

Perinatal Epigenome Associated with Birth Outcomes and Exposure

by

Hwa Jin Lee

A dissertation submitted to Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

November 14, 2013

© 2013 Hwa Jin Lee
All Rights Reserved

Abstract

Birth outcomes such as gestational age, or environmental exposures like mercury, serve either as one of the strongest predictors for neonatal, adolescent, and adult morbidity and mortality or associated with common diseases such as cancer, cardiovascular disease, and neurological disorders through unknown mechanisms. Identification of genomic loci undergoing epigenetic changes, specifically DNA methylation, would increase our understanding of these unknown mechanisms. To address this, we performed CHARM 2.0, a genome-wide array-based analysis of DNA methylation, in 141 newborns collected in Baltimore, Maryland using bump-hunting based novel statistical methodology to identify genomic regions associated with gestational age. Through this analysis, we identified three DMRs at genome-wide significance levels associated with gestational age near three genes (*NFIX*, *RAPGEF2* and *MSRB3*) and one DMR commonly associated with total and methyl mercury exposure (*TCEANC2*). All of the three regions associated with gestational age were validated, and the region associated with both mercury exposure types were replicated by bisulphite pyrosequencing. Of the genes near or containing the DMRs, *RAPGEF2* and *TCEANC2* gene showed an inverse correlation between DNA methylation level and its expression level. For all of the gestational age DMRs, the DNA methylation levels at these regions appear similar or more extreme than those of the latest gestational ages in a heterogeneous population of adults. Together, the existence of gestational age DMRs suggests that epigenetic changes can occur not only during embryogenesis, but also during later stages of gestation. Also, The existence of mercury DMRs raise the

possibility that environmental exposures, particularly heavy metals during pregnancy, would serve as inducing/mediating factors for epigenetic changes in neonates.

Readers:

Margaret Daniele Fallin, Ph.D

Andrew P. Feinberg, M.D., M.P.H.

Thesis Advisor:

Andrew P. Feinberg, M.D., M.P.H.

Acknowledgments

During the journey of my graduate school years, there are many people who helped me go through the process with precious advice, guidance, support and comments. Although it's impossible to mention everybody here, I would like to show my best kind to recognize people who made this work possible.

First, and foremost, I would like to thank Dr. Andrew P. Feinberg as a great thesis advisor and mentor. After joining the lab, he provided wonderful research opportunities and encouraged me to fully commit on learning how to conduct rigorous, fun and rightful science. Also, he put his efforts on training me as an independent scientist, including how to read and write scientific literatures, and in particular, how to present scientific work. Without Dr. Feinberg's guidance, training, and care, I would never have been able to grow as a trained scientist and graduate before my military deadline. I will always remember his guidance and advice and in my future career.

I would also like to acknowledge Dr. M. Daniele Fallin and Dr. Lynn R. Goldman as closely working collaborators, mentors for publication and grants, and as thesis committee members, particularly Dr. Fallin as a committee chair and thesis reader. To be frank, epidemiology was a new field to me, and I would have never completed my thesis work without their gracious support and academic guidance.

In addition, I would like to thank Dr. Sean Taverna and Dr. Michael Wolfgang as

thesis committee member. Their scientific suggestions helped developing my thesis as a more complete piece of work.

I would also like to thank Dr. Andrew Jaffe, who closely worked with me as a biostatistician and co-first author for identification of gestational age DMRs. I enjoyed discussing about our work, and will remember his enthusiasm towards science.

Plus, I would like to thank all of our current and former lab members for all the helps they provided me to be established in the lab, and open to both scientific and non-scientific discussions. Especially, I thank Dr. Kelly Bakulski, Carolina Montano, and Jason Feinberg for their contribution to this thesis work.

I would also want to thank Colleen Graham, Leslie Lichter, Dr. Rajini Rao, my Cellular and Molecular Medicine program colleagues, and the basketball members for their support, kindness, and friendship.

I would like to thank all of the members who contributed to coordinate and establish all the epigenetic and epidemiological data in THREE and NCS Vanguard study. Without their work and commitment, this thesis work would never have even started.

I would like to thank my parents, family members and friends, in particular Luke Hwang, who always supported me with all of their heart and soul. After I born as a pre-term and low birthweight with various health issues, their unconditional love and support enabled me to become a healthy scientist.

Also, I would like to thank Samsung Scholarship for the funding support throughout my graduate school and the human networks among the scholarship recipients.

Lastly, I thank God that I am graduating before the military deadline. I will remember everybody's support during the journey of graduate school.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iv
Table of Contents.....	vii
List of Tables.....	viii
List of Figures.....	x
Chapter 1: Background.....	1
Chapter 2: Introduction.....	10
Chapter 3: Materials and Methods.....	17
1. Study Samples.....	18
2. Laboratory Analyses.....	20
a. CHARM DNA methylation.....	20
b. Infinium HumanMethylation450 assay.....	21
c. Bisulphite Pyrosequencing.....	22
d. Quantitative real-time PCR.....	23
3. Statistical Analyses.....	24
Chapter 4: Results – Identification of Gestational Age DMRs.....	37
Chapter 5: Results – Identification of mercury associated DMRs.....	57
Chapter 6: Discussion.....	76
Curriculum Vitae.....	88

List of Tables

Table 3.1 : Bisulfite Sequencing Primers.....	28
Table 3.2 : Real-time PCR Primers for Expression Analyses.....	34
Table 4.1 : Characteristics of THREE study newborns included in this epigenetics project.....	49
Table 4.2 : Top 30 list of DMRs associated with gestational age at birth identified via CHARM 2.0.....	50
Table 4.3 : Candidate significant DMRs associated with gestational age identified via CHARM 2.0.....	52
Table 4.4 : Co-efficient (95% CIs) of linear relationship between potential confounders and gestational age at birth or average methylation at each of the identified DMRs.....	53
Table 4.5 : Comparison of regression coefficients [95% CI] for relationship between methylation and gestational age with and without adjustment for potential confounders.....	54

Table 4.6 : Results for univariate and multivariate regression analyses of methylation on birthweight and/or gestational age.....	55
Table 5.1 : Characteristics of THREE and NCS Vanguard study newborns included in this project.....	69
Table 5.2 : Distribution of total/methyl mercury from THREE and NCS Vanguard study newborns included in this project.....	70
Table 5.3 : Candidate significant DMRs associated with total (a) and methyl (b) mercury exposure in THREE study.....	71
Table 5.4 : Linear relationship coefficient [95%CI] between potential confounders and total (a) and methyl (b) mercury exposure or average methylation at DMR inside <i>TCEANC2</i> gene.....	72
Table 5.5 : Comparison of regression coefficients [95% CI] for association between DNA methylation and total (a) and methyl (b) mercury exposure with and without adjustment for race.....	73
Table 5.6 : Correlation between estimated blood cell counts and total mercury exposure in NCS Vanguard study.....	74

List of Figures

Figure 4.1 : Distribution of gestational age at birth among 141 newborns in the THREE Study.....	44
Figure 4.2 : Methylation plots for three identified DMRs for gestational age at birth.....	45
Figure 4.3 : Bisulfite pyrosequencing results for each DMR.....	46
Figure 4.4 : Methylation plots for three identified DMRs for gestational age at birth with adult methylation results included.....	47
Figure 4.5 : Correlations between gene expression and DNA methylation for each DMR and its nearest gene.....	48
Figure 5.1 : Methylation plots for DMR inside <i>TCEANC2</i> associated with methyl and total mercury exposure.....	63
Figure 5.2 : Methylation plots for other identified DMRs associated with total mercury exposure.....	64
Figure 5.3 : Bisulfite pyrosequencing results for DMR inside <i>TCEANC2</i>	65
Figure 5.4 : Bisulfite pyrosequencing results for DMR inside <i>ANGPT2</i> , <i>PRPF18</i> and near <i>FOXD2</i>	66

Figure 5.5 : DNA methylation levels at four Infinium HumanMethylation450

probe sets located near/inside the DMR inside *TCEANC2*.....67

Figure 5.6 : Correlation between DNA methylation level at *TCEANC2* DMR and

TCEANC2 gene expression level.....68

Chapter 1: Background

Epigenetics is defined as the study of heritable marks other than the primary DNA sequence itself. Although DNA methylation is the only epigenetic mark proven to meet the traditional definition as a heritable mark through cell division, epigenetics now cover far more marks due to the complex nature and interactions between different marks including DNA methylation. Epigenetic marks now cover DNA cytosine modifications (hydroxylation, formylation and carboxylation), histone post-translational modifications (methylation, acetylation, phosphorylation, ubiquitination, sumonylation, crotonylation), and histone variants (H3.3, H2AZ, macroH2A, γ H2AX). Together, these marks determine the chromatin environment inside nucleus such as nucleosome occupancy, high order chromatin structures like Large Organized Chromatin K modifications (LOCKS) (Wen et al., 2009), hypomethylated blocks (Hansen et al., 2011), Partially Methylated Domains (PMDs) (Lister et al., 2009) and Lamin Associated Domains (LADs) (Guelen et al., 2008). Although not an epigenetic mark itself, non-coding RNAs such as long intergenic non coding RNAs (lincRNAs) and enhancer RNAs (eRNAs) in mammalian system are included due to their role of mediating changes in epigenetic marks and chromatin structure (Mousavi et al., 2013; Tsai et al., 2010). Collectively, epigenetic marks, chromatin environment, and non-coding RNAs are important in both establishment of cellular identity and in disease context, such as cancer or common diseases.

One of the epigenetic marks is DNA methylation, which occurs at the fifth carbon position of the DNA cytosine nucleotide in vertebrates. Generally, DNA methylation occurs in the context of CpG dinucleotides, but non-CpG methylation has also been observed in stem cells (Lister et al., 2009; Lister et al., 2011; Ramsahoye et al., 2000).

Methylation of DNA is catalyzed by DNA methyltransferases (DNMTs) by transferring methyl group from S-Adenosyl Methionine (SAM) to cytosine. Of the DNMTs, DNMT1 serves as a maintenance methyltransferase by methylating hemimethylated DNA during the DNA replication process. DNMT3A and B are *de novo* methyltransferases. Demethylation of DNA can occur passively by not replicating methylation patterns to daughter strand, or actively through combination of base excision repair (BER), nucleotide excision repair (NER) pathway and Ten Eleven Translocation (TET) enzymes (Wu and Zhang, 2010). Specifically, TET enzymes catalyze the oxidation of 5-methylcytosine to 5-hydroxymethylcytosine (Tahiliani et al., 2009), and further catalyze to 5-formylcytosine and 5-carboxycytosine, and finally excised by Thymine-DNA Glycosylase (TDG) (He et al., 2011; Ito et al., 2011). These demethylation intermediate marks were observed in many tissue types, but most abundantly in early embryos, embryonic stem cells, primordial germ cells, Purkinje neurons, and hippocampal dentate gyrus (Guo et al., 2011; Kriaucionis and Heintz, 2009).

One of the recent technological developments relevant for DNA methylation research is the establishment of genome-scale DNA methylation assay, either utilizing array-based or sequencing based systems. One of the array based genome-scale methylation technology used in this thesis study and other published studies is called Comprehensive High-throughput Array-based Relative Methylation (CHARM) assay, which utilized MspI, a methyl-sensitive restriction enzyme. The original CHARM array design (custom designed NimbleGen HD2) interrogates 4.6 million CpGs across the genome, and covers almost all of the CpG islands, promoters, and non-repetitive lower CpG dense regions (Irizarry et al., 2008; Ladd-Acosta et al., 2010). Through this unique

array system, investigators not only found methylation signature patterns across different cell and tissue types but also identified new discoveries which could not be found from using CpG-island focused arrays, including the finding that most methylation changes observed between different tissues (brain, liver and spleen), colon cancer, hematopoiesis and iPSC system are not at CpG islands but at CpG island shores, defined as sequences up to 2kb distant from CpG islands (Doi et al., 2009; Irizarry et al., 2009; Ji et al., 2010). Other array based genome-scale assays such as Infinium HumanMethylation450 BeadChip assay are also widely used in various DNA methylation research areas. One of the sequencing based DNA methylation assays called Whole Genome Bisulfite Sequencing is by far the most comprehensive genome-scale DNA methylation assay. This assay is based on treating bisulfite on genomic DNA, which converts unmethylated cytosine to uracil, whereas methylated cytosine remains unconverted (Krueger et al., 2012). Thus, bisulfite treatment creates sequence differences based on cytosine methylation status of DNA. After shearing genomic DNA, bisulfite treatment, the prepared sequence library is subjected to next generation sequencer. This approach enabled us to create genome-wide, single-base resolution DNA methylation map on various organisms including *Arabidopsis thaliana* (Cokus et al., 2008; Lister et al., 2008), and in human system (Lister et al., 2013; Lister et al., 2009; Lister et al., 2011). In addition, this technology directed us to new discoveries such as identification of non-CpG methylation and large methylation structures (PMDs and blocks), which blocks found to display stochastic variation in several types of tumor tissues with large-scale hypomethylation (Hansen et al., 2013; Hansen et al., 2011).

One of the rising epigenetic research fields along with the described technological advancements in genome-scale DNA methylation assay is epigenetic epidemiology, or Epigenome-Wide Association Studies (EWAS). These population-level genome scale studies are important since it's possible that epigenetic marks influence disease phenotypes by affecting the expression of target genes independent of or interacting with any sequence variation within or nearby the gene and the environmental factors (Bjornsson et al., 2004). EWAS started to replace traditional approaches to conduct epigenetic assays targeting selected candidate genes previously known to be involved in diseases, and they hold great promise for systematically dissecting out the role of epigenetic variation in health and disease (Michels et al., 2013). One of the several challenges for EWAS is to establish statistical methodology to properly deal with large-scale epigenetic data combined with epidemiological database. Although recent EWAS adopted GWAS-like approach to identify association between the phenotype of interest and DNA methylation at individual CpG level, more advanced strategies to analyze DNA methylation data by identification of Differentially Methylated Regions (DMRs) can be more productive (Bock, 2012). Recent work suggested novel statistical analysis tool called bump-hunting in order to identify DMRs based on techniques that borrows statistical power from adjacent CpGs to produce estimates that are substantially more precise than methods focusing on individual CpGs for EWAS (Jaffe et al., 2012b).

In the subsequent chapters, two perinatal EWAS to identify DNA methylation differences associated with birth outcomes and environmental exposure, particularly gestational age and mercury exposure levels are described. Experimentally, CHARM 2.0, an upgrade version of CHARM assay, was used to obtain genome-scale methylation data,

which further described in Laboratory Analyses part in Materials and Methods chapter. For biostatistical analysis, bump-hunting algorithm was used to identify DMRs associated with our variables of interest, which further described in Statistical Analyses part in Materials and Methods chapter.

References

- Bjornsson, H.T., Fallin, M.D., and Feinberg, A.P. (2004). An integrated epigenetic and genetic approach to common human disease. *Trends in genetics : TIG* 20, 350-358.
- Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nature reviews. Genetics* 13, 705-719.
- Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215-219.
- Doi, A., Park, I.H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., *et al.* (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature genetics* 41, 1350-1353.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., *et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948-951.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423-434.
- Hansen, K.D., Sabunciyan, S., Langmead, B., Nagy, N., Curley, R., Klein, G., Klein, E., Salamon, D., and Feinberg, A.P. (2013). Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome research*.
- Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., *et al.* (2011). Increased methylation variation in epigenetic domains across cancer types. *Nature genetics* 43, 768-775.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., *et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303-1307.
- Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddloh, J.A., Wen, B., and Feinberg, A.P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome research* 18, 780-790.
- Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., *et al.* (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics* 41, 178-186.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300-1303.

Jaffe, A.E., Murakami, P., Lee, H., Leek, J.T., Fallin, M.D., Feinberg, A.P., and Irizarry, R.A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *International journal of epidemiology* 41, 200-209.

Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., *et al.* (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 467, 338-342.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929-930.

Krueger, F., Kreck, B., Franke, A., and Andrews, S.R. (2012). DNA methylome analysis using short bisulfite sequencing data. *Nature methods* 9, 145-151.

Ladd-Acosta, C., Aryee, M.J., Ordway, J.M., and Feinberg, A.P. (2010). Comprehensive high-throughput arrays for relative methylation (CHARM). *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.] Chapter 20*, Unit 20 21 21-19.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., *et al.* (2013). Global epigenomic reconfiguration during mammalian brain development. *Science* 341, 1237905.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133, 523-536.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., *et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315-322.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., *et al.* (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471, 68-73.

Michels, K.B., Binder, A.M., Dedeurwaerder, S., Epstein, C.B., Grealley, J.M., Gut, I., Houseman, E.A., Izzi, B., Kelsey, K.T., Meissner, A., *et al.* (2013). Recommendations for the design and analysis of epigenome-wide association studies. *Nature methods* 10, 949-955.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Molecular cell* *51*, 606-617.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America* *97*, 5237-5242.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., *et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* *324*, 930-935.

Tsai, M.C., Manor, O., Wan, Y., Mosammamaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* *329*, 689-693.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature genetics* *41*, 246-250.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nature reviews. Molecular cell biology* *11*, 607-620.

Chapter 2: Introduction

Part of this chapter is reproduced from published article in International Journal of
Epidemiology (Lee et al. 2012)

Childhood disorders such as autism spectrum disorders (ASD) present lifetime challenges for children and their families with huge impact on public health. For example, ASDs may affect as many as 1 in 150 children with showing symptoms such as absent or delayed non-verbal communication, impaired social interactions, and rote or repetitive behaviors (Rice et al., 2007). The identification of causes and mechanisms for the progression of such disorders would have profound impact on the diagnosis, prevention, and possible treatment of these public health problems. The causes of childhood disorders could be environment (either *in utero* or post-birth), inherent genetics and epigenetic signatures such as DNA methylation. Thus, integration of the environmental exposure status, birth outcomes, genetics, and epigenetics of neonates would be the most comprehensive way to find the ultimate cause for the disorders.

Gestational age is the age of fetus *in utero*, and serves as one of the birth outcomes. Gestational age is the most important indicator of perinatal mortality in developed countries (Goldenberg et al., 2008), and also contributes to childhood and adult morbidity and mortality (Crump et al., 2011; Saigal and Doyle, 2008; Swamy et al., 2008). In 2005, approximately 13% of infants in the USA were born pre-term (<37 weeks), a rise from <10% in 1990 (Martin et al., 2007). The mechanism by which pre-term birth (PTB) increases morbidity and mortality is largely unknown. Recognition of specific genes that are still undergoing regulatory change prior to birth would not only increase our understanding of the developmental changes that are occurring during late pregnancy, but also it would aid in identifying genetic, epigenetic and environmental factors that could lead to PTB. The risks of negative public health consequences of PTB are many, including mortality, learning disabilities and respiratory illnesses (Beck et al.,

2010). Identification of epigenetic factors has the potential to prevent or ameliorate these adverse impacts.

Mercury, one of the toxic heavy metals, act as an environmental factor causing severe health consequences with symptoms such as sensory disturbance, visual field constriction, ataxia and deafness (Edwards, 1865). Of the several forms of inorganic and organic mercury, the major form of the mercury exposure in human is through methyl mercury (WHO, 2010). The known major sources of methyl mercury exposure are diet, particularly large fish and seafood, and socioeconomic status (Golding et al., 2013). The toxicity of high level methyl mercury exposure has been shown by multiple historical episodes such as Minamata disease and Iraq poison grain syndrome, affecting more on fetus due to *in utero* exposure with similar health consequences comparing to adults (Bakir et al., 1973; Social Scientific Study Group on Minamata Disease, 1999). Furthermore, a case from Swedish family with infants suffering from mental retardation and severe deficient in motor development without any effect on the mother emphasized the fetal effect of methyl mercury exposure *in utero* (Engleson and Herner, 1952). Not only high-level methyl mercury exposures show these fetal disorders, low-level methyl mercury exposure can also affect fetus *in utero*, showing association with brain function deficits (Grandjean et al., 1997). Identification of epigenetic factors associated with the exposure levels would reveal potential mechanisms inducing the adverse health outcomes and possibly prevent or ameliorate them.

From the point of view of a developmental change that is associated with health risk and environmental mediators, epigenetic changes in the fetus are potentially important, since epigenetic information affects gene expression, and its function varies

within an individual across developmental stages. A significant challenge in understanding the role of epigenetic changes in epidemiology is integrating novel molecular, epidemiological and biostatistical tools at a genome-scale level. Unlike classical genome sequence analyses, the methods and study designs for whole genome epigenetic epidemiology are not yet well established. The approach we have taken here is to design a genome-scale epidemiological analysis a priori from this joint conceptual perspective. We focused on DNA methylation because it is a key primary epigenetic process, with a well-established mechanism for propagating non-sequence-based information during cell division. The DNA methylation analysis presented here can serve as a paradigm for other epidemiological studies intending to characterize epigenetic profiles in specimen repositories, in which DNA methylation but not other epigenetic marks (e.g. histone modifications) are preserved. We have applied a significant technological extension of our previously described comprehensive high-throughput array-based relative methylation (CHARM) approach (Irizarry et al., 2008) that can now detect 5.2 million cytosine–guanine dinucleotide (CpG) sites which can be subject to DNA methylation. We also formally define an epigenetic variable, termed differentially methylated region (DMR), which we have used previously, but now have advanced its genome-wide detection to include novel statistical strategies to improve signal to noise detection, as well as the concept of regional methylation detection (Jaffe et al., 2012b).

While one would expect large-scale epigenetic changes to occur between early embryogenesis and the end of gestation, at present nothing is known about epigenetic changes in the fetus that occur relatively late in pregnancy, covering intervals relevant to the variation in gestational ages at birth that represent dramatic changes in health

outcomes. Epigenetic changes in placental samples across gestation have been observed, implying the importance of such modifications for support of a growing fetus (Novakovic et al., 2011), but genome-scale and site-specific methylation data on the fetus itself, and with respect to the late gestational ages associated with most births, have not yet been reported before our publication. Also, no genome-scale study has been yet performed to discover the association between metal exposures, particularly mercury. The integration of the findings from these two genome-scale studies and other studies would enhance our knowledge of how epigenetics would be involved in the birth outcomes and exposures, and direct possible changes in the medical policy for pregnant women. For these reasons, we performed a genome-scale comprehensive analysis of DNA methylation on 141 newborns to identify regions of the genome with DNA methylation levels correlated to gestational age at birth and mercury exposure. We then validated and replicated these microarray results via bisulphite sequencing. For gestational age DMRs, we further characterized the relationship between developmental age and DNA methylation at the DMRs by comparing these newborn results to the same regions among adult DNA samples.

References

- Bakir, F., Damluji, S.F., Amin-Zaki, L., Murtadha, M., Khalidi, A., al-Rawi, N.Y., Tikriti, S., Dahahir, H.I., Clarkson, T.W., Smith, J.C., *et al.* (1973). Methylmercury poisoning in Iraq. *Science* *181*, 230-241.
- Beck, S., Wojdyla, D., Say, L., Betran, A.P., Merialdi, M., Requejo, J.H., Rubens, C., Menon, R., and Van Look, P.F. (2010). The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. *Bulletin of the World Health Organization* *88*, 31-38.
- Crump, C., Sundquist, K., Sundquist, J., and Winkleby, M.A. (2011). Gestational age at birth and mortality in young adulthood. *JAMA : the journal of the American Medical Association* *306*, 1233-1240.
- Edwards, G. (1865). Two cases of poisoning by mercuric methide. *Saint Bartholomew's Hosp Rep*, 141-150.
- Engleson, G., and Herner, T. (1952). Alkyl mercury poisoning. *Acta paediatrica* *41*, 289-294.
- Goldenberg, R.L., Culhane, J.F., Iams, J.D., and Romero, R. (2008). Epidemiology and causes of preterm birth. *Lancet* *371*, 75-84.
- Golding, J., Steer, C.D., Hibbeln, J.R., Emmett, P.M., Lowery, T., and Jones, R. (2013). Dietary Predictors of Maternal Prenatal Blood Mercury Levels in the ALSPAC Birth Cohort Study. *Environmental health perspectives* *121*, 1214-1218.
- Grandjean, P., Weihe, P., White, R.F., Debes, F., Araki, S., Yokoyama, K., Murata, K., Sorensen, N., Dahl, R., and Jorgensen, P.J. (1997). Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. *Neurotoxicology and teratology* *19*, 417-428.
- Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddloh, J.A., Wen, B., and Feinberg, A.P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome research* *18*, 780-790.
- Jaffe, A.E., Murakami, P., Lee, H., Leek, J.T., Fallin, M.D., Feinberg, A.P., and Irizarry, R.A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *International journal of epidemiology* *41*, 200-209.
- Martin, J.A., Hamilton, B.E., Sutton, P.D., Ventura, S.J., Menacker, F., Kirmeyer, S., Munson, M.L., Centers for Disease, C., and Prevention National Center for Health Statistics National Vital Statistics, S. (2007). Births: final data for 2005. *National vital*

statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System 56, 1-103.

Novakovic, B., Yuen, R.K., Gordon, L., Penaherrera, M.S., Sharkey, A., Moffett, A., Craig, J.M., Robinson, W.P., and Saffery, R. (2011). Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors. *BMC genomics* 12, 529.

Saigal, S., and Doyle, L.W. (2008). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* 371, 261-269.

Social Scientific Study Group on Minamata Disease (1999). In the Hope of Avoiding Repetition of Tragedy of Minamata Disease. Minamata, Japan:National Institute for Minamata Disease. Available <http://www.nimd.go.jp/syakai/webversion/SSSGMDreport.html> [accessed 8 July 2010].

Swamy, G.K., Ostbye, T., and Skjaerven, R. (2008). Association of preterm birth with long-term survival, reproduction, and next-generation preterm birth. *JAMA : the journal of the American Medical Association* 299, 1429-1436.

WHO (2010). Children's Exposure to Mercury Compounds.

Chapter 3: Materials and Methods

Part of this chapter is reproduced from published article in International Journal of
Epidemiology (Lee et al. 2012)

1. Study Samples

Cord blood clot samples were obtained from the Baltimore Tracking Health Related to Environmental Exposures [THREE] Study (Apelberg et al., 2007). THREE is a cross-sectional sample of newborns born at the Johns Hopkins Hospital in Baltimore, MD, between November 2004 and March 2005. Of the 603 children delivered during that time window, 300 were eligible (24 twin births removed, 291 did not have any or ample cord blood available). Of these, 187 contributed a cord blood clot from which DNA could be isolated for the epigenetic project. Clots were saved during the second half of the data collection period.

Cord blood buffy coat samples and maternal first and third trimester blood buffy coat samples were obtained from the Vanguard Study, which is a pilot study of National Children's Study [NCS]. Briefly, a total of 384 (147 maternal first and third trimester period peripheral blood, 90 cord blood) buffy coat samples were collected from seven different locations across U.S (Duplin County, NC; Queens County, NY; Orange County, CA; Waukesha County, WI; Salt Lake City, UT; Montgomery County, PA; composite location of four adjacent counties in South Dakota and Minnesota). Of those, 147 maternal first trimester period peripheral blood buffy coat samples and 90 cord blood buffy coat samples were subjected to DNA isolation and Infinium HumanMethylation450 assay (Illumina). 85 cord blood buffy coat samples were subjected to bisulfite pyrosequencing assay.

For both of the birth cohorts, similar distribution for total and methyl mercury exposure, gestational age, race, birthweight, maternal age, smoking status, body mass

index and n-3 fatty acid level were shown between the population subjected to this study and the rest of the study population. Study personnel abstracted data from maternal and infant medical records and study clinicians reviewed a 10% random sample for accuracy; gestational age was taken as the best obstetrical estimate. Information on potential confounders was based on clinical records. Women who reported smoking during pregnancy or had an umbilical cord serum cotinine measurement > 10 ng/ml were considered active smokers; the remainder were considered passive smokers or non-smokers (not reporting smoking and cotinine < 1 ng/ml) (Bernert et al., 1997). Copper (previously found to be associated with gestational age in this population) and selenium were measured in umbilical cord serum using inductively coupled plasma dynamic reaction cell mass spectrometry (ICP-DRC-MS) (CDC, 2008) at Centers for Disease Control and Prevention (CDC) laboratories, with 4 mg/dl as the limit of detection. Total and methyl mercury and lead levels were measured by high performance liquid chromatography linked with inductively coupled plasma mass spectrometry (HPLC-ICP-MS). For analysis of n-3 fatty acid levels, cord serum samples were transferred to United States National Institute of Alcohol Abuse and Alcoholism (NIAAA) via automated fast gas chromatography (Wells et al., 2011). The THREE study was reviewed and approved by the Johns Hopkins School of Medicine Institutional Review Board, and the NCS Vanguard Study samples were collected under the authority of the NIH and study center IRBs.

For comparison of newborn methylation results with adult samples, CHARM 2.0 data were available on 156 adult samples obtained as unrelated controls for a schizophrenia case-control epigenetics consortium (Aliyu et al., 2006; Calkins et al.,

2007; Gur et al., 2007). This sample was 40% male and had a broad age range of between 16 and 89 years (interquartile range 31–55 years). DNA was obtained from the Rutgers University Cell & DNA Repository (RUCDR). DNA had been isolated from whole blood using Qiagen Autopure LS and pellets were hydrated in 1% Tris-EDTA (TE) buffer. Sample concentration and integrity were verified locally using NanoDrop and gel electrophoresis. DNA methylation was measured using the CHARM 2.0 assay.

2. Laboratory Analyses

a. CHARM DNA methylation

DNA was isolated from cord blood clot samples using the DNeasy[®] Blood & Tissue kit (Qiagen), following the manufacturer's instructions. From the 187 fetal cord blood clot samples available, 167 (89.3%) yielded enough DNA for methylation array analysis. DNA methylation was measured via the CHARM 2.0 assay, a customized microarray method extended from our previous CHARM procedure, a genomescale microarray technique for DNA methylation that identifies differential DNA methylation without assumptions regarding where such changes would be, interrogating all CpG islands, as well as CpG island 'shores' (Irizarry et al., 2009). CHARM 2.0 now includes 2.1 million probes, which cover 5.2 million CpGs arranged into probe groups (where consecutive probes are within 300 bp of each other) that tile regions of at least moderate CpG density. It includes all annotated and non-annotated promoters and microRNA sites on top of the features that are present in the original CHARM method. The design

specifications are freely available on our website (rafalab.jhu.edu). For the CHARM 2.0 assay, 5 mg of purified genomic DNA was sheared, digested, purified, amplified, labeled as described (Ladd-Acosta et al., 2010), but hybridized onto our new CHARM 2.0 array. We dropped 26 arrays with <80% of their probes above background intensities, resulting in 141 samples for DNA methylation analysis. We then filtered probes where signal was below background in <25% of arrays (542,055) and removed sex chromosomes (39,454) to improve the batch correction methods, leaving 1,569,888 autosomal probes covering 4,254,946 CpGs spread across 114,984 probe groups. Subsequent pre-processing, normalization and correction for batch effects are described in the Statistical Methods subsection. CHARM hybridization and processing for these samples were performed across 5 separate days, with the following numbers of samples per day: 40, 36, 38, 21, 6, reflecting a potential source of batch effects that was addressed through the surrogate variable analysis (SVA) described in the Statistical Methods subsection.

b. Infinium HumanMethylation450 assay

For estimating the blood cell type distribution, we performed Infinium HumanMethylation450 assay on 147 maternal first trimester period peripheral blood buffy coat samples and 90 cord blood buffy coat samples from NCS Vanguard study. Isolation of DNA from NCS Vanguard study samples was performed using Agencourt Genefind v2 (Beckman Coulter) on the Biomek NXp laboratory automation workstation according to manufacturer's instruction. Genomic DNA from each sample was quantified via Picogreen (Invitrogen). Agarose gel electrophoresis (1% agarose) was performed to

check the quality of extracted DNA. Genomic DNA was normalized/aliquotted to 1ug per each sample and sent to Center for Inherited Disease Research (CIDR) for further bisulfite conversion of genomic DNA and subjected to Infinium HumanMethylation450 (Illumina) assay.

c. Bisulphite Pyrosequencing

Individual CpGs inside the DMRs meeting our significance threshold were chosen for validation based on MethPrimer software (Li and Dahiya, 2002). Of the 141 samples for which CHARM data were generated, 139 had ample DNA for subsequent pyrosequencing validation for DMRs associated with gestational age. For replication of mercury-associated DMRs, 85 (94%) out of 90 cord blood buffy coat samples from NCS Vanguard study, had enough DNA for pyrosequencing. Also, 35 DNA samples extracted from a subset of THREE study cord blood clot samples containing *TCEANC2* gene expression data underwent pyrosequencing inside *TCEANC2* DMR. Genomic DNA (200 ng) from each sample was bisulphite treated using an EZ DNA Methylation-Gold™ Kit (Zymo research) according to the manufacturer's instructions. Bisulphite-treated genomic DNA was PCR amplified using unbiased nested primers, and DNA methylation was subsequently assessed quantitatively by pyrosequencing using a PSQ HS96 (Biotage). Quantitative measurements (percentage methylation at each CpG) from the pyrosequencing results were determined using the Q-CpG methylation software (Biotage). Control titration standards of 0, 25, 50, 75 and 100% methylated samples were generated

using appropriate mixtures of Whole Genome Amplified (WGA) Human Genomic DNA: Male (Promega) using a REPLI-g Mini Kit (Qiagen) and SSI-treated WGA DNA. Primer sequences used for the bisulphite pyrosequencing reactions can be found in Table 3.1.

d. Quantitative real-time PCR

To examine the correlation between DNA methylation and gene expression in cord blood clots for each of the top three DMRs associated with gestational age and a DMR associated with both methyl and total mercury exposure, we performed real-time PCR assays. Primers were designed to determine the mRNA expression of the gene closest to each DMR. Since this analysis required isolation of mRNA from cord blood clots, we were only able to perform these expression analyses on a subset of newborns with cord blood clots available. For the genes near gestational age DMRs, there were 10 babies with gestational age at birth <35 weeks, 15 with gestational age at 40 weeks and 17 with gestational ages 541 weeks, with a total of 42 samples. For *TCEANC2* gene containing both methyl and total mercury DMR, we performed the assay on 42 samples and removed four babies based on the criteria of Ct standard deviation less than 0.25, which reflects maximum Ct difference between the technical replicates is within 1. Thus, a total of 38 babies with similar distribution of total and methyl mercury exposure levels comparing to the THREE study samples subjected to CHARM 2.0 analysis (methyl mercury exposure: 0.92 (0.62-1.34), total mercury exposure: 1.37 (0.85-1.68)). For isolation of RNA, fetal cord blood clot samples were treated with TRIzol (Invitrogen) and

RNA was purified using a PureLink™ RNA Mini Kit (Invitrogen) according to the manufacturer's instructions. cDNA was synthesized using a QuantiTect Reverse Transcription Kit (Qiagen) and random hexamers. Real-time PCR amplification was performed by using a Fast SYBR_ Green Master Mix (Applied Biosystems), and transcript levels were quantified using an ABI 7900 Sequence Detection Systems (Applied Biosystems). Relative expression level for each gene was calculated based on the standard curve and normalized by the relative expression of beta-actin. Primer sequences used for the real-time PCR reactions are in Table 3.2.

3. Statistical Analyses

Descriptive statistics (median and interquartile range (IQR)) of THREE study and NCV Vanguard study samples were calculated for variables such as sex, maternal age, maternal race, gestational age, birthweight, total and methyl mercury levels, and n-3 fatty acids. Also, descriptive statistics (median or percentage) for gestational age at birth and potential confounders were calculated and compared using chi-squared tests for categorical variables and Mann–Whitney U-tests for continuous variables.

The CHARM microarray data were pre-processed and normalized as previously described (Aryee et al., 2011; Jaffe et al., 2012a). We employed a novel statistical approach for identifying regions of the epigenome associated with gestational age in days. Briefly, we fit a linear model predicting methylation at each probe as a function of either gestational age at birth or natural log-transformed methyl/total mercury exposure levels, adjusted for surrogate variables estimated via SVA (Leek and Storey, 2007) to account

for unmeasured potential confounding often due to batch effects. SVA identifies combinations of probes in the data associated with heterogeneity of DNA methylation, conditioned on the covariate of interest, in this case, gestational age or methyl/total mercury level, and then constructs a ‘surrogate variable’ for each set. A value for each individual based on each surrogate variable can then be used for adjustment in subsequent regression. Measured variables in our data set most associated with these surrogate variables (assessed through pruned regression trees of all possible variables) were array quality control score and hybridization date/batch. We did not adjust for sex, but did remove sex chromosome probes from the initial genome-wide screen. The estimated regression coefficients from these linear models for gestational age at each probe were then smoothed within the CHARM array’s pre-defined probe groups. Consecutive smoothed slopes above a fixed cut-off of either 99.5th percentile (for identification of gestational age DMRs) or 99.995th percentile (for identification of mercury level associated DMRs) of all smoothed slopes were summed into a region-level statistic reflecting the area of the DMR. We then ranked DMRs by their areas and calculated two measures of statistical uncertainty, a P-value and q-value, for each DMR by permutation that accounts for genome-wide testing. Either Gestational ages or methyl/total mercury exposure levels were permuted 1000 times, and each time, the above regression, smoothing, and thresholding procedure was repeated exactly as on the observed data to get 1000 sets of declared DMRs that occurred solely by chance. Empirical P-values, defined as the fraction of the maximum areas from each permutation greater than the observed area, were calculated (P_{\max}) to compare with a specified family-wise error rate control of 10% (for gestational age DMRs) or 20% (methyl and total

mercury DMRs). False discovery rate (FDR) q-values were obtained by pooling all areas across all permutations, calculating the proportion of these ‘null’ areas greater than the observed area, then converting this to a q-value for comparison to an FDR control of 5% (for gestational age DMRs) or 10% (methyl or total mercury DMRs) (Storey and Tibshirani, 2003). DMRs with an empirical $P_{\max} < 0.10/0.20$ or an FDR q-value $< 0.05/0.1$ were examined visually via plots of the methylation curve within the DMR. Average methylation for each newborn across all probes within a DMR was plotted against gestational age at birth with slopes and P-values estimated via linear regression and Wald statistics.

Univariate relationships between potential confounders and methylation at DMRs were also estimated via linear regression. Although some potential confounding due to these variables may already be addressed via the SVA adjustment, we also explicitly estimated relationships between average DNA methylation for each DMR and each confounder through linear models adjusted for the same surrogate variables used in our discovery. To do this, we applied SVA analysis to the methylation data first, and then took SVA-adjusted methylation as the methylation metric for linear regression with the covariate, to ensure the same SVA adjustment was applied in each analysis. Also, as a sensitivity analysis to assess the influence of sex on our list of identified DMRs, we repeated the original DMR identification procedure adjusting for sex. For gestational age DMRs, we further performed the original discovery procedure after omitting samples with mothers who had pregnancy-induced hypertension (PIH), intrapartum fever or diabetes, separately, to assess influence of these variables on our results. To assess the influence of race on *TCEANC2* DMR, we repeated the original analysis procedure after

adjusting for race after combining the African Americans and Asians as one race category.

For analysis of DNA methylation data from pyrosequencing, we fit a linear model at every CpG inside gestational age DMRs predicting DNA methylation as a function of gestational age. For methyl/total mercury DMRs, we fit a linear model at every CpGs predicting DNA methylation as a function of log-transformed total or methyl mercury exposure. We assessed the functional implications of differential methylation at *NFIX*, *RAPGEF2*, and *MSRB3* gene by fitting linear models at each CpG assessing the linear association between DNA methylation and gene expression. For *TCEANC2* gene, we used spearman's rank order correlation between DNA methylation and gene expression with respect to each CpG. Heavily skewed gene expression values were transformed to log₂ scale.

For assessing the relationship between blood cell type distribution and DNA methylation at *TCEANC2* DMR, we used a publicly available database containing DNA methylation signatures for 6 sorted blood cell types (Reinius et al., 2012). Four probe sets from the Reinius et al. Illumina 450k dataset (cg01109333, cg01986665, cg02270108, and cg02626873), which positioned inside *TCEANC2* DMR, were chosen. We plotted the DNA methylation levels at each of these probe set across the sorted blood cell types (B cells, CD4⁺ T cells, CD8⁺ T cells, granulocytes, monocytes, natural killer cells, and whole blood). For assessing the relationship between blood cell type distribution and total mercury exposure levels from NCS Vanguard study samples, we inferred relative cell type proportions in this population using an epigenetic signature prediction algorithm (Houseman et al., 2012).

Table 3.1 Bisulfite Sequencing Primers

Gene	Primer	Sequence (5' → 3')	Chr	CG1	CG2	CG3
NFI1	Forward	TTTTAATTTTTTGTGTTTGGGAAAG	19	13131545	13131565	13131573
	Reverse	AAAATAAAAACAACAACAATCCCAC				
	Nested_forward	TTTTTTTGTTAAGAGAGTTTTGAGG				
	Nested_reverse	/5Biosg/ATTAAAAAACAAAACAAAATACAC				
	Sequencing 1(F)	GTAGGTTTTTGAGGTTTTATTGAGA				
	Sequencing 2(F)	TTGATTTTTAATTTTTTTTTAGGAG				
MSRB3	Forward	AAATTTAAGTATTTTTGTTGTGAAAAATTAT	12	65671893	65671906	65671912
	Reverse	AACCAAAAACCTATAAAAAAAC				
	Nested_forward	TGTTTGGGTTTTATATATGGTGTTTAA				
	Nested_reverse	/5Biosg/AACAAAAACAAAACCTACCAATTAATTACTT				
	Sequencing 1(F)	TATTTATTTTTTTTTGTTAGAGAGG				
RAPGEF2	Forward	TTTTAAGAATATTGTTTTAAGTGTAAAGT	4	160026608	160026685	160026983
	Reverse	TTAATAAAAACAATAAACTACCTTCC				
	Nested_forward	TTTTTTTGGTTGTTTTTGGATAAGT				
	Nested_reverse	CCATAATCTCCCAAATATAACAACCTC				
	Forward2	TTTTTTTGTGTTGATTATTATTTAT				
	Reverse2	AAAATCTACTTTTCCTTCACTAAAAC				
	Nested_forward2	TTGTGGGGAGAGTATAATAAAATAGATTT				
	Nested_reverse2	/5Biosg/AAAACAAAATCAATACAAAACATTCCT				
	Sequencing 1(F)	GATAAGTTTTAAGAGTGGTATTTGGT				
	Sequencing 2(F)	TGTTTTTGGATAAGTTTTAAGAGTG				
	Sequencing 3(F)	TATTTGGGAGATTATGGATAGATTG				
TCEANC2	Forward	TTTGGGTTAGGTAGAGAGGAAAAT	1	54562272	54562295	54562346

	Reverse	ATCCTTAACATATTCACAATAAAAT				
	Nested_forward	TTGATGTTGGTTGTGTTAGATTTTT				
	Nested_reverse	/5biosg/CATCACCCCTTTCACCATATTAATAATA				
	Sequencing 1(F)	GTAGGAATGTGTTTTATTAGTTGTA				
	Sequencing 2(F)	GTTGGTTGTGTTAGATTTTTTTTATG				
ANGPT2	Forward1	AGAATTGTGTTGTTGTTTTTTGTGT	8	6418033	6418040	6418044
	Reverse1	AACTCCACACCTATTCTCCCAA				
	Nested_forward1	TTATAGGTGATAAAAATATAGGAGAAAAATA				
	Nested_reverse1	/5biosg/CCTAAAAAAAACAATTTACTCTCTC				
	Forward2	GTGTAAATTGTTTTTTTTAGGTGTA				
	Reverse2	ATAAACAAAACATCATATTTCTTTCTTAAAT				
	Nested_forward2	TTGGGAGAATAGGTGTGGAGTT				
	Nested_reverse2	/5biosg/TAAACAACAATAACACAATAATATTTCAA				
	Forward3	TAGTTTTGTATTTGGTTTAGTATTT				
	Reverse3	ATCACTATAATTTACTTTTAAACATCTTCT				
	Nested_forward3	TTTTTATGTTTTTTGGTTGTTTTTT				
	Nested_reverse3	/5biosg/TAATTAATAACTTTCCACATCCAC				
	Sequencing 1(F)	TTTGTTTTATTTTTAAGTTAGAAG				
	Sequencing 2(F)	GATAAATATTAAGTTATTTTTGGGG				
	Sequencing 3(F)	GTAGATTAGAATATTTATTGTAGTG				
PRPF18	Forward1	TTGTTGATTTTTTATTTTTGGGAATA	10	413684055	413684060	413684275
	Reverse1	CCAAAATACCATTTCAAAAAAACTC				
	Nested_forward1	TTTTTATAAATGTGGTTGTGTTT				
	Nested_reverse1	/5biosg/AATCTCCAACAAAATCATTTCACTAC				
	Forward2	TTATTATAATTTTATGGGTTTTTGAAATTA				
	Reverse2	CAAACAATAAATCTCCAACAAAATC				
	Nested_forward2	TAGGTTTTTTAGTTTTGTTGGAGTA				

	Nested_reverse2	/5biosg/CACACAACCACATTTATAAAAAAAA				
FOXD2	Forward1	TTTTTTGGGTATTTTTTTTATTTGTTT	1	48059802	48059934	48060162
	Reverse1	CACCCTCATCTTAAAACTAAAACTCA				
	Nested_forward1	GGAGTTTTTATGAAGGTTTAGGGTT				
	Nested_reverse1	/5biosg/TAAACAACAAAACAAAATTTACACCA				
	Forward2	GGTTTTTAAATTTATTTTTTATGTTTTTAT				
	Reverse2	ACCACTCCTTCTAACAATTA ACT				
	Nested_forward2	TTTTAAAAGTATTAGAGAGAGGTTGGAATA				
	Nested_reverse2	/5biosg/AAAAAACACCAAACAAAACCTTAC				
	Forward3	AAGGTTTTGTTTGGTGTTTTTTTTTA				
	Reverse3	CACCAACACTAAAAATCAAATTTCA				
	Nested_forward3	TTTTTAGTTGGGTAATGGGGTTAG				
	Nested_reverse3	/5biosg/CCA ACTCTTAAAAAATACCCACTTC				
	Sequencing 1(F)	GATAGAGATTGAGTAGTAGTGGAGG				
	Sequencing 2(F)	GAGTAAGATTAGTTATATATTTGGG				

Table 3.1 Bisulfite Sequencing Primers
continued

Gene	Primer	CG4	CG5	CG6	CG7	CG8	CG9	CG10
NFIX	Forward	13131612						
	Reverse							
	Nested_forward							
	Nested_reverse							
	Sequencing 1(F)							
	Sequencing 2(F)							
MSRB3	Forward	65671924	65671942					
	Reverse							
	Nested_forward							
	Nested_reverse							
	Sequencing 1(F)							
RAPGEF2	Forward							
	Reverse							
	Nested_forward							
	Nested_reverse							
	Forward2							
	Reverse2							
	Nested_forward2							
	Nested_reverse2							
	Sequencing 1(F)							
	Sequencing 2(F)							
	Sequencing 3(F)							
TCEANC2	Forward	54562368						

	Reverse							
	Nested_forward							
	Nested_reverse							
	Sequencing 1(F)							
	Sequencing 2(F)							
ANGPT2	Forward1	6418111	6418147	6418172	6418449	6418469	6418511	6418530
	Reverse1							
	Nested_forward1							
	Nested_reverse1							
	Forward2							
	Reverse2							
	Nested_forward2							
	Nested_reverse2							
	Forward3							
	Reverse3							
	Nested_forward3							
	Nested_reverse3							
	Sequencing 1(F)							
	Sequencing 2(F)							
	Sequencing 3(F)							
PRPF18	Forward1							
	Reverse1							
	Nested_forward1							
	Nested_reverse1							
	Forward2							
	Reverse2							
	Nested_forward2							

	Nested_reverse2		
FOXD2	Forward1	48060221	48060268
	Reverse1		
	Nested_forward1		
	Nested_reverse1		
	Forward2		
	Reverse2		
	Nested_forward2		
	Nested_reverse2		
	Forward3		
	Reverse3		
	Nested_forward3		
	Nested_reverse3		
	Sequencing 1(F)		
	Sequencing 2(F)		

Table 2. Real-time PCR Primers for Expression Analyses

Gene	Primer	Sequence (5' → 3')
NFIX	Forward	AGGAGATGCGGACATCAAA
	Reverse	TACTCTCACCAGCTCCGTCA
MSRB3	Forward	AGTAGCCCTTCGAGCCTGT
	Reverse	GTTAGCCGCTTCCTCAGTTC
RAPGEF2	Forward	CAGACAAAGCACATCCCAAC
	Reverse	TGGCAAGTCAGGAGTAGCAC
TCEANC2	Forward	CAGCTACCAGGGCTGCTT
	Reverse	GGGACTCCGACTACCTTGAC

References

- Aliyu, M.H., Calkins, M.E., Swanson, C.L., Jr., Lyons, P.D., Savage, R.M., May, R., Wiener, H., McLeod-Bryant, S., Nimgaonkar, V.L., Ragland, J.D., *et al.* (2006). Project among African-Americans to explore risks for schizophrenia (PAARTNERS): recruitment and assessment methods. *Schizophrenia research* 87, 32-44.
- Apelberg, B.J., Goldman, L.R., Calafat, A.M., Herbstman, J.B., Kuklennyik, Z., Heidler, J., Needham, L.L., Halden, R.U., and Witter, F.R. (2007). Determinants of fetal exposure to polyfluoroalkyl compounds in Baltimore, Maryland. *Environmental science & technology* 41, 3891-3897.
- Aryee, M.J., Wu, Z., Ladd-Acosta, C., Herb, B., Feinberg, A.P., Yegnasubramanian, S., and Irizarry, R.A. (2011). Accurate genome-scale percentage DNA methylation estimates from microarray data. *Biostatistics* 12, 197-210.
- Bernert, J.T., Jr., Turner, W.E., Pirkle, J.L., Sosnoff, C.S., Akins, J.R., Waldrep, M.K., Ann, Q., Covey, T.R., Whitfield, W.E., Gunter, E.W., *et al.* (1997). Development and validation of sensitive method for determination of serum cotinine in smokers and nonsmokers by liquid chromatography/atmospheric pressure ionization tandem mass spectrometry. *Clinical chemistry* 43, 2281-2291.
- Calkins, M.E., Dobie, D.J., Cadenhead, K.S., Olincy, A., Freedman, R., Green, M.F., Greenwood, T.A., Gur, R.E., Gur, R.C., Light, G.A., *et al.* (2007). The Consortium on the Genetics of Endophenotypes in Schizophrenia: model recruitment, assessment, and endophenotyping methods for a multisite collaboration. *Schizophrenia bulletin* 33, 33-48.
- CDC (2008). CfDCaP. Serum multi-element DLS method ITS005A. In: Division of Laboratory Science CfDCaP, editor. Atlanta, GA, 2008. pp. 1-63.
- Gur, R.E., Nimgaonkar, V.L., Almasy, L., Calkins, M.E., Ragland, J.D., Pogue-Geile, M.F., Kanes, S., Blangero, J., and Gur, R.C. (2007). Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. *The American journal of psychiatry* 164, 813-819.
- Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics* 13, 86.
- Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., *et al.* (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics* 41, 178-186.
- Jaffe, A.E., Feinberg, A.P., Irizarry, R.A., and Leek, J.T. (2012). Significance analysis and statistical dissection of variably methylated regions. *Biostatistics* 13, 166-178.

Ladd-Acosta, C., Aryee, M.J., Ordway, J.M., and Feinberg, A.P. (2010). Comprehensive high-throughput arrays for relative methylation (CHARM). *Current protocols in human genetics* / editorial board, Jonathan L. Haines ... [et al.] *Chapter 20*, Unit 20 21 21-19.

Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics* 3, 1724-1735.

Li, L.C., and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18, 1427-1431.

Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S.E., Greco, D., Soderhall, C., Scheynius, A., and Kere, J. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PloS one* 7, e41361.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* 100, 9440-9445.

Wells, E.M., Jarrett, J.M., Lin, Y.H., Caldwell, K.L., Hibbeln, J.R., Apelberg, B.J., Herbstman, J., Halden, R.U., Witter, F.R., and Goldman, L.R. (2011). Body burdens of mercury, lead, selenium and copper among Baltimore newborns. *Environmental research* 111, 411-417.

Chapter 4: Results –
Identification of gestational age DMRs

This chapter is reproduced from published article in International Journal of
Epidemiology (Lee et al. 2012)

To identify epigenetic changes that occur throughout later stages of gestation in an unselected population of newborns, we performed the CHARM 2.0 assay, which now includes approximately one-third of all single-copy CpG sites including all islands and shores, as well as all annotated promoters and microRNAs. Bisulphite pyrosequencing and real-time PCR were performed to validate DNA methylation levels and functional significance of the DMRs associated with gestational age at birth. Of the 141 newborns with CHARM data, there were 18 PTBs (<37 weeks) and the range of gestational ages in days was 208–292 (Figure 4.1 for full distribution). The pre-term newborns did not differ in the distributions of sex or maternal age, race or diabetes status compared with newborns born after 37 weeks (Table 4.1). Birthweight differed strongly between the two groups, as did smoking and serum copper levels, which had been previously reported for the full study sample of 300 newborns (Wells et al., 2011).

Previous research indicates that increasing gestational ages at birth through 39–41 weeks is advantageous for neurodevelopment (Davis et al., 2011; Yang et al., 2010) and confers a lower risk of respiratory morbidity (Hansen et al., 2008), suggesting the need to study gestational age on a continuum. Thus, treating gestational age as a continuous variable in linear regression, compared with pre-term and term birth categories can be useful. Using this approach, we identified 8611 candidate DMRs associated with gestational age at birth (top 30 showed in Table 4.2), of which the top three DMRs met our genome-wide threshold of protecting family-wise error rates <10% and false discovery rates <5% (Table 4.3). The first of these DMRs was found to be positioned in the first intron of the nuclear factor I/X (*NFIX*) gene, encoding a transcription factor known to be responsible for fetal-specific transcription regulation during skeletal muscle

development (Messina et al., 2010). Another was positioned in the first intron of the alternative transcript of the Rap guanine nucleotide exchange factor (*RAPGEF2*) gene, which encodes one of the RAS protein family activators that maintains the GTP-bound state of RAS. Although this DMR was not located at the promoter of the canonical gene, the DMR contains strong DNase I hypersensitive sites and a number of strong transcription factor-binding sites including Gata-2 and PU.1, which are the critical transcription factors in haematopoiesis (Ramirez et al., 2010; Vicente et al., 2012). The third DMR was located next to the promoter region of the methionine-Ssulphoxide reductase 3 (*MSRB3*) gene, which encodes the enzyme involved in the methionine cycle and is responsible for antioxidant repairing by converting methionine sulphoxide to methionine (Marchetti et al., 2005). Two of the three DMRs are located at the CpG island shore, suggesting that these DMRs may be associated with alternative transcription or splicing.

The methylation values at each probe for each of these DMRs are shown in Figure 4.2 according to gestational age in weeks (calculated from days). Smoothed lines indicate the average methylation curve for each week of gestational age at birth, and show a dose-response trend between gestational age and methylation levels across all weeks for each DMR. To further illustrate this point, Figure 4.2 also shows the relationship between the average methylation across all probes in the DMR and gestational age, and the linear fit of this relationship (see insets in each panel). For the DMR near *NFIX*, DNA methylation levels of each probe are greater in high gestational age neonates when compared with low gestational age neonates (Figure 4.2a), and the average DNA methylation level of each sample in the DMR exhibits a linear correlation with gestational age, with an estimated

increase of 1.57% DNA methylation per week of gestation [95% confidence interval (CI) 1.02–2.12], or an increase of 7.85 between Weeks 35 and 40, roughly corresponding to late pre-term vs term births ($P=8.6 \times 10^8$ for Wald statistic; see Figure 4.2a, inset). In contrast, the DMRs at *RAPGEF2* and *MSRB3* show lower DNA methylation levels of each probe in higher gestational age neonates when compared with lower gestational age neonates (Figure 4.2b and c), and the average DNA methylation levels of each sample in these DMRs exhibit inverse linear correlation with gestational age. For *RAPGEF2*, there is a 1.33 decrease in %DNA methylation (95% CI -1.76 to -0.9) per week of gestation or a decrease of 6.65 between Weeks 35 and 40; (Wald $P=9.9 \times 10^9$) and for *MSRB3*, a 2.08 decrease (95% CI -2.51 to -1.64) per week or 10.4 between Weeks 35 and 40 (Wald $P=1.3 \times 10^{16}$; see Figure 4.2b and c insets). Also note the progressive change in DNA methylation within each gestational age bin, a dose–response relationship consistent with a functional relationship between methylation and gestational age.

To validate these findings on a separate platform, we designed bisulphite pyrosequencing assays for CpGs within each DMR (indicated as red blocks in Figure 4.2). The individual CpG results within each DMR were correlated (average pair-wise correlation for neighboring CpG methylation was 0.85 for *NFIX*, 0.68 for *RAPGEF2* and 0.82 for *MSRB3*) and confirmed the CHARM differences in methylation by gestational age. For *NFIX*, four CpGs were assayed (see Figure 4.2 for locations), each showing an incremental increase in methylation with increase in gestational age at birth, consistent with the pattern detected in CHARM (Figure 4.3a). All three of the CpGs assayed in *RAPGEF2* (see Figure 4.2 for locations) showed greater methylation with early gestational age at birth, consistent with the CHARM results (Figure 4.3b). For *MSRB3*,

all five CpGs assayed showed greater methylation in earlier gestational age samples as seen in CHARM (Figure 4.3c). Thus, these DNA methylation analyses on an independent measurement platform confirmed the differential methylation by gestational age for each of the three genes identified via CHARM.

Since the three DMRs we identified reflect variability in methylation corresponding to late-stage development in utero, we considered whether adult DNA methylation at these same sites would show any variability and whether adult levels would be similar to those of full-term births. We compared CHARM 2.0 data for each DMR among healthy adult blood DNA samples with our newborn sample results. Although the three DMRs appear very dynamic and progressive with gestational age in the newborn sample, these exact same regions have little variability in the adult population. Given the span of adult ages represented, this suggests that these sites are stable in adulthood. The magnitude of adult DNA methylation levels is similar to or more extreme than those of the latest gestational ages in a direction consistent with the newborn sample correlations to gestational age (Figure 4.4). These results provide compelling support for maturation-related changes in DNA methylation at these loci, and also indicate that the process continues beyond birth, but reaches a maximum at some time at or before adult life.

To address potential confounding by sex, maternal age, race, maternal smoking, presence of PIH, intrapartum fever, maternal smoking and serum copper levels, we estimated the linear relationships between each of these variables and gestational age at birth. Consistent with the general characteristics comparing pre-term babies to the rest of the newborns, maternal smoking, PIH and serum copper were associated with gestational

age (Table 4.4). To further address whether these potential confounders were associated with methylation at the identified DMRs, we estimated the linear relationship between these variables and the average methylation value per DMR as well. PIH and serum copper were also associated with methylation at each of these DMRs (Table 4.4), suggesting the potential for confounding. However, the strong association between methylation and gestational age remained even after adjusting for PIH and copper in both CHARM and pyrosequencing data. For example, in the CHARM data, the coefficient for gestational age at birth in linear models predicting average methylation at each DMR with and without adjustment for copper (which had a stronger effect than PIH) changed from 1.57 to 1.37 for *NFIX*, -1.33 to -1.17 for *RAPGEF2* and -2.08 to -1.87 for *MSRB3*, and all remained statistically significant. We in fact examined the potential influence of each potential confounder on the detected associations with these three DMRs and saw no substantial change in effect sizes after adjustment for any of these covariates (Table 4.5).

Birthweight was also correlated with both gestational age and with methylation at each of the three DMRs. This is expected given the strong relationship between gestational age and birthweight. Gestational age is the best indicator of maturation of the newborn including growth parameters. Since birthweight is largely a consequence of gestational age, removing birthweight variability would almost completely restrict variability for gestational age in our analyses, so we did not condition on birthweight for these analyses. When we considered birthweight for gestational age as a separate phenotype, we saw no relationship to methylation at the three DMRs (Table 4.6).

To explore the functional significance of the differential methylation, we measured the expression of the *NFIX*, *RAPGEF2* and *MSRB3* using real-time PCR. *RAPGEF2* showed an inverse linear correlation between expression and DNA methylation levels in two of the three CpGs at this DMR (CpG1: $P=0.37$; CpG2: $P=0.013$, CpG3: $P=0.014$, Figure 4.5).



Figure 4.1 Distribution of gestational age at birth among 141 newborns in the THREE Study.

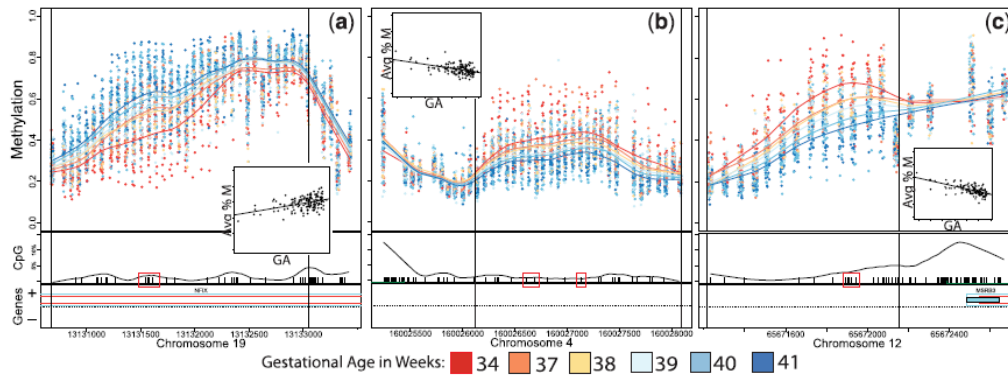


Figure 4.2 Methylation plots for three identified DMRs for gestational age at birth.

(a) *NFIX*, (b) *RAPGEF2*, (c) *MSRB3*. Top half of panels show individual methylation levels at each probe by genomic position, with coloured lines reflecting the average methylation curve for samples binned by gestational age—gestational ages in weeks were split into equal sized bins, and the average age for each bin is shown in the legend. Bottom half of panels show location of CpG dinucleotides (black tick marks) and CpGs validated by bisulphite pyrosequencing (black tick marks contained in red box) as well as the CpG density by position (black curve) and the location of refseq annotated genes (bar, + and - represent the direction of the gene, green bar indicates CpG island). Vertical lines represent boundaries of the DMR. Inset box: linear regression plot of average methylation across the DMR (Avg %M) per sample by gestational age (GA)

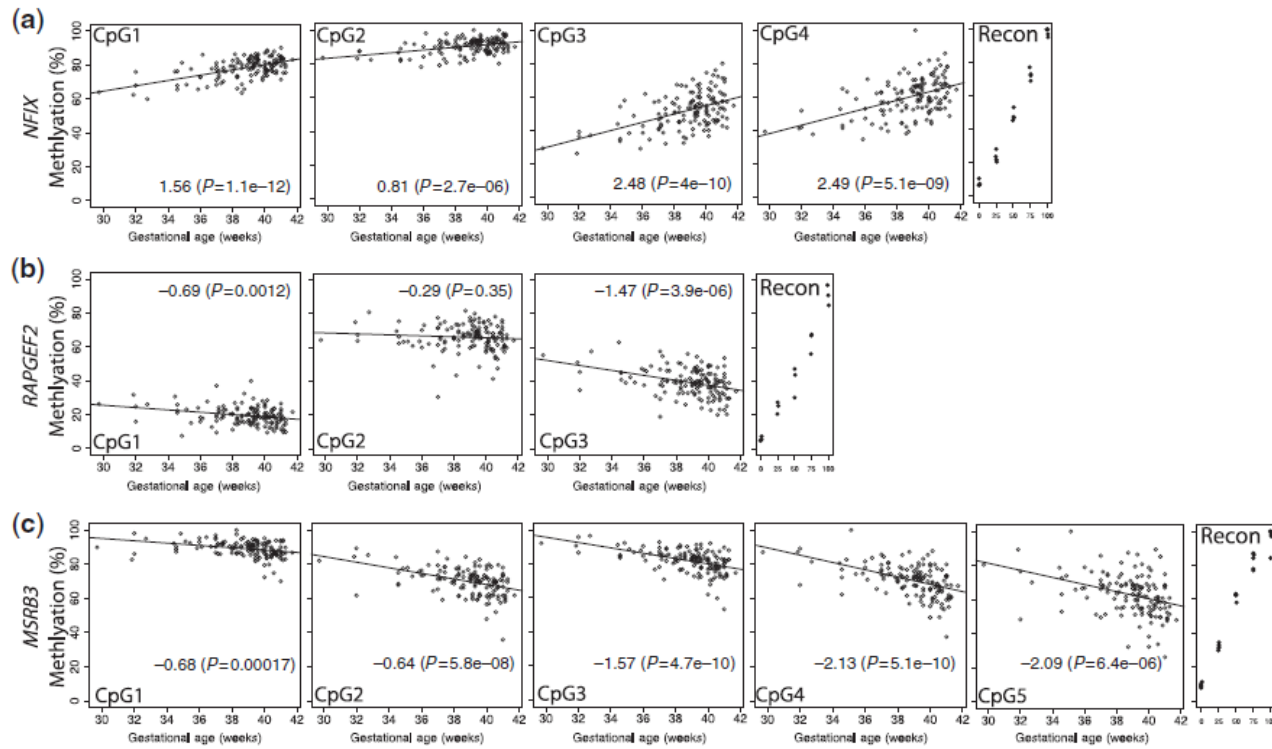


Figure 4.3 Bisulphite pyrosequencing results for each DMR.

(a) *NFIX*, (b) *RAPGEF2*, (c) *MSRB3*. Circles represent methylation values (y-axis) at individual CpGs for their corresponding gestational age in weeks (x-axis). Lines represent predicted values from linear regression. Reconstitution controls (represented as black dots) with explicitly designed % methylation (x-axis) are located at the right of each panel (Recon). The numbers on the bottom of each figure represent effect size/slope estimate from the regression of methylation on gestational age and P-value for a Wald test of this slope

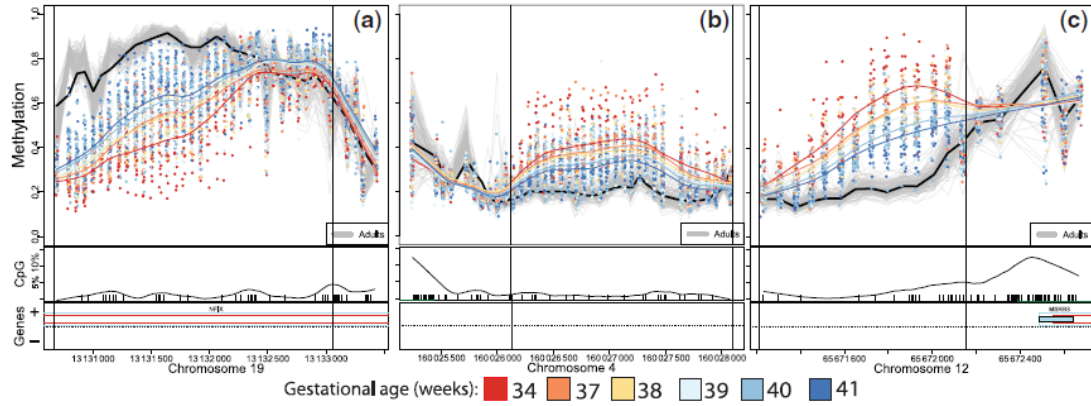


Figure 4.4 Methylation plots for three identified DMRs for gestational age at birth with adult methylation results included.

Individual adult methylation levels are represented as grey lines, and the black line represents mean adult methylation level. (a) NFIX, (b) RAPGEF2 and (c) MSRB3

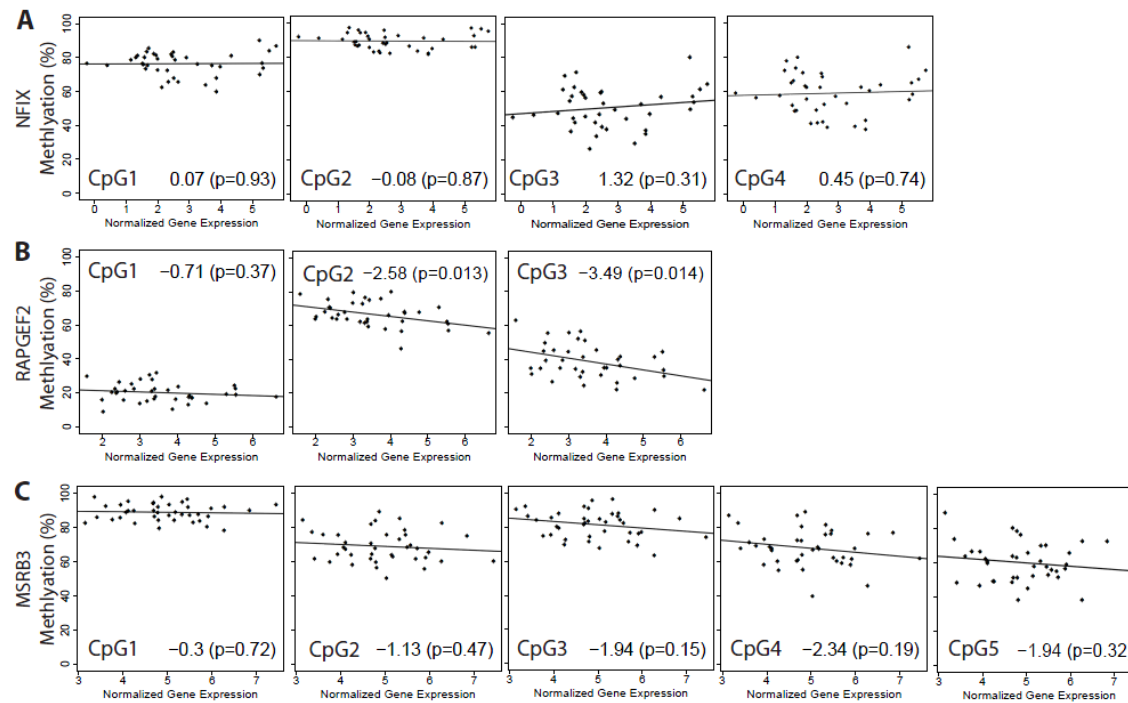


Figure 4.5 Correlations between gene expression and DNA methylation for each DMR and its nearest gene.

Panel A: *NFIX*, panel B: *RAPGEF2*, panel C: *MSRB3*. Black dots represent methylation values (y-axis) at individual CpGs and their corresponding normalized expression level (x-axis). Lines represent fit of a linear regression. The numbers on the bottom of each figure represent effect size/slope estimate from linear regression and p-value for a Wald statistic for this slope estimate

Table 4.1. Characteristics of THREE study newborns included in this epigenetics project

	Pre-term (<37 weeks), $N = 18$	Term/post-term (≥ 37 weeks), $N = 123$	<i>P</i> -value*
Male sex (%)	56	52	0.98
Maternal age, median (IQR) (years)	28 (22–30)	24 (20–29)	0.14
Maternal race (%)			0.32
Caucasian	33	21	
AA	67	73	
Asian	0	6	
Maternal smoking (%)			<0.01
Non-smoker	56	75	
Passive smoker	0	11	
Active smoker	44	14	
Birthweight, median (IQR) (g)	2422 (2102–2689)	3279 (2906–3648)	<0.01
Pregnancy-induced hypertension (%)	22	6	0.04
Intrapartum fever (%)	6	8	1.00
Serum copper, median (IQR) ($\mu\text{g}/\text{dl}$)	26 (22–34)	41 (30–55)	<0.01
Elected delivery (%) ^a	44	37	0.59

^aCaesarean section or induced delivery.

**P*-values based on chi-squared tests for categorical variables and Mann–Whitney tests for quantitative variables.

IQR = interquartile range; AA = African American.

Table 4.2. Top 30 list of DMRs associated with gestational age at birth identified via CHARM 2.0

Chr	name	DMR area	Mean Slope	Pval_max	Qval	DMR Start Position	DMR end position	description
19	NFIX	0.343	0.010	1.0E-03	1.2E-03	13130686	13133039	inside intron
4	RAPGEF2	0.223	-0.008	4.7E-02	2.9E-02	160026138	160028079	upstream
12	MSRB3	0.197	-0.014	9.8E-02	4.1E-02	65671230	65672140	promoter
8	INTS10	0.160	-0.010	2.3E-01	8.8E-02	19615409	19616461	upstream
10	C10orf140	0.147	0.008	3.2E-01	9.4E-02	21803713	21804972	inside exon
3	IQSEC1	0.144	-0.006	3.5E-01	9.4E-02	12937650	12939468	overlaps 3'
14	BCL11B	0.138	0.007	4.1E-01	9.8E-02	99725056	99726396	inside intron
4	LEF1	0.132	0.006	4.7E-01	1.0E-01	109085956	109087500	covers exon(s)
6	PREP	0.130	0.007	5.1E-01	1.0E-01	106034980	106036264	upstream
14	BCL11B	0.122	0.006	5.8E-01	1.0E-01	99706752	99708241	inside intron
12	LEMD3	0.120	-0.010	6.1E-01	1.0E-01	65564532	65565296	overlaps exon downstream
1	C1orf83	0.118	-0.005	6.4E-01	1.0E-01	54561298	54563187	overlaps exon upstream
19	UQCRFS1	0.116	-0.006	6.7E-01	1.0E-01	29696666	29697995	downstream
4	TRIM2	0.115	0.005	6.8E-01	1.0E-01	154125212	154126966	overlaps 5'
17	TNFSF13	0.114	-0.009	7.0E-01	1.0E-01	7462633	7463465	covers exon(s)
14	BCL11B	0.112	0.005	7.2E-01	1.0E-01	99709752	99711218	inside intron
14	FBLN5	0.112	-0.006	7.2E-01	1.0E-01	92412243	92413573	overlaps exon downstream
5	CRHBP	0.107	-0.012	7.8E-01	1.2E-01	76248520	76249150	overlaps 5'
8	DEFA4	0.101	-0.006	8.6E-01	1.2E-01	6792265	6793315	close to 3'
14	APEX1	0.100	-0.007	8.6E-01	1.2E-01	20924708	20925624	covers exon(s)
6	WRNIP1	0.100	-0.007	8.6E-01	1.2E-01	2769840	2770750	covers exon(s)
1	ITPKB	0.098	0.005	8.8E-01	1.2E-01	226900271	226901597	inside intron
3	IFT80	0.097	0.007	8.9E-01	1.2E-01	159941894	159942728	downstream
9	SLC24A2	0.097	-0.007	8.9E-01	1.2E-01	19493250	19494306	downstream

14	RNASE10	0.096	-0.006	9.0E-01	1.2E-01	20978625	20979745	covers
3	CTDSPL	0.095	-0.006	9.0E-01	1.2E-01	37904056	37905258	inside intron
14	NPAS3	0.095	0.006	9.0E-01	1.2E-01	33400717	33401757	upstream
13	DLEU2	0.095	0.008	9.0E-01	1.2E-01	50702836	50703608	upstream
12	CCND2	0.094	-0.006	9.1E-01	1.2E-01	4324191	4325241	upstream

Table 4.3. Candidate significant DMRs associated with gestational age identified via CHARM 2.0

Chr	Nearest gene	DMR area	<i>P</i> -value	<i>q</i> -value	DMR start position	DMR end position	Location relative to gene
19	<i>NFIX</i>	0.343	0.001	0.012	13 130 686	13 133 039	Inside intron
4	<i>RAPGEF2</i>	0.223	0.047	0.029	160 026 138	160 028 079	Upstream, intron of alt. transcript
12	<i>MSRB3</i>	0.197	0.098	0.041	65 671 230	65 672 140	Promoter

P-values and *q*-values based on comparison of observed DMR area ranks to ranks among 1000 permutations of gestational age values. All coordinates are based on hg19/build 37.

Chr, Chromosome; Alt., Alternative.

Table 4.4. Co-efficient (95% CIs) of linear relationship between potential confounders and gestational age at birth or average methylation at each of the identified DMRs

	<i>GA (in weeks)</i>	<i>NFIX^a</i>	<i>RAPGEF2^a</i>	<i>MSRB3^a</i>
Male sex	-0.12 (-0.84 to 0.60)	-5.72 (-8.16 to -3.27)	1.85 (-0.21 to 3.91)	1.07 (-1.32 to 3.46)
Maternal age	-0.06 (-0.13 to -0.00)	-0.17 (-0.40 to 0.05)	0.17 (-0.01 to 0.35)	0.24 (0.04 to 0.44)
Caucasian race (vs African American)	-0.38 (-1.26 to 0.49)	-1.09 (-4.27 to 2.10)	0.53 (-2.01 to 3.07)	0.14 (-2.76 to 3.04)
Maternal smoking ^b	-0.62 (-1.06 to -0.18)	0.86 (-0.83 to 2.54)	0.39 (-0.95 to 1.73)	-0.21 (-1.75 to 1.33)
PIH	-2.25 (-3.54 to -0.96)	-8.41 (-13.09 to -3.73)	5.01 (1.23 to 8.79)	6.77 (2.46 to 11.09)
Intrapartum fever	0.60 (-0.7 to 1.91)	-1.93 (-6.76 to 2.9)	-0.24 (-4.1 to 3.62)	0.13 (-4.28 to 4.55)
Serum copper (µg/dl)	0.04 (0.02 to 0.06)	0.13 (0.06 to 0.20)	-0.10 (-0.16 to -0.05)	-0.15 (-0.21 to -0.08)
Birthweight (kg)	2.07 (1.63 to 2.51)	2.86 (0.87 to 4.84)	-2.89 (-4.43 to -1.34)	-3.97 (-5.71 to -2.22)
Elected delivery	-0.12 (-0.84 to 0.61)	-1.84 (-4.50 to 0.82)	1.10 (-1.03 to 3.22)	-0.56 (-3.00 to 1.88)

Bold indicates statistical significance at $P < 0.05$.

^aAverage residual DNA methylation across DMR adjusted for the same surrogate variables from SVA used in the primary analysis.

^bOrdinal variable: 0 = non-smoker, 1 = passive smoker, 2 = active smoker.

Table 4.5. Comparison of regression coefficients [95% CI] for relationship between methylation and gestational age with and without adjustment for potential confounders

Model: $M^* = b_1GA + b_2Z$

	NFIX	RAPGEF2	MSRB3
Z	b_1	b_1	b_1
N/A (unadjusted)	1.57 [1.02,2.12]	-1.33 [-1.76,-0.90]	-2.08 [-2.51,-1.64]
Copper	1.37 [0.79,1.94]	-1.17 [-1.62,-0.72]	-1.87 [-2.32,-1.42]
Male sex	1.54 [1.03,2.05]	-1.32 [-1.74,-0.89]	-2.07 [-2.51,-1.63]
PIH	1.39 [0.82,1.95]	-1.25 [-1.70,-0.80]	-2.00 [-2.45,-1.54]
Smoking	1.75 [1.17,2.34]	-1.47 [-1.93,-1.00]	-2.25 [-2.71,-1.79]
Birthweight	1.70 [0.99,2.40]	-1.28 [-1.83,-0.74]	-2.18 [-2.73,-1.62]

Table 4.6. Results for univariate and multivariate regression analyses of methylation on birthweight and/or gestational age

DMR	Model	b₁	pval	b₂	pval
NFIX	M* = b ₁ GA	1.57	8.0E-08		
	M* = b ₁ BW	0.0029	5.0E-03		
	M* = b ₁ BW + b ₂ GA	-0.0007	0.58	1.69	4.70E-06
	M* = b ₁ (BW GA)	-0.0007	0.62		
RAPGEF2	M* = b ₁ GA	-1.33	9.9E-09		
	M* = b ₁ BW	-0.0029	3.0E-04		
	M* = b ₁ BW + b ₂ GA	-0.0002	0.80	-1.28	8.7e-06
	M* = b ₁ (BW GA)	-0.0002	0.83		
MSRB3	M* = b ₁ GA	-2.07	1.3E-16		
	M* = b ₁ BW	-0.0040	1.4E-05		
	M* = b ₁ BW + b ₂ GA	0.0005	0.57	-2.20	1.9e-12
	M* = b ₁ (BW GA)	0.0005	0.65		

References

- Davis, E.P., Buss, C., Muftuler, L.T., Head, K., Hasso, A., Wing, D.A., Hobel, C., and Sandman, C.A. (2011). Children's Brain Development Benefits from Longer Gestation. *Frontiers in psychology* 2, 1.
- Hansen, A.K., Wisborg, K., Uldbjerg, N., and Henriksen, T.B. (2008). Risk of respiratory morbidity in term infants delivered by elective caesarean section: cohort study. *Bmj* 336, 85-87.
- Marchetti, M.A., Pizarro, G.O., Sagher, D., Deamicis, C., Brot, N., Hejtmancik, J.F., Weissbach, H., and Kantorow, M. (2005). Methionine sulfoxide reductases B1, B2, and B3 are present in the human lens and confer oxidative stress resistance to lens cells. *Investigative ophthalmology & visual science* 46, 2107-2112.
- Messina, G., Biressi, S., Monteverde, S., Magli, A., Cassano, M., Perani, L., Roncaglia, E., Tagliafico, E., Starnes, L., Campbell, C.E., *et al.* (2010). Nfix regulates fetal-specific transcription in developing skeletal muscle. *Cell* 140, 554-566.
- Ramirez, J., Lukin, K., and Hagman, J. (2010). From hematopoietic progenitors to B cells: mechanisms of lineage restriction and commitment. *Current opinion in immunology* 22, 177-184.
- Vicente, C., Conchillo, A., Garcia-Sanchez, M.A., and Otero, M.D. (2012). The role of the GATA2 transcription factor in normal and malignant hematopoiesis. *Critical reviews in oncology/hematology* 82, 1-17.
- Wells, E.M., Jarrett, J.M., Lin, Y.H., Caldwell, K.L., Hibbeln, J.R., Apelberg, B.J., Herbstman, J., Halden, R.U., Witter, F.R., and Goldman, L.R. (2011). Body burdens of mercury, lead, selenium and copper among Baltimore newborns. *Environmental research* 111, 411-417.
- Yang, S., Platt, R.W., and Kramer, M.S. (2010). Variation in child cognitive ability by week of gestation among healthy term births. *American journal of epidemiology* 171, 399-406.

Chapter 5: Results – Identification of
mercury associated DMRs

To identify candidate genomic regions showing epigenetic differences associated with total and methyl mercury exposure, we utilized previously obtained CHARM 2.0 DNA methylation data originally used for identifying gestational age DMRs (Lee et al., 2012). Descriptive statistics for Baltimore THREE and NCS Vanguard cohort samples used for this study are given in Table 5.1. Distribution of sex and gestational age between the two cohorts were similar. Higher maternal age and birthweight were observed in NCS Vanguard cohort. Total and methyl mercury exposures are higher in THREE cohort samples, which further described in Table 5.2. Although the distribution is different, both cohorts only contain two samples with methyl mercury exposure levels higher than the reference dose (5.8ug/L), suggesting the exposure levels in both cohorts are within the normal exposure range. Both cohorts also display different race distribution, which African Americans are dominant in THREE cohort, whereas Caucasian contributes as a major population in NCS Vanguard cohort. Uniquely, n-3 fatty acids (EPA and DHA), an essential nutrient for healthy neurological development and may counter the toxic effect from mercury exposure (National Research Council, 2007), were measured in THREE cohort to assess n-3 fatty acids as a potential confounder.

Since methyl mercury exposure as a most relevant mercury species related to adverse health outcomes and site-specific DNA methylation changes shown to be associated with methyl mercury exposure levels (Basu et al., 2013; Goodrich et al., 2013), we sought to first identify DMRs associated with methyl mercury exposure after natural log transformation of the exposure data due to its lognormal distribution. By this approach, we identified 130 candidate DMRs associated with methyl mercury exposure, of which one candidate DMR associated with methyl mercury exposure passed our

genome-wide threshold of family wise error rates < 20% and false discovery rate < 10% (Table 5.3). The candidate DMR with genome-wide significance was located in the exon of Transcription Elongation Factor A (SII) N-Terminal And Central Domain Containing 2 (*TCEANC2*) gene, which encodes a putative transcription elongation factor.

Figure 5.1 demonstrates the methylation plots for *TCEANC2* DMR negatively associated with either methyl (Figure 5.1a) or total (Figure 5.1b) mercury exposure. Each dot represents the methylation values at each probe for each neonate sample. Smoothed line represents the average methylation curve for each exposure level quartile, which show dose-dependent negative trend for both types of the exposure levels. Methylation plots for the other three candidate DMRs associated with total mercury exposure are shown in Figure 5.2 a-c, and Figure 5.2 d-f represents the overlapping regions plotted by quartiles of methyl mercury exposure. The DMR inside *ANGPT2*, which encodes Angiopoietin 2, represents positive correlation between the methylation and both of the exposure levels. Negative association between the methylation and exposure levels for both total and methyl mercury were shown in DMR inside *PRPF18* gene, encoding Pre mRNA Processing Factor 18, and DMR near *FOXD2* gene, which encodes a transcription factor Forkhead Box D2.

To replicate these findings on a separate methylation assay platform, bisulphite pyrosequencing assays were performed on four CpGs within the *TCEANC2* DMR (red blocks in Figure 5.1). DNA methylation at all of the four CpGs displayed decrease in DNA methylation level along with increase in both methyl mercury (Figure 5.3a) and total mercury (Figure 5.3b) exposure level, consistent with the pattern detected in CHARM 2.0. Thus, the association between DNA methylation levels at *TCEANC2* DMR

and exposure levels identified by CHARM 2.0 is replicated using an independent measurement platform with independent set of cohort samples. Other three DMRs which displayed association between DNA methylation and total mercury exposure were also tested for replication via bisulfite pyrosequencing (Figure 5.4). The DNA methylation trend identified from the CpGs inside three DMRs was not consistent with the pattern detected from CHARM 2.0, meaning that the other three DMRs weren't able to replicate.

To address potential confounding by sex, maternal age, race, birthweight, gestational age and lead, selenium, copper and n-3 fatty acid levels, we estimated the linear association between each of these variables and total or methyl mercury exposure level obtained from THREE study samples. Consistent with the previous reports (Wells et al., 2011), race and n-3 fatty acid levels were associated with both types of mercury exposure levels (Table 5.4). We then also estimated the linear relationship between these potential confounders and the average methylation level at *TCEANC2* DMR to further address for potential confounding. Race, not n-3 fatty acid level, was also associated with methylation at *TCEANC2* DMR, suggesting the potential for confounding. However, the association between average methylation at the DMR and total or methyl mercury exposure levels after adjusting for race (Table 5.5) remained statistically significant even after adjusting.

Different cell type distribution has been suggested as a potential confounder, in particular for DNA methylation studies using whole blood samples due to cell type distribution shifts due to the nature of the disease or the exposure, and has been addressed in several publications (Joubert et al., 2012; Liu et al., 2013). To examine whether the relationship between blood DNA methylation and total mercury level was confounded by

cell type, we used a publicly available database containing DNA methylation signatures for sorted blood cells (Reinius et al., 2012). Within or near our DMR (chr1: 54562102 – 54562548), there were four probe sets from the Illumina 450k dataset (cg01109333, cg01986665, cg02270108, and cg02626873). We plotted the DNA methylation levels at each of these probe sets across the sorted blood cell types (B cells, CD4+ T cells, CD8+ T cells, granulocytes, monocytes, natural killer cells, and whole blood) (Figure 5.5). There were six replicates of each cell type. Across three of the probe sets, DNA methylation was not heterogeneous and there was no relationship between cell type and DNA methylation. For one of the probe sets (cg01986665), DNA methylation levels were heterogeneous and there was some variability across cell type. It is possible that DNA methylation varies by cell type at this locus.

To address whether the relationship between DNA methylation and mercury to be confounded by cell type, we looked to our Infinium HumanMethylation450 assay data from NCS Vanguard study samples. We were able to infer relative cell type proportions in this population using an epigenetic signature prediction algorithm (Houseman et al., 2012). Maternal blood total mercury levels measured during first trimester were not associated with proportions of cell types inferred by DNA methylation levels in maternal first trimester whole blood buffy coat samples, as well as by cord blood buffy coat samples (Table 5.6). Since there was no relationship observed between estimated blood cell distribution and total mercury levels, cell type thus is not a confounder of the DNA methylation and mercury association in our study.

To examine the functional significance of the methylation difference at TCEANC2 DMR, we measured the expression level of the TCEANC2 gene using real-

time PCR assay. All of the four assessed CpGs show an inverse correlation between expression and DNA methylation levels, although not statistically significant (spearman correlation p-value of 0.29 for CG1, 0.16 for CG2, 0.11 for CG3, and 0.061 for CG4, Figure 5.6).

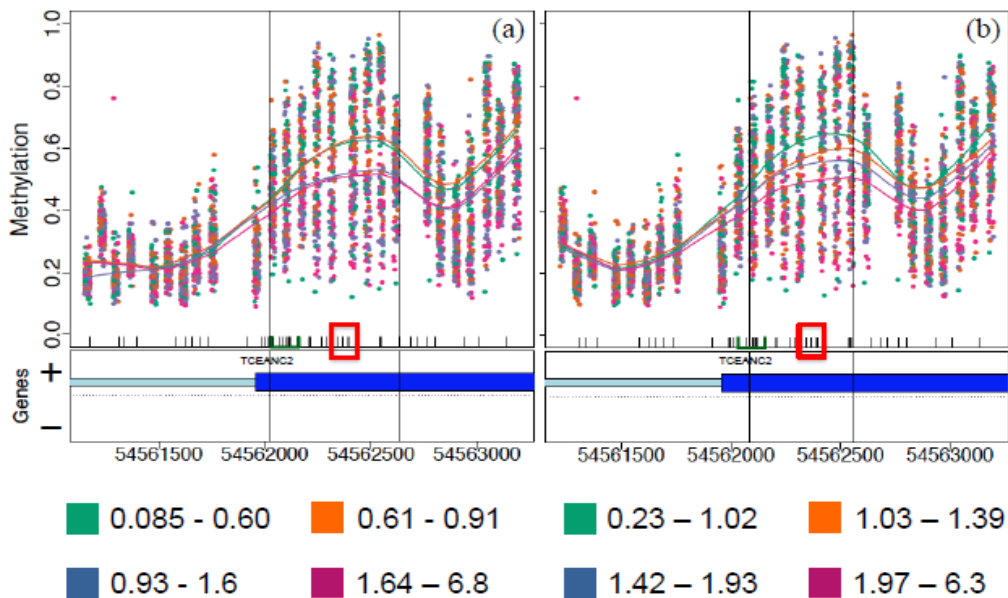


Figure 5.1. Methylation plots for DMR inside *TCEANC2* associated with methyl and total mercury exposure

(a) *TCEANC2* DMR with color codes representing quartiles of methyl mercury exposure and (b) total mercury exposure. Top half of panels show individual methylation levels at each probe by genomic position, with colored lines reflecting the average methylation curve for samples. The median exposure level for each quartile is shown in the legend. Bottom half of panels show location of CpG dinucleotides (black tick marks) and CpGs validated by bisulphite pyrosequencing (black tick marks contained in red box) as well as the location of refseq annotated genes (bar, + and - represent the direction of the gene, green bar indicates CpG island).

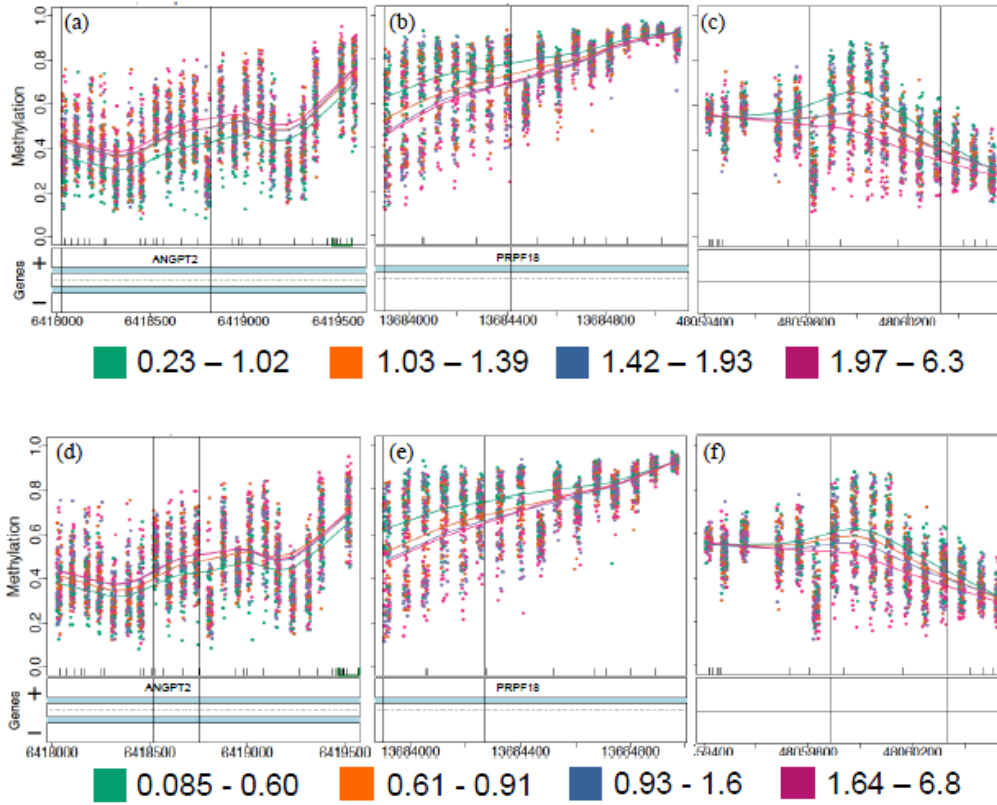


Figure 5.2. Methylation plots for other identified DMRs associated with total mercury exposure

(a), (b), and (c) represent DMR inside or near *ANGPT2*, *PRPF18* and *FOXD2* with color codes representing quartiles of total mercury exposure. (d), (e), and (f) represents same DMRs with color codes representing quartiles of methyl mercury exposure.

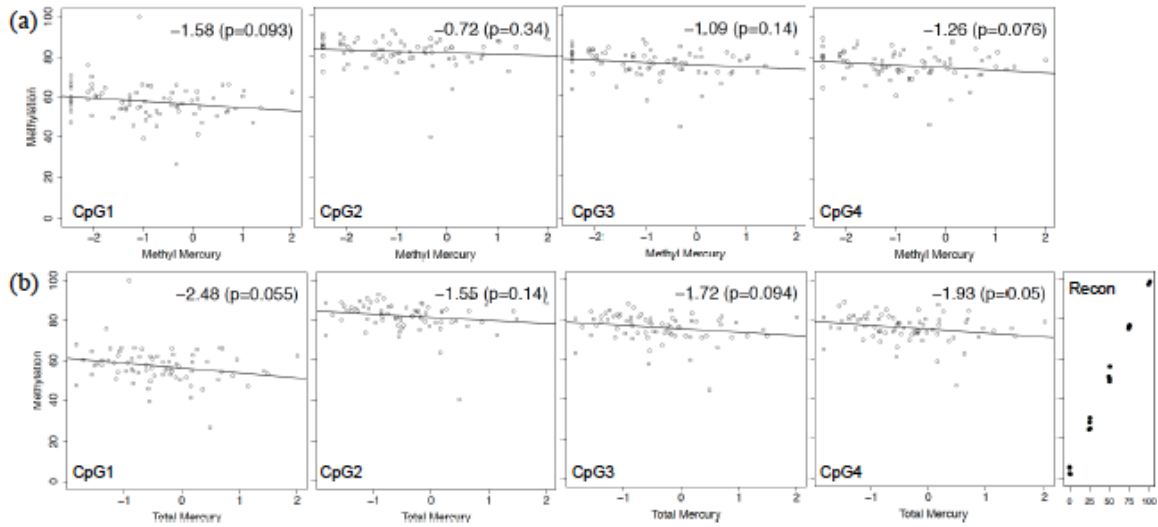


Figure 5.3. Bisulfite pyrosequencing results for DMR inside *TCEANC2*

Circles represent methylation values (y-axis) at individual CpGs for their corresponding methyl mercury exposure levels (a) or total mercury exposure levels (b) on the x-axis. Fitted lines represent predicted values from linear regression. Reconstitution controls (represented as black dots) with predicted % methylation (x-axis) are located at the right bottom corner (Recon). The numbers on the top of each figure represent effect size/slope estimate from the regression of methylation on mercury exposure and P-value for a Wald test of this slope

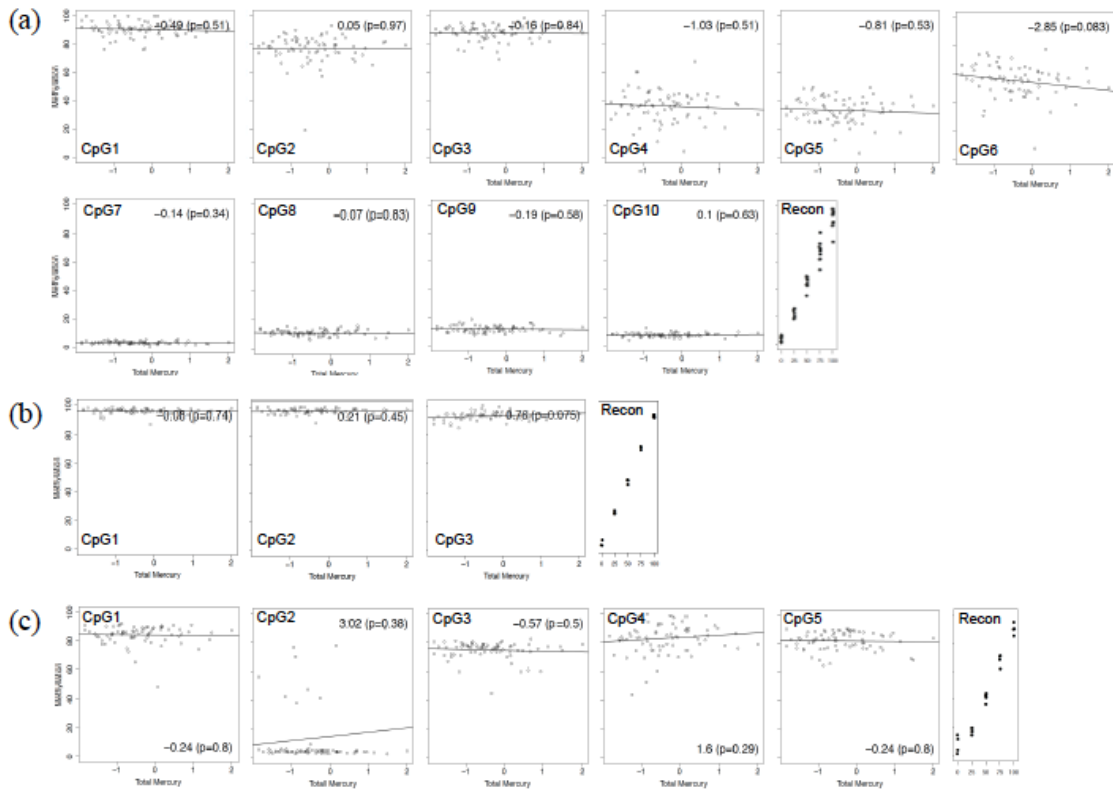


Figure 5.4. Bisulfite Pyrosequencing Results for DMR inside *ANGPT2*, *PRPF18* and near *FOXD2*

Pyrosequencing results for (a) *ANGPT2*, (b) *PRPF18*, and (c) *FOXD2*. Circles represent methylation values (y-axis) at individual CpGs for their corresponding total mercury exposure levels on the x-axis.

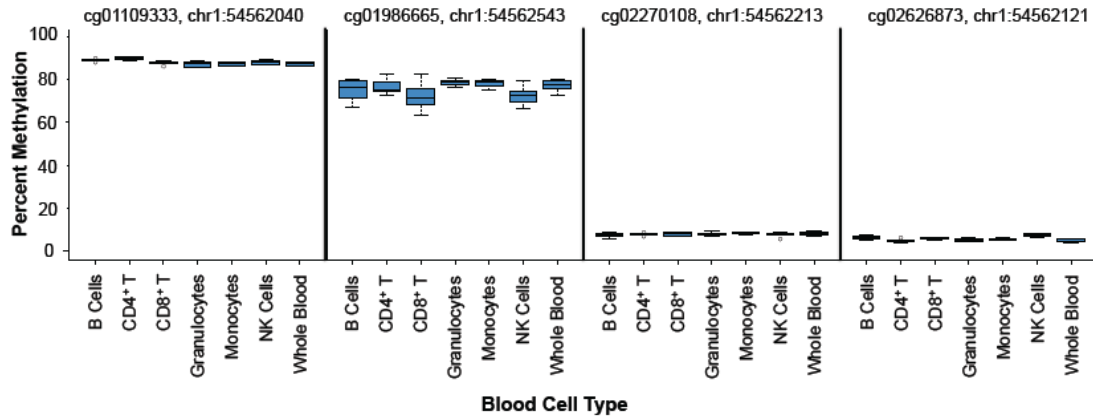


Figure 5.5. DNA methylation levels at four Infinium HumanMethylation450 probe sets located near/inside the DMR inside *TCEANC2*

Each box represents the four probes near or inside the *TCEANC2* DMR. Boxplots represent percent methylation levels (y axis) for each given blood cell types (x axis). The labels on the top of each box represent the probe ID assigned on the HumanMethylation450 array and the chromosomal location of the CpG inside the probe.

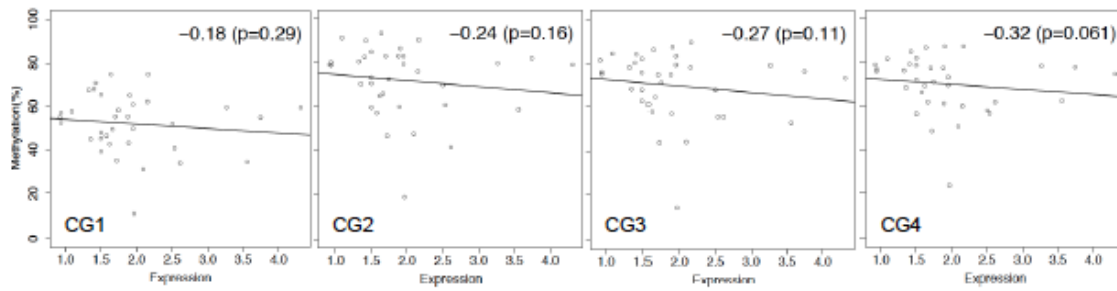


Figure 5.6. Correlation between DNA methylation level at *TCEANC2* DMR and *TCEANC2* gene expression level

Circles represent the normalized expression level of *TCEANC2* (x axis) with respect to DNA methylation level (y axis) for each CpG subjected to pyrosequencing. Fitted line represents predicted values from linear regression. The numbers on the top of each figure represent spearman's rank correlation coefficients with p-values derived from spearman's rank test

Table 5.1. Characteristics of THREE and NCS Vanguard study newborns included in this project

	THREE, <i>N</i> =141	NCS, <i>N</i> =77
Male sex (%)	52	51
Maternal age, median (IQR) (years)	24 (20-29)	28 (25-33)
Maternal race (%)		
Caucasian	23	75
AA	72	6
Asian	5	1
American Indian/Native	NA	1
Other	NA	10
2 or more races	NA	5
Gestational age, median (IQR) (days)	275 (266-282)	275 (270-281)
Birthweight, median (IQR) (g)	3183 (2794-3566)	3402 (3102-3691)
Total Mercury, median (IQR) (ug/L)	1.4 (1.0-2.0)	0.6 (0.4-1.1)
Methyl Mercury, median (IQR) (ug/L)	0.9 (0.6-1.6)	0.4 (0.2-0.9)
EPA+DHA, median (IQR) (ug/mL)	48.1 (40.8 - 56.9)	NA

Table 5.2. Distribution of total/methyl mercury from THREE and NCS Vanguard study newborns included in this project

	Total mercury in THREE (ug/L)	Methyl mercury in THREE (ug/L)	Total mercury in NCS (ug/L)	Methyl mercury in NCS (ug/L)
<i>N</i>	140	141	75	77
Limit of detection (LOD)	0.33	0.12	0.12	0.12
N (%) below LOD	4 (2.8%)	10	2	10
Minimum	0.233	0.0848	0.085	0.085
10th percentile	0.71	0.35	0.22	0.085
25th percentile	1.02	0.61	0.35	0.15
Median	1.39	0.93	0.59	0.38
75th percentile	1.97	1.64	1.13	0.87
90th percentile	2.51	2.18	1.97	2.01
Maximum	6.3	6.8	7.4	7.48
Arithmetic mean (95% CI)	1.60 (1.43,1.77)	1.23 (1.05,1.41)	0.96 (0.77,1.15)	0.78 (0.53,1.04)
Geometric mean (95% CI)	1.35 (1.22,1.49)	0.87 (0.753,1.02)	0.64 (0.55,0.74)	0.41 (0.32,0.53)

Table 5.3. Candidate significant DMRs associated with total (a) and methyl (b) mercury exposure in THREE study

(a)

Chr	Nearest gene	DMR area	FWER	q-value	DMR start position	DMR end position	Location relative to gene
8	<i>ANGPT2</i>	3.198	0.03	0.045	6418036	6418808	Inside Intron
10	<i>PRPF18</i>	2.982	0.05	0.045	13683914	13684400	Inside Intron
1	<i>FOXD2</i>	2.512	0.13	0.078	48059831	48060319	Downstream
1	<i>TCEANC2</i>	2.257	0.19	0.079	54562102	54562548	Inside Exon

(b)

Chr	Nearest gene	DMR area	FWER	q-value	DMR start position	DMR end position	Location relative to gene
1	<i>TCEANC2</i>	2.124	0.05	0.088	54562036	54562618	Inside Exon

Table 5.4. Linear relationship coefficient [95%CI] between potential confounders and total (a) and methyl (b) mercury exposure or average methylation at DMR inside *TCEANC2* gene

(a)	Mercury (Total)	TCEANC2 methylation
Sex	0.02 [-0.19,0.23]	-1.79 [-8.8,5.22]
Maternal Age (yrs)	0.01 [-0.01,0.02]	0.08 [-0.52,0.68]
Race	-0.47 [-0.82,-0.12]	12.88 [4.54,21.23]
Birthweight (kg)	0 [0,0]	0 [-0.01,0]
Gestational Age (days)	0 [-0.01,0]	-0.11 [-0.36,0.14]
Lead	0.15 [-0.03,0.32]	-2.48 [-8.43,3.48]
Selenium	-0.01 [-0.01,0]	-0.16 [-0.43,0.12]
Copper	-0.19 [-0.43,0.05]	-0.87 [-9.1,7.36]
EPA+DHA	0.01 [0,0.02]	-0.19 [-0.44,0.06]

(b)	Mercury (Methyl)	TCEANC2 methylation
Sex	0.01 [-0.31,0.33]	-1.8 [-8.07,4.47]
Maternal Age (yrs)	0.02 [0,0.05]	0.09 [-0.45,0.63]
Race	-0.47 [-0.84,-0.09]	11.63 [4.19,19.07]
Birthweight (kg)	0 [0,0]	0 [-0.01,0]
Gestational Age (days)	-0.01 [-0.02,0.01]	-0.12 [-0.34,0.11]
Lead	0.19 [-0.09,0.46]	-1.81 [-7.16,3.54]
Selenium	-0.01 [-0.02,0]	-0.14 [-0.39,0.1]
Copper	-0.41 [-0.78,-0.04]	-1.53 [-8.87,5.81]
EPA+DHA	0.01 [0,0.02]	-0.16 [-0.39,0.06]

Table 5.5. Comparison of regression coefficients [95% CI] for association between DNA methylation and total (a)/methyl (b) mercury exposure with and without adjustment for race

Model: $M = \beta_1(\text{Total mercury}) + \beta_2C$

		TCEANC2	
(a)	C	β_1	β_2
	N/A (unadjusted)	-8.9 [-14.85,-2.96]	
	Race	-6.51 [-12.61,-0.41]	-11.8 [-20.15,-3.46]

Model: $M = \beta_1(\text{Methyl mercury}) + \beta_2C$

		TCEANC2	
(b)	C	β_1	β_2
	N/A (unadjusted)	-5.92 [-9.29,-2.56]	
	Race	-4.89 [-8.25,-1.53]	-11.01 [-19.55,-2.47]

Table 5.6. Correlation between estimated blood cell counts and total mercury exposure

Cell Type	Pearson's Correlation	95% CI	Mean (IQR)
Maternal Trimester 1			
CD8 ⁺ T cells	-0.0512	(-0.22, 0.12)	1.74 (0,2.45)
CD4 ⁺ T cells	0.0465	(-0.12, 0.21)	10.39 (7.77,12.32)
Natural killer cells	-0.0012	(-0.17, 0.17)	4.7 (2.46,6.62)
B cells	0.1116	(-0.06, 0.27)	4.89 (3.45,5.96)
Monocytes	-0.1019	(-0.26, 0.07)	9.74 (8.22,11.09)
Granulocytes	-0.0020	(-0.17, 0.17)	61.63 (57.42,67.39)
Fetal Cord Blood			
CD8 ⁺ T cells	-0.1250	(-0.33, 0.10)	2.14 (0,3.77)
CD4 ⁺ T cells	-0.0803	(-0.29, 0.14)	14.15 (10.27,17.67)
Natural killer cells	-0.1783	(-0.38, 0.04)	5.73 (1.57,9.31)
B cells	-0.0540	(-0.27, 0.17)	13.05 (9.93,15.35)
Monocytes	0.1833	(-0.04, 0.39)	11.8 (9.71,13.04)
Granulocytes	0.1152	(-0.11, 0.33)	49.33 (44.26,55.09)

References

- Basu, N., Head, J., Nam, D.H., Pilsner, J.R., Carvan, M.J., Chan, H.M., Goetz, F.W., Murphy, C.A., Rouvinen-Watt, K., and Scheuhammer, A.M. (2013). Effects of methylmercury on epigenetic markers in three model species: mink, chicken and yellow perch. *Comparative biochemistry and physiology. Toxicology & pharmacology : CBP* *157*, 322-327.
- Goodrich, J.M., Basu, N., Franzblau, A., and Dolinoy, D.C. (2013). Mercury biomarkers and DNA methylation among Michigan dental professionals. *Environmental and molecular mutagenesis* *54*, 195-203.
- Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics* *13*, 86.
- Joubert, B.R., Haberg, S.E., Nilsen, R.M., Wang, X., Vollset, S.E., Murphy, S.K., Huang, Z., Hoyo, C., Middtun, O., Cupul-Uicab, L.A., *et al.* (2012). 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environmental health perspectives* *120*, 1425-1431.
- Lee, H., Jaffe, A.E., Feinberg, J.I., Tryggvadottir, R., Brown, S., Montano, C., Aryee, M.J., Irizarry, R.A., Herbstman, J., Witter, F.R., *et al.* (2012). DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. *International journal of epidemiology* *41*, 188-199.
- Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., *et al.* (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature biotechnology* *31*, 142-147.
- National Research Council (2007). *Seafood Choices: Balancing Benefits and Risks*. National Institute of Medicine, National Research Council.
- Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S.E., Greco, D., Soderhall, C., Scheynius, A., and Kere, J. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PloS one* *7*, e41361.
- Wells, E.M., Jarrett, J.M., Lin, Y.H., Caldwell, K.L., Hibbeln, J.R., Apelberg, B.J., Herbstman, J., Halden, R.U., Witter, F.R., and Goldman, L.R. (2011). Body burdens of mercury, lead, selenium and copper among Baltimore newborns. *Environmental research* *111*, 411-417.

Chapter 6: Discussion

Part of this chapter is reproduced from published article in International Journal of
Epidemiology (Lee et al. 2012)

Using a genome-wide custom DNA methylation array technology and novel statistical methods, we have identified three differentially methylated regions associated with gestational age at birth and one DMR associated with both methyl and total mercury exposure level. Array-based methylation results for all three regions were validated or replicated via bisulphite pyrosequencing.

DMRs associated with gestational age target areas of the genome likely to be under developmental regulation in late gestation, which may have implications for understanding the reasons for immediate as well as long-term health effects of gestational age at birth. The observed incremental progression between methylation and gestational age at birth is further supported by the observation that adults are not variable at these DMRs, but rather appear to be stable at levels similar to or more extreme than newborns with the latest gestational ages at birth. The genes nearest the identified DMRs may play important roles in late-stage fetal development. *NFIX* is known to be responsible for regulating skeletal muscle (Messina et al., 2010), brain and bone development (Campbell et al., 2008; Driller et al., 2007; Mason et al., 2009), which show substantial growth during late gestation. This finding offers face validity that our approach can identify epigenomic regions relevant to late gestational development. *RAPGEF2* plays a critical role in embryonic haematopoiesis (Satyanarayana et al., 2010) and brain development (i.e. commissures) (Bilasy et al., 2011). Although this DMR was not located at the promoter of the canonical gene, the DMR contains strong DNase I hypersensitive sites and a number of strong transcription factor-binding sites including *Gata-2* and *PU.1*, which are the critical transcription factors in haematopoiesis (Ramirez et al., 2010; Vicente et al., 2012). *In utero*, a fetus has a higher haematocrit (given lower available oxygen in utero),

low B-cell function (given ready transplacental passage of maternal antibodies) as well as lower platelet counts than seen in babies (born at term). This methylation change with gestational age could be involved in the ontogeny of the haematopoietic system and the switch from production of erythrocytes to increased production of B-lymphocytes and megakaryocytes in preparation for birth and, respectively, secretion of antibodies in response to antigenic assaults as well as production of platelets to prepare for possible birth trauma. Furthermore, anaemia of prematurity is known to cause morbidity in pre-term infants; disruption of regulation of this system may contribute to anaemia of prematurity, due to higher haematocrits and restricted erythropoiesis at birth. The differential methylation detected in our newborn sample did correlate with expression of *RAPGEF2* in cord blood cells, lending support for involvement in development of the haematopoietic system. Finally, *MSRB3* encodes a methionine sulphate reductase enzyme involved in antioxidant repair, converting methionine sulphoxide to methionine. This specific reductase has been found to be present in many tissues including the human lens and the cochlea and has been suspected to be involved in cataracts caused by oxidative damage to lens cells (Marchetti et al., 2005). Most congenital cataracts are idiopathic; however, PTB and the administration of certain drugs in utero have been identified as risk factors (Rahi and Dezateux, 2000), pointing to a possible role for oxidative stress for cataract formation in infants as well as adults. Generally, a number of morbid conditions associated with term birth have been tied to oxidative stress, from administration of oxygen, including retinopathy of prematurity, bronchopulmonary dysplasia, necrotizing enterocolitis and intraventricular haemorrhage (Walsh et al., 2009). *MSRB3* and other Methionine Sulfoxide Reductases (MSRs) may play a role in this sensitivity to oxidative

stress. Mutations in *MSRB3* also cause hereditary deafness (Ahmed et al., 2011) and variants in this gene have been associated with primary tooth development during infancy in a recent genome-wide association study (Pillas et al., 2010).

The DMR associated with both methyl and total mercury target *TCEANC2* gene which encodes Transcription Elongation Factor A (SII) N-Terminal And Central Domain Containing 2, a hypothetical transcription elongation factor based on the sequence domain homology. Although no publication in regards to the function of *TCEANC2* protein itself has yet been reported, methyl mercury can increase RNA synthesis *in vitro* at least partly through stimulating chain elongation by RNA Polymerase II (Chao and Frenkel, 1983; Frenkel and Ducote, 1987; Frenkel and Randles, 1982). Together with the negative association between *TCEANC2* gene expression and methylation changes at the DMR from our result, there is a possibility of potential function for *TCEANC2* by mediating between the methyl mercury exposure and increase in RNA transcription.

These results do not appear to be sensitive to confounding by measured variables. Furthermore, it is possible that methylation may be part of the mechanism relating factors to gestational age at birth or mercury exposure levels. In these cases, one would not want to adjust for such factors in analysis. Thus, we were conservative in our approach to adjustment. Nonetheless, inclusion of potential confounders in our models did not attenuate the relationship between methylation and gestational age or mercury exposure levels at these DMRs. Our use of SVA to reduce the impact of measurement issues, such as batch effects, may also have adjusted for potential residual confounding not captured by a measured variable. It is worth noting that serum copper levels have previously been shown in this sample to be related to gestational age at birth and, therefore, a potential

confounder. Nonetheless, the relationship between DMR methylation and gestational age did not attenuate after adjustment for copper. We did, however, observe a relationship between copper levels and methylation in these adjusted models, suggesting an independent effect of copper on methylation, consistent with the growing interest in environmental impacts on the epigenome and their implications for human health (Dolinoy and Jirtle, 2008; Sutherland and Costa, 2003). We also observed a relationship between race and methylation in adjusted model, also suggesting an independent effect of race on methylation at *TCEANC2* DMR. Although we saw a relationship between birthweight and gestational age DMRs, this appeared to be a function of the relationship between gestational age and birthweight, rather than specific to birthweight itself. Although a recent report did see a relationship between global DNA methylation and birthweight for gestational age (Michels et al., 2011), we did not see an association with these particular DMRs when considering birthweight adjusted for gestational age (Supplementary Table 5).

An important caveat in this study is that we measured DNA methylation from a surrogate tissue, blood, for which methylation changes may not reflect those of tissues undergoing developmental epigenetic changes. Despite this, one of the genes near gestational age DMRs, *RAPGEF2*, and *TCEANC2* gene containing mercury-associated DMR showed the expected inverse relationship of DNA methylation and gene expression. Consistent with this idea, *RAPGEF2* regulates embryonic haematopoiesis (Satyanarayana et al., 2010), and *TCEANC2* is expressed in blood, whereas *NFIX* and *MSRB3* play in the development of organs such as brain, tooth, skeletal muscle and bone (Campbell et al., 2008; Mason et al., 2009; Messina et al., 2010; Pillas et al., 2010). Thus, differential

expression by methylation patterns of the latter two genes may not be detectable in cord blood, or these DMRs may regulate the enhancer function of distal genes or focally modify the high-order chromatin structure and thus not manifest a change in cord blood expression of *NFIX* or *MSRB3*. These results are quite encouraging for epigenetic epidemiology in general, since they indicate that DNA methylation differences may be widespread, and methylation profiles in blood may be a useful indicator of developmental change even in tissues that do not utilize the differentially methylated genes in normal developmental processes.

Overall, the results obtained here by genome-wide DNA methylation analysis are encouraging for the field of epigenetic epidemiology, since they indicate that DNA methylation differences are detectable with this strategy. Specifically, this work identifies epigenetic changes associated with gestational age at birth and mercury exposure. The underlying reason for this correlation cannot be determined in this cross-sectional study or the replication study, but there are several implications of these findings for the epidemiology of PTB and mercury exposure.

First, regions of the genome that are still undergoing DNA methylation variation late in gestation may be functionally related to the health consequences of PTB, and our findings can inform new epidemiologic research and biological mechanisms towards understanding the reasons for negative outcomes in premature babies and lessening these negative infant, childhood or even adult health consequences related to gestational age at birth.

Secondly, it is possible that these results reflect involvement of DNA methylation in the aetiology of PTB. There are a number of mechanisms (including infections leading

to inflammation (Leitich and Kiss, 2007), preeclampsia (Moldenhauer et al., 2003) and stress (Smith, 2007)) and risk factors (African American race, bacterial vaginosis, cigarette smoking and low maternal pregnancy body mass index (Behrman, 2007; Kramer et al., 2011)) associated with PTB, which could be associated with epigenetic changes themselves, although this explanation is less consistent with the function of the particular genes identified in our study. In addition, the use of assisted reproductive technology and nutritional deficiencies have been identified as possible risk factors for PTB (Dunlop et al., 2011; Henderson et al., 2012) and also have the potential to alter the epigenome (Chmurzynska, 2010; DeBaun et al., 2003). Identification of epigenetic changes associated with PTB potentially could be useful for identifying, among the many factors associated with PTB, which are most likely to be causal factors, although our design did not contain a large number of spontaneous PTBs and thus the relationship between methylation and causes of PTB may be best suited for subsequent studies in different samples.

Thirdly, the findings in our study further implicates about the biological effect of low level mercury exposure, particularly methyl mercury, in neonates. Although historical cases such as Minamata disease or Iraq poison grain syndrome showed negative health outcomes due to high level methyl mercury exposure on both adults and children (Bakir et al., 1973; Social Scientific Study Group on Minamata Disease, 1999), low level methyl mercury exposure *in utero* is also associated with brain function deficits in childhood (Grandjean et al., 1997), and the negative effect of the exposure on cognitive development in children is more apparent when adjusting with maternal fish intake, which contains beneficial nutrients for dampening mercury toxicity (Budtz-

Jorgensen et al., 2007; Strain et al., 2008). Our results provide further evidences along with these reports by showing that mercury exposure mostly within normal exposure range is associated with epigenetic differences in neonates. Also, methylation changes associated with mercury exposure is not confounded by n-3 fatty acid levels, which serves as one of the indicators for maternal fish consumption, suggesting that the methylation differences we observed would be more associated with negative health outcomes. Thus, our findings would help understanding more about the potential mechanisms of negative health outcomes on neonates by mercury exposure, particularly methyl mercury, *in utero*.

Further work is required to determine whether the detection of DNA methylation in non-primary proxy tissues (in this instance, blood) indeed is a useful indicator of developmental change in the primary tissue for expression of affected genes or functional changes induced by mercury exposure. However, the work presented here shows that DNA methylation changes progressively during late fetal development or changes associated with mercury exposure, thus opening the door to studies of the epigenetic epidemiology, and possibly helpful by providing reference data which would be part of integrating environmental exposure status, birth outcomes, genetics, and epigenetics of neonates. Together with other findings, the findings presented in this thesis would help to find ultimate cause for childhood disorders.

References

- Ahmed, Z.M., Yousaf, R., Lee, B.C., Khan, S.N., Lee, S., Lee, K., Husnain, T., Rehman, A.U., Bonneux, S., Ansar, M., *et al.* (2011). Functional null mutations of MSRB3 encoding methionine sulfoxide reductase are associated with human deafness DFNB74. *American journal of human genetics* 88, 19-29.
- Bakir, F., Damluji, S.F., Amin-Zaki, L., Murtadha, M., Khalidi, A., al-Rawi, N.Y., Tikriti, S., Dahahir, H.I., Clarkson, T.W., Smith, J.C., *et al.* (1973). Methylmercury poisoning in Iraq. *Science* 181, 230-241.
- Behrman, R.E.B., A. S. (2007). In *Preterm Birth: Causes, Consequences, and Prevention*, R.E. Behrman, and A.S. Butler, eds. (Washington (DC)).
- Bilasy, S.E., Satoh, T., Terashima, T., and Kataoka, T. (2011). RA-GEF-1 (Rapgef2) is essential for proper development of the midline commissures. *Neuroscience research* 71, 200-209.
- Budtz-Jorgensen, E., Grandjean, P., and Weihe, P. (2007). Separation of risks and benefits of seafood intake. *Environmental health perspectives* 115, 323-327.
- Campbell, C.E., Piper, M., Plachez, C., Yeh, Y.T., Baizer, J.S., Osinski, J.M., Litwack, E.D., Richards, L.J., and Gronostajski, R.M. (2008). The transcription factor Nfix is essential for normal brain development. *BMC developmental biology* 8, 52.
- Chao, E.S., and Frenkel, G.D. (1983). Studies on the mechanism of the stimulation of polymerase II-catalyzed RNA synthesis by mercury compounds. *The Journal of biological chemistry* 258, 9861-9866.
- Chmurzynska, A. (2010). Fetal programming: link between early nutrition, DNA methylation, and complex diseases. *Nutrition reviews* 68, 87-98.
- DeBaun, M.R., Niemitz, E.L., and Feinberg, A.P. (2003). Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. *American journal of human genetics* 72, 156-160.
- Dolinoy, D.C., and Jirtle, R.L. (2008). Environmental epigenomics in human health and disease. *Environmental and molecular mutagenesis* 49, 4-8.
- Driller, K., Pagenstecher, A., Uhl, M., Omran, H., Berlis, A., Grunder, A., and Sippel, A.E. (2007). Nuclear factor I X deficiency causes brain malformation and severe skeletal defects. *Molecular and cellular biology* 27, 3855-3867.

- Dunlop, A.L., Kramer, M.R., Hogue, C.J., Menon, R., and Ramakrishnan, U. (2011). Racial disparities in preterm birth: an overview of the potential role of nutrient deficiencies. *Acta obstetrica et gynecologica Scandinavica* *90*, 1332-1341.
- Frenkel, G.D., and Ducote, J. (1987). The enhanced rate of transcription of methyl mercury-exposed DNA by RNA polymerase is not sufficient to explain the stimulatory effect of methyl mercury on RNA synthesis in isolated nuclei. *Journal of inorganic biochemistry* *31*, 95-102.
- Frenkel, G.D., and Randles, K. (1982). Specific stimulation of alpha-amanitin-sensitive RNA synthesis in isolated HeLa nuclei by methyl mercury. *The Journal of biological chemistry* *257*, 6275-6279.
- Grandjean, P., Weihe, P., White, R.F., Debes, F., Araki, S., Yokoyama, K., Murata, K., Sorensen, N., Dahl, R., and Jorgensen, P.J. (1997). Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. *Neurotoxicology and teratology* *19*, 417-428.
- Henderson, J.J., McWilliam, O.A., Newnham, J.P., and Pennell, C.E. (2012). Preterm birth aetiology 2004-2008. Maternal factors associated with three phenotypes: spontaneous preterm labour, preterm pre-labour rupture of membranes and medically indicated preterm birth. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet* *25*, 642-647.
- Kramer, M.R., Hogue, C.J., Dunlop, A.L., and Menon, R. (2011). Preconceptional stress and racial disparities in preterm birth: an overview. *Acta obstetrica et gynecologica Scandinavica* *90*, 1307-1316.
- Leitich, H., and Kiss, H. (2007). Asymptomatic bacterial vaginosis and intermediate flora as risk factors for adverse pregnancy outcome. *Best practice & research. Clinical obstetrics & gynaecology* *21*, 375-390.
- Marchetti, M.A., Pizarro, G.O., Sagher, D., Deamicis, C., Brot, N., Hejtmancik, J.F., Weissbach, H., and Kantorow, M. (2005). Methionine sulfoxide reductases B1, B2, and B3 are present in the human lens and confer oxidative stress resistance to lens cells. *Investigative ophthalmology & visual science* *46*, 2107-2112.
- Mason, S., Piper, M., Gronostajski, R.M., and Richards, L.J. (2009). Nuclear factor one transcription factors in CNS development. *Molecular neurobiology* *39*, 10-23.
- Messina, G., Biressi, S., Monteverde, S., Magli, A., Cassano, M., Perani, L., Roncaglia, E., Tagliafico, E., Starnes, L., Campbell, C.E., *et al.* (2010). Nfix regulates fetal-specific transcription in developing skeletal muscle. *Cell* *140*, 554-566.

- Michels, K.B., Harris, H.R., and Barault, L. (2011). Birthweight, maternal weight trajectories and global DNA methylation of LINE-1 repetitive elements. *PloS one* *6*, e25254.
- Moldenhauer, J.S., Stanek, J., Warshak, C., Khoury, J., and Sibai, B. (2003). The frequency and severity of placental findings in women with preeclampsia are gestational age dependent. *American journal of obstetrics and gynecology* *189*, 1173-1177.
- Pillas, D., Hoggart, C.J., Evans, D.M., O'Reilly, P.F., Sipila, K., Lahdesmaki, R., Millwood, I.Y., Kaakinen, M., Netuveli, G., Blane, D., *et al.* (2010). Genome-wide association study reveals multiple loci associated with primary tooth development during infancy. *PLoS genetics* *6*, e1000856.
- Rahi, J.S., and Dezateux, C. (2000). Congenital and infantile cataract in the United Kingdom: underlying or associated factors. British Congenital Cataract Interest Group. *Investigative ophthalmology & visual science* *41*, 2108-2114.
- Ramirez, J., Lukin, K., and Hagman, J. (2010). From hematopoietic progenitors to B cells: mechanisms of lineage restriction and commitment. *Current opinion in immunology* *22*, 177-184.
- Satyanarayana, A., Gudmundsson, K.O., Chen, X., Coppola, V., Tessarollo, L., Keller, J.R., and Hou, S.X. (2010). RapGEF2 is essential for embryonic hematopoiesis but dispensable for adult hematopoiesis. *Blood* *116*, 2921-2931.
- Smith, R. (2007). Parturition. *The New England journal of medicine* *356*, 271-283.
- Social Scientific Study Group on Minamata Disease (1999). In the Hope of Avoiding Repetition of Tragedy of Minamata Disease. Minamata, Japan:National Institute for Minamata Disease. Available <http://www.nimd.go.jp/syakai/webversion/SSSGMDreport.html> [accessed 8 July 2010].
- Strain, J.J., Davidson, P.W., Bonham, M.P., Duffy, E.M., Stokes-Riner, A., Thurston, S.W., Wallace, J.M., Robson, P.J., Shamlaye, C.F., Georger, L.A., *et al.* (2008). Associations of maternal long-chain polyunsaturated fatty acids, methyl mercury, and infant development in the Seychelles Child Development Nutrition Study. *Neurotoxicology* *29*, 776-782.
- Sutherland, J.E., and Costa, M. (2003). Epigenetics and the environment. *Annals of the New York Academy of Sciences* *983*, 151-160.
- Vicente, C., Conchillo, A., Garcia-Sanchez, M.A., and Odero, M.D. (2012). The role of the GATA2 transcription factor in normal and malignant hematopoiesis. *Critical reviews in oncology/hematology* *82*, 1-17.

Walsh, B.K., Brooks, T.M., and Grenier, B.M. (2009). Oxygen therapy in the neonatal care environment. *Respiratory care* 54, 1193-1202.

Curriculum Vitae

Name: Hwa Jin Lee

Sex: Male

Address: 111 West Centre Street APT 1808, Baltimore MD 21201, USA

Contact number: 607- 351-7915

Email: hwajin2k@gmail.com, hlee133@jhmi.edu

Education:

Johns Hopkins University, Baltimore, MD

- Cellular and Molecular Medicine Ph.D program
- Advisor: Dr. Andrew P.Feinberg
- Research Area: Perinatal Epigenome, Cancer Epigenetics
- Ph.D candidate

Cornell University, Ithaca, NY

- Bachelor of Arts Degree in Biological Sciences, May 2007
- Concentration: Molecular and Cell Biology
- Graduated as Honors (*cum laude*)

Kyoungnam Science High School, JinJu, Kyoungnam-Province, South Korea

- Graduated in February 2003

Awards and Scholarships:

Samsung Scholar

- Received from Samsung scholarship foundation in October 2007

Presidential Science scholar in South Korea

- Received from Korea Science and Engineering foundation from 2003 to 2007

National Society of Collegiate Scholars

- Selected to Dean's list for 2 semesters**

Research Experience:

Feinberg Lab in Center for Epigenetics, Johns Hopkins School of Medicine

- PhD candidate
- Projects: Perinatal epigenetic signatures associated with birth characteristics and metal exposure, Distinctive genome-wide epigenetic patterns in Cancer Stem Cells, Chromatin dynamics during cell transition

Felding Lab in Molecular and Experimental Medicine building, The Scripps Research Institute

- Worked as a research assistant during 2007-2008
- Project: Regulation of breast cancer brain metastasis by modulating invasion through Blood Brain Barrier

Gary Whittaker Lab in Veterinary Medical Center, Cornell University

- Worked as an undergraduate researcher during September 2005 till May 2007

- Projects:

- 1) Sequencing analysis of Avian Coronavirus Infectious Bronchitis Virus (IBV) M41 strain and Beaudette strain
- 2) Mutagenesis of IBV M41 Spike protein gene to determine the importance of trypsin cleavage site and furin cleavage site in different tropism and viral assembly signaling

In-sub Han Lab in Immunomodulation Research Center, Ulsan University

- Worked as an intern from January 2003 to June 2003

- Project: Identifying potential novel signal transduction protein involved in hematopoietic stem cell (HSC) differentiation inhibition by Ck β 8-1 chemokine

Publications:

1. Heparan sulfate is a selective attachment factor for the avian coronavirus infectious bronchitis virus Beaudette.

Madu IG, Chu VC, Lee H, Regan AD, Bauman BE, Whittaker GR.

Avian Dis. 2007 Mar;51(1):45-51.

2. Comprehensive methylome map of lineage commitment from haematopoietic progenitors.

Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, Rossi DJ, Inlay MA, Serwold T, Karsunky H, Ho L, Daley GQ, Weissman IL, Feinberg AP.

Nature. 2010 Sep 16;467(7313):338-42. Epub 2010 Aug 15.

3. Comparison of in vitro and in vivo approaches to studying brain colonization by breast cancer cells.

Lorger M, Lee H, Forsyth JS, Felding-Habermann B.

J Neurooncol. 2011 Sep;104(3):689-96. Epub 2011 Feb 27.

4. DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth.

Lee H, Jaffe AE, Feinberg JI, Tryggvadottir R, Brown S, Montano C, Aryee MJ, Irizarry RA, Herbstman J, Witter FR, Goldman LR, Feinberg AP, Fallin MD.

Int J Epidemiol. 2012 Feb;41(1):188-99.

5. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies.

Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA.
Int J Epidemiol. 2012 Feb;41(1):200-9.