide significant association with birthweight ($\beta = 1.68$, standard error (SE) = 0.31, FDR = 0.04; **Fig. S4-1A)**. For the EWAS using 5mC+5hmC data, we identified methylation at

617 CpGs associated with birthweight (**Fig. S4-1B**). Among these 617 CpGs, we found a higher proportion trending positively with birthweight (n = 571, 92.5%). 5mC EWAS identified 41 significant CpGs (**Fig. S4-1C**), with 28 of these (68.3%) positively associated with birthweight.

Comparison of effect estimates between the 5mC+5hmC EWAS and the 5hmC EWAS revealed very little correlation ($r = 0.23$, $p < 0.0001$) (**Fig. 4-1C**). 5hmC EWAS estimates were also far more variable, ranging from -1.76 to 1.81, while estimates resulting from 5mC+5hmC EWAS showed less variability, ranging from -0.53 to 0.49 (**Fig. 4-1C**). Comparing the effect estimates from the 5mC+5hmC EWAS to the 5mC EWAS showed greater concordance ($r = 0.77$, $p < 0.0001$; **Fig. 4-1D**), and similarity in the ranges of effect estimates with 5mC EWAS estimates ranging from -0.49 to 0.55, similar to 5mC+5hmC range of -0.53 to 0.49 (**Fig. 4-1D**). The one CpG with 5hmC significantly associated with birthweight did not reach significance in the 5mC+5hmC EWAS (FDR = 0.57; **Fig 4-1C**). We identified 15 CpGs whose methylation was significantly associated with birthweight in both the 5mC+5hmC EWAS and 5mC EWAS (**Fig. 4-1D**).

The single CpG showing differential 5hmC associated with birthweight was annotated to the body of the *CUBN* gene on chromosome 10. The CpG was located in a CGI "open sea", and was annotated to an enhancer region and within a DNAse hypersensitivity site. Investigation of the relationship between 5hmC proportions at this CpG and RNA transcript counts of *CUBN* revealed no significant association between 5hmC and expression of *CUBN* ($\beta = 5.98$, $p = 0.4$). Assessment of *CUBN* transcript counts with birthweight Z-score, though, did reveal a significant positive association between *CUBN* expression and birthweight ($\beta = 0.51$, $p < 0.001$; **Fig. 4-2**).

Discussion

In this study, we assessed the relationships between DNA hydroxymethylation with birthweight in 213 placentae from a US mother-infant cohort, as well as examined how correction for 5hmC in 5mC data could impact EWAS results. We used data from parallel bisulfite- and oxidative bisulfite-treated placental samples to estimate the proportion of 5hmC at more than 600,000 individual CpG sites for each sample, and a rigorous statistical approach to assess associations of individual CpG 5-hydroxymethylcytosine proportion with infant birthweight, independent of infant sex, birth length Z-score, gestational age, parity, and STB proportions.

We found 5hmC at one CpG (cg12867894) to be significantly associated with birthweight ($\beta = 1.68$, standard error (SE) = 0.31, FDR = 0.04) (**Fig. S4-1A**), with increased 5hmC proportions at this locus associated with increased birthweight. This CpG is found on chromosome 10 in the body of the *CUBN* gene. *CUBN* encodes Cubilin, a plasma membrane receptor that aids in vitamin D metabolism[62, 63]. *CUBN* has been implicated in previous placenta-birthweight studies; one study[63] assessed the associations between genetic variants in placenta-derived DNA and birthweight, and identified three variants at the *CUBN* locus to be associated with low birthweight, independent of sex, race, and gestational age. A murine model[64] showed that Cubilin acts as a binding partner for galectin-3 (encoded by Lgals3) in the placental-fetal membrane, and dysregulation of placental Lgals3 has been shown to be associated with a number of pregnancy complications including intrauterine growth restriction[65]. This would suggest that reduced expression of CUBN could impact its ability to effectively bind Lgals3, which in turn could lead to growth restriction and lower birthweight, and our study adds to that evidence demonstrating that placental samples with higher CUBN expression were associated with larger birthweight infants (**Fig. 4-2**).

The lack of significant birthweight-associated CpGs in the 5hmC EWAS may indicate that 5hmC does not play a role in controlling the function of the placenta for its role in fetal growth. This lack of association could also be driven by the low proportions of 5hmC and its limited variability in the placenta. Previous studies have shown that 5hmC is about 10-fold less abundant than 5mC across the genome[66], with a prior placental 5hmC study[41] showing limited levels of 5hmC at each CpG. Additionally, the larger estimates we observed in the 5hmC EWAS (in comparison to estimates from the 5mC+5hmC and 5mC EWAS's) are also in line with the limited variability of 5hmC and thus limited power to detect effects of a smaller magnitude. Studies with a larger sample size would be needed to more comprehensively assess associations between placental 5hmC and birthweight, as well as other outcomes.

As a secondary analysis, we also assessed the influence of 5hmC in an EWAS using BS-converted methylation data alone (5mC+5hmC data). We identified 617 CpGs with differential methylation associated with birthweight in the 5mC+5hmC EWAS (**Fig. S4-1B**), and CpG sites of differential methylation in the 5mC EWAS (**Fig. S4-1C**). Significant differential methylation of fifteen CpGs overlapped between the 5mC+5hmC and 5mC EWAS analyses (**Fig. 4-1D**), although generally the effect estimates were highlight correlated (r=0.77). The limited identified associations between 5hmC and the high correlation of effect estimates and beta values of 5mC data coming from either traditional bisulfite modification or oxidative bisulfite modification taken together suggests that the majority of the signal in 5mC+5hmC EWAS in the placenta is from 5mC.

The findings of this study should be interpreted in the context of its limitations. This is an observational study where placental methylation and birthweight were both measured at birth. Thus, it is difficult to determine the directionality of this relationship. Additionally, methylation

and birthweight were measured at term, and thus we cannot conclude that our results represent this observed relationship throughout development. Lastly, the RICHS cohort consists mostly of healthy, white mothers from the New England region of the United States, and thus these findings may not be generalizable to more diverse populations.

To our knowledge, this study provides one of the only examinations of 5hmC in the placenta and its relationship to birthweight. Utilizing parallel bisulfite and oxidative bisulfite methods and array-based assessments allowed us to also elucidate an exclusive 5mC signal at each CpG, and to perform a robust multi-EWAS analysis to explore not only 5hmC's association with birthweight, but also if it is a significant artifact in EWAS of DNA methylation.

Conclusion

DNA methylation has been linked to the development and function of the placenta throughout gestation, as well as long-term health outcomes, and this prenatal programming may manifest as the phenotype of birthweight. Here, we have shown that placental hydroxymethylation at the *CUBN* gene is associated with birthweight. While 5hmC is not absent in the placenta, we have provided evidence that correction of 5mC data through the removal of the 5hmC signal does not substantially impact the results obtained from using only BS-converted placenta DNA. Though additional studies are required to develop a more complete understanding of the associations between placental hydroxymethylation, methylation, and birthweight, these findings serve as a good stepping stone for researchers who aim to further investigate this relationship.

Statements and Declarations

Data Availability

Raw data were generated at Emory University. Derived data supporting the findings of this study are available through the Gene Expression Omnibus (GEO, accession number: GSE144129).

Funding
This work was supported by grants from the National Institutes of Health, National Institute of Environmental Health Sciences (P30ES 019776, U24 ES028507, R21 ES028226) and Eunice Kennedy Shriver National Institute of Child Health and Development (R01 HD108310).

Conflicts of Interest
The authors report there are no competing interests to declare.

Ethical Standards
All enrolled participants provided written informed consent and study protocols were approved by the institutional review boards (IRB) at WIH and Emory University.

Tables

	RICHS (n = 213)
Maternal characteristics	
Age in years (mean, SD)	30.9 (4.8)
Educational attainment (n, %)[a]	
High school or less	32 (15.0)
Post-high school or junior college	104 (48.8)
College	77 (36.2)
Self-reported race/ethnicity (n, %)	
Asian	10 (4.7)
Black	12 (5.6)
Indian	2 (0.9)
More than one race	3 (1.4)
Unknown/not reported	19 (8.9)
White	167 (78.4)
Primigravida (n, %)	51 (23.9)
Infant characteristics	
Age in weeks (mean, SD)	39.4 (1.0)
Birthweight in grams (mean, SD)	3,540.5 (663.7)
Birthweight category (n, %)[b]	
SGA	32 (13.0)
AGA	117 (54.9)
LGA	64 (30.0)
Birthlength in cm (mean, SD)	50.6 (2.7)
Sex (n, %)	
Male	109 (51.2)
Female	104 (48.8)

Table 4-1: RICHS participant demographics

[a]Two subjects with missing maternal education level.

[b]Infants born with birthweight percentile ≤ 10% (small for gestational age [SGA]), 10-90% (appropriate for gestational age [AGA]), and ≥ 90% (large for gestational age [LGA])

Figures

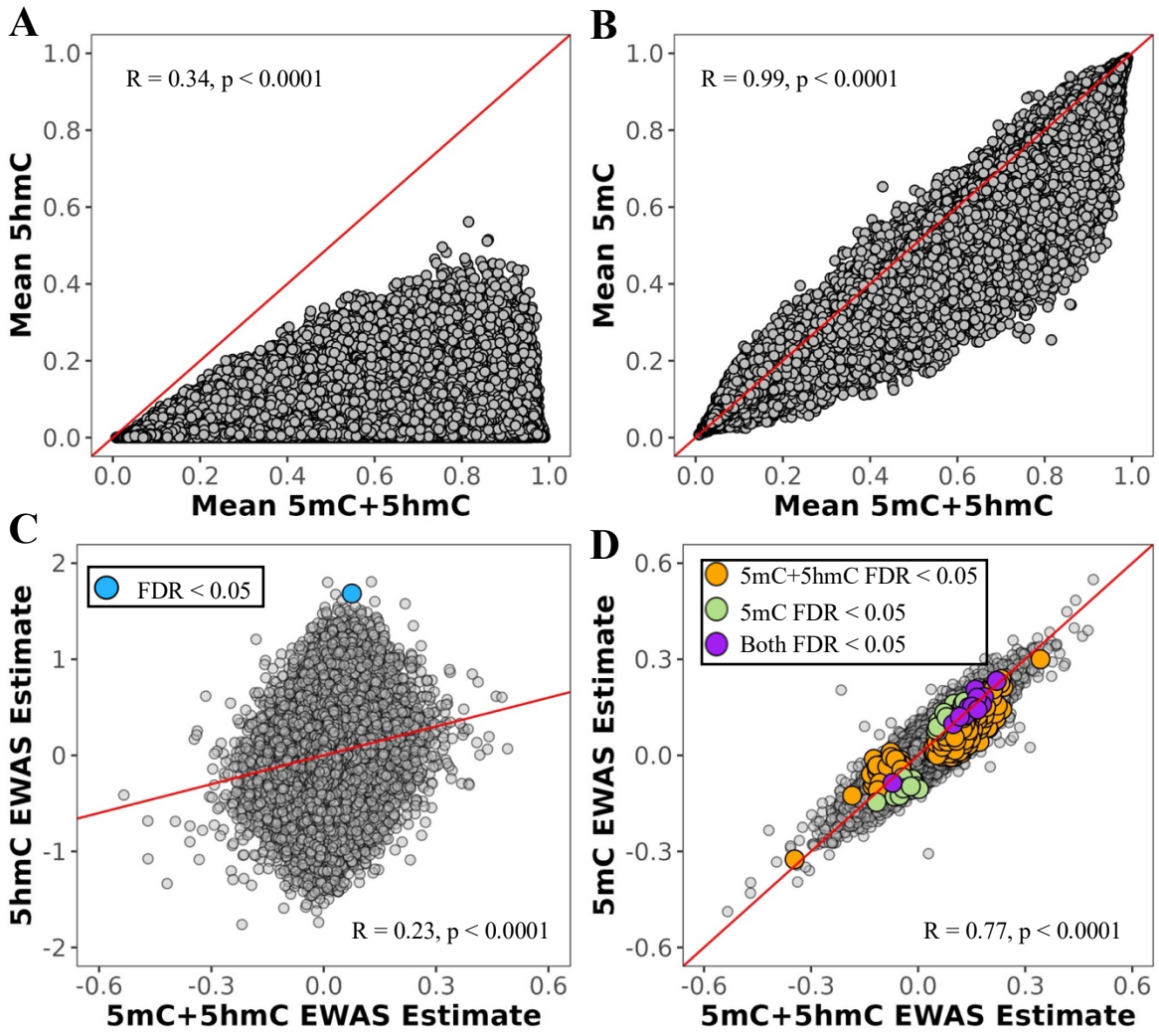


Figure 4-1: Comparison of placental hydroxymethylation and methylation β-values and resulting EWAS effect estimates. Comparison of **A)** 5hmC and **B)** 5mC β-values with 5mC+5hmC β-values. 5mC+5hmC data was obtained through BS-conversion of placental-derived DNA, while 5mC data was obtained through oxBS conversion. 5hmC data was obtained from maximum likelihood estimation of combined BS and oxBS data. β-values are represented as mean β-value for each CpG across all placental samples. Red line represents 1:1 correlation between β-values. Pearson's correlation coefficient (R) and corresponding p-value from correlation between β-values are shown. Comparison of **C)** 5hmC EWAS estimates and **D)** 5mC EWAS estimates with 5mC+5hmC EWAS estimates. Each EWAS was fit with the following model: $(5h/mC) = \beta_0 + \beta_1(\text{Birthweight Z-score}) + \beta_2(\text{Infant sex}) + \beta_3(\text{Birthlength Z-score}) + \beta_4(\text{Gestational Age}) + \beta_5(\text{Parity}) + \beta_6(\text{STB Proportions}) + \varepsilon$. Blue point in panel C represents CpG that met significance (FDR < 0.05) in 5hmC EWAS, while orange, green, and purple points in panel D represent CpGs that meet significance in 5mC+5hmC EWAS, 5mC EWAS, and in both EWAS's, respectively.

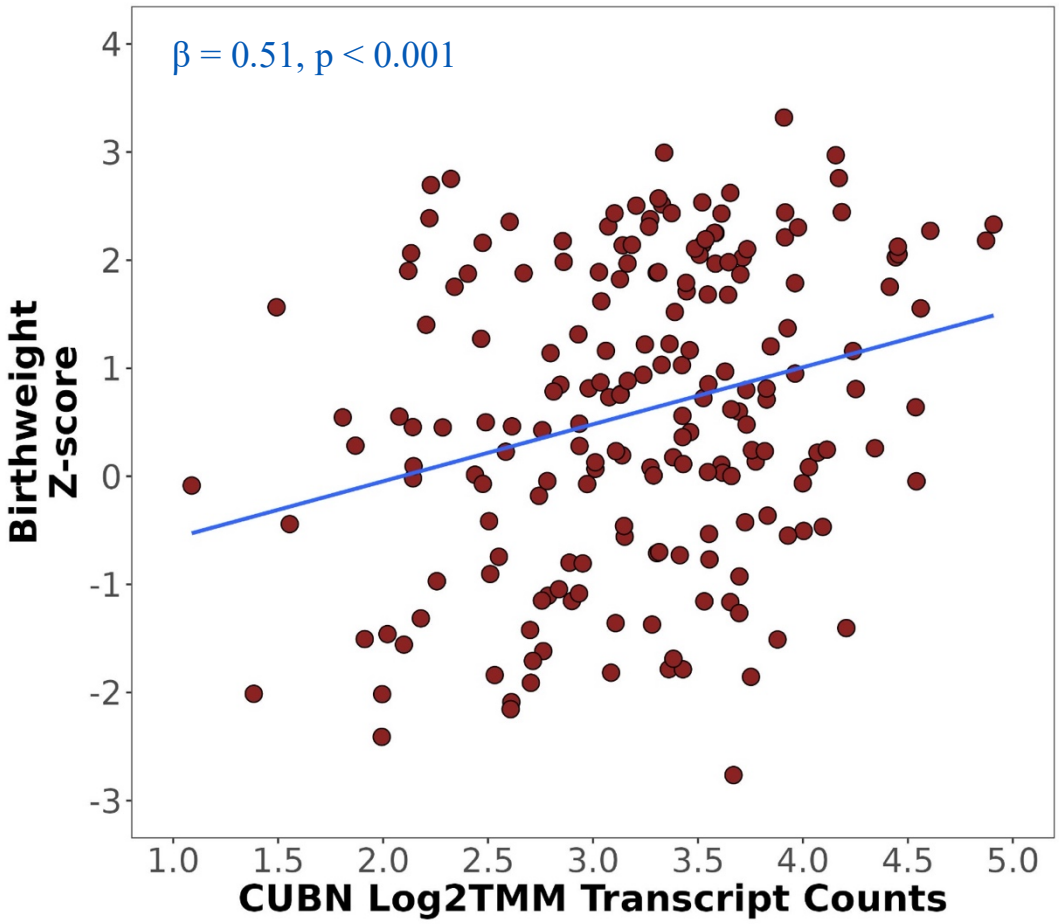


Figure 4-2: Association between placental CUBN expression and infant birthweight Z-score. CUBN expression is represented as $\log_2$TMM transcript counts. Blue line represents linear regression model: (Birthweight Z-score) = $\beta_0 + \beta_1$(CUBN $\log_2$TMM) + β_2(Gestational Age) + ε. Resulting effect estimate (β) and p-value are shown.

Supplemental Figures

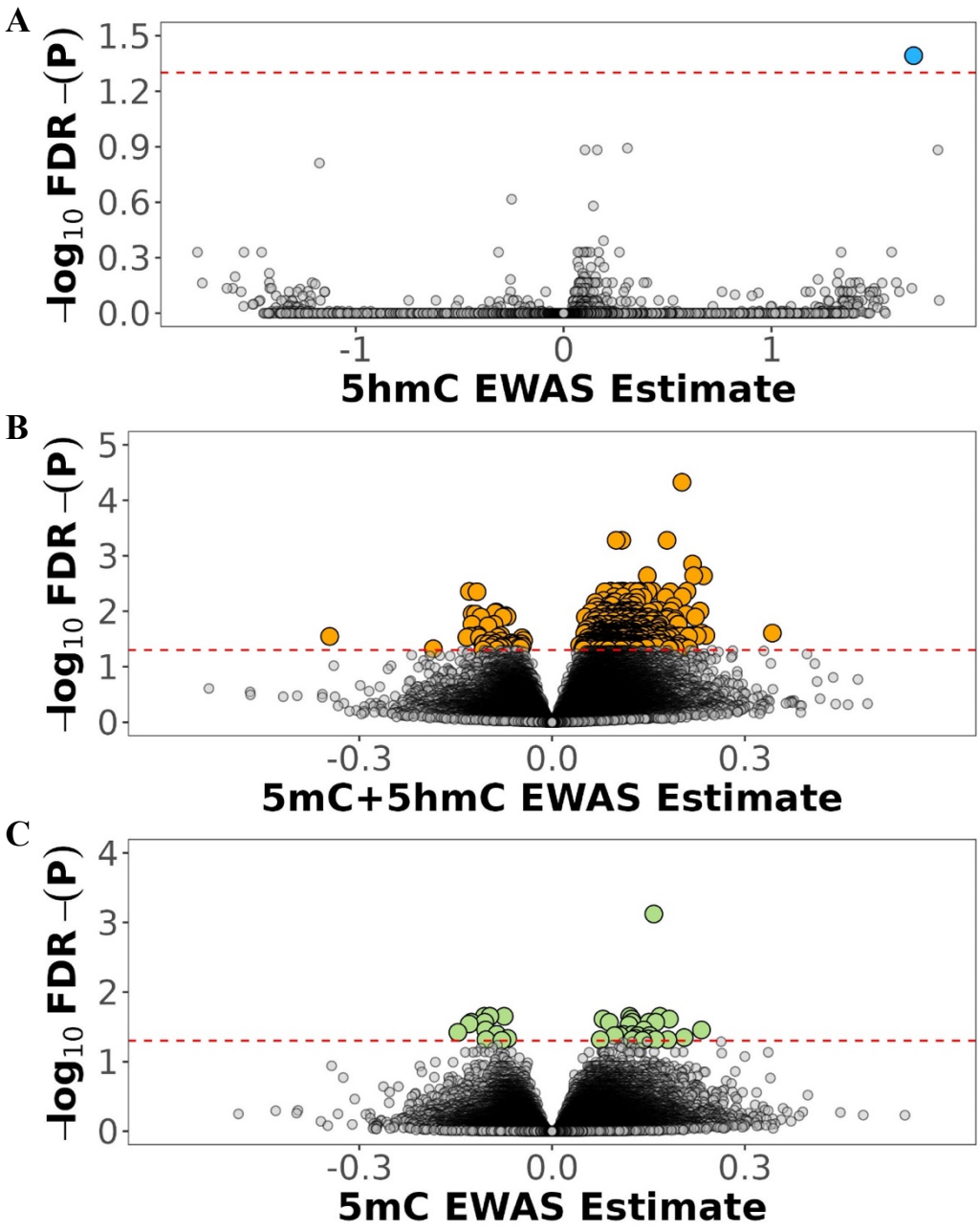


Figure S4-1: EWAS's between placental A) 5hmC, B) 5mC+5hmC, and C) 5mC with birthweight Z-score. On the y-and-x-axes, -$\log_{10}$(FDR) in the association of each CpG with birthweight Z-score and effect estimates, or the change in methylation or hydroxymethylation M-value per every one unit increase in birthweight Z-score, are shown respectively. Dashed red line represents FDR < 0.05 (Benjamini-Hochberg correction method), with colored points above lines representing CpGs that met significance cutoff.

References

1. Argente, J., O. Mehls, and V. Barrios, *Growth and body composition in very young SGA children.* Pediatr Nephrol, 2010. **25**(4): p. 679-85.
2. Filiberto, A.C., et al., *Birthweight is associated with DNA promoter methylation of the glucocorticoid receptor in human placenta.* Epigenetics, 2011. **6**(5): p. 566-72.
3. Ng, S.K., et al., *Risk factors and obstetric complications of large for gestational age births with adjustments for community effects: results from a new cohort study.* BMC Public Health, 2010. **10**: p. 460.
4. Bloom, M.S., M. Varde, and R.B. Newman, *Environmental toxicants and placental function.* Best Pract Res Clin Obstet Gynaecol, 2022. **85**(Pt B): p. 105-120.
5. Gude, N.M., et al., *Growth and function of the normal human placenta.* Thromb Res, 2004. **114**(5-6): p. 397-407.
6. Kaufmann, P., T.M. Mayhew, and D.S. Charnock-Jones, *Aspects of human fetoplacental vasculogenesis and angiogenesis. II. Changes during normal pregnancy.* Placenta, 2004. **25**(2-3): p. 114-26.
7. Red-Horse, K., et al., *Trophoblast differentiation during embryo implantation and formation of the maternal-fetal interface.* J Clin Invest, 2004. **114**(6): p. 744-54.
8. Marsit, C.J., *Placental Epigenetics in Children's Environmental Health.* Semin Reprod Med, 2016. **34**(1): p. 36-41.
9. Mortillo, M. and C.J. Marsit, *Select Early-Life Environmental Exposures and DNA Methylation in the Placenta.* Curr Environ Health Rep, 2023. **10**(1): p. 22-34.
10. Peng, S., et al., *Genetic regulation of the placental transcriptome underlies birth weight and risk of childhood obesity.* PLoS Genet, 2018. **14**(12): p. e1007799.
11. Illingworth, R.S. and A.P. Bird, *CpG islands--'a rough guide'.* FEBS Lett, 2009. **583**(11): p. 1713-20.
12. Klungland, A. and A.B. Robertson, *Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine.* Free Radic Biol Med, 2017. **107**: p. 62-68.
13. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function.* Neuropsychopharmacology, 2013. **38**(1): p. 23-38.
14. Greenberg, M.V.C. and D. Bourc'his, *The diverse roles of DNA methylation in mammalian development and disease.* Nat Rev Mol Cell Biol, 2019. **20**(10): p. 590-607.
15. Butler, M.G., *Genomic imprinting disorders in humans: a mini-review.* J Assist Reprod Genet, 2009. **26**(9-10): p. 477-86.
16. Payer, B. and J.T. Lee, *X chromosome dosage compensation: how mammals keep the balance.* Annu Rev Genet, 2008. **42**: p. 733-72.
17. Heard, E., P. Clerc, and P. Avner, *X-chromosome inactivation in mammals.* Annu Rev Genet, 1997. **31**: p. 571-610.
18. Brooks, S.C., et al., *5-methylcytosine recognition by Arabidopsis thaliana DNA glycosylases DEMETER and DML3.* Biochemistry, 2014. **53**(15): p. 2525-32.
19. Deniz, O., J.M. Frost, and M.R. Branco, *Regulation of transposable elements by DNA modifications.* Nat Rev Genet, 2019. **20**(7): p. 417-431.
20. de Mendoza, A., R. Lister, and O. Bogdanovic, *Evolution of DNA Methylome Diversity in Eukaryotes.* J Mol Biol, 2019.

21. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes Dev, 2011. **25**(10): p. 1010-22.
22. Long, H.K., et al., *Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates.* Elife, 2013. **2**: p. e00348.
23. Zhang, L., X. Xiao, and Z.C. Xu, *iPromoter-5mC: A Novel Fusion Decision Predictor for the Identification of 5-Methylcytosine Sites in Genome-Wide DNA Promoters.* Front Cell Dev Biol, 2020. **8**: p. 614.
24. He, B., et al., *Tissue-specific 5-hydroxymethylcytosine landscape of the human genome.* Nat Commun, 2021. **12**(1): p. 4249.
25. Ito, S., et al., *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.* Science, 2011. **333**(6047): p. 1300-3.
26. Robertson, J., A.B. Robertson, and A. Klungland, *The presence of 5-hydroxymethylcytosine at the gene promoter and not in the gene body negatively regulates gene expression.* Biochem Biophys Res Commun, 2011. **411**(1): p. 40-3.
27. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* Science, 2009. **324**(5929): p. 930-5.
28. Song, C.X., et al., *Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.* Nat Biotechnol, 2011. **29**(1): p. 68-72.
29. Schutsky, E.K., et al., *Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase.* Nat Biotechnol, 2018.
30. Ficz, G., et al., *Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation.* Nature, 2011. **473**(7347): p. 398-402.
31. Hahn, M.A., P.E. Szabo, and G.P. Pfeifer, *5-Hydroxymethylcytosine: a stable or transient DNA modification?* Genomics, 2014. **104**(5): p. 314-23.
32. Pfeifer, G.P., *5-hydroxymethylcytosine stabilizes transcription by preventing aberrant initiation in gene bodies.* Nat Genet, 2023. **55**(1): p. 2-3.
33. Bachman, M., et al., *5-Hydroxymethylcytosine is a predominantly stable DNA modification.* Nat Chem, 2014. **6**(12): p. 1049-55.
34. Banister, C.E., et al., *Infant growth restriction is associated with distinct patterns of DNA methylation in human placentas.* Epigenetics, 2011. **6**(7): p. 920-7.
35. Chen, P.Y., et al., *Prenatal Growth Patterns and Birthweight Are Associated With Differential DNA Methylation and Gene Expression of Cardiometabolic Risk Genes in Human Placentas: A Discovery-Based Approach.* Reprod Sci, 2018. **25**(4): p. 523-539.
36. Tekola-Ayele, F., et al., *DNA methylation loci in placenta associated with birthweight and expression of genes relevant for early development and adult diseases.* Clin Epigenetics, 2020. **12**(1): p. 78.
37. Lester, B.M., E. Conradt, and C.J. Marsit, *Are epigenetic changes in the intrauterine environment related to newborn neurobehavior?* Epigenomics, 2014. **6**(2): p. 175-8.
38. Rong, C., et al., *DNA methylation profiles in placenta and its association with gestational diabetes mellitus.* Exp Clin Endocrinol Diabetes, 2015. **123**(5): p. 282-8.
39. Zhang, Y., et al., *Aberrant hydroxymethylation of ANGPTL4 is associated with selective intrauterine growth restriction in monochorionic twin pregnancies.* Epigenetics, 2020. **15**(8): p. 887-899.
40. Fenton, T.R. and J.H. Kim, *A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants.* BMC Pediatr, 2013. **13**: p. 59.

41. Green, B.B., et al., *Hydroxymethylation is uniquely distributed within term placenta, and is associated with gene expression.* FASEB J, 2016. **30**(8): p. 2874-84.
42. Fortin, J.P., T.J. Triche, Jr., and K.D. Hansen, *Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi.* Bioinformatics, 2017. **33**(4): p. 558-560.
43. Maksimovic, J., L. Gordon, and A. Oshlack, *SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips.* Genome Biol, 2012. **13**(6): p. R44.
44. Houseman, E.A., K.C. Johnson, and B.C. Christensen, *OxyBS: estimation of 5-methylcytosine and 5-hydroxymethylcytosine from tandem-treated oxidative bisulfite and bisulfite DNA.* Bioinformatics, 2016. **32**(16): p. 2505-7.
45. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
46. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.* BMC Bioinformatics, 2010. **11**: p. 587.
47. Deyssenroth, M.A., et al., *Whole-transcriptome analysis delineates the human placenta gene network and its associations with fetal growth.* BMC Genomics, 2017. **18**(1): p. 520.
48. Yuan, V., et al., *Accurate ethnicity prediction from placental DNA methylation data.* Epigenetics Chromatin, 2019. **12**(1): p. 51.
49. Venables, W. and B. Ripley, *Modern Applied Statistics with S*. 4th ed. 2002.
50. Zeileis, A. and T. Hothorn, *Diagnostic checking in regression relationships.* 2002.
51. Zeileis, A., *Econometric computing with HC and HAC covariance matrix estimators.* 2004.
52. Zeileis, A., *Object-oriented computation of sandwich estimators.* Journal of Statistical Software, 2006. **16**(1): p. 1-16.
53. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.
54. R Core Team (2021), *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.*
55. Agha, G., et al., *Birth weight-for-gestational age is associated with DNA methylation at birth and in childhood.* Clin Epigenetics, 2016. **8**: p. 118.
56. Dwi Putra, S.E., et al., *Being Born Large for Gestational Age is Associated with Increased Global Placental DNA Methylation.* Sci Rep, 2020. **10**(1): p. 927.
57. Novakovic, B., et al., *Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors.* BMC Genomics, 2011. **12**: p. 529.
58. Wilson, R.L., et al., *Characterization of 5-methylcytosine and 5-hydroxymethylcytosine in human placenta cell types across gestation.* Epigenetics, 2019. **14**(7): p. 660-671.
59. Hansen, K., *IlluminaHumanMethylationEPICanno.ilm10b2.hg19: Annotation for Illumina's EPIC methylation arrays.*, in *R package version 0.6.0*. 2016.
60. Ernst, J. and M. Kellis, *Chromatin-state discovery and genome annotation with ChromHMM.* Nat Protoc, 2017. **12**(12): p. 2478-2492.
61. Roadmap Epigenomics, C., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.

62. Ashley, B., et al., *Placental uptake and metabolism of 25(OH)vitamin D determine its activity within the fetoplacental unit.* Elife, 2022. **11**.
63. Workalemahu, T., et al., *Placental genetic variations in vitamin D metabolism and birthweight.* Placenta, 2017. **50**: p. 78-83.
64. Crider-Pirkle, S., et al., *Cubilin, a binding partner for galectin-3 in the murine utero-placental complex.* J Biol Chem, 2002. **277**(18): p. 15904-12.
65. Jeschke, U., et al., *Expression of galectin-1, -3 (gal-1, gal-3) and the Thomsen-Friedenreich (TF) antigen in normal, IUGR, preeclamptic and HELLP placentas.* Placenta, 2007. **28**(11-12): p. 1165-73.
66. Ruzov, A., et al., *Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development.* Cell Res, 2011. **21**(9): p. 1332-42.

Chapter 5 – Summary, Limitations, Future Directions, and Conclusions

Summary

Investigating the epigenetic landscape of placental 5hmC and its associations with gene expression and birthweight may provide insight into the epigenetic underpinnings of the DOHAD hypothesis. Fetal stressors resulting from maternal health and behaviors, and environmental exposures during pregnancy have been implicated in placental dysfunction[1-3]. However, abnormal placental pathologies are rare within the general population[4], and may not account for many of the gestational and longitudinal health outcomes that are susceptible to developmental programming. Epigenetic mechanisms in the placenta may be attributed to the developmental programming of long-term health outcomes in offspring and may represent more subtle but potentially important changes to placental function[5-8]. This work addressed how placental epigenetic alterations, particularly methylation and hydroxymethylation, associate with gene expression and birthweight, thereby providing one possible mechanism for how long-term health outcomes are developmentally programmed.

In chapter two, we reviewed the literature on placental methylation and early-life environmental exposures to understand how various early-life environmental exposures, including air pollution, maternal smoking, chemicals, trace metals, socioeconomic adversity, and circadian disruption, impact DNA methylation in the placenta. From this review, various themes emerged, which included how variability in methylation tends to work in an exposure-specific framework. However, comparing methylation levels across exposures from these various studies helped identify several candidate genes that exhibit differential methylation. Many of the reviewed studies that assess exposures in a trimester-specific framework found larger effects on methylation when the exposure occurred during the 1st trimester. We also identified some themes

related to temporal changes across studies; earlier studies generally assessed methylation in a targeted, gene-specific framework, while later studies have utilized epigenome-wide, array-based technology. Results from this chapter shed light on the link between the maternal-placenta environment, epigenetic variation, and long-term health.

In chapter three, using methylation and expression data from term placenta samples, we examined the association of 5-hydroxymethylcytosine (5hmC) with gene expression. We found that proportions of 5hmC and transcript abundance were largely CpG location-specific; among CpGs in the body of the gene, 5hmC increased as transcription of the gene increased. Using expression quantitative trait hydroxymethylation (eQTHM) analysis, revealed an overall positive association between 5hmC and expression. We observed significant enrichment and depletion of eQTHMs in enhancers and active transcription start sites (TSS), respectively. Finally, we identified a number of differentially hydroxymethylated regions (DHMRs), contiguous regions of the genome where 5hmC may be playing a more functionally relevant role. This chapter further elucidated the limited functional capacity and distribution of 5hmC in the placenta.

In chapter four, we examined the association between placental hydroxymethylation, methylation, and birthweight using an epigenome-wide association study (EWAS) design. For methylation, we used 5mC+5hmC data (derived from bisulfite (BS) conversion), as well as 5mC data (derived from the removal of the 5hmC signal through oxidative bisulfite (oxBS) conversion). For hydroxymethylation, we used a maximum-likelihood technique to obtain only a 5hmC signal at each CpG. Together, these methods allowed us to developed a better understand of the effect BS and oxBS conversion has on methylation signals. Using the 5hmC data, we identified hydroxymethylation at one CpG to be significantly associated with birthweight. This CpG lies in the body of the CUBN gene, which has been shown to play a role in fetal growth.

5mC+5hmC and 5mC data identified a larger number of CpGs where methylation was significantly associated with birthweight, with the majority of these signals being positively associated with birthweight. Though we only identified a few relationships between 5hmC and birthweight, we did show that the characterization of 5mC is not overtly biased by 5hmC content, as evidenced by the strong correlation in the 5mC+5hmC and 5mC EWAS results. Results from this chapter shed light on a potential link between placental hydroxymethylation, methylation, and birthweight, and how correction of the 5mC signal through the removal of the 5hmC signal impacts the observed results from traditional epigenome-side studies that rely on BS-conversion to assess methylation.

This work adds to a growing body of scientific literature that sheds light on the placental epigenome, and suggests epigenetic modifications in the placenta may be playing a role in birth outcomes. We also contribute to the emerging field of environmental exposure influence of placental methylation, by reviewing how environmental exposures placed on the fetus throughout gestation may affect methylation. While previous work within this field has shown that various environmental perturbations during pregnancy can explain changes in the epigenetic signatures of the placenta[9-13], the review we conducted is among the first to examine these associations in a cross-environmental framework. Though recent work into the functionality of 5hmC has elucidated its role in promoting gene expression through the blockage of transcriptionally repressive methyl CpG-binding (MBD) proteins [14-18], few studies have empirically examined its association with gene expression. A previous study[19] from our lab leveraged publicly available placental gene expression data to evaluate associations with 5hmC, though the work described in this thesis is unique in that it is the first to utilize paired sample RNA-sequencing data to characterize this association empirically. Additionally, previous work

has shown placental methylation may represent the link between genetic and environmental influence on birthweight[20-22], though these studies generally assess these associations using BS-converted methylation data, which does not discriminate between 5mC and 5hmC. By performing a parallel oxBS conversion method, coupled with a maximum-likelihood estimation to generate a separate 5hmC readout, this study is the first to separately assess the associations between placental 5hmC and 5hmC with birthweight.

The findings from this study play into a larger relevance in the field of placental epigenetics and clinical outcomes. The functional annotation performed throughout this work highlights the critical role 5hmC plays in placental gene expression, as well as the roles of 5mC and 5hmC in fetal birthweight. As a central vascular organ that oversees fetal growth and development throughout gestation[23-25], proper function of the placenta is essential to ensuring successful gestational outcomes. Thus, perturbations to the molecular processes that occur within the placenta, including and especially gene expression, may result in placental dysregulation and subsequent programming of long-term health effects. As fetal growth restriction has been linked to the development of chronic diseases throughout life[26], the findings from this study have far-reaching clinical implications. Additionally, placental development and fetal growth occur concurrently, and thus disruptions in either process may pose adverse effects to the other. Finally, many of the environmental exposures that we assessed (and their subsequent impact on placental methylation) are common exposures within the general population, and thus the findings from this study are applicable to the larger population.

Given the relationships between early-life environmental exposures, placental methylation and function, and long-term health outcomes, it is reasonable to suggest that epigenetic mechanisms such as placental hydroxmethylation and methylation may serve as

potential mechanisms for the developmental programming of chronic diseases. This work also aids in clinical biomarker discovery, as placental hydroxymethylation and gene expression are molecular processes that were detected at birth. Thus, this may allow for development of early intervention methods to help alleviate long-term health effects.

Overall Limitations

The findings from this study should be analyzed and interpreted within the context of its limitations. Firstly, the placental tissue samples that were utilized were limited to term placenta, which presents challenges in characterizing epigenetic associations that are prevalent during pre-term periods, yet may still play a role in health outcomes. Given existing federal policies and health practices, as well as the fact that the placenta needs to be delivered, samples cannot be taken during pregnancy without significant risk to the mother and infant. Thus, placenta tissue collected at birth is usually the only safe, legal, and ethical way to study molecular mechanisms associated with fetal growth and development.

Environmental exposure assessment, hydroxymethylation, methylation, gene transcript abundance, and birthweight were all measured at birth, making it difficult to elucidate a temporal or causal relationships between these exposures, processes, and outcomes. However, given the aforementioned federal policies and practices, it is difficult to obtain placental biomarker readouts prior to birth, and thus measuring these molecular hallmarks at birth is often the only available option.

The placenta is a highly complex tissue composed of various cell types[27, 28], and thus many of the observed associations may be cell-type specific. To overcome this, we employed a reference panel and method through R package *planet*[29], to estimate sample-specific cellular compositions. Recent single-cell sequencing studies within the placenta have shown a large representation of trophoblasts[30], and *planet* estimates using RICHS data have validated these

results, with more than 90% of cells in the placental samples being cytotrophoblasts and synctiotrophoblasts (STBs). To account for this cell-type heterogeneity, we included STB proportions as a covariate in our eQTHM and birthweight models.

The RICHS cohort is largely composed of white mothers of European descent, and the majority of these mothers are from the New England area of the United States. Due to the limited racial and ethnic diversity of the cohort, these findings may not be applicable to more diverse populations.

Future Directions

The field of placental epigenetics is ripe for additional mechanistic understanding, and emerging studies utilizing newer, more robust techniques for measuring methylation will only aid in this promise. A placental epigenetic clock can now be used to estimate epigenetic gestational age, and the difference between this and the actual gestational age can be used as a relevant outcome[31]. Additionally, as sequencing technology is being used more often in placental studies, this can potentially increase coverage and reveal distinct methylation patterns in genomic regions that may be missed using array-based and targeted approaches.

Additional methods can also be employed to better understand the causality related to the developmental origins of health and disease. As mentioned earlier, environmental exposure assessment, 5hmC and 5mC, and gene expression were all measured at birth, thereby limiting any causal inference. Incorporating repeated measures of exposure throughout the entire pregnancy period will allow for a better understanding of the impacts of exposure on these molecular processes, as well as help craft a more concrete link between exposures, placental methylation and function, and long-term health outcomes.

In order to further our understanding of the molecular mechanisms underlying the developmental origins of health and disease, more work is needed to characterize the

relationships between placental methylation and additional early life health outcomes, such as blood pressure, BMI, and neurodevelopmental and psychiatric phenotypes that may manifest in late childhood or early adulthood. Conducting longer, longitudinal-based studies using screening and diagnostic tools would allow for detection of some of these later-developing phenotypes. Recruitment of larger sample sizes would also yield sufficient statistical power to conduct mediation analyses, where molecular processes such as gene expression can be seen as causal mediators on the path to developmental outcomes. Utilizing larger sample sizes would also help rectify the aforementioned diversity problem in this study, as more diverse samples will allow for interaction between genetic, epigenetic, and environmental sources of variation throughout the gestational period.

Conclusions

The characterization of placental 5hmC and its association with gene expression suggests a role for this mark in the placenta, though its lack of association with birthweight indicates it still has unknown phenotypic consequences. Changes in the placental epigenome in response to gestational conditions has been implicated in placental dysregulation, which in turn can lead to adverse pregnancy outcomes including preeclampsia and fetal growth restriction, highlighting the potential role of these epigenetic mechanisms in the developmental programming of long-term health outcomes. Here, we have characterized associations between placental 5hmC and expression, as well as 5hmC and 5mC with birthweight. Considering these modifications are detectable at birth, it is hopeful that the findings of this study may be utilized to aid in the development of early intervention methods to help alleviate long-term health effects. Combined, this will aid in the ongoing fight to reduce the global prevalence of chronic diseases.

References

1. Bronson, S.L. and T.L. Bale, *The Placenta as a Mediator of Stress Effects on Neurodevelopmental Reprogramming.* Neuropsychopharmacology, 2016. **41**(1): p. 207-18.
2. Gagnon, R., *Placental insufficiency and its consequences.* Eur J Obstet Gynecol Reprod Biol, 2003. **110 Suppl 1**: p. S99-107.
3. Pintican, D., et al., *Effects of maternal smoking on human placental vascularization: A systematic review.* Taiwan J Obstet Gynecol, 2019. **58**(4): p. 454-459.
4. Jim, B. and S.A. Karumanchi, *Preeclampsia: Pathogenesis, Prevention, and Long-Term Complications.* Semin Nephrol, 2017. **37**(4): p. 386-397.
5. Conradt, E., et al., *The roles of DNA methylation of NR3C1 and 11beta-HSD2 and exposure to maternal mood disorder in utero on newborn neurobehavior.* Epigenetics, 2013. **8**(12): p. 1321-9.
6. Maccani, J.Z., et al., *Placental DNA Methylation Related to Both Infant Toenail Mercury and Adverse Neurobehavioral Outcomes.* Environ Health Perspect, 2015. **123**(7): p. 723-9.
7. Marsit, C.J., et al., *Placental 11-beta hydroxysteroid dehydrogenase methylation is associated with newborn growth and a measure of neurobehavioral outcome.* PLoS One, 2012. **7**(3): p. e33794.
8. Stroud, L.R., et al., *Prenatal Major Depressive Disorder, Placenta Glucocorticoid and Serotonergic Signaling, and Infant Cortisol Response.* Psychosom Med, 2016. **78**(9): p. 979-990.
9. Vlahos, A., et al., *Human placental methylome in the interplay of adverse placental health, environmental exposure, and pregnancy outcome.* PLoS Genet, 2019. **15**(8): p. e1008236.
10. Saenen, N.D., et al., *Air pollution-induced placental alterations: an interplay of oxidative stress, epigenetics, and the aging phenotype?* Clin Epigenetics, 2019. **11**(1): p. 124.
11. Lapehn, S. and A.G. Paquette, *The Placental Epigenome as a Molecular Link Between Prenatal Exposures and Fetal Health Outcomes Through the DOHaD Hypothesis.* Curr Environ Health Rep, 2022. **9**(3): p. 490-501.
12. Jedynak, P., et al., *Pregnancy exposure to synthetic phenols and placental DNA methylation - An epigenome-wide association study in male infants from the EDEN cohort.* Environ Pollut, 2021. **290**: p. 118024.
13. Everson, T.M., et al., *Placental DNA methylation signatures of maternal smoking during pregnancy and potential impacts on fetal growth.* Nat Commun, 2021. **12**(1): p. 5095.
14. Smith, Z.D. and A. Meissner, *DNA methylation: roles in mammalian development.* Nat Rev Genet, 2013. **14**(3): p. 204-20.
15. Baubec, T. and D. Schubeler, *Genomic patterns and context specific interpretation of DNA methylation.* Curr Opin Genet Dev, 2014. **25**: p. 85-92.
16. Bachman, M., et al., *5-Hydroxymethylcytosine is a predominantly stable DNA modification.* Nat Chem, 2014. **6**(12): p. 1049-55.
17. Nestor, C.E., et al., *Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes.* Genome Res, 2012. **22**(3): p. 467-77.
18. Pfeifer, G.P. and P.E. Szabo, *Gene body profiles of 5-hydroxymethylcytosine: potential origin, function and use as a cancer biomarker.* Epigenomics, 2018. **10**(8): p. 1029-1032.

19. Green, B.B., et al., *Hydroxymethylation is uniquely distributed within term placenta, and is associated with gene expression.* FASEB J, 2016. **30**(8): p. 2874-84.
20. Banister, C.E., et al., *Infant growth restriction is associated with distinct patterns of DNA methylation in human placentas.* Epigenetics, 2011. **6**(7): p. 920-7.
21. Chen, P.Y., et al., *Prenatal Growth Patterns and Birthweight Are Associated With Differential DNA Methylation and Gene Expression of Cardiometabolic Risk Genes in Human Placentas: A Discovery-Based Approach.* Reprod Sci, 2018. **25**(4): p. 523-539.
22. Tekola-Ayele, F., et al., *DNA methylation loci in placenta associated with birthweight and expression of genes relevant for early development and adult diseases.* Clin Epigenetics, 2020. **12**(1): p. 78.
23. Gude, N.M., et al., *Growth and function of the normal human placenta.* Thromb Res, 2004. **114**(5-6): p. 397-407.
24. Maltepe, E. and S.J. Fisher, *Placenta: the forgotten organ.* Annu Rev Cell Dev Biol, 2015. **31**: p. 523-52.
25. Woods, L., V. Perez-Garcia, and M. Hemberger, *Regulation of Placental Development and Its Impact on Fetal Growth-New Insights From Mouse Models.* Front Endocrinol (Lausanne), 2018. **9**: p. 570.
26. Audette, M.C. and J.C. Kingdom, *Screening for fetal growth restriction and placental insufficiency.* Semin Fetal Neonatal Med, 2018. **23**(2): p. 119-125.
27. Yuan, V., et al., *Cell-specific characterization of the placental methylome.* BMC Genomics, 2021. **22**(1): p. 6.
28. Campbell, K.A., et al., *Placental cell type deconvolution reveals that cell proportions drive preeclampsia gene expression differences.* Commun Biol, 2023. **6**(1): p. 264.
29. Yuan, V., et al., *Accurate ethnicity prediction from placental DNA methylation data.* Epigenetics Chromatin, 2019. **12**(1): p. 51.
30. Pique-Regi, R., et al., *Single cell transcriptional signatures of the human placenta in term and preterm parturition.* Elife, 2019. **8**.
31. Lee, Y., et al., *Placental epigenetic clocks: estimating gestational age using placental DNA methylation levels.* Aging (Albany NY), 2019. **11**(12): p. 4238-4253.